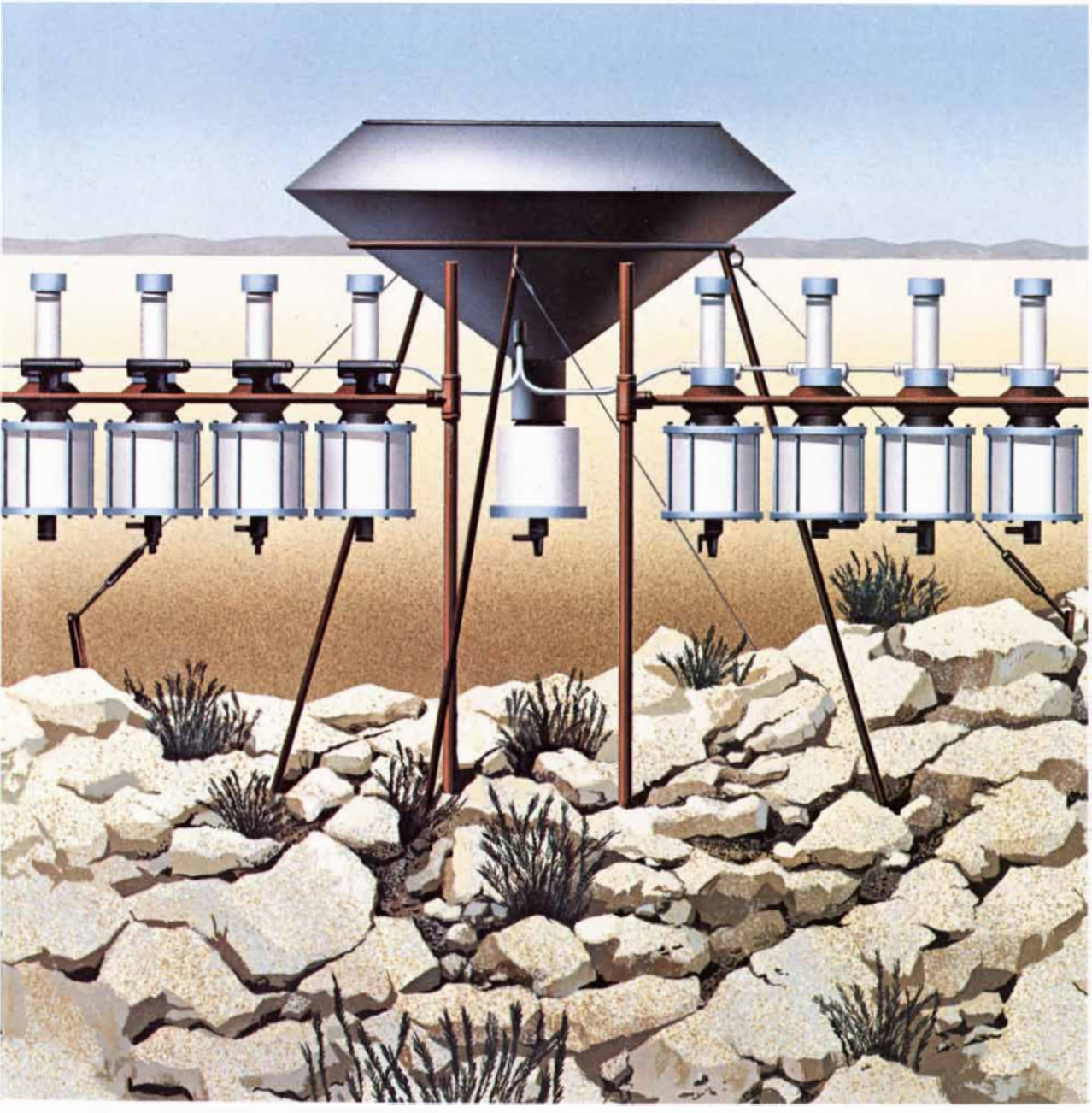
SCIENTIFIC AMERICAN



SINAI-NEGEV PALEOWATER

$2.50

July 1985

...10...9...8...

Superior technology should simplify things. Not complicate them. So years ago, when Honeywell supplied computer systems to support the checkout and launch of the space shuttle, we kept things simple. Working with NASA, Honeywell developed several systems that perform pre-launch, launch and inventory functions. Key tasks that help slingshot multi-ton space ships into orbit. For a white paper on the technology involved, write Shuttle, c/o Honeywell, Honeywell Plaza, MN12-4164, Minneapolis, MN 55408. Or call 1-800-328-5111, ext. 2762.

Together, we can find the answers.

Honeywell

The Straightaway.

The Hairpin.

The S's.

The Loop.

The 1985 Honda Accord Hatchback has a powerful 12-valve engine to speed you down the Straightaway at the Daytona Speedway. It has front-wheel drive to pull you through the Hairpin at Sears Point Raceway. A special sport suspension that can straighten out the S's at Lime Rock Park. And a long list of standard features to take you in comfort through The Loop in Chicago.

HONDA

The Accord Hatchback

Established 1845 **SCIENTIFIC AMERICAN** July 1985 Volume 253 Number 1

ARTICLES

DEPARTMENTS

SCIENTIFIC AMERICAN (ISSN 0036-8733), PUBLISHED MONTHLY BY SCIENTIFIC AMERICAN, INC., 415 MADISON AVENUE, NEW YORK, N.Y. 10017. SECOND-CLASS POSTAGE PAID AT NEW YORK, N.Y., AND AT ADDITIONAL MAILING OFFICES. AUTHORIZED AS SECOND-CLASS MAIL BY THE POST OFFICE DEPARTMENT, OTTAWA, CANADA, AND FOR PAYMENT OF POSTAGE IN CASH. SUBSCRIPTION RATE: $24 PER YEAR IN THE U.S. AND ITS POSSESSIONS; $33 PER YEAR IN ALL OTHER COUNTRIES. POSTMASTER: SEND ADDRESS CHANGES TO SCIENTIFIC AMERICAN, 415 MADISON AVENUE, NEW YORK, N.Y. 10017.

SCIENTIFIC AMERICAN

CORRESPONDENCE

Offprints of more than 1,000 selected articles from earlier issues of this magazine, listed in an annual catalogue, are available at $1.25 each. Correspondence, orders and requests for the catalogue should be addressed to W. H. Freeman and Company, 4419 West 1980 South, Salt Lake City, UT 84104. Offprints adopted for classroom use may be ordered direct or through a college bookstore. Sets of 10 or more Offprints are collated by the publisher and are delivered as sets to bookstores.

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

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

Subscription correspondence should be addressed to Subscription Manager, SCIENTIFIC AMERICAN, P.O. Box 5969, New York, NY 10017. The date of the last issue on your subscription is shown in the upper right-hand corner of each month's mailing label. For change of address notify us at least four weeks in advance. Please send your old address (if convenient, on a mailing label of a recent issue) as well as the new one.

Name

New Address

Street

City

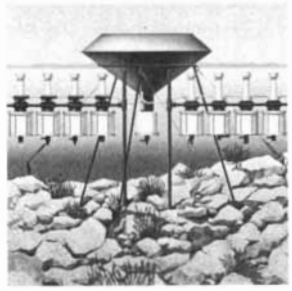
State and ZIP

Old Address

Street

City

State and ZIP



THE COVER

The painting on the cover depicts a water-sampling device astride a barren desert of the Sinai-Negev peninsula. The arid surface conditions belie the subterranean geology: vast quantities of fossil water fill the pores of a sandstone layer underlying the peninsula. Rainwater collected by the sampler provided evidence bearing on the history of the underground reservoir (see "Fossil Water under the Sinai-Negev Peninsula," by Arie Issar, page 104). The sampler was designed to protect water samples from evaporation and contamination in the harsh desert environment. The large funnel collects rainwater and routes it to the horizontal row of containers under it. A magnetic device hermetically seals each container when it is filled, and the additional flow is subsequently diverted to the next container in the series.

THE ILLUSTRATIONS

Cover painting by Ted Lodigensky

Page	Source	Page	Source
8	SCIENTIFIC AMERICAN	87	James Kilkelly
15	The Learning Company	88–91	Ilil Arbel
16–17	Ilil Arbel	92	James Kilkelly
20	National Aeronautics and Space Administration	95	J. Wilson Myers, Eleanor Emlen Myers and Gerald Cadogan
21	William K. Hartmann	96	Andrew Tomko (*top*); Raymond V. Schoder, S.J., Loyola University (*bottom*)
27	Brian J. Ford		
32–37	Ilil Arbel		
39	Jet Propulsion Laboratory, California Institute of Technology	98	J. Wilson Myers, Eleanor Emlen Myers and Gerald Cadogan
40–42	Ian Worpole	100	Alan D. Iselin
43	Christian Veillet, Centre d'Études et de Recherches Géodynamiques et Astronomiques	101–102	Andrew Tomko
		105	Earth Survey Corporation
		106	Andrew Tomko
44–45	Ian Worpole	107	Arie Issar, The Jacob Blaustein Institute for Desert Research
46	U.S. Naval Observatory (*top*), Ian Worpole (*bottom*)	109–110	Andrew Tomko
47	Ian Worpole	113	Lynn M. Stone, Earth Scenes
49	Quesada/Burke (*top*), International Business Machines Corporation (*bottom*)	114–116	Enid Kotschnig
		117	Lynn M. Stone, Animals, Animals
50–56	Andrew Christie	118–119	Ilil Arbel
73	The British Museum	122	Worth Booth
74–82	Patricia J. Wynne	123–126	Michael Goodman

PRINTER BARGAIN

We guarantee that you'll never find a bargain as sensational as the letter-quality daisy wheel printer you see in this advertisement.

We covered up the name of the printer soas not to embarrass anyone who may be selling it at $1795.

I realize that what you are about to read may seem incredible. I can understand. But occasionally there are indeed bargains and opportunities that only come once in a lifetime. I'm convinced that this is one of them.

By Joseph Sugarman, President

The letter-quality printer you see above has a suggested retail price of $1795. It prints bi-directionally at 40 characters per second using a daisy wheel print element, comes with a parallel interface and prints a 13.6 inch line.

A similar printer might be found at some discount computer dealers for as low as $1200. And at that price it is indeed a bargain.

AUTOMATIC FEATURES

If we'd offer it to you at $599—a unitthat we could be selling for $1795 and that would be a bargain at $1200—you'd probably think that there was a catch. But guess again. As unusual a bargain as this may appear, and despite all its quality features, this printer can be yours for the incredible low price of only $599—below wholesale, below dealer cost and without question, the lowest priced high-speed letter-quality printer sold today.

JS&A bought out an entire warehouse full of these printers, promised that we would not display the name on the unit, in our photo, mention the name in our ad nor reveal the name over the phone to avoid embarrassing the manufacturer or his dealers.

I'm so convinced that this is one of the greatest values I've ever offered, I am making a bet and a commitment. First, thecommitment. I'm giving you 30 days to test it out. If it's not the best printer value in the country, return it for a full refund including the $25 postage and handling charge.

Finally, I'll bet that you'll immediately recognize the name of the billion dollar company whosenameis on the product or you don't deserve to own a computer. The company is a solidly successful computer company whose products you may even nowown. They'll back the product with service through its 90-day limited warranty and for years to come with hundreds of national factory service centers throughout the United States.

EXCEPTIONAL VALUE

The printer is not an outdated model with old technology but a brand new unit with the latest electronics and the most advanced technology. For example, a precision print wheel produces outstanding print quality. There's an automatic print pressure control which automatically varies the printing pressure according to the shape of the character. This single feature produces an incredibly clean impression while prolonging the life of the daisy wheel. But there's more.

An aluminum diecast integral-constructed frame gives the printer a solid home for its advanced electronics. And with a weight of 30 pounds, you know there's built-in commercial-quality construction. Thecontrols include: 'line feed' which advances the carriage by one line, 'page advance' which advances the document to the next page when using continuous form paper and a 'set page' button that tells the printer where the start of the form is located. A lighted condition panel tells you the printer status with red and green LEDs. You can use single sheets or continuous form fan-fold paper and with the 'paper out sensor' the printer detects the last sheet of the fan fold paper and automatically stops

PRINTING FLEXIBILITY

There are also features that give you enormous printing flexibility. You can underscore words, double print each character which creates a bold look oryou can use shadow print which moves the print head 1/120th of an inch between strikes. With the proper printhead, you can also set the printer for proportional printing which gives your documents a professional—almost printed look.

There's a self-test mode which lets you print out all the characters on your daisy wheel continually until you stop. And the system uses standard Diablo daisy wheels and ribbons which you can get from JS&A or any computer store.

With the Pica pitch, you can print up to 136 columns and with the Elite pitch, up to 163. The 15.5'' carriage will take a print area of 13.6 inches. It measures 6 x 16 x 24'', comes with a 10-pitch daisywheel, one ribbon and complete instructions. The unit has provisions for a tractor feed and one is available from JS&A.

You can select either 10, 12 or 15 for the print pitch or even use the 10-pitch daisywheel supplied with the unit at the 12-pitch seetting for large and tight letter spacing. There are dip switches which let you customize each printer to any computer with a parallel printer interface. Setting recommendations are supplied for IBM, Apple and other popular computers.

What happened? How can JS&A obtain and then sell these printers—products that are brand new with the latest state-of-the-art technology and from a major manufacturer at a price that at first is hard to believe? Quite frankly, it wasn't easy.

With our low overhead we can efficiently sell these units in tremendous quantities without the high markups that many stores must have to make a profit. And we can do this on a mass-market national scale. Since many manufacturers know this and know that we could quickly move tremendous amounts of merchandise without upsetting many of the retailers (because we keep the name of the manufacturer confidential) they are willing to let us buy their product direct, often at foreign, export prices. The savings are sensational. In this case however, not only did we get a sensational price but we also managed to scoop even the manufacturer by introducing the product first.

We'll be happy to supply companies with several printers for their computer departments to upgrade their printing speed and quality. There is no limit to the number you can order although we only have a few thousand available so we reserve the right to return your order should we run out.

ACT QUICKLY

There's no quantity discount and please don't call our operators for the name of the manufacturer. And finally, I urge you to act quickly. The last computer promotion we ran sold out so quickly that we eneded up returning thousands of dollars in checks only a few short weeks after we announced the sale.

There are bargains available that are indeed too good to be true and often end up to be much less than you expected. But here's an example that is not only too good to be true but that we guarantee you'll find better than you expected. Order one at no obligation, today.

Simply send your check or money order using the order numbers shown in parentheses (IL residents add 7% sales tax) plus $25 for postage and handling for each printer ordered. Credit card buyers call toll-free number below. Send your order to: JS&A Special Printer Bargain Offer at the address below.

Printer (6083)	**$599**
Tractor (6084)	**189**
3-Pack Ribbons (6085)	**19**
Daisy Wheels: (List supplied with unit)	
Standard Parallel Cable (6086)	**39**

JS&A® PRODUCTS THAT THINK®

One JS&A Plaza, Dept SA
Northbrook, Illinois 60062 (312) 564-7000

ORDERS ONLY: 800 228-5000

IL residents add 7% sales tax. ©JS&A Group, Inc.,1985

LETTERS

To the Editors:

"The Growth of Core Regions in the Third World," by Daniel R. Vining, Jr. [SCIENTIFIC AMERICAN, April], is a remarkable illustration of the difference between science and ideology. As a study of development it demonstrates that economic growth in the Third World is strongly linked with urbanization and a non-Communist government. This is interesting, although not remarkable, for the same has been true in the developed world. Vining chooses, however, to address the data as though urban growth were a problem requiring explanation and correction rather than a natural companion (indeed a virtually necessary condition) of desired economic growth.

Your headlines, pictures and apparatus take this strange approach several steps further by apparently lauding the "techniques" Communist governments employ to control their populations. Thus your headline says urban population increase poses "grave social and economic problems," and an illustration presents P'Yŏngyang as the result of "careful" control, of building "in a planned way" and of "strictly limited" growth, resulting in streets that are "uncrowded." A truly idyllic scene. (Perhaps symbolically, the people in the photograph are totally indistinguishable as well as virtually invisible.)

In contrast, the picture of Seoul, which to the average eye appears to show a cheerful commercial scene, strikes you otherwise. The scene is the result of "uncontrollable" growth, an indication of systems that are "not yet complete" and of a city that faces multiple "shortages." If these "shortages" are so bad, one might ask why people continue to come to Seoul not only from the periphery in South Korea but also from North Korea if they can. The answer is that conditions there are substantially better than in the competing areas and than in the past.

Communist successes at demographic control have ranged from the Ukrainian famine and the liquidation of the kulaks in the Soviet Union through the Communist conquest of China and the famines of 1961–63. Vining's exceedingly modest description of the Cambodian holocaust as "the most disturbing example of the stringent control exerted by a powerful state" raises the question of how he might have characterized German "population control policies" in Poland and Russia in the early 1940's.

In reality the relation of urbanization and economic systems to economic growth has been clear for many years. No country in the bottom third of countries by per capita income is *more than* 38 percent urban, and only seven of 39 are more than 30 percent urban. Of the top 28 countries in per capita income, only two are *less than* 54 percent urban (both countries have populations of less than two million); on the average such wealthy countries are more than 75 percent urban.

I continue to find it interesting that intellectuals, in the face of massive evidence to the contrary, choose to focus on urbanization as though it were a problem that needs to be solved, even by totalitarian means.

DANNY J. BOGGS

Deputy Secretary of Energy
Washington, D.C.

To the Editors:

The achievements of Sri Lanka in providing for the well-being of its people, as reported by Daniel R. Vining, Jr., seem to contradict his own statement about the development of nations. He states: "Economic growth is, of course, essential to providing a better life for a country's citizenry." In the paragraphs preceding that statement, he describes the glittering record of Sri Lanka in providing that better life: the government supplies "virtually equal access to food, shelter, education and health care" everywhere in the country. As a result Sri Lanka enjoys "exceptionally high life expectancy, a high literacy rate, a low infant mortality rate and very little undernourishment" compared with other Third World countries. He then states that the price of this achievement has been a reduced rate of economic growth.

Either conditions not mentioned by Vining make the case of Sri Lanka unique or Sri Lanka's experience demonstrates that high rates of economic growth are not essential to providing a better life for a country's citizenry.

JOHN S. WARREN

Brattleboro, Vt.

To the Editors:

As Deputy Secretary Boggs says, a high degree of urbanization (the fraction of people living in settlements of, say, 5,000 or more) appears to be a necessary condition for economic development. Economic development entails the shift of labor from agricultural to nonagricultural sectors. Since nonagricultural activities are most efficiently organized in urban agglomerations, the spatial concentration of population always accompanies economic development.

What is less determinate in this process is the regional distribution of these agglomerations and their comparative sizes. In most developing countries urbanization and hence development are very unevenly distributed across regions. Such unbalanced regional development causes stresses that are, to Third World governments at least, of serious and ongoing concern, as urbanization per se is not. Boggs makes light of this concern. We shall have to see if his laissez-faire attitude is justified.

It is fashionable nowadays to be cynical about the motives and effectiveness of central governments. I know of no evidence, however, that they are not reasonably good at sensing problems that emerge in their countries and threaten those countries' unity, security and well-being. I would imagine that even Boggs would agree, for example, that the South Korean government's alarm at having half of that nation's very dynamic economy and virtually all its higher-order activities concentrated in a single city and its environs only 30 kilometers from the army of North Korea is well intentioned and amply justified by history. And I suspect that Boggs would entertain the possibility that spending more than half of a country's capital budget in order to keep the primary city functioning could cause resentment elsewhere strong enough to threaten the unity of the country and that this is a legitimate concern of the central government, as is the fact that agglomerations of 20 to 30 million people may not be ecologically sustainable.

Unfortunately Boggs does not address the problems associated with the size and regional distributions of Third World urban agglomerations, which were the subject of my article, but rather urbanization per se. Urbanization in general has not attracted the same degree of policy concern in the Third World as its skewed distribution has. I was probably not as careful as I should have been to stress the distinction between the two. I am grateful to Boggs for forcing me to stress it here.

Boggs objects to my dispassionate description of another of the most striking universals in spatial demography: the slow growth of large cities in Communist nations. I tried to explain, in the space available to me, the mechanisms by which this is accomplished, to wit the destruction by the state of

competing institutions. This enables the state to penetrate the society and lay down a particularly fine-grained "grid of intelligibility," as it has been called, making it possible to monitor and control population movements quite carefully. As the sociologist Paul Hollander observed: "It is easier for Americans to go to the moon than for the Chinese to get around the strict residential control system." The failure of contemporary right-wing regimes to control population movements in their countries is consistent with the view that such regimes are not totalitarian to the same degree as Communist states.

Mr. Warren's letter likewise points to an ambiguity in my text but one that is easier to clarify. The case of Sri Lanka proves that a country can better the lives of its citizens in a number of ways although its population remains relatively dispersed. What has not been shown is how a country can raise itself to economic parity with the developed world without setting in motion large-scale territorial transformations and creating regional imbalances. Warren may ask why economic parity with the West (which is really what we mean by economic development) should be the goal of any country or people, but in fact it is. In addition to good health, long lives, freedom from hunger, and adequate shelter (the so-called basic needs) people also want the abundance of material goods and the variety of occupational and personal choices that are found only in the developed world. They are not yet available to the Sri Lankan people, and it is not apparent that Sri Lanka has even embarked on a path that will lead to such abundance. I do not denigrate Warren's apparent impatience with those who would make such a goal primary. I would simply suggest that for most people and in most places, including Sri Lankans and Sri Lanka, this goal is in fact primary.

DANIEL R. VINING, JR.

University of Pennsylvania
Philadelphia

ERRATUM

In "Anesthesiology," by Peter M. Winter and John N. Miller (SCIENTIFIC AMERICAN, April), the diagram of the chemical structure of nitrous oxide (N_2O) was in error. It depicted nitrogen dioxide (NO_2), which is not an anesthetic.

50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

JULY, 1935: "Recent reports from California announce the discovery of four white-dwarf stars, doubling the known number of these remarkable bodies. Why should they be so unlike other stars and so like one another, and what stage in a star's possible history do they represent? A theory generally accepted is that these tiny bodies are stars in the process of going out. The theory applies only to masses of stellar order of magnitude, for which the central pressures are very great."

"Surprising discoveries made in the Near East by archaeologists from the Oriental Institute of the University of Chicago have thrown new light on the technical progress of mankind in the third millennium B.C. and lead to the conclusion, expressed by Henri Frankfort of the Institute, that 'man's mastery over matter progressed further in early dynastic and Akkadian times than is often believed.' Among the results of excavations in the Tigris valley near Baghdad, Frankfort reports (1) the discovery of glass dating from 2600 or 2700 B.C., (2) the discovery of evidence that terrestrial iron was employed for weapons before 2700 B.C. and (3) the discovery in a private house of four arched doorways."

"There is every indication of an awakened interest in rotating-wing aircraft. Designers are devoting increased attention to the inherent possibilities of machines with lifting surfaces that revolve about either horizontal or vertical axes. A prime object is to ensure ascent and descent in an almost vertical path and thus to make usable areas now restricted for landing and takeoff. The helicopter, long considered an aeronautical dream if not an actual nightmare, gives promise of rapid emergence into a more practical realm."

"'Brain waves' tapped electrically are providing a new clue to the mystery of epilepsy. Frederic A. Gibbs, Hallowell Davis and E. Lovett Garceau of the Harvard Medical School have reported to the American Physiological Society that an electrical hookup to the brain producing wavy lines traced on paper gives a new clue to what goes wrong in this malady. They find that epilepsy is probably a neurological storm that results in a great piling up of electrical discharges."

"All-wave reception has become a universal hobby. Every day and hour there are literally thousands of people in all parts of the world occupied in the diversion of tracking down distant stations, or merely enjoying the varied programs, educational features and unexpected thrills the wave bands of radio have to offer. It is the greatest show on earth."

SCIENTIFIC AMERICAN

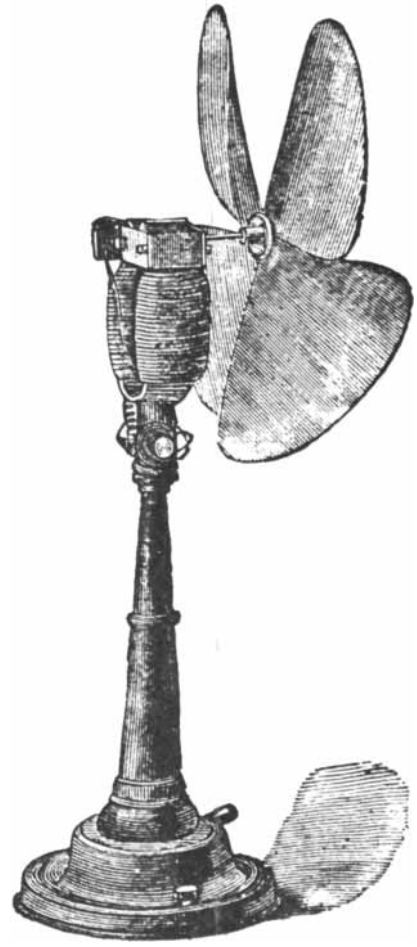
JULY, 1885: "There is nothing to which chemistry has been applied that is more wonderful than the results that have followed the utilization of common gas or coal tar. Thirty years ago the refuse of the retorts in gas works was utterly useless. Practical chemists were then applied to, and one of their first achievements was to discover that naphtha could be extracted from this refuse. After the naphtha was extracted the tar was left, in the form of a heavy oil, and this was still more of a nuisance than the original compound. Faraday next awoke interest in coal tar by his discovery of benzine as a product of the tar oil. In 1857 William Henry Perkin made the wonderful discovery that coal tar has aniline properties. This discovery has almost revolutionized the trade in dyestuffs."

"The course in electrical engineering in Cornell University has now been established for two years and is already well patronized. It requires four years of study for its completion, the object being to turn out, not electricians or electrical engineers merely, but educated men. The students pursue mathematics through calculus, study the French or the German language, give some time to the study of English, devote several terms to the theory and practice of machine drawing and study mechanics as applied to engineering—all besides the work in general physics and electricity."

"Several attempts have been made to utilize electrical energy through the medium of secondary batteries for propelling tricycles and light vehicles, but so far we have not seen anything beyond the experimental stage. Many inventors have also striven to produce a mechanical motor depending for its movement upon the explosion of a gaseous mixture composed of petroleum and compressed air. Such a vehicle can be seen in the Inventions Exhibition. It is stated that by the consumption of from three to five pints of common petroleum oil per hour in the 'velocycle,' as it is called, sufficient power is generated to give the vehicle together with its rider a speed of from 10 to 15 miles per hour."

"This is the season for showers, rainbows and thunder storms. We suggest that a photograph of a first-class rainbow might be an interesting subject for experiment with the camera. Another possibility is a night exposure of a plate when the lightning is vivid. Photographs of lightning strokes have been made, but we call to mind none of the rainbow."

"A very refreshing invention, especially for the hot-weather season, is the electric fan shown in our engraving. It consists of an ornamental standard about a foot high on which is mounted a screw-propeller fan. When the user connects the wires of a battery with the standard, the fan revolves rapidly and delivers a cool breeze in any direction desired."



The electric fan

A quick, close-ratio 5-speed. Tight, rack-and-pinion power steering. Power brakes. Radio with digital clock. Dual reclining bucket seats. Mileage ratings of 34 est. hwy. mpg and 23 est. city mpg.**

And a standard 5 year/ 50,000 mile Protection Plan† that'll even cover an optional turbocharged engine.

See your Dodge Dealer about buying or leasing† a new Daytona. For just $8505* you won't just get outfitted for fun, you'll be dressed to thrill.

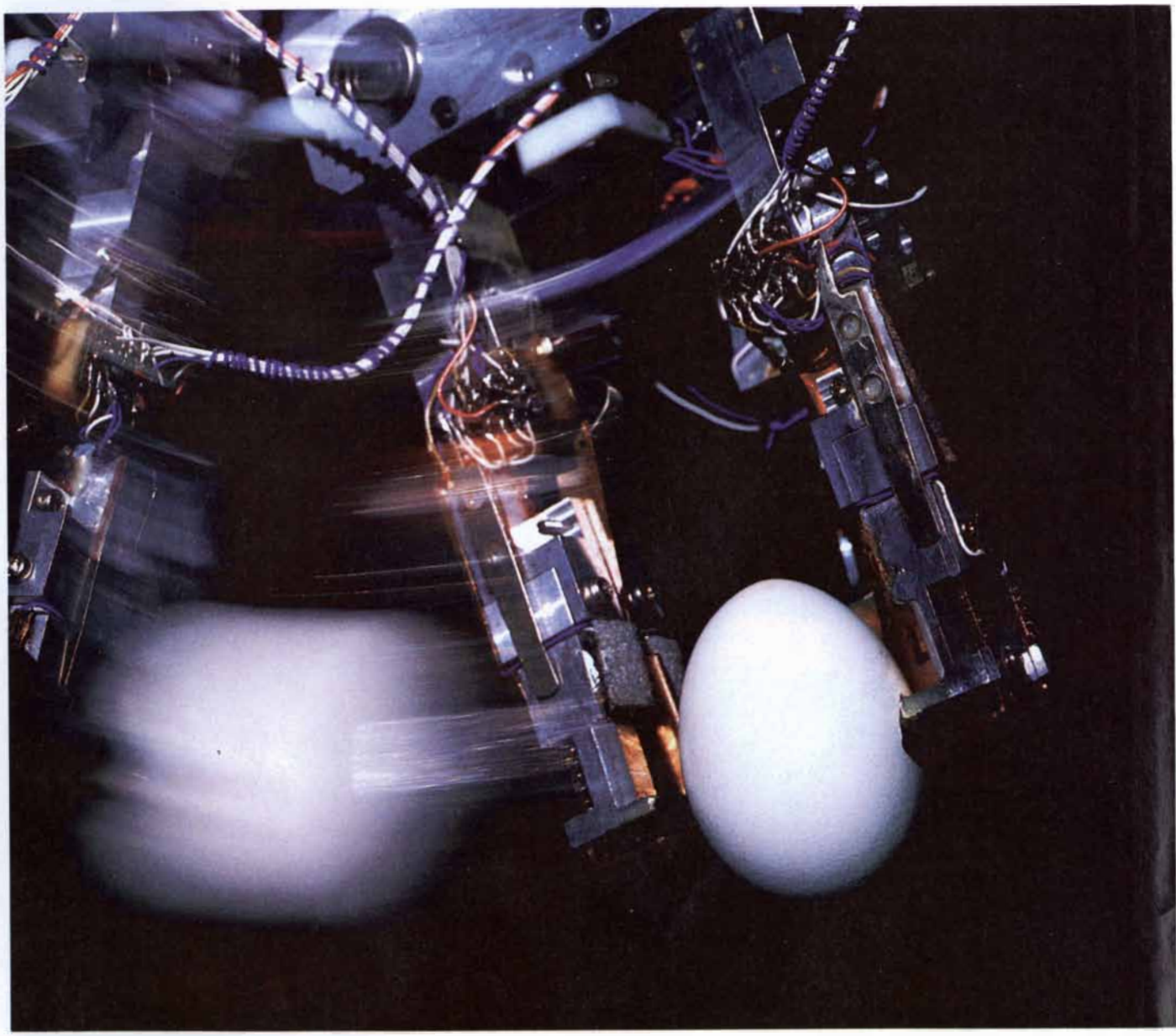
*Sticker price excludes tax and destination charges. Aluminum wheels shown, $322 extra. **EPA ests. for comparison. Your mileage may vary depending on options, driving conditions and habits, and vehicle's condition. †Whichever comes first. Limited warranty on engine, powertrain, and outer body rust-through. Excludes leases. Deductible applies. Ask for details. BUCKLE UP FOR SAFETY.

Dodge
DIVISION OF CHRYSLER

AN AMERICAN REVOLUTION

DRESSED TO THRILL.

Dodge Daytona. Fast. Fuel-injected. Fuel efficient. Front-wheel drive. 5/50 protection. Just $8505.



Hitachi's visual-tactile robot can handle objects as fragile as eggs, because its sensors detect size, shape and required pressure to attain sensitivity almost equal to that of a human hand.

ROBOT

Nearly two decades ago, Hitachi began turning common science fiction into startling industrial fact. The device: The company's first servo-manipulator, a key component in the development of real robots to eliminate the monotony, danger and dirty work of manufacturing.

Your mechanical right-hand man

Today, the results of Hitachi research are in use all around you. Robot welders using microcomputers and built-in sensors to detect weld lines automatically. Spray-painting robots capable of remembering up to 2,000 instructions and performing 99 different painting tasks. Process robots that can be programmed for new job functions through a simple teaching box. Robots on wheels for transporting parts and warehouse stock.

Our electronics and mechanical engineering experts have joined their talents to give robots the benefits of high technology. They have created models with expanded memory capacities and advanced sensing systems. And they have applied them in Hitachi's own factories, where our production specialists suggest further refinements.

In fact, we are constantly coming up with innovations and new applications. One of the latest: A visual-tactile sensing robot with multiple arms and seven camera eyes, developed to independently assemble home appliances such as vacuum cleaners.

These examples demonstrate a few of the ways in which Hitachi is improving upon basic technology. Then using it to create practical tools that meet your needs...and those of professionals in marine exploration, aerospace, and virtually every other field you can name.

The best of worlds is yet to come

Our vision of the future includes robots with artificial intelligence that will learn from their own experiences. Flexible manufacturing systems where robots handle every step of production. Personal robots that will take the drudgery out of household chores. And much, much more.

We'd like you to share in the benefits of our scientific research, covering the next generation of lasers, sensors and other electronic devices. For improved business efficiency. For a higher quality of life. Two goals we've pursued for 75 years as part of our commitment to a better world through electronics.

WE BELIEVE ROBOTS FREE MINDS TO CREATE BY FREEING BODIES FROM TOIL

HITACHI

Robotics Group, Industrial Components Sales & Service Division, Hitachi America, Ltd., 50 Prospect Avenue, Tarrytown, NY 10591-4698, Tel: 914-332-5800

THE AUTHORS

C. ARDEN MILLER ("Infant Mortality in the U.S.") is chairman of the department of maternal and child health in the school of public health at the University of North Carolina in Chapel Hill, where he is also professor. A 1944 graduate of Oberlin College, he got his M.D. at Yale University in 1948. In 1951 he joined the department of pediatrics at the University of Kansas School of Medicine. He remained there until 1966, eventually becoming dean and provost of the University of Kansas Medical Center. Miller then moved to Chapel Hill.

ROBERT HAMILTON BROWN and DALE P. CRUIKSHANK ("The Moons of Uranus, Neptune and Pluto") are planetary astronomers. Brown is on the scientific staff of the Jet Propulsion Laboratory of the California Institute of Technology. He was educated at Purdue University and the University of Hawaii at Manoa, receiving a Ph.D. in astronomy in 1982. He then became assistant researcher in planetary science at the Hawaii Institute of Geophysics, where he did telescopic studies of satellites in the outer solar system. At the beginning of this year Brown moved to the J.P.L., where he works with the Voyager Imaging Science Team, which is currently preparing for the encounter of *Voyager 2* with Uranus. Cruikshank is a graduate of Iowa State University and the University of Arizona, which granted him a Ph.D. in astronomy in 1968. He joined the Institute of Astronomy at the University of Hawaii, where he holds the position of astronomer, in 1970. Recently Cruikshank's research has focused on the chemistry of small bodies in the solar system: asteroids, planetary satellites and comets.

CHARLES H. BENNETT and ROLF LANDAUER ("The Fundamental Physical Limits of Computation") are respectively a research staff member and an IBM Fellow at the IBM Thomas J. Watson Research Center in Yorktown Heights, N.Y. Bennett, a graduate of Brandeis University and Harvard University, where he got a Ph.D. in chemical physics in 1970, spent two years at the Argonne National Laboratory before joining IBM in 1972. His enduring research interest has been in the mathematical theory of randomness and the relation between statistical mechanics and the theory of computation. Bennett is currently taking a sabbatical leave from IBM to teach computer science at Boston University. Landauer earned a B.S. in 1945 and a Ph.D. in 1950, both from Harvard. In 1952, after a stint at NACA, the National Advisory Committee for Aeronautics (the forerunner of NASA), he joined IBM, where he took part in early research on semiconductors. Landauer served as director of physical sciences at the IBM Yorktown research laboratory and assistant director of IBM's research division before being appointed IBM Fellow in 1969.

G. LEDYARD STEBBINS and FRANCISCO J. AYALA ("The Evolution of Darwinism") are geneticists at the University of California at Davis, where Stebbins is professor emeritus and Ayala is professor. After getting his doctorate from Harvard University in 1931, Stebbins taught first at Colgate University and then at the University of California at Berkeley. In 1950 he moved to Davis, where he founded the department of genetics. Since his retirement in 1973 he has taught at universities in the U.S., Australia, Chile and France as a visiting professor. Stebbins has discovered five new plant species in California, three of which are named for him. Ayala, a native of Spain, studied at Columbia University, where he got a Ph.D. in 1964. In 1971, after teaching at Providence College and Rockefeller University, he moved to Davis. There, in addition to his teaching and research, he was from 1977 to 1981 associate dean for environmental studies and director of the Institute of Ecology.

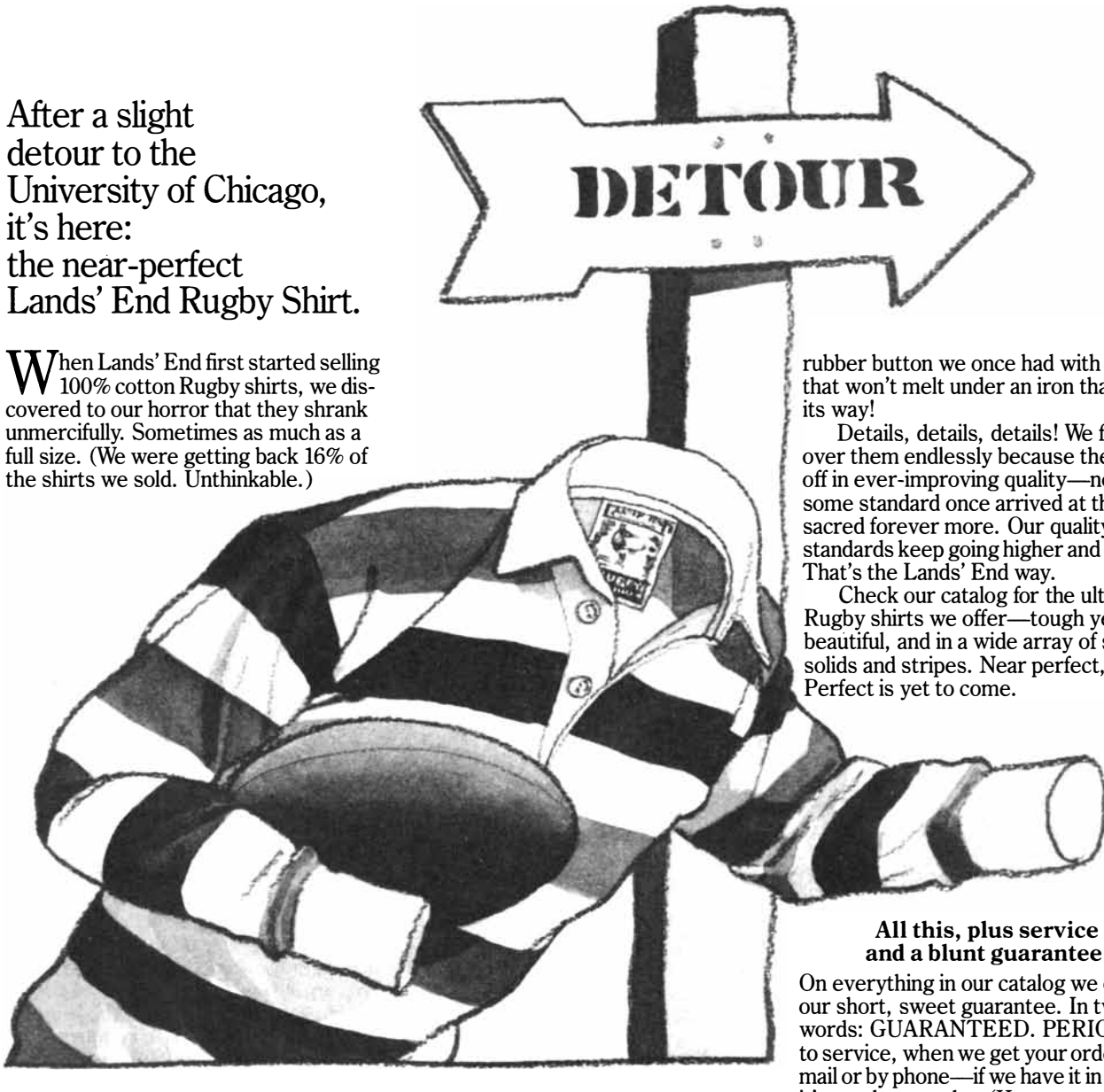
GARY K. BEAUCHAMP, KUNIO YAMAZAKI and EDWARD A. BOYSE ("The Chemosensory Recognition of Genetic Individuality") are respectively member and associate member of the Monell Chemical Senses Center (MCSC) of the University of Pennsylvania and member of the Sloan-Kettering Institute for Cancer Research. Beauchamp was educated at Carleton College and the University of Chicago, where he received a Ph.D. in biopsychology in 1971. He then took a position at the MCSC; he also teaches as adjunct associate professor at the University of Pennsylvania School of Veterinary Medicine. Yamazaki was graduated from the University of Tokyo in 1970 with a doctorate in biology. In 1974 he went to Sloan-Kettering as visiting investigator to work with Boyse and Lewis Thomas on the immunogenetics of self-identification. Since 1980 Yamazaki has been at the MCSC. Boyse earned his degree in medicine at the University of London in 1952 and held various hospital appointments in clinical medicine, pathology and microbiology before turning to full-time research in 1957. He did research in immunogenetics at Guy's Hospital in London until 1960, when he moved to the New York University School of Medicine. In 1962 Boyse took a post at Sloan-Kettering; he holds a concurrent professorship at Cornell University.

PETER M. WARREN ("Minoan Palaces") is professor of ancient history and classical archaeology at the University of Bristol. After getting a doctorate in Minoan archaeology at the University of Cambridge he held teaching posts at the University of Cambridge, the University of Durham, the British School at Athens and the University of Birmingham. In 1977 he took his present position in Bristol. He has directed excavations that uncovered an Early Bronze Age village at Myrtos in southern Crete; he has supervised other digs at Debla in western Crete and at Knossos.

ARIE ISSAR ("Fossil Water under the Sinai-Negev Peninsula") is professor of geology at Ben-Gurion University of the Negev and a member of its Institute for Desert Research. After receiving a master's degree in geology at the Hebrew University of Jerusalem he joined the hydrogeologic division of the Geological Survey of Israel. He took part in a water exploration project in the Negev desert before returning to Hebrew University to complete work for his Ph.D., which he received in 1961. In that year he went to Iran, where he took part first in a program to develop alternative water resources and later in a regional development project. On his return to Israel in 1965 he was appointed head of the water-resources division of the geologic survey and a consultant on international development to the Ministry of Agriculture and the Foreign Office. In 1974 Issar became one of the founders of the Institute for Desert Research.

DOUGLASS H. MORSE ("Milkweeds and Their Visitors") is professor of biology and director of the graduate program in ecology and evolutionary biology at Brown University. He studied biology at Bates College, the University of Michigan and Louisiana State University, where he got his doctorate in 1965. He taught in the department of zoology at the University of Maryland at College Park and worked on the foraging ecology of birds before moving to Brown in 1979.

After a slight detour to the University of Chicago, it's here: the near-perfect Lands' End Rugby Shirt.



When Lands' End first started selling 100% cotton Rugby shirts, we discovered to our horror that they shrank unmercifully. Sometimes as much as a full size. (We were getting back 16% of the shirts we sold. Unthinkable.)

So, we went to work. First, to locate a new manufacturer that pre-shrank his heavyweight 100% cotton jersey knit so it eliminated virtually all shrinkage.

This particular model also sported underarm gussets, to keep the shirt from levitating out of your trousers when you raised your arm in action. A really neat feature.

Still, we felt, there was more we could do.

Typically, we were out to field the toughest Rugby shirt available—an obsession with quality that hounds us constantly whatever item we're pondering. So, thinking to submit our shirt to the acid test, we submitted it to the University of Chicago Rugby Team.

It performed exceptionally well. The players liked it. But we noted a certain amount of premature ripping in the seams and the placket stitching. So, it was back to the drawing board.

Now, though, we're out to face all comers.

We've conquered the seam problem by going to a safety seam stitch. And on the placket problem we designed a continuously looped placket that leaves all stress with the placket itself, not the stitching. We've even changed the rubber button we once had with one that won't melt under an iron that loses its way!

Details, details, details! We fuss over them endlessly because they pay off in ever-improving quality—not just some standard once arrived at that is sacred forever more. Our quality standards keep going higher and higher. That's the Lands' End way.

Check our catalog for the ultimate Rugby shirts we offer—tough yet beautiful, and in a wide array of spirited solids and stripes. Near perfect, so far. Perfect is yet to come.

All this, plus service and a blunt guarantee.

On everything in our catalog we offer our short, sweet guarantee. In two words: GUARANTEED. PERIOD. As to service, when we get your order—by mail or by phone—if we have it in stock, it's out the next day. (How many catalog companies can say that and mean it?)

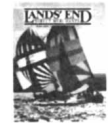
Lands' End Direct Merchants—what pleasant shopping is coming to. Send for our catalog or call 800-356-4444. A friendly, helpful operator will serve you.

We specialize in quality people, too!

LANDS' END®
DIRECT MERCHANTS

of fine wool and cotton sweaters, Oxford button-down shirts, traditional dress clothing, snow wear, deck wear, original Lands' End soft luggage and a multitude of other quality goods from around the world.

Please send free catalog.
Lands' End Dept. Q-32
Dodgeville, WI 53595



Name _____

Address _____

City _____

State _____ Zip _____

Or call Toll-free:
800-356-4444

COMPUTER RECREATIONS

A circuitous odyssey from Robotropolis to the electronic gates of Silicon Valley

by A. K. Dewdney

Like colliding galaxies, computer games and computer education appear to be on the point of a grand merge. A harbinger was the appearance last year of Robot Odyssey, a computer game that is also an education in circuit design.

Many computer-assisted instruction programs that employ a game format have been developed, but none, to my knowledge, offers the natural and engaging intensity of Robot Odyssey. Many computer games have evolved from the primeval chaos of the video arcade onto firm cerebral land. None of them has the educational value of Robot Odyssey.

Robot Odyssey leaves the player no choice but to design logic circuits if he or she is to escape from Robotropolis, an urban dreamscape inhabited entirely by robots. Most of these automatons display open hostility to a human presence. When the player (in the form of a small homunculus on the screen) attempts to run a maze or retrieve an object from a room, a guard robot snatches up the homunculus and dumps it unceremoniously back on the threshold. Three robots turn out to be helpful companions in the escapade. The circuitry of the companion robots, which guard robots ignore, enables them to circumvent the automated inhabitants of Robotropolis by negotiating mazes or retrieving such objects as subway tokens and energy crystals.

Can the player, after making a robot odyssey, design a computer? Not quite, but the exercise is sound. The game begins in a place called the robot room. Here the homunculus, operated through the keyboard, meets three companion robots cheerfully bouncing about in a rectangular area. Each robot behaves differently. One glides about the room hugging the walls. Another bounces horizontally back and forth. The third roams up and down the right-hand wall as though seeking a way through. The player begins by typing an *R* and touching the space bar. This action has the surrealistic effect of converting the homunculus into a remote-control unit, which then commands the robots to halt. At this point the player can restore the remote-control unit to the homuncular form. The homunculus climbs on board any one of the robots. When it does so, a view of the robot's cavernous interior replaces the scene of the room. The homunculus appears to have shrunk to one-tenth its original size. It rides in a robot only when negotiating a maze or in some other situation in which the robot's pattern of locomotion is helpful.

Each companion robot is equipped to sense its environment and to move about in it. Bumpers detect contact with walls. Thrusters point up, down, left and right. A grabber claw can pick objects up and release them. An antenna allows communication with the other two companion robots. Inside each robot this equipment is represented by connectors that can be wired directly into the onboard logic circuit. They are of two types: input connectors, which communicate the status of a sensor (bumper, grabber claw, antenna) to the circuit, and output connectors, which relay circuit commands to an effector (thruster, grabber claw, antenna). Thus the grabber claw and the antenna have both input and output connectors. The grabber claw's output connector activates it, and the input connector determines whether the claw currently holds an object.

The operation of the simplest of the three companion robots shows how the circuitry works. This robot, named Ping-Pong, is controlled by a single flip-flop element wired to horizontal bumpers and thrusters. In Robot Odyssey a flip-flop is a two-sided memory cell in which one side is on while the other is off. When the input to the off side is turned on, even momentarily, the off side flips to on and the on side flops to off. In the absence of any change to its inputs, the flip-flop maintains its state. It remembers. Nothing more complicated than a flip-flop is needed to produce Ping-Pong's behavior: the output of one side is wired to the robot's right-hand thruster, and the input of the same side is wired to the right-hand bumper. If that side of the flip-flop is off and the robot touches a wall on the right, the bumper input turns on and the wire relays the signal to the flip-flop. The off side immediately turns on. The right-hand thruster fires and the robot begins moving to the left. So much for the ping. Pong is arranged by wiring the other side of the flip-flop in the same way to the left-hand bumper and thruster.

The full set of standard logic elements is needed to realize more complicated behavior. This set includes not only the flip-flop memory element but also decision elements called *or*-gates, *and*-gates and *not*-gates, or inverters. Briefly, the output of an *or*-gate is on if at least one of its inputs is on, and the output of an *and*-gate is on if both its inputs are on. The output of an inverter is on if its input is off.

To get beyond the robot room the Robot Odyssey player must retrieve a key from inside one of the robots and open a doorway that leads into a region known as the city sewer. An object is picked up by superposing the homunculus on it and touching the space bar. A beep announces success; thereafter the object moves with the homunculus. Without this facility Robot Odyssey would be impossible: the homunculus must carry the three companion robots as it explores the game's five levels. But how is it to carry the robots when only one object at a time can be carried? Since any object can be taken into or out of a robot and since a robot is itself an object, the player does not take long to think of placing one robot inside another. It works. In fact, the homunculus can drag a robot inside a second robot that is already inside a third. This tactic saves space inside the third robot; as the adventure proceeds its interior gets cluttered with subway tokens, sensors, energy crystals and more circuit elements.

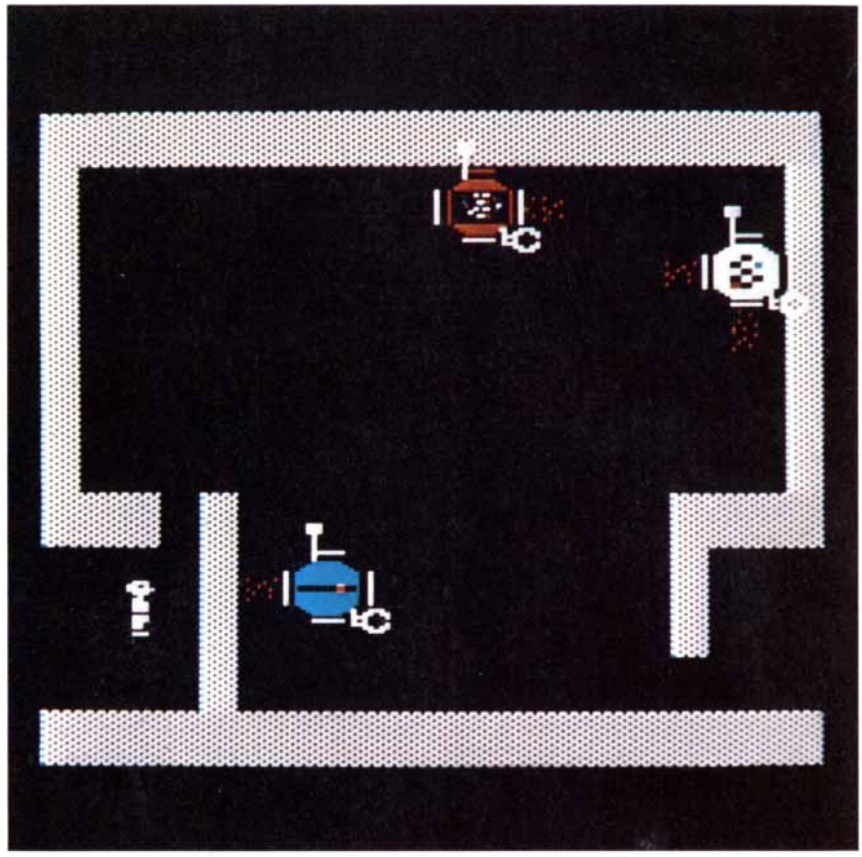
My own game of Robot Odyssey began as a seemingly endless journey through an array of rooms adjacent to the city sewer. On entering the sewer I came to a room zealously guarded by one of Robotropolis' meaner denizens. At the rear of the room against one wall was an energy crystal. Coveting it, I retrieved the wall-follower robot from inside Ping-Pong and set it in motion. As the wall-follower glided along the walls of the chamber, the guardian robot took no notice. When the wall-

follower encountered the energy crystal, it scooped up the crystal in the grabber claw, emitting a beep as it did so, and continued following the walls out to the entrance. There I turned the wall-follower off and tried to remove the energy crystal from the grabber. The robot would not let go. Setting that problem aside to be solved later, I put the wall-follower back inside Ping-Pong and was about to continue my odyssey when there came a grinding noise followed by the whining decrescendo that in the game signals a draining battery. A vampire robot called Ampire'Bot had just sunk its electric probe into Ping-Pong and sucked out half of its power. One must always be on the alert in Robotropolis.

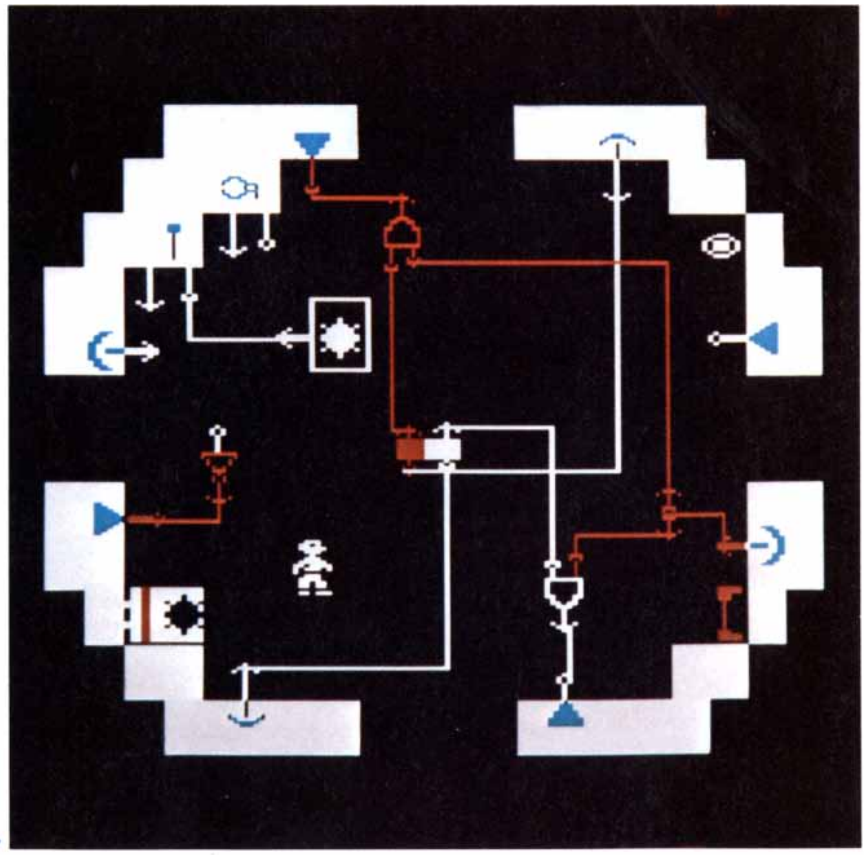
After wandering about picking up useful-looking objects here and there I came on a passageway bearing the message DO YOU HAVE EVERYTHING? I did not. I had forgotten to take a magnet from a room next to the sewer because I feared Ampire'Bot, which apparently patrolled the sewer. My homunculus made its way past the sign and came to a maze called the sewer grate. This consisted of vertical walls interfingering from above and below [*see bottom illustration on next page*]. It was guarded by another mean robot.

One of the companion robots appeared to be wired for the job. When the robot I call right-seeker is turned on, it always moves to the right. If a wall blocks rightward movement, the right-seeker glides up or down until it strikes a horizontal wall, where it immediately reverses direction. The electronic reasons for this behavior are apparent from a glance at its wiring [*see bottom illustration at right*]. First, the left thruster is always on because it receives input from an inverter that is not connected to anything. (Since the inverter's input is off, its output is on.) Control of the vertical thrusters is achieved by a flip-flop element wired in the same way as the circuit governing Ping-Pong. There is, however, one simple difference: outputs to the vertical thrusters pass through *and*-gates. The right-hand bumper provides the second input to each *and*-gate. This means that a command to fire either the up- or the down-thruster will have no effect unless the right-hand bumper is touching something.

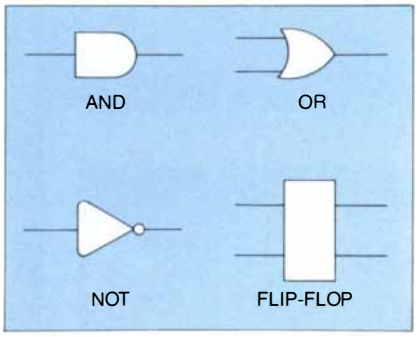
I extracted the right-seeker robot from inside the robot I had been carrying and sent my homunculus on board to ride it. By superposing the homunculus on an eye-shaped icon in the right-seeker robot I obtained a periscopic view of the environment: the robot interior was replaced on my computer screen by a view of the maze. There on the left was the right-seeker



The entrance to Robotropolis and the three robots found there



The bumpers, thrusters, grabber, antenna, connectors and circuitry of the right-seeker robot



Logic elements in Robot Odyssey

with a periscope protruding. It was not clear to me how the scene could include the viewing device, but I let the question pass.

While it was perched on the periscope eye there was nothing to stop my homunculus from becoming a control unit. I pressed *R* and then the space bar. The right-seeker, still bearing my homunculus, glided into the grate and struck a wall. Exploring upward, we encountered the top of the maze and bounced downward until we came to a gap. Here, below the maze, was a collection of unfortunate homunculuses that had not been able to pass through the grate. We glided by the guardian robot, which seemed not to know that a human presence was riding the right-seeker. We traversed the gap, encountered another vertical wall and proceeded upward. When we finally arrived at the other side of the maze, I realized I had blundered badly: I had left my other two companions behind. My fear of Ampire'Bot had kept me from thinking clearly. How could I return for the two robots? The right-seeker would have to be rewired. Readers who reflect on its circuitry for a moment will probably see how the rewiring was done.

In order to execute the mechanical task of rewiring a robot, one types *S,* and the homunculus is turned into a soldering iron. In this mode the space bar is used to break old connections and make new ones. If a part is lacking, type *T;* a tool kit then appears on the screen. The homunculus can walk inside and pick up the needed element.

When I had finally gathered all three robots on the other side of the sewer grate, my odyssey continued. Next came a locked chamber. I used the key to unlock it, entered and found the room had no other exits. Where could one go from here? A wall bore an ominous legend: PREPARE FOR AN EXPERIENCE THAT WILL TRANSPORT YOU TO A NEW DIMENSION. In the middle of the room a ghostly replica of my homunculus danced within a stationary box.

At any point in Robot Odyssey it is possible to save the game on a spare disk. If the next move turns into a disaster, the game can be replayed from the position at which it was saved. This is a feature of many computer adventure games. One hopes young aficionados do not grow up under the impression that the adventure we call life can also be saved for replay.

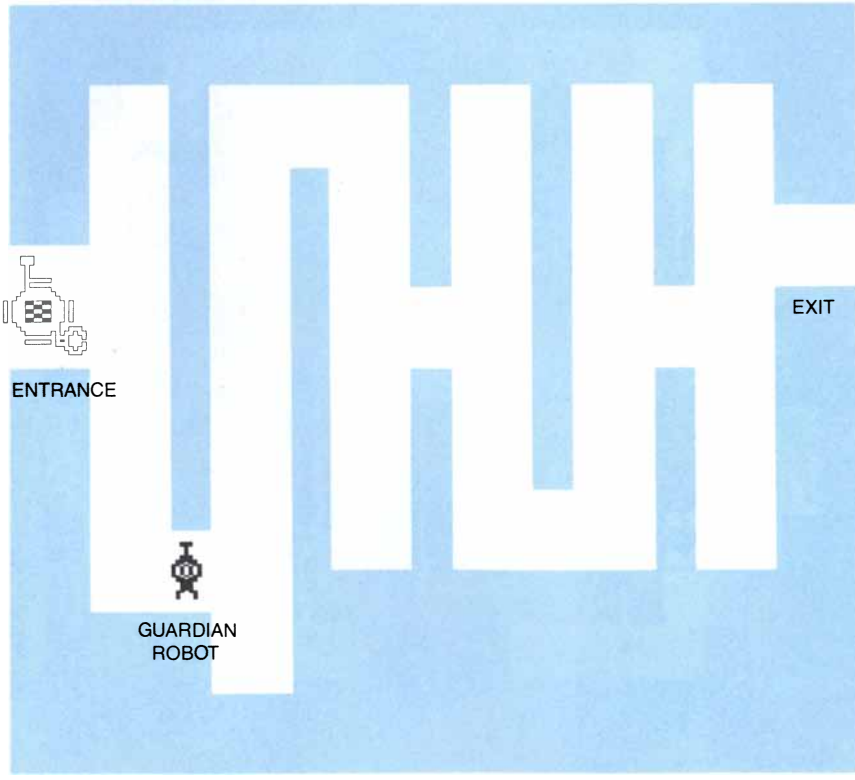
Having saved the game, I superposed my homunculus onto its duplicate and pressed the space bar. Everything seemed to explode in a shower of sparks, and a strange sound accompanied my move to a new dimension. Actually I had landed in a new room on the second level of Robotropolis—minus my companions. Apparently I was supposed to bring them along. Using the saved game, it was possible to replay the dimensional transport this time with a firm grasp on my robots.

To describe my further adventures in Robotropolis might well spoil the game for readers who would like to play it themselves. Robot Odyssey is sold by the Learning Company (545 Middlefield Road, Suite 170, Menlo Park, Calif. 94025). In addition to the game, buyers of Robot Odyssey receive disk tutorials on robot anatomy, use of the tool kit and chip design. This last facility is embodied in a place called the innovation lab. At any point in the game a player may visit the lab to test a new circuit design on robots kept there. As circuits grow in size and complexity it is useful to be able to transfer them to chips. This is done in the burner room of the lab.

One of the companion robots, the wall-follower, was already equipped with an advanced chip. The connectors of the chip are shown in schematic form on the opposite page. Readers might enjoy the challenge of designing a wall-following circuit to be burned into such a chip. Use only the flip-flops and gates described. I shall be happy to evaluate circuits sent to me; some readers, however, will want to see for themselves how well their circuits work.

The Robot Odyssey game was designed and written by Michael Wallace, who was then a student at Stanford University, and Leslie Grimm of the Learning Company. The game has predecessors in the Learning Company's collection. Gertrude's Puzzles and Rocky's Boots both require the player to design logic circuits to solve problems. Warren Robinett, formerly of the Learning Company, has provided all these games with challenging environments. The rooms and passageways of Robotropolis, for example, are finite and yet they go on forever, at least until the player realizes that leaving through the right side of one room might well involve entering a room already visited from the left.

Designing a circuit for Robot Odys-

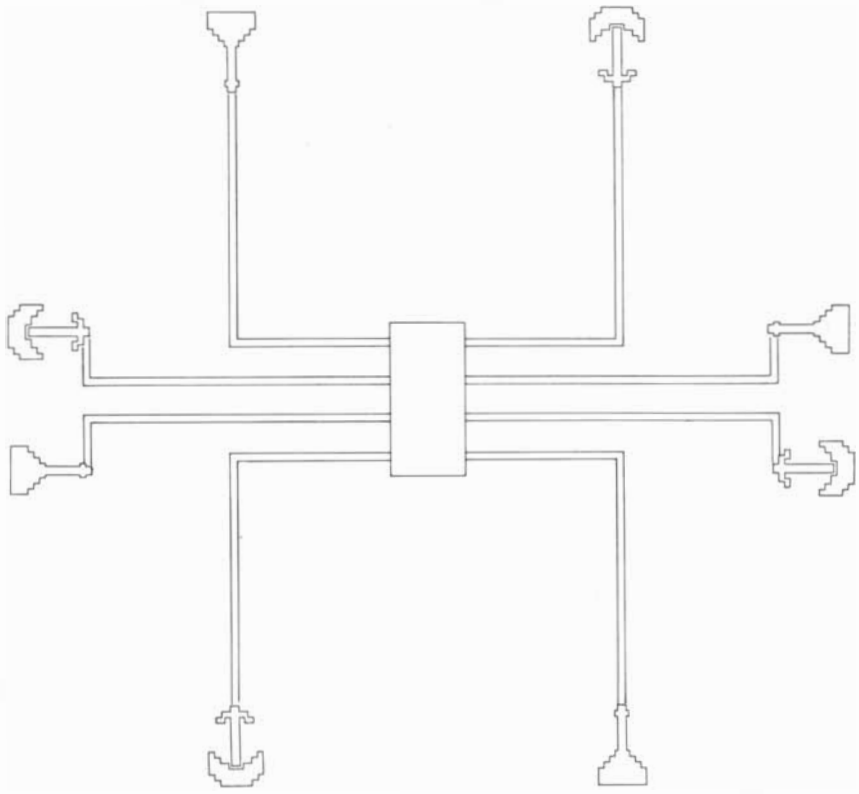


The sewer grate: a maze to be solved by the right-seeker robot

sey is a very different task from designing one for a modern computer. Still, it may be possible to build a primitive computer using components supplied in the game. Those who become expert may want to find out whether a computer can be designed that will fit into a single robot. The computer might be programmed by inputting a sequence of 0's and 1's just before the robot is activated. This sequence would amount to an ultrasimple language in which subsequences represent instructions such as "Move right until wall" or "Activate grabber."

Designing a logic circuit for a modern computer consists in laying out a vast array of silicon-based transistors, capacitors and other components on a large plan stored in a computer. Once it has been checked and debugged by simulation, the plan can be transferred to a photographic film and etched onto a silicon chip. At first glance this intricate, highly complicated process seems only distantly related to the gates and flip-flops of Robot Odyssey or to the basic logic-design courses taught at most universities. The connection was explained to me some years ago by Christopher R. Clare, who is now at CTX International in Sunnyvale, Calif. According to Clare, the point of the introduction to circuit design through standard logic is to encourage abstract thinking by the would-be engineer. Faced with a problem to be solved, the trick is to elevate it into the abstract realm of *and*-, *or*- and *not*-gates and flip-flop elements in order to solve it there. The habit of abstract thought is then transferable to the more complicated realm of field-effect transistors and other advanced components that correspond to other logic-gate types.

According to Peter Ashkin, a design manager at Apple Computers, Robot Odyssey serves that purpose. Ashkin says the game teaches two skills essential to a successful computer designer. The first skill is the ability to discover just what problem one is trying to solve; the second is the ability to implement the solution in logic, of whatever type. For example, when I last played Robot Odyssey, I was faced with retrieving a subway token from the middle of a guarded room. It was plain that the problem was to get one of the companion robots to retrieve the token. How could the goal be achieved? The wall-follower would be no help since it would never get near enough to grab the token. A token sensor equipped with directional outputs was available that would certainly get the robot to the token; unfortunately the device would also prevent the robot from leaving with its prize. At any rate the analysis yielded the identification of two discrete tasks: to design a token-seeking circuit and to find an easily programmed behavior to get the robot out of the room. Although some problems are very nebulous, Ashkin notes, bite-size pieces of them can be solved. In this way the overall problem becomes more definite. The two processes, defining the problem and implementing parts of it in logic, are both parts of the design cycle.



The bumpers, thrusters and connectors of the wall-follower robot. What circuitry lies therein?

By teaching a player how to design circuits Robot Odyssey also offers a significant psychological benefit. It is the sense of mastery that develops as a player gains confidence in wiring the robot companions. Confidence is an antidote to the sense of cold unapproachability that computers can inspire. To learn standard logic and to implement it in Odyssey's robots is to understand that the same relation exists between the basic logic elements of a computer and its ultimate function. One has learned that there is no barrier to exploring further; the sense of mystery within the box is replaced by a sense of mastery over it.

Mastery is one of the phases through which the computer neophyte may pass, according to Sherry Turkle, author of *The Second Self*. Her book documents the developing relations between people and computers. There is an ominous quality in some stories of those who have come to think of themselves as machinelike. It seems to me this self-opinion is possible at all levels of mastery. I know of scientists and workers in artificial intelligence who regard themselves in the same way, albeit on a more sophisticated level. It may be true in one sense that we are all machines, but the insight provides little guidance in the matter of behavior and may encourage the adoption of disastrous role models. This is why, metaphorically speaking, it might be useful to acquire the skill of escaping from Robotropolis.

Readers' responses to the "five easy pieces" described in my April column lead to a histogram much like the negative-exponential distribution. Hundreds of readers wrote the first piece, a program called PINT for approximating pi. Only a few readers wrote the fifth piece, a program that simulates the behavior of queues.

In PINT simulated cannonballs are fired into a simulated field surrounding a simulated pond. I apologize to the thousands of neighbors who were kept awake by the all-night salvos.

Once one has written a pi-generating program of this type the temptation to go well beyond 1,000 shots is apparently irresistible. The champion of this piece is Joshua Simons of the Harvard Medical School. By running PINT for a

OPEN WIDE

AND SAY ENCORE

If you want European styling, say the word. Encore. An amazing perspective on hatchback design. The panoramic rear view lets you see and be seen in a new dimension. Ergonomic instrumentation provides a bank of driver data. Renault's unique pedestal seats offer an extra degree of style and rear legroom. The fold-down rear seat is designed to extend cargo space convenience. Encore's 4-wheel independent suspension with twin coaxial rear torsion bars and power front disc/rear drum brakes results in impressive performance and handling. Your choice of engines, including the optional new 1.7 litre, for increased performance. Electronic fuel injection produces fuel ratings that excel: **40** MPG HWY, **34** MPG CITY.* The front wheel drive Renault Encore. *Built in America.* **$5895.****

The word is out: Amazing!

*Base model 1.4-litre engine. Use EPA est. for comparison. Your results may differ.
**List price. Tax, license, destination charges, optional and regional equipment extra. Rear spoiler $74, wheel covers $93. All options not available on all models.
Buy or Lease. From American Motors. Safety Belts Save Lives.

RENAULT

THE ONE TO WATCH

week on a mainframe computer Simons was able to take nearly four billion shots at the pond. His estimate for pi is 3.14157. Unfortunately the fifth decimal digit of pi is 9, not 7.

At this writing only 134 readers have submitted estimates of pi generated by 1,000-shot runs. When readers submitted multiple runs, I selected only the first for inclusion in the grand average. Here it is:

3.14137

The first three digits after the decimal point are correct. The fourth decimal digit of pi is 5, not 3.

I computed another average, weighted by the length of each run, for all runs longer than 1,000 shots. A few runs of several million shots may have distorted this estimate:

3.13948

In all of this I never once had to look up the value of pi. Instead I used a simple mnemonic: "How I wish I could enumerate pi easily." Count the letters of each word in that phrase. Can any readers suggest an easily remembered extension to the mnemonic?

Joseph W. McKean, a statistician at Western Michigan University in Kalamazoo, Mich., points out that one can establish a 95 percent confidence interval around one's estimate. The estimate is assumed to lie in the middle of a range of numbers that make up the confidence interval. The interval is defined in such a way that the true value of pi lies somewhere in the interval with a 95 percent probability. The size of the interval varies inversely with the square root of *n*, the number of shots fired. Hence to increase the accuracy of the estimate by a factor of 10, or one extra digit, *n* must be increased by a factor of 100. Such diminishing returns on the computational investment account for the frustration expressed by several readers.

Everyone who tried ZOMBIE had no difficulty generating histograms that mimic the shape of the negative-exponential curve, which was illustrated in the April column.

The GALTON program intrigued Edgar F. Coudal of Park Ridge, Ill. His simulation of marbles rolling down a triangular array of pegs led to the expected distribution at the bottom of the Galton board: the binomial distribution. The binomial distribution has a discretely belled shape, approximating its famous continuous cousin, namely the normal distribution.

In an amazing instance of simultaneous invention, two readers thought of generalizing GALTON to a three-dimensional array of pegs. At the bottom of their simulated boards J. Michael Matuszewicz of Columbus, Neb., and James Nugent of Peoria, Ill., found a softly rounded mound of marbles. Can the reader guess its shape?

Some readers who tried VOTERS had disappointing results. In one case even an all-night run failed to produce the expected collapse of democracy. To such experimenters I can only suggest extended runs. Mine have taken the better part of a day. The patterns are the things to watch.

Finally, only two readers attempted QWING, the simulation of the line of customers at a bank with a single teller. Victor H. Auerbach of Philadelphia wrote a time-slice version of the program. He watched the length of the queue drift steadily upward, although there were many short-term decreases. Joshua Goodman of Quincy, Mass., also modified QWING, but much more drastically: he replaced the queuing system with two zombie doors in a single wall. Readers might enjoy figuring out how Goodman's program works with only that hint to go on.

I remarked that the expected length of the queue is neither finite nor infinite if both the service times and the times between arrivals have the same long-term average. Such a queue grows at first, but then it shrinks to zero. Again it grows, even more this time, before it shrinks to zero again. The length of the queue continues oscillating ever more wildly. The explanation is that when there are no customers in the queue, potential service time is lost forever.

Many readers complained of suspected faults in their random-number generators. Ronald L. Guye and Gerald V. Post of Oakland University in Rochester, Mich., have confirmed the suspicion for the BASIC random-number generator (RND) on the IBM PC. This program, which will run only in the graphics mode, is diagnostic:

```
10 SCREEN 2
20 X = RND * 640
30 Y = RND * 200
40 PSET (X,Y)
50 GO TO 20
```

If RND were perfect, a random spattering of dots would appear on the screen, but instead one sees speckled bands. No doubt a similar effect can be prompted from other computers.

The display of the WORM program for the Apple II-C computer in my March column caught a slight typographical infection of its own. The second number in the DATA statement labeled 100 should be 255. As things stand, the worm creeps through memory at a painfully slow pace.

BOOKS

Stormy Mars, magnanimous space, flagellate scullers, troubled air, an old Dutch amateur

by Philip Morrison

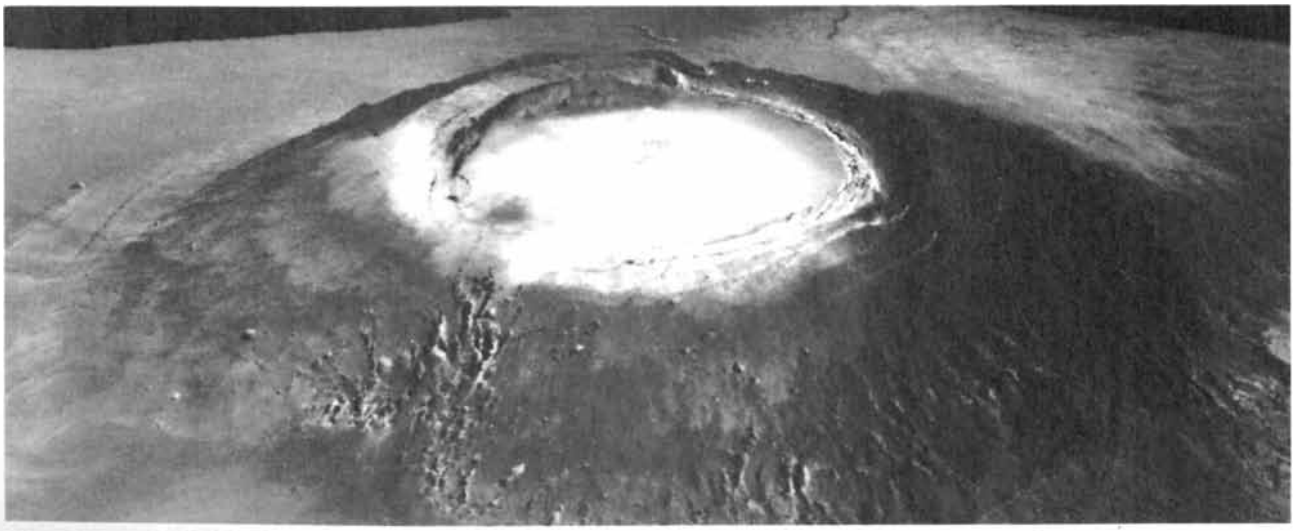
THE GEOLOGY OF THE TERRESTRIAL PLANETS, by Michael H. Carr, R. Stephen Saunders, Robert G. Strom and Don E. Wilhelms. NASA Special Publication SP-469, Superintendent of Documents, Washington, D.C. 20402 ($16; overseas, $20). OUT OF THE CRADLE: EXPLORING THE FRONTIERS BEYOND EARTH, by William K. Hartmann, Ron Miller and Pamela Lee. Workman Publishing Company, Inc. ($19.95; paperbound, $11.95). Geology is no longer earthbound; nowadays it is fully Copernican. The study of the events and processes that have shaped the five rocky planets (the earth's planet-size companion belongs to the list) can claim some mastery for all these bodies, each one the topic of a specialist's chapter in the first of these two books.

The authors draw on a lode of research, amassed in the course of the past two decades of geologic investigation. During that time well-financed virtuoso engineers extended the range of comparative geology to an entire class of kindred bodies. Mercury we know as a place from information gleaned through three flybys a decade back; Mars and Venus have each been explored in repeated visits by robot experts; the moon has yielded plenty of samples for laboratory study to skillful and curious human visitors as well. At the same time, here on the earth, new results and insights brought a revolutionary vigor to geology and yielded the unexpected unity of plate tectonics.

Surely the last word has not been written about Mars or the moon; we must expect surprises out there too. The narratives here are well told, although condensed and technical. They are strongly supported by images, maps and data in variety. Nowhere very fine, the net of understanding is also full of holes. Even though some fine big fish have been caught in it, many of the most splendid specimens elude us still. For instance, we cannot yet say whether the moon is sister, daughter or wife to the earth, nor are we sure why hot Venus seems geologically static like cold Mars rather than mobile like the temperate earth it matches so well in size.

From each principal chapter consider a point or two, intellectual hand samples. One graph shows the central iron cores marked to scale. More than 40 percent of the volume of little Mercury is core, in contrast to only one-sixth of the earth and about 4 percent of the moon. *Pioneer 10* found that a magnetic field duly surrounds the dense little planet. Its origin is probably in fluid flow within the metallic core. Yet such a supposition is a little problematical: it is not easy to sustain the molten state within so small a sphere even at depth. The cratered surface is moonlike, although with understandable differences. Long scarps, some forming cliffs a kilometer high and hundreds of kilometers in extent, are a global feature of Mercury, not to be seen on Mars or the moon. They are overthrusts, compressive faults, a sign of the general contraction of the planet during the cooling process that followed the formation of that outsize core.

Venus, its waterless surface red-hot under perpetual clouds penetrated only by earthbound radar and a number of heroic probes, is still a challenge. One graph is eloquent: it compares relative area at each elevation for the earth and Venus. Two nearly equal peaks describe the earth, the wide continental platforms five kilometers above the mature ocean floor. Venus shows a single peak of mean elevation; there are no deep basins, although a few large uplands skew the distribution toward the heights. We can make out some ridges and rifts, but the two-story world of plate tectonics appears to be absent. The theorists have argued that the high surface temperature has resulted in a thick and buoyant crust, which cannot be drawn under even if plates collide. Has a mere gaseous blanket ruled the deep



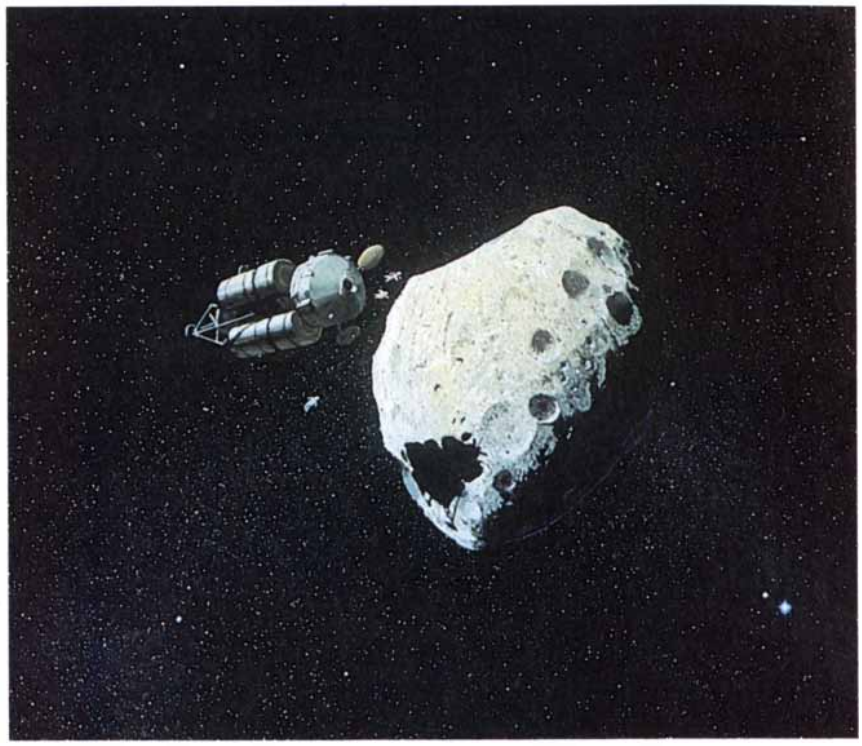
Eruptions from vents on its flanks have built a spur on the Martian volcano Arsia Mons. The image is photogrammetrically derived

rocks, Pluto held under the stilling sway of Aeolus?

The moon is the subject of the longest and most detailed chapter. Its pocked surface is the best-known relic of the ancient epoch when crashing impacts ruled the surface of every inner planet. Floods of molten basalt have here and there extruded out among the impact debris that covers the roughened uplands, and frozen into "seas." This silent world once knew explosive change indeed, but there geologic time has stopped. One table lists the measured ages of the principal lunar rock samples retrieved across the orbital gap. Fully half of them predate any rock ever found on the earth, and nearly all are more than three billion years old. The crater named Tycho is counted as new; its formation sent those rays still fresh and bright across the entire near face of the moon while big reptiles gazed incuriously from a cloudy earth. The colorful geologic maps of the lunar surface close their record during the dimmest past of earth history. "Man has yet to see the Moon's surface change except by his own hand," although seismometers emplaced there have heard a few small distant impacts and samples confirm that some small craters are geologically truly recent.

Mars is more like home. Wind, water and ice have all marked it; spectacular deserts, dusty dunes, great canyons, river valleys and grand volcanoes are well mapped. Liquid water is gone now; it was never copious, although it was transiently powerful here and there. The surface of Mars does not shift. The shield volcanoes of Hawaii build up over only millions of years, then drift away, carried by plate motion, from the hot plume that shapes them. On Mars a slower ooze of lava can accumulate over billions of years; its peaks overtop Everest threefold. Its super-Tibetan upwarp and its Grander Canyons, however formed, do not erode away; at most they yield slowly under the load of their own weight. It is a geologic museum: like the changeless moon, it recalls the time of impacts; it also records all those active local processes that do not depend on turning and sliding plates.

A brief chapter on earth geology is largely a clear account of plate tectonics, aided by helpful diagrams. The earth's record of the epoch of terrible impacts is all but lost to us. For since that antique time two great cyclical processes have been hard at work: plate motion engulfs much of the crust down to magmatic depths and back again, and swift water-driven weathering followed by slow uplifts perpetually shuffles the surface rocks. These two processes make the surface of the third planet more mobile by orders of magnitude than those of Mars, the moon or Mercury. The case of Venus is less clear; the craters there seem ancient, yet the slabby rocks shown in the Venera pictures do not appear to be as heavily weathered by the hot atmosphere as a number of geochemists have supposed. Are the surfaces somehow made anew?



Astronauts explore an asteroid; the earth and the moon form a binary system **(bottom right)**

These authors have reason for satisfaction, even if more novelties were to turn up than they signal in this broad outline. Their text is too compact to be easily readable, but it repays effort. The visual materials are a treasury. A useful set of maps of all these planets is presented, albeit at a scale too small for easy reading. (The sources of the maps are somehow not cited.) The substantial volume is a real GPO bargain.

The second book is as much a dream of the future as the first is a sober-sided summation of the decade past. Around a meticulous yet imaginative collection of paintings of visionary spacescapes and explorations the three artists, all of whom are astronomically and geologically well prepared (William Hartmann is a professional planetologist), have woven an engaged and engaging nontechnical text, at once a personal plea and an optimistic projection.

Hartmann tells of his part in the 1962 recognition of the ringed basins of the moon, those largest of craters, and of the radar views of the polar uplands of Venus returned by Soviet spacecraft only last year. The nontechnical text details hopes and plans for unmanned probes ahead, provides rationales for a lunar colony to come and even describes the classes of asteroids presumed by analogy with the reflectivity of known types of meteorites.

See paintings of the return to the moon: an astronaut explores among boulders on the floor of Tycho; across the page some cosmonauts visit the 1959 craterlet made by hard-landing Lunik, first artifact to reach another world. A lunar dome shelters a tropical garden. The earth overhead in the black sky, winged weekend athletes glide in the sealed air against the gentle moon pull. Look a little further into the future: a spacecraft has carried its crew to an earth-crossing asteroid; the vessel's distorted shadow on the irregular rock evokes the old Apollo landing module. Here is the parachuted resupply of an anxious team, just in time before they are lost under a fierce orange Martian dust storm. A plausible aircraft flies over the Martian landscape, and another readies its approach to the atmosphere of far Titan, smoggy with the dust of organic polymers. The ring plane of Saturn, crowded with colliding icy spheres, is convincingly rendered as a too bold visitor might see it from within.

One may hold all of these as dreams alone. Yet not long ago the compara-

WHAT SHOULD YOU
WORKING "PARTNERSHIP"

"Crown Zellerbach's ECZEL Corporation uses an HP Office Information Network and speeds response to customers, saving $1.5 million in just 10 months.

"A network of 23 HP 3000 computer systems links our 22 distribution centers across the country. HP 3000 Software for Distributors automates data processing for marketing, sales, warehouse functions and management planning."

ECZEL produces and distributes supplies for computer and word processing systems.

In San Francisco, California, Linda Bos, Vice President of Sales, says: "We're growing in excess of 10% a month. Ten months ago we made a major investment in HP. The return on that investment is already apparent. Our HP systems can expand with us as we grow. By networking the systems and using HP's word processing and electronic mail software, we've streamlined communications. HP's reliability has enabled us to improve productivity throughout our company saving $1.5 million this year alone.

"HP responds quickly and effectively to our needs. Based on this response, we are confident in HP as an office system supplier and plan to expand our distribution network with additional HP 3000 computer systems and software."

EXPECT FROM A WITH HEWLETT-PACKARD?

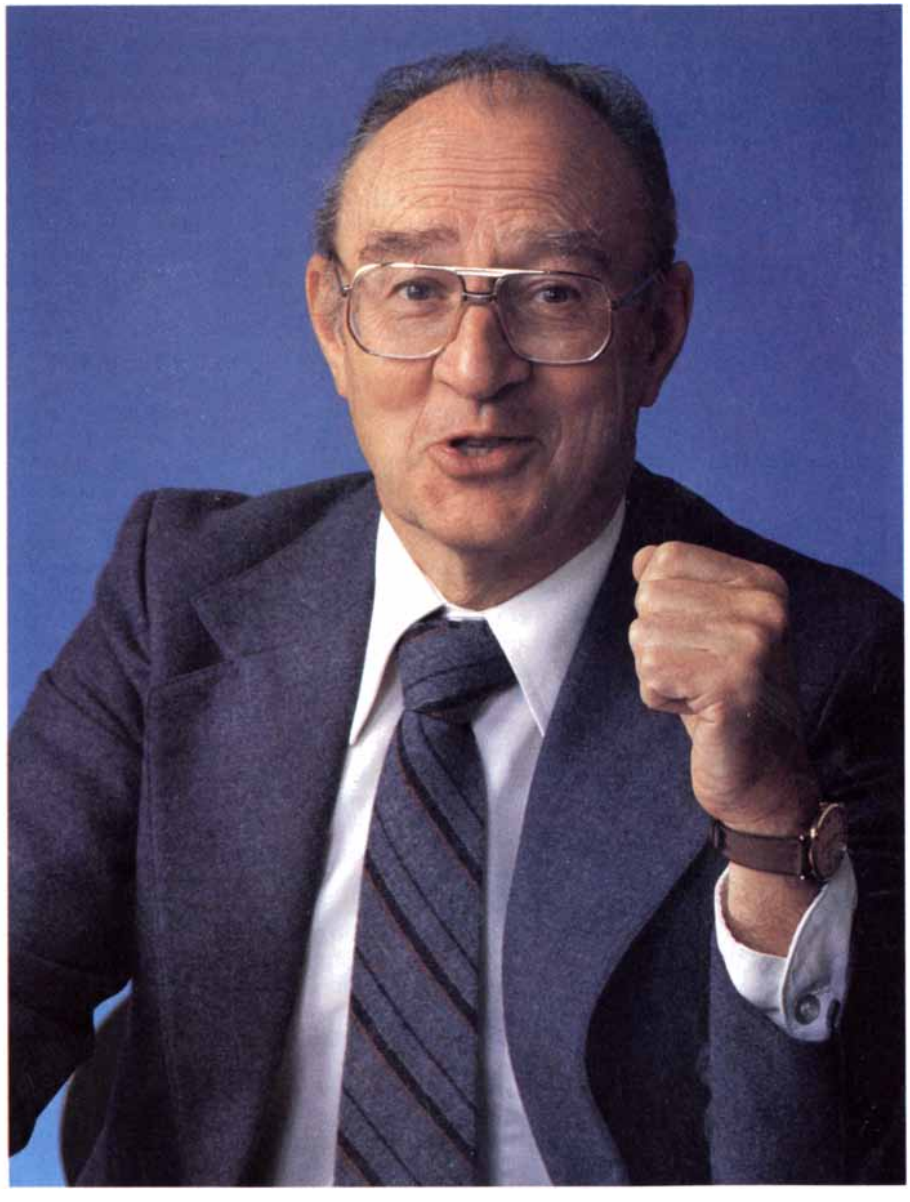
"At Westinghouse Canada, Inc., an HP Automatic Test System improves quality control and tests PC boards 30 times faster than before.

"The precision testing of the HP 3060A automatic board test system enables us to pick up defects early in the production cycle, thereby improving quality control and speeding production."

Westinghouse, Canada, Inc., in Burlington, Ontario, manufactures and distributes data terminals to over 100 major airlines. David Havart, Supervisor of Test Engineering, says: "Westinghouse monitored the performance of the HP test system and found it identified 98.5% of all faults. That represents an 18% improvement in accuracy over our prior test method.

"Because the HP 3060A is so simple to operate, we were able to realize the benefits of our HP system almost immediately—without investing significant time or dollars in training.

"HP's reputation for dependable test equipment was a major reason behind our decision to purchase the HP 3060A. Once installed, the speed and accuracy of the HP system enabled us to streamline quality control procedures, saving us more than $200,000 a year in operating costs."



Isn't Your Business Worth HP?

HEWLETT PACKARD

0002504

tive geology of the terrestrial planets was itself an improbable vision. It is clear that plans for national military control over the high ground of space, or claims of exclusive rights here and there over its resources, are dreams as well—or rather nightmares. They are accomplishments no more probable than these cheerful fancies, but they are acutely threatening. In counterpoint these artists have embedded their spacefaring dreams in an appeal for magnanimity out there. They project a world of nations engaged not in deadly star wars but in joint orbital endeavors, a genuine Strategic Initiative in space, one oriented much closer to the true direction of our national security than are the extravagant current promises of salvation by zapping.

A Functional Biology of Free-Living Protozoa, by Johanna Laybourn-Parry. University of California Press ($27.50; paperbound, $14.95). Bacteria and the smaller unicellular algae, all of them cells without nuclei, are of course the patricians of life, their pedigrees Precambrian by three billion years. Their biochemical wonders are worked by structures on a macromolecular scale, well below the resolving power of the light microscope. But ever since Antony van Leeuwenhoek first described his "animalcules," microscopists have been fascinated by the visibly complex and interacting creatures whose forms, changes and motions are not so hidden as those of the smaller ancestral biochemical specialists. Some of these very diverse forms, found in every pond, are catchwords even to those deprived students who have never seen them alive through the lens: amoeba, paramecium, volvox.

The protozoa are not a well-defined group among those larger cells that have true nuclei; they represent a level of organization rather than a set of close relatives. Most of those discussed in this brief, fresh review appear to have induced a certain empathy in the biologists who study them. Some swim about, flagellate scullers or ciliate galleys. Some instead ooze slowly along, pseudopods sent out and retracting in turn, on their hunt for bacterial prey. Some hang like attentive trumpets crowned with moving hairs to fan in what floats their way. The teeming marine forms include those famous for intricately radiating crystalline skeletons and tiny multichambered shells that are abundant enough to form entire hills and cliffs.

All are examples of evolved complexity reached by packing distinct organelles largely fashioned of thready proteins into one single cellular compartment. Each organism is more generally competent than any of our own individual cells, whose later evolutionary solution was specialization and cohesion, a mutual division of labor within the society that is every multicellular organism. These cells instead are generalists, no factory hands, signalers or teamsters but sturdy smallholders; each one can do it all, from secretion to reproduction. It is easier to see analogies to mouth or arm or tireless hunting among these visibly intricate little forms than to recognize the deeper enzymatic identity between ourselves and some tiny bacterial rod. That two distinct heliozoans may temporarily fuse to digest mutually some large morsel shows the gap between their protistan and our animal individuality.

Dr. Johanna Laybourn-Parry, of the University of Lancaster, is a real physiologist of protozoa, with ecological interests. She has measured the oxygen consumption of many of them, using an ingenious microrespirometer. Little cells must still be measured a dozen at a time, but big *Stentor,* a hefty full microgram dry weight, can be reported individually. The results do not fully conform to the famed scaling law clear from mouse to elephant, but then comparable resting metabolic rates are not so easy to obtain. Small predators rush about chasing some fleeing *Paramecium,* whereas *Stentor* sways languidly on its holdfast.

Such laboratory studies, together with the examination of ultrastructure, are the basis of most of what we know about these organisms. This book, the first of a series aiming to review how and why organisms work as a whole in their natural contexts, tries to look at the forms under less constraint, as adapting populations in the changing natural setting. That task is far from complete.

All protozoa are aquatic: they live in freshwater ponds, in the water films around soil particles, in the salt sea or on its bottom. Many dwell as symbionts or parasites in watery cavities or tissues of larger forms. Pond forms share a remarkable adaptation akin to the quiescent state of insect pupa or hibernating bear. In hard times, cold or dry, they shrink, simplify and encyst themselves within a number of complex new cell walls.

There is good evidence that the presence of some protozoa increases the growth rate of the bacteria they graze on, and hence the overall rate of substrate use, say in nitrogen fixation. There seem to be several reasons for the effect, ranging from microturbulence as the big fellows stir about to the actual secretion of a chemical probiotic that stimulates bacterial growth. The protozoa must not be neglected, as they often are, in any study of the ecological web. The majority of them are microbivores. On them in turn feed fiercer protozoa and the somewhat larger animal filter feeders.

Protozoa are efficient filterers, although they do not appear to maximize their feeding activities by adjusting their intake rate to the density and kind of particles present. For them the normal environment seems to be rich enough in food; they are better prepared for quite different hazards during their short life cycles. A definite sequence of changing protozoan species appear, rise and are replaced in a remarkable succession of changing forms within the artificial but vigorous freshwater microbial environment of working sewage-treatment plants. Clear effluents of high quality carry a variety of ciliated protozoa; plants that are working poorly discharge no ciliates at all. The effect of these predators on the bacterial throng is only sketchily understood.

Individual properties are not slighted, although the book does not intend to be a monograph on any well-known form. The key role of the microtubule protein fibers that weave most of the organelles is made plain, and the nature of their movement is well summarized. Protozoa clearly presage but do not deploy the detailed mechanisms of animal muscular contraction. It is now held that those tiny cilia, beating together in wavy patterns better timed than the oars of any trireme, are not synchronized by any protoneurological signals at all; hydrodynamical coupling through the tiny layer of water they mutually engage accomplishes the task.

The reproductive division of these undying cells has been a manifest puzzle for evolutionary design. Symmetrical forms manage to cleave lengthwise or transversely to taste. Others bud off, supplying the parts for an enclosing armor, say, to be assembled scale by scale around the daughter bud before it splits. The intricately framed radiolarians have no such possibility; they bear swarmers alive, tiny hopefuls each sent off with its small endowment of a coded nucleus, a seed of skeletal crystal, and a couple of flagellar oars to seek their way in the world. Unfortunately no one has yet been able to follow the development of such a crystal swarmer into the adult jewel. What a film that will make one day!

Indoor Air Quality and Human Health, by Isaac Turiel. Stanford University Press ($24.95). We dwell in caves still, albeit artificial ones; 80 or

TAKE THE 3-VOLUME HANDBOOK OF ARTIFICIAL INTELLIGENCE (A $141.95 VALUE) FOR ONLY $4.95

when you join the Library of Computer and Information Sciences.

You simply agree to buy 3 more books—at handsome discounts—within the next 12 months.

Just completed, the massive, 3-volume HANDBOOK OF ARTIFICIAL INTELLIGENCE promises to become the standard reference work in the growing AI field.

Conceived and produced by leading scientists and researchers at Stanford University, with contributions from universities and laboratories across the nation, the *Handbook* makes available to scientists, engineers, students, and hobbyists who are encountering AI for the first time the techniques and concepts in this rapidly expanding computer universe.

The 200 articles cover the emerging issues, technical problems, and design strategies which have been developed during the past 25 years of research. The *Handbook* has been written for people with no background in AI; jargon has been eliminated; and, the hierarchical organization of the book allows the reader to delve deeply into a particular subject or browse the articles which serve as overviews of the various subfields.

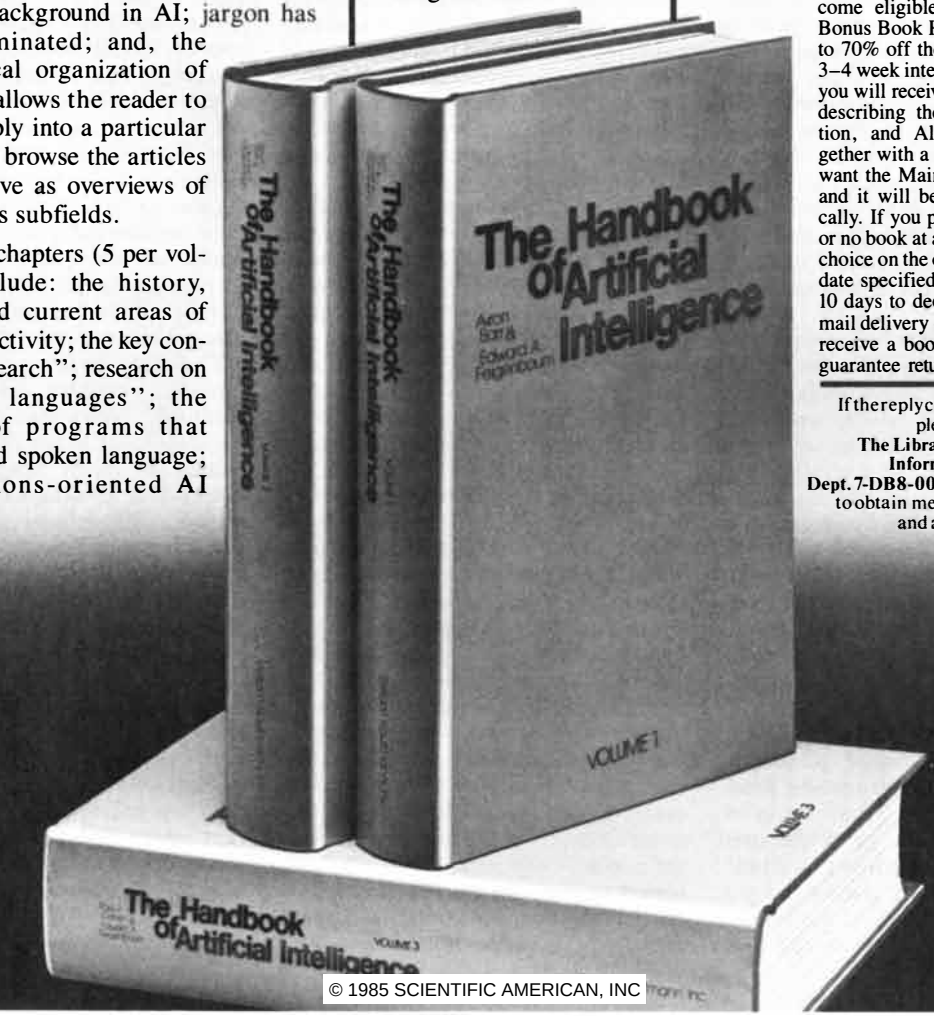
The 15 chapters (5 per volume) include: the history, goals, and current areas of research activity; the key concept of "search"; research on "natural languages"; the design of programs that understand spoken language; applications-oriented AI research in science, medicine, and education; automatic programming; models of cognition; automatic deduction; vision and learning research; and, planning and problem solving.

The comprehensive HANDBOOK OF ARTIFICIAL INTELLIGENCE answers questions like:

- What is a "heuristic problem-solving program?"
- How do computers understand English?
- Can computer programs outperform human experts?

AND INCLUDES:

- Over 1,450 pages.
- More than 200 articles in 15 chapters.
- With numerous charts, tables, and schematics.
- Edited by Avron Barr, Paul Cohen and Edward Feigenbaum.



The Library of Computer and Information Sciences is the oldest and largest book club especially designed for the computer professional. In the incredibly fast-moving world of data processing, where up-to-date knowledge is essential, we make it easy for you to keep totally informed on all areas of the information sciences.

Begin enjoying the club's benefits today!

MEMBERSHIP BENEFITS: In addition to getting the 3-volume Handbook of Artificial Intelligence for only $4.95, when you join, you keep saving substantially on the books you buy. Also, you will immediately become eligible to participate in our Bonus Book Plan, with savings of up to 70% off the publishers' prices. At 3–4 week intervals (16 times per year) you will receive the Book Club News, describing the coming Main Selection, and Alternate Selections, together with a dated reply card. If you want the Main Selection, do nothing and it will be sent to you automatically. If you prefer another selection, or no book at all, simply indicate your choice on the card, and return it by the date specified. You will have at least 10 days to decide. If, because of late mail delivery of the News, you should receive a book you do not want, we guarantee return postage.

If the reply card has been removed, please write to:
The Library of Computer and Information Sciences
Dept. 7-DB8-00614, Riverside, N.J. 08075
to obtain membership information and an application.

90 percent of all the air that humans in the developed countries breathe is indoor air. That air of course derives from the ambient atmosphere, clean or polluted, but in practice its confinement modifies its composition. Just about everyone knows from time to time the need to seek air fresh from the outdoors.

Indoors there are significant sources and sinks for many airborne substances; the composition of indoor air is fixed by those rates and by the average holding time within. Isaac Turiel, a staff scientist at the Lawrence Berkeley Laboratory, knows his subject thoroughly. Room air is exchanged with the open atmosphere through open doors and windows and a variety of leaks, as well as by ventilating fans and ducts. A practical measure of exchange is the time required for full air replacement. Average air turnover times for residential buildings range between half an hour and five hours; most conventionally built houses show values of 50 to 100 minutes, apart from those drafty times of unusual winds or extreme temperature contrast. Those rates are basic to the topic. Larger buildings require forced ventilation by design.

Our homes hold fires as regularly as did the caves. For we too eat our food cooked, and we still require heat in cold seasons. Such fires now generally burn gas or oil in furnaces and kitchen stoves, one or less per household, and most combustion products are vented directly outdoors by way of chimney or hood. Curiously, even more fires, although tiny ones, burn unvented just at nose end: cigarettes. They are consumed at a mean rate of about one new little fire per hour per U.S. household, night and day.

Combustion is the ancient source of particulate and gaseous additions to the indoor air, for where there is fire, there are gases and usually smoke. In the myriad small kitchens of India wood and dung are primary fuels for the hardworking housewife or cook; in this setting one study documented mean particulate concentrations more than two dozen times the U.S. open-air standard for 24-hour exposure; carcinogen exposures amounted to smoking many packs of cigarettes a day. Even the poorest natural-gas stove is a great deal less noxious than that.

Nevertheless, our modern houses hold a wide selection of substances whose vapors and degradation products can contaminate the air. They include everything from pesticides used to preserve wood structure to a daunting variety of organic sprays and solvents. Direct analysis in a very small sample of houses has shown low concentrations of scores of organic air contaminants. These seem more abundant in summer, in spite of the seasonally more open houses, than in the sealed-up winter; evidently the higher vapor pressures and reaction rates of warm weather more than make up for the better summer ventilation. It is no pleasure to see the name of a carcinogen such as benzene leading the list; still, the levels of contamination are so low that they argue only for the prudent support of more research.

About a million Americans live in houses insulated by wall spaces filled with an air-blown foam of urea-formaldehyde resin. This is quite an effective insulator, first developed in Germany in the 1930's, and for decades installed in northern Europe, Canada and the U.S. The same resin is the most common adhesive for bonding plywood and groundwood boards. All these materials slowly release formaldehyde, a pungent vapor that can induce both acute and chronic respiratory illness in an important fraction of all people, at levels as low as less than a part per million in air. Such amounts are commonly measured in UF-insulated houses. There have been thousands of complaints and many well-documented cases of serious illness. Installation of the foam has effectively stopped, but the technique is again legal in the U.S. since a court decision of 1983 reversed a Federal ban. The government of Canada, where the foam remains banned, will grant $5,000 to the owner of any home insulated with UF to help pay for replacement.

Two paths may be followed toward judging such hazards when they are not evident. One way—call it modeling—is to go from measured cause to inferred effect: measure contaminant concentrations widely and infer the overall risks to health from a few specific cases, usually involving accidental or occupational exposures at a much higher level. The second way is epidemiological: work back from observed effect to probable cause. Find enough examples of injury and count effects within that sample compared with control groups formed with reference to suspected causes.

The effect of cigarettes in raising the incidence of fatal lung cancer among smokers is by now firmly established. It amounts to a factor of 10 or 12 increase among pack-a-day smokers. But what of the side-stream smoke that rises directly from the distal end of the cigarette, perhaps to be inhaled involuntarily by the smoker's neighbors in the room? Both paths have been followed to seek an answer. Respirable particulate concentrations have been measured in many houses and in a wide variety of public places. On this basis it appears that the presence of a couple of smokers in a modest room can raise particle counts close to or beyond the legal standards for outside air. Measured lung function seems poorer among Boston children and adult Frenchwomen when they live in a household among smokers than it is otherwise. True, that is hardly the same as cancer. Several studies of nonsmoking women married to heavy smokers show doubling and more of their lung-cancer rate; however, some samples are small and the controls have come into question. In Japan, for instance, nonsmokers may be more affluent and have more kitchen space, and therefore be exposed less to cooking fire. The cancer risk to involuntary smokers remains uncertain.

Clearly, energy-efficient house heating means no drafts. But cutting air infiltration, which in typical situations demands about a third of the energy of house heating, implies longer air-turnover times and some health risk. That is the question behind a couple of chapters of this book, which is a model of balanced simplification of a complicated and embattled topic. Swedish standards require that ventilation in residential buildings be adequate to allow air turnover in at most two hours. Very tightly built houses approach a 10-hour turnover; a heat exchanger can be designed to recover much of the heat otherwise lost with the exhausted air. Cutting the sources of pollutants and introducing explicit sinks for particles, perhaps by collecting them within electrostatic precipitators, are also quite practical measures. Small fan filters do not suffice. Tests of many commercial devices are cited here, useful for those who, like fireplace users, maintain strong sources.

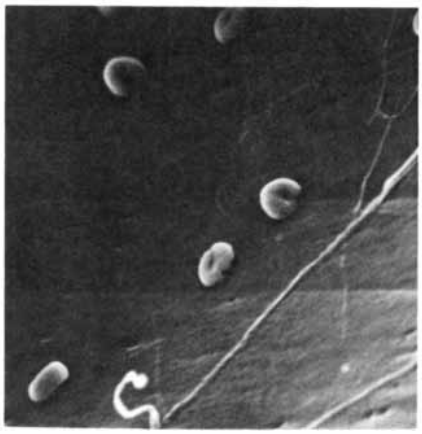
The cave dwellers lived mainly in limestone country, where caves are plentiful. Nowadays we build our homes on all kinds of geologic substrates. In many places the natural radioactive substances in rock and its product soil release radon, an inert gas that can diffuse through thick layers of subsoil into basements by way of openings and cracks, and thence through walls and floors into living spaces. It decays to deposit active daughter products on dust particles or even directly within the lungs. Radon is a proved cause of lung cancer in heavily exposed uranium miners. This natural contaminant is easily measurable. A graph shows its concentration rising and falling neatly as the test-house ventilation was altered; the source was constant but the turnover time changed. Worst-case estimates suggest that radon may cause a large

fraction of the lung-cancer fatalities among nonsmokers in this country; more cautious estimates would still put the number up in the thousands per year. The principal source is the radon fed in from the subsoil; the quantity is a function both of overall geochemistry and local pathways. Too-tight houses in radon country economize on heat but risk health, unless wise engineering provisions are made.

This factual, nontechnical and even-handed book is a fine place to encounter at a simple level the methods of and the quandaries around public health in a complex and changing society. The Lawrence Berkeley expert has sequestered equations into one appendix page. Practical references include manufacturers of a variety of useful devices and government offices for information, complaint and help.

SINGLE LENS: THE STORY OF THE SIMPLE MICROSCOPE, by Brian J. Ford. Harper & Row, Publishers, Inc. ($14.95). The young microscopist had been studying the work of Antony van Leeuwenhoek, trying to decide how the simple microscopes that pioneer had built could have worked as well as they apparently did. Although in 1981 professional historians were halfway through the editing and publication of the large number of Leeuwenhoek's letters long kept in the vaults of the Royal Society of London, Brian Ford wanted to look at the originals for himself. "As I lifted the final leaf of Leeuwenhoek's letter of 1 June 1674, it felt heavy. Pasted to the back of the last blank page was a white paper envelope..."

On the book's next page we see a photograph of the contents of that long-unseen envelope, taken after the envelope was opened, the scholar's breath firmly held. They were four small packets of folded paper annotated in the hand of the old Dutch amateur. Each had held a specimen prepared with care by Leeuwenhoek himself for the microscopists of London, and three of the specimens were still in place. A little farther on we see the first specimen found, a section of cork, sliced skillfully with a sharp open razor, the image made at a magnification of 266 diameters through a surviving Leeuwenhoek microscope still held in Utrecht. (A few of Leeuwenhoek's own red blood cells are shown in an electron micrograph. They were found on the surface of another specimen he had cut; no one can avoid leaving some red cells on the blade after a shave. Their unlysed condition implies that their source had shaved with a dry blade, although just possibly it was a cough that sent them there.) Others had seen some of the packets before (nine turned up in the files) but no one had examined them with any care. One set was missed because the translator, working from microfilm copies, mistook the images of the labeled and pasted packets of dissected cottonseed specimens for "drawn rectangles" at the end of the old letter.



Van Leeuwenhoek's red blood cells, ca. 1674

This informal, iconoclastic, personal book is both persuasively argued and well supported by modern expertise engaged in firsthand experiences with evidence from the past. The author seeks out and refurbishes the old instruments to see how they work, a powerful method open to historians of science and technology. The author and a Dutch lens expert, J. van Zuylen, measured the magnifying power of all nine of the simple microscopes currently ascribed to Leeuwenhoek. (He may have made as many as 500 of them during his long life.) Five of them magnify clearly by more than 100 times; the most powerful, the lens now in Utrecht, is in fact somewhat nonspherical, which improves its field of view. It was made without any signs of polishing; van Zuylen and J. Nieuwland have since made a similar lens, like a tiny bead, in their own glass-blowing flame. Leeuwenhoek indeed claimed that he had found "a useful way of blowing lenses which were not round."

Some unknown attic in London may hold today 26 of Leeuwenhoek's earlier microscopes, "ground by myself, and mounted in silver... almost all of them in silver that I extracted from the ore." Silver is costly, but it is softer and easier to work than brass. His daughter sent those instruments to the Royal Society as her father's legacy, arranged and labeled in a "small Indian cabinet" carefully prepared long before by their fabricator. The microscopist Henry Baker, drawing and describing them 20 years after the legacy had come,

The last time she saw England.

The newlyweds were off to live in America. She had put off saying good-bye until the very last moment.

As always, he had something for her: an antique cameo brooch. "It was your mother's," he said. "I know she'd want you to have it." Then she tried to put her feelings into words.

But how do you begin to tell the person who single-handedly raised you how much you love him?

Call England. Ten minutes can average just 76¢ a minute.*

Saying good-bye is never easy—but saying hello is, with AT&T. A ten-minute phone call to England can average as little as 76¢ a minute.

Just dial the call yourself any night from 6 pm until 7 am.

If you don't have International Dialing in your area, you'll still get the same low rate as long as special operator assistance is not required.

AT&T International Long Distance Service.

England		
Rate Level	**Average Cost Per Minute For a 10-Minute Call****	**Hours**
Economy	**.76**	6pm–7am
Discount	.95	1pm–6pm
Standard	$1.26	7am–1pm

Average cost per minute varies depending on length of call. First minute costs more; additional minutes cost less. All rates are for calls dialed direct from continental U.S. during hours listed. Add 3% Federal Excise Tax. For further information, call our International Information Service, toll free **1 800 874-4000.

*During Economy time periods.

AT&T

The right choice.

remarked in a footnote: "At the Time I am writing this, the Cabinet of Microscopes left by that famous Man... is standing upon my Table." No more. That cabinet and its contents are lost. Their last known borrower, in about 1820, was an eminent London physician, a man who published many microscopic illustrations made by able technicians in his employ. An unsuccessful search by the Royal Society in 1855 left a tangled tale. The silver devices remain a tantalizing lost treasure, although less of a marvel than the unmatched specimens Ford came on a few years ago.

The botanist and microscopist Robert Brown, for whom we name the motion of all tiny particles recoiling from the random rain of molecular bombardment, carefully studied that phenomenon in 1828 under his simple microscope. He called the moving bits "active Molecules." Since he noticed them first in flower pollen less than a year old, it seemed natural to him that they might be some microscopic form of life. But after a very careful study of dead and fossil organic specimens, including "a fragment of the Sphinx" within a long list of mineral and rock powders, he found that any substance he could powder finely enough to suspend for a while in water shared the incessant motion. His paper was firm in its conclusion that the motion was neither animate nor the result of currents or evaporation. He never claimed original discovery; others he cited had reported the motion before him.

The author has found Brown's microscope too, dirty, blackened and damaged, and has made it workable. Quite likely it was the very one peered through by young Charles Darwin when he visited the secretive old botanist in London while Darwin waited for the readying of H.M.S. *Beagle*. Darwin later wrote of what he saw, "I... believe now that it was the marvellous currents of protoplasm in some vegetable cell," but all the old man would answer to Darwin's questions was, "That is my little secret." Years later Brown published an account of streaming protoplasm in the giant stamen hairs of the spiderwort. Brown ought to be still better known for his recognition and naming of "a single circular areola,... or nucleus of the cell as perhaps it might be termed" in each cell. He reported them in many cells of many monocot plants and sometimes in the epidermis of dicots.

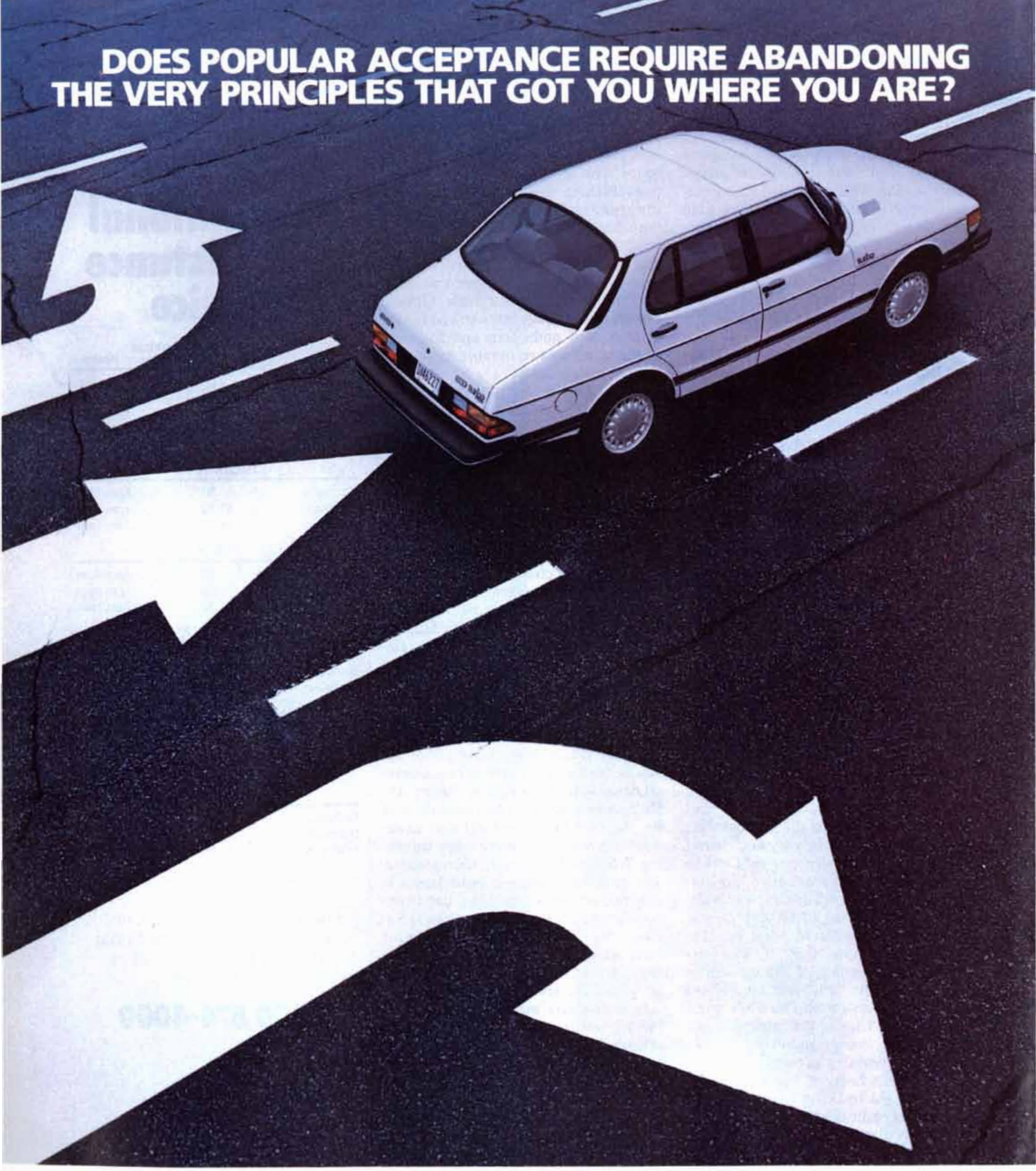
Again the discovery is convincingly re-created: through Brown's simple microscope we see cell nuclei quite clearly. In his day the best simple microscopes of the London makers were indeed excellent; they had workable fine adjustments, convenient stands and optics of good quality that could achieve magnification of about 200 powers. This high state of the art brings us to the book's thesis: that the mid-19th-century compound microscope with its optical corrections and conveniences was no precondition for the grand rise of microbiology and the "mushrooming awareness" of microstructure. In fact, it was the simple lens that had made the great earlier breakthroughs, and an instrument as good as van Leeuwenhoek's little Utrecht beauty might easily have served for the discovery of pathogens and for most of the structures resolvable by today's optical microscopes.

Ford's case is strong that the single lens has been underestimated. Its harvest has been rich, and its best performance is demonstrable to a resolution of about one micrometer at an enlargement of close to 300 times, or a bit better. Over the past century the use of well-corrected high-power oil-immersion lenses has brought the resolution down to less than a third of a micrometer. The change is not large, but it may well be important; a clear gain in resolving power by a factor of two or three is not to be brushed aside. Very possibly the arcane art of staining and the refinement of sectioning and fixing techniques were even more important.

The book has much more to say. Ford pretty well clears up the story of the alleged lens from Assyria: not optically usable, it is a striated rock-crystal ornament of the eighth century B.C. He makes a good if rather subtle case for the fact that merchant van Leeuwenhoek began his microscopy on seeing Hooke's best-seller *Micrographia*. The making of simple bead lenses is outlined by Hooke, and van Leeuwenhoek's specimen gifts to the Royal Society recapitulate Hooke's choice of early samples. Not all citations in this book are as complete as they might be, and there are no footnotes.

It is artifacts more than texts that Ford is pleased to read. It would be splendid to see him compare his fine single lenses with the compound microscopes of high Victorian days. Bacteria may be shown well enough by the single-lens view; perhaps it was more the pure-culture techniques of Koch and Pasteur than fine new microscopes that led to classical bacteriology. The dance of the chromosomes is certainly one major performance to which the lenses of the past, single and compound, should be brought in Ford's adept and critical spirit. Clio needs to be made fully at home at the laboratory bench, and this microscopist and writer has already welcomed her there.

DOES POPULAR ACCEPTANCE REQUIRE ABANDONING THE VERY PRINCIPLES THAT GOT YOU WHERE YOU ARE?



It happens in every field.

A novelist compromises to broaden his appeal and get on the best seller list. A musician, instead of following *her* muse, tries to divine what the crowd likes. And industry, as you would suspect, is hardly different.

So: when will Saab curb its performance characteristics to appeal to the broadest possible market segment?

And: when will the researchers tell the engineers to make the suspension a little more squishy because the 25-44-year-olds in the focus groups say they prefer it?

And: when will Saab start adding those oh-so-swell designer touches?

The answer to all those questions? NEVER!

Saab may sell out. But not in the sense implied in all those queries. Only in the

Saabs range in price from $11,850 for the 900 3-door 5-speed to $18,620 for the 4-door 5-speed 16-valve Turbo. Manufacturer's suggested retail price. Not including taxes, license, freight, dealer charges or options.

SAAB 900

The most intelligent car ever built.

Established 1845 SCIENTIFIC AMERICAN July 1985 Volume 253 Number 1

Infant Mortality in the U.S.

After many years of a steady drop in the infant mortality rate in the U.S. the rate of decline has abruptly slowed. The change coincides with cutbacks in programs for mothers and children

by C. Arden Miller

In 1979 the U.S. Public Health Service established nine infant deaths per 1,000 live births as a goal the nation should be able to meet by 1990. Early last year a Government official testified in Congress that the goal could be met if previous favorable trends continued. In December of that year, reviewing provisional data for 1983 and the first nine months of 1984, the same official acknowledged that the trends were less favorable and that the goal would not be reached by 1990.

The infant mortality rate had been declining steadily for many years, dropping from 47 in 1940 to 13.1 in 1979. The decline continued at an average of 4.6 percent per year until 1983, reaching a rate of 10.9 infant deaths per 1,000 live births. Then, according to the provisional data, the speed of decline diminished markedly, to 2.7 percent (10.6 deaths) in 1984. Although the Public Health Service asserted in May that the goal of an infant mortality rate of nine will still be reached by 1990, the likelihood that it will not be is strong.

What happened? As is usual in complex social issues, a number of possible contributing factors can be cited, and it would be difficult to prove that any one of them by itself tipped the scales. Still, one definite change that took place not long before the rate of decline flattened out was the reduction by the Reagan Administration in the funding of several programs for children, mothers of young children and pregnant women. I and many other observers think these cutbacks have contributed significantly to the change of trend in the infant mortality rate by weakening national policies (which had never been more than marginal) for the care and protection of pregnant women. Senior officials of the Department of Health and Human Services deny the connection. They point instead to such factors as the high rate of teen-age pregnancy, the use of tobacco, alcohol and drugs by many pregnant women and the complex racial mixture of the U.S. population; they also cite the possibility that high-technology medicine is merely postponing the death of some infants who earlier would have appeared in the statistics relating to naturally aborted pregnancies. The Administration has declined a proposal to study the effect of the cutbacks.

The statistic at issue here—the infant mortality rate—is officially defined as deaths (per 1,000 live births) in the first year of life. (More people in the U.S. die then than in any other single year of age up to 65.) Neonatal deaths, involving babies less than 28 days old, account for 70 percent of infant deaths, and two-thirds of the neonatal toll is attributable to low birth weight.

The risk of low birth weight is increased both among black mothers and among women who give birth when they are younger than 16 or older than 35. It is also higher for women who have poor prenatal care or none, whose diet is inadequate and who gain less than 20 pounds during pregnancy. Smoking, abuse of drugs and excessive consumption of alcohol are factors; so are stress, frequent childbearing and previous miscarriages. The relative weight that should be given these factors is not well documented. By far the most impressive correlation is with poverty and minority status. Recent emphasis has been laid too on the adverse influence of cigarette smoking before and during pregnancy.

The postneonatal infant mortality rate (deaths from 28 days through 12 months) is less substantially correlated with low birth weight but is heavily influenced by environmental circumstances that contribute to accidents and contagious diseases. The postneonatal rate is high among populations that have low socioeconomic status, poor sanitation, unsafe housing and limited water supply. It is also high in households where the mother has had little education. The sudden-infant-death syndrome (the abrupt death of an apparently healthy baby) is the largest single cause of postneonatal infant mortality. Recent reports identify a particularly high rate of sudden infant death in households headed by unmarried mothers.

The connection between infant mortality and poverty and related social factors is not a new finding. The Federal Children's Bureau, established in 1912, undertook as its first inquiry the subject of infant mortality. The findings, based on surveys in 10 cities and a number of rural areas, have a familiar ring: "The coincidence of a high infant mortality rate with low earnings, poor housing, the employment of the mother outside the home and large families was indicated in all these studies. They all showed that there is great variation in infant mortality rates, not only in different parts of the United States but also in differ-

ent parts of the same state and in the same city, town or rural district. These differences were found to be caused by different population elements, widely varying social and economic conditions and differences in appreciation of good prenatal and infant care and the facilities available for such care."

Another parallel with 1912 is concern with the comparison between the U.S. and other countries in infant mortality. Data from the census of 1910 indicate that the infant mortality rate in the U.S. was 124 per 1,000 live births. Many other nations had better records, but it is noteworthy that the rate was exceedingly high in several countries (including France, the Netherlands, Switzerland and the U.K.) that in later decades came to be among the world's leaders in infant survival.

Since 1910 the record has improved, albeit unequally, in every country. The U.S. has tended to slip down in the ranking of industrialized nations that have acceptable reporting systems and a population exceeding two million. In the 1950–54 period the U.S. was in seventh place; in 1960–64, 11th; in 1970–74, 16th, and in 1982, 17th. Finland and Japan, with rates of 6.0 and 6.6 respectively in 1982, are now the leaders, having passed Sweden.

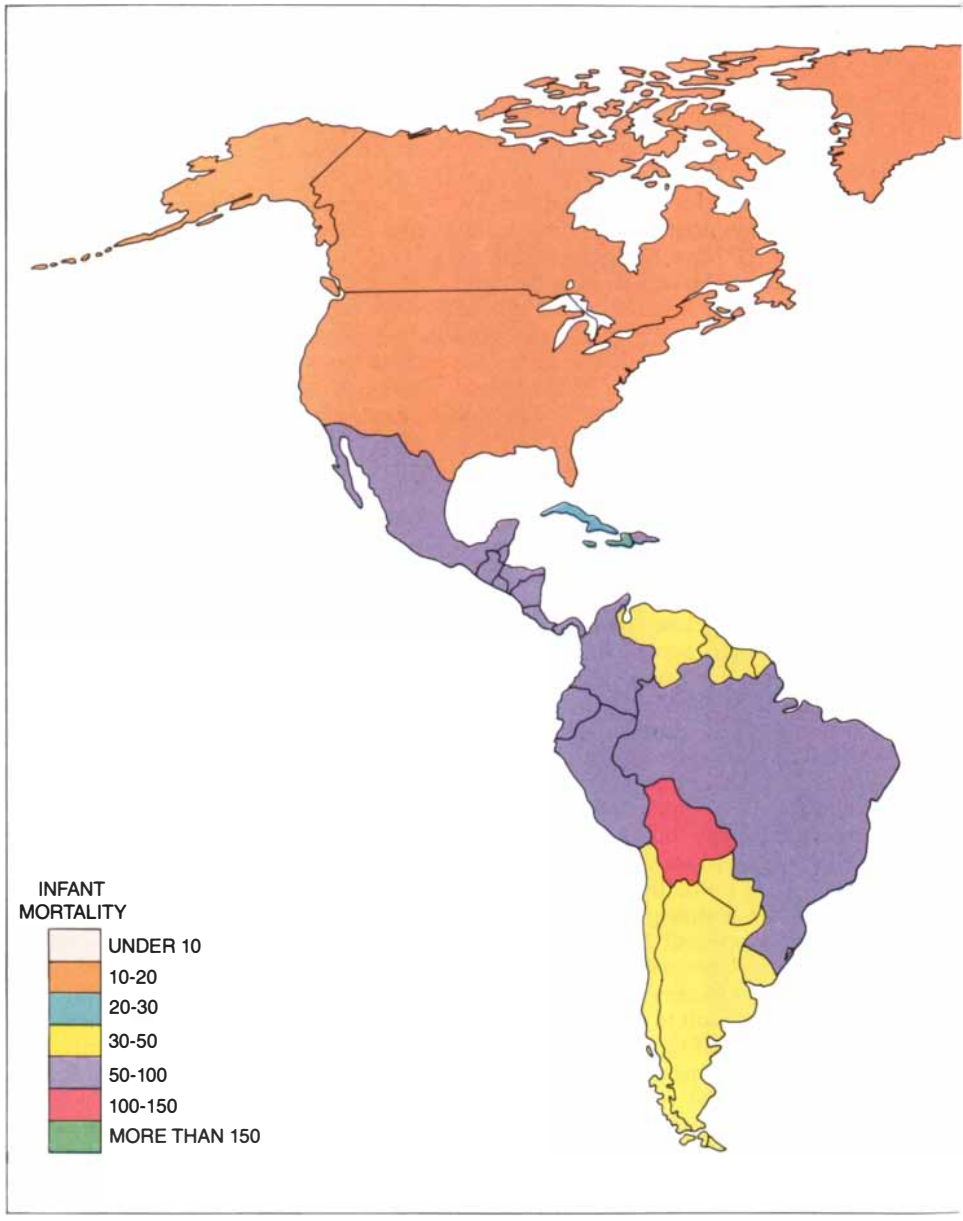
Every nation's social policies derive from its unique cultural traditions; still, the infant mortality figures from other countries help to define more sharply the influence on U.S. statistics of such variables as the age of childbearing, socioeconomic status and participation in health services. One influence that is often cited to explain the relatively high rate in the U.S. is the racial and cultural diversity of the population, but other factors are at work. For example, one can point out that the mortality rate for black infants is twice as high as the rate for whites in the U.S., but one must also account for the fact that the rate for white infants in the U.S. is higher than the rate for European whites of comparable socioeconomic status and national origin.

The major difference between U.S. babies and their age peers in other countries is that U.S. babies are smaller. If one compares babies of the same birth weight (say 2,000-gram babies in the U.S. with 2,000-gram babies in Europe), the differences in mortality disappear. Indeed, the chances for survival among babies of comparable birth weight are even better in the U.S. than they are in countries that have better overall infant mortality rates. The problem is that a disquieting number of newborns in the U.S. (6.8 percent of all babies and 12.4 percent of black babies) weigh less than 2,500 grams (5.5 pounds), the official measure of low birth weight.

The conditions contributing to low birth weight are the subject of study. One of them is either lack of or poor participation in a program of prenatal care that begins early in pregnancy and continues throughout it. Ideally such a program should do a good deal more than monitor the health of the mother and fetus and provide appropriate treatment. It should also include genetic screening and counseling; risk assessment; counseling on the hazards of cigarette smoking, drug use and alcohol consumption; education on and arrangements for delivery; advice on breast-feeding and parenting; enrollment in appropriate programs such as Medicaid and Aid to Families with Dependent Children, and counseling on family planning.

In this context the experience of France is noteworthy. France has recently made impressive advances in reducing the infant mortality rate. Part of that success has been attributed to a scheme of cash payments to pregnant women. The payments serve as an incentive for the women to engage in prenatal care that features early identification of risks and programs of stress reduction.

Another factor that is occasionally proposed as a contributor to in-

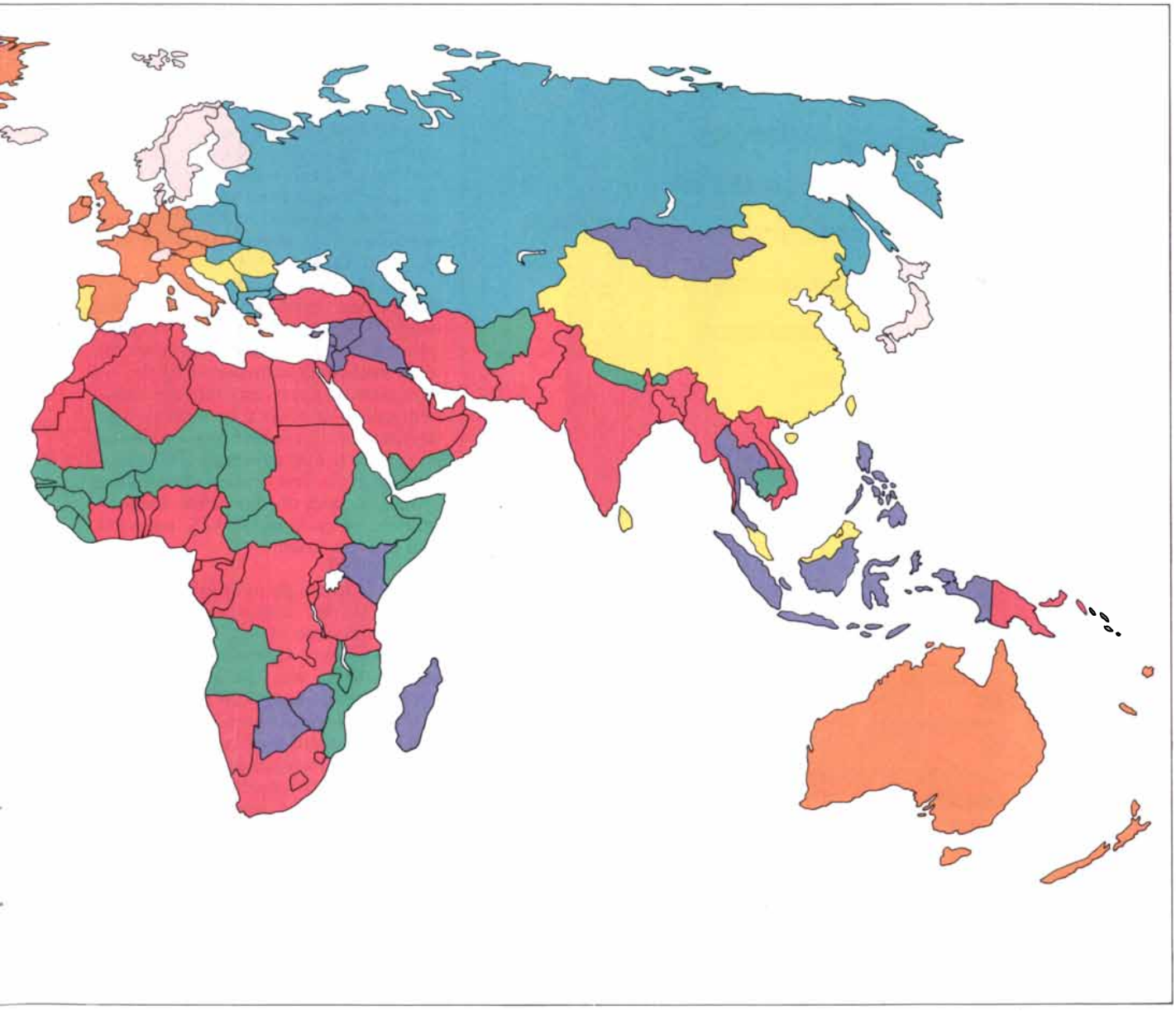


PATTERN OF INFANT MORTALITY around the world in the period 1975–80 highlights the fact that the U.S. is not among the leading nations. Since 1980 the U.S. has slipped a bit farther down on the list of industrialized nations in this statistic. The infant mortality

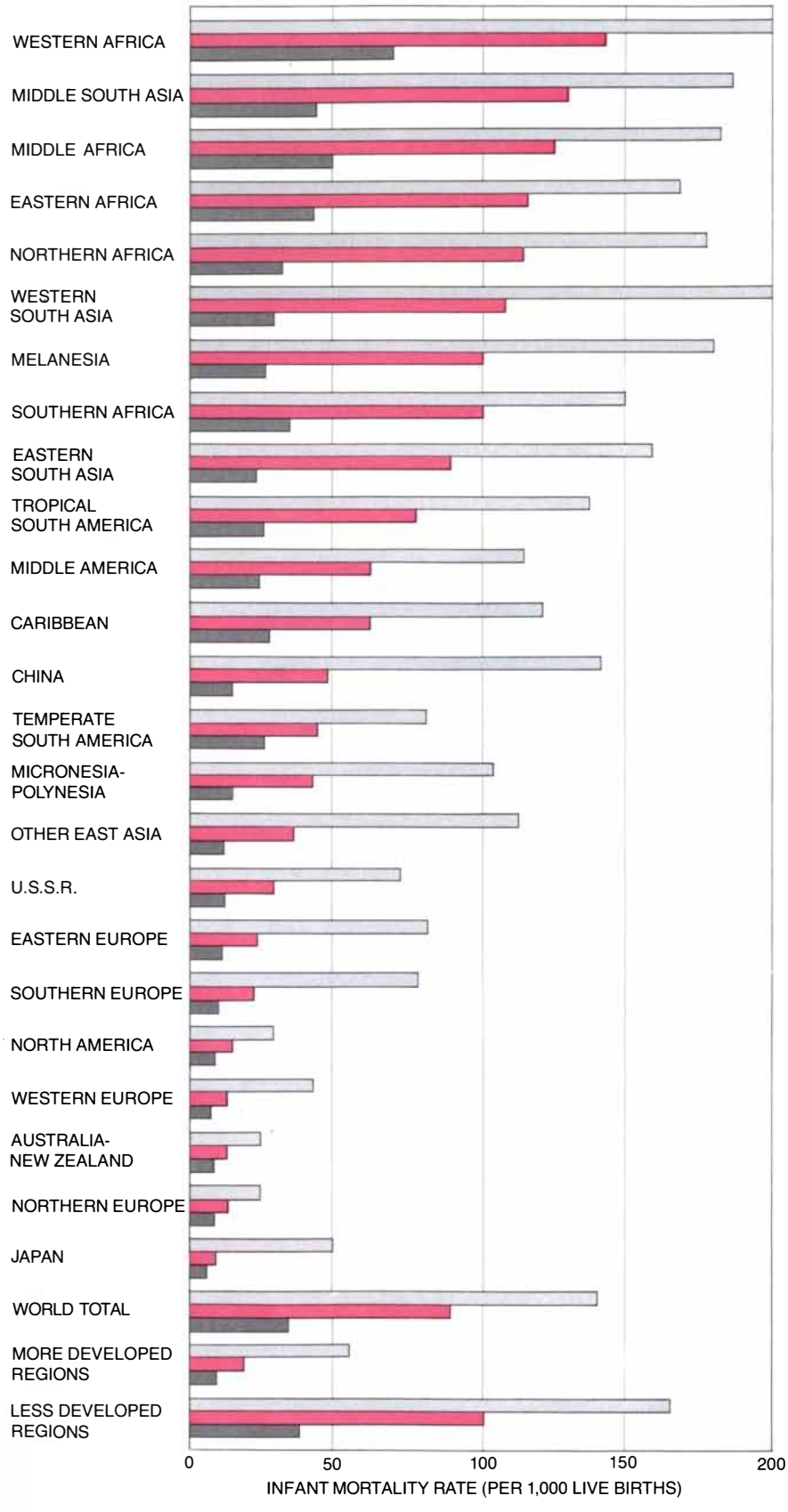
fant mortality is biological differences among racial or ethnic subgroups in the heterogeneous U.S. population. One need only look at the example of Sweden to be made skeptical of this argument. Sweden, which is consistently among the leading nations in low infant mortality rates, has recently had many immigrants from southeastern Europe, where infant mortality rates are relatively high. The people come for menial service jobs and are generally worse off than other people in Sweden in income and living conditions. Nevertheless, the record of infant mortality among the immigrant families is somewhat better than the record among Swedes. The immigrant women appear to avail themselves fully of the extensive Swedish health services, which include prenatal care. In other nations too the descendants of immigrants tend to acquire patterns of infant mortality more like the ones in the country of residence than the ones in the country of origin.

A careful analysis of neonatal mortality among racial groups was done in 1979 by Richard L. Naeye of the Pennsylvania State University College of Medicine. His finding on perinatal mortality (late fetal deaths—after 20 or more weeks of pregnancy—together with deaths in the first 28 days of life per 1,000 births) was: Orientals 23 per 1,000 pregnancies, whites 34, Puerto Ricans 41 and blacks 51. When the data were corrected for such factors as prepregnancy weight, prenatal care, weight gain, height, work outside the home, cigarette smoking, family income, education and marital status, the so-called racial differences were eliminated or greatly reduced. A small unexplained gap between blacks and other groups persisted, associated with a slightly higher infection rate among black people.

Speculations about racial biological differences as a cause of increased infant mortality should be regarded with great caution, particularly in view of the fact that the survival of black infants has been shown to increase as the



rate is stated in terms of deaths during the first year of life per 1,000 live births. For some years until 1982 the nation with the lowest infant mortality rate was Sweden; the leaders now are Finland and Japan, at 6.0 and 6.6 respectively. The rate in the U.S. that year was 11.5 per 1,000 live births. The map is based on data assembled by the United Nations Secretariat and published by it in 1983.



REGIONAL PATTERNS for three five-year periods show that the infant mortality rate has declined throughout the world but in an uneven way. The data cover 1950–55 (*light gray*) and 1975–80 (*color*); the chart also includes projections for the period 2020–25 (*dark gray*). The data are from the Population Division of the UN Secretariat's Department of International Economic and Social Affairs as published in the UN's *Population Bulletin*.

parents improve their socioeconomic circumstances. Important though racial differences may be for people undertaking the statistical analysis of variables, the environmental determinants for the disproportion of infant mortality among blacks appear to provide an ample base for considerations of social policy.

Also deserving skepticism are the easy assumptions that a wealthy nation should have a low infant mortality rate and a poor nation a high one. Among low-income nations in 1981 the rate ranged from 130 (Bangladesh) to 43 (Sri Lanka), among middle-income nations the range was from 120 (Turkey) to 26 (Portugal) and among high-income nations the range was from 130 (Oman) to seven (Japan).

Sri Lanka provides an impressive example of how extreme national poverty does not necessarily entail a high infant mortality rate. The government of that nation tries to make sure that several basic services—education, nutritional supplements, family planning and prenatal care—are provided to all the people who need them. The programs are not necessarily elaborate or sophisticated, but everyone is assured access to them.

The success of this approach in Sri Lanka and some other parts of the world suggests that neglect of people may have a devastating effect on their well-being as measured by the infant mortality rate, and that the situation can be improved by services that are modest in content but ambitious in outreach and coverage. The experience of a few other developing countries and of numerous demonstration projects in the U.S. supports this interpretation.

In the U.S. the period of sharpest decline in the infant mortality rate was the decade of the 1970's. Many influences were at work. Conspicuous among them was the expansion of social-support programs such as Aid to Families with Dependent Children, food stamps and Medicaid and the development of special programs to improve access to maternity-related services. Also at work were changes both upward and downward in the predominant ages of childbearing; a decrease in the proportion of unwanted childbearing through the provision of easier access to family-planning and abortion services; a general improvement in the economy with reductions in the poverty rate; the development of nutritional supplements specifically for pregnant women, and dramatic advances in medical technology for the care and improved survival of infants born at extremely low weight.

As one might expect, few of these factors came with built-in mechanisms for assessing their relative contribution to the decline in the infant mortality rate. That kind of information would be useful in the formulation of public policies to further reduce the rate and to diminish the worsening disparities among different segments of the population. The imprecision of the record makes it possible for experts to put forward quite divergent policies.

In the absence of statistical data on how different factors independently affect the infant mortality rate it is useful to discuss certain of the influences qualitatively.

Demographic changes constitute one of the influences. Childbearing by unmarried women has increased dramatically (from 2.3 percent to 10 percent of all newborns among whites between 1960 and 1980 and from 21.6 percent to 48.5 percent among nonwhites). Among teen-agers the pregnancy rate is far higher than it is in such other industrialized countries as Sweden, Canada and England (96 per 1,000 in the U.S., 35 in Sweden, 44 in Canada and 45 in England).

Unmarried motherhood carries a high risk against survival of the infant. The reasons are not well documented. If unmarried mothers can be presumed to suffer more than others from stress, arising from such factors as social stigma and difficult economic circumstances, that might be part of the explanation. To the extent that women are without help from social supports such as a loving companion or helpful relatives, that too is a known adverse influence.

Social policy in the U.S. provides few supports for pregnant women and new mothers. Sheila B. Kamerman of Columbia University and her colleagues studied maternity policies for working women in 75 countries, including all the industrialized nations and many of the developing ones. The U.S. was the only one without a law ensuring for working women the basic health services and social supports consonant with the unique developmental and nurtural requirements of pregnancy, childbearing and infant care. Indeed, the policy in the U.S. is to treat pregnancy and childbearing as if they were disabilities. Benefits are provided to such women in the same way and to the same limited extent as they are for illnesses. Even that assurance has prevailed only since 1978, when Congress spelled it out in the Pregnancy Discrimination Act.

It seems plausible that the survival of infants born to unmarried mothers would increase under certain alternative policies. They include better job-related maternity benefits, more accessible perinatal health services, more readily available day care and assured access to services aimed at averting unwanted childbearing.

INFANT MORTALITY RATE in the U.S. has declined steadily since 1950 for both whites (*color*) and blacks (*broken black*) as has the overall rate (*solid black*). The flattening curves in about 1982 reflect a slowing of the rate of decline in infant mortality that began then.

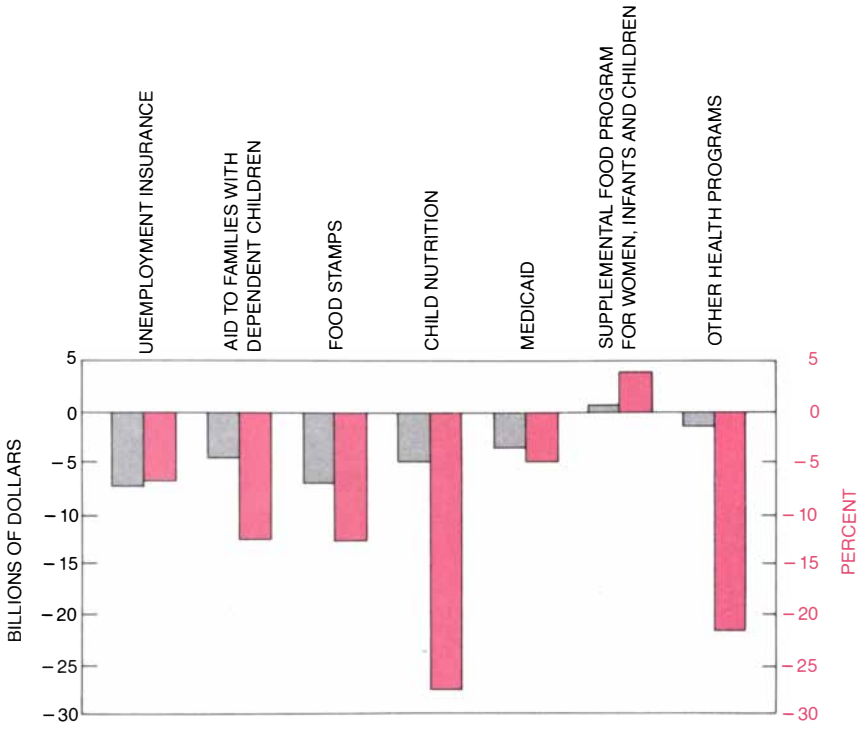
A second influence is provided by services dealing with family planning and abortion. Michael Grossman of the National Bureau of Economic Research and Steven Jacobowitz of Long Island University analyzed a group of programs initiated between 1964 and 1977 in order to measure their relative effect in reducing infant mortality rates. The programs included Medicaid, Maternal and Infant Care projects (Federally funded community programs, often abbreviated MIC, to deliver comprehensive perinatal services in underserved areas), Federally subsidized family-planning services for low-income women, the legalization of abortion and the widespread adoption of oral and intrauterine contraceptive techniques.

What Grossman and Jacobowitz found was that the increase in legal abortions was the single most important factor in reducing neonatal mortality. The second factor in impact was the use of organized family-planning services by low-income women.

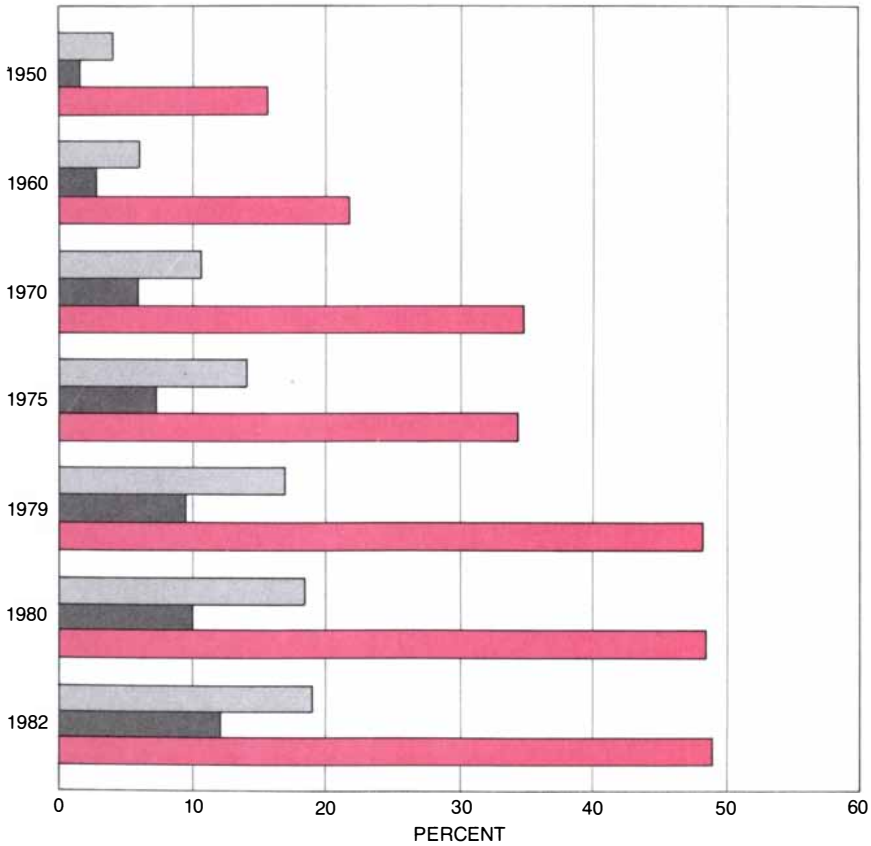
Since then most states have stopped funding abortion. Federal funding of family planning has not increased in real dollars since 1973. Support for both programs has been sharply reduced since 1981. Indeed, the Reagan Administration pursues a policy that attempts to criminalize abortion, with the president himself formally endorsing and pledging "my full support" to three separate legislative initiatives that would overturn the decision by the U.S. Supreme Court in 1973 establishing women's constitutional right to abortion.

A third major influence is poverty. Although the correlation between poverty and high infant mortality rates is undisputed, the best strategies for reducing infant mortality among poor populations are not self-evident. Countries that have had a social revolution (China and Cuba more notably than the U.S.S.R.) resulting in a redistribution of wealth and an emphasis on equal participation in health services show impressive successes in reducing infant mortality. Even better success has been achieved in capitalist nations that have established extensive welfare and health-benefit systems (the examples include the Scandinavian countries, the Netherlands and the United Kingdom).

I have already mentioned the impor-



DECLINING SUPPORT of programs that help pregnant women, infants, new mothers and families with dependent children is reflected in the changes in Federal outlays between 1982 and 1985. They are the result of policies carried out by the Reagan Administration between 1981 and 1983. The changes are shown in dollars (*gray*) and percent (*color*).



RISE IN BIRTHS OUT OF WEDLOCK has also affected the infant mortality rate; for a variety of reasons many infants born to unmarried mothers do not survive. Births to unmarried women as a percentage of all births are shown in light gray. The out-of-wedlock births for whites are represented by the dark gray bars, for nonwhites by the colored bars.

tance of prenatal care. It is noteworthy that the women in the U.S. who get the most and best prenatal care are the ones who are at the lowest risk of losing an infant. For such women the measurable contribution of the care to the survival of the infants may not be highly significant.

Most of the information about prenatal care comes from birth certificates, which record the date of the first visit and the number of visits. In the statistical analysis of these data the care is presumed to be inadequate if it did not begin until the third trimester (the last three months of pregnancy) or consisted of fewer than five visits. All recommended standards call for at least 10 visits and for more than that if special risks are identified. Unfortunately little information is available on the adequacy of care in terms of the content of the visits.

David L. Kessner of the Institute of Medicine of the National Academy of Sciences analyzed data on New York City women and found that social risks (the woman was single, of an unfavorable age or poorly educated) appeared oftener than medical ones. Furthermore, the infant mortality rates more than doubled in the presence of either kind of risk. An important finding was that 75 percent of the risks could be identified in the first five minutes of the first prenatal visit.

The number of American women receiving prenatal care rose during the 1970's, but a gap persisted between blacks and whites. By 1979 some 40 percent of black women were not seen for prenatal care during the first three months of pregnancy, whereas nearly 80 percent of white women were. Overall participation in prenatal care appears to have declined since 1981 in areas with high unemployment or large minority populations.

The strongest influence in reducing infant mortality in recent years has been the growth of programs for care of newborn babies who are at high risk of death. Some 600 hospitals in the U.S. have established units capable of applying the advanced technology now available. About 6 percent of newborn babies are admitted to such units, often having been brought from a distance by special rapid-transport facilities.

For newborn infants weighing more than 1,500 grams these systems have brought about an impressively improved rate of survival, without an appreciable risk of later problems of health or development. For infants weighing less than 1,500 grams the survival rates are also enormously improved, but the long-term prospects for these babies are more doubtful.

Several reports indicate that the improved survival of such infants has been accompanied by a relative increase in postneonatal mortality. The observation suggests that elaborate neonatal technology delays rather than prevents some infant deaths.

In a study done for the Office of Technology Assessment, Peter Budetti estimated the average cost of neonatal intensive care in 1978 to be $13,616 per patient. Routine prenatal care has been estimated to cost about $350 per patient. On the basis of these figures a group advocating comprehensive prenatal care for all low-income women has calculated that by providing such a program the Federal Government would save $360 million per year in outlays for neonatal intensive care and the rehospitalization of babies born at low weight.

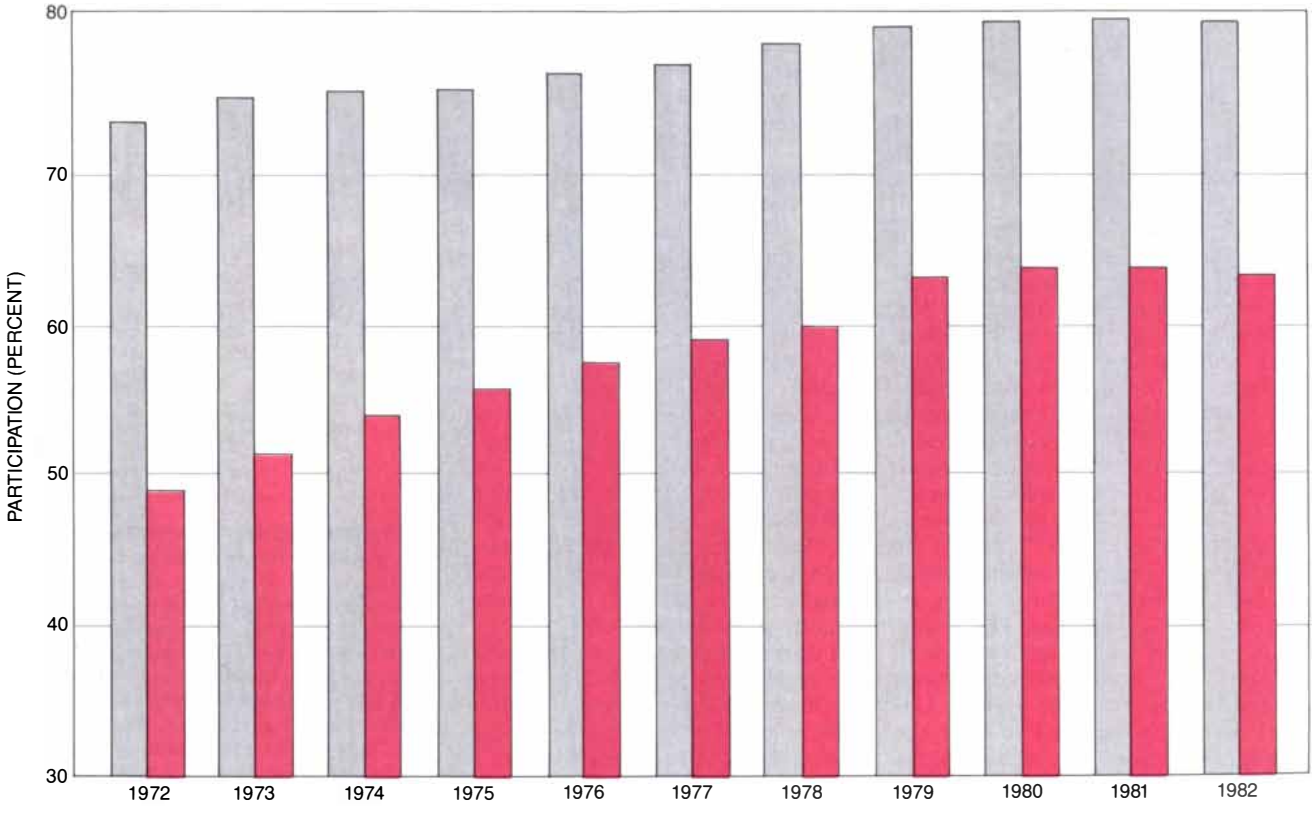
Against this background one can examine the effects of Federal policy on health services relating to infant mortality. For this purpose it is instructive to compare the recession of 1974–75, when publicly sponsored social supports and health services were expanded, with the recession of 1981–82, when programs of that nature were curtailed.

To make the comparison my colleagues and I examined data from a group of eight states that were seriously affected by the earlier recession (Georgia, Michigan, Missouri, Nebraska, New York, Ohio, South Carolina and Virginia) and eight that were not (California, Iowa, Kentucky, Louisiana, Maryland, Montana, Oklahoma and Texas). For each group we analyzed the rates for infant mortality, neonatal mortality and low birth weight for whites and nonwhites in the period 1973–77. We found that infant-health trends improved equally in both groups of states and among all populations. No reversal of favorable trends could be identified.

A strictly parallel analysis for the later recession is not yet possible, but the preliminary data indicate a very different experience. The states hardest hit by the recession reported a decline in the participation of pregnant women in prenatal care, consistent with the loss of Medicaid and insurance benefits and with the curtailment of support for public clinics. A number of states reported a reversal of previously favorable trends for rates of infant mortality and birth weight, particularly among minority populations.

When the Public Health Service established in 1979 the goal of an infant mortality rate of nine per 1,000 live births by 1990, it also proposed some related objectives to be achieved within the same period. One was that the infant mortality rate should not exceed 12 per 1,000 live births for any racial or ethnic group. Another was that low-weight births should constitute no more than 5 percent of all live births, with no racial or ethnic group exceeding 9 percent. A third was that 90 percent of pregnant women would begin prenatal care within the first trimester.

Analyzing reported rates of progress by state, the Children's Defense Fund has concluded that none of these goals are likely to be met by 1990 on a nationwide basis. The means whereby the goals could be met are not mysterious, but they require the implementation of certain public policies that are not even being seriously considered. The means include assured access to comprehensive perinatal care, guaranteed maternity leaves for all working pregnant women and recent mothers, job protection during the leave and cash benefits equal to a significant portion of wages during the leave. These measures can be promoted on the basis of humanitarian concern, social equity, cost-effectiveness and even national security to the extent that it will depend on a coming generation both vigorous and productive.



PARTICIPATION IN PRENATAL CARE beginning during the first three months of pregnancy is a factor in reducing the infant mortality rate. Participation by black women (*color*) is lower than participation by white women (*gray*), although the rate of participation by blacks rose notably in 1970–79. The chart is based on information assembled by the National Center for Health Statistics.

The Moons of Uranus, Neptune and Pluto

When Voyager 2 flies by Uranus in January and by Neptune in 1989, it will send back closeup pictures of their icy moons. Ground-based studies of the outer solar system suggest what the probe might find

by Robert Hamilton Brown and Dale P. Cruikshank

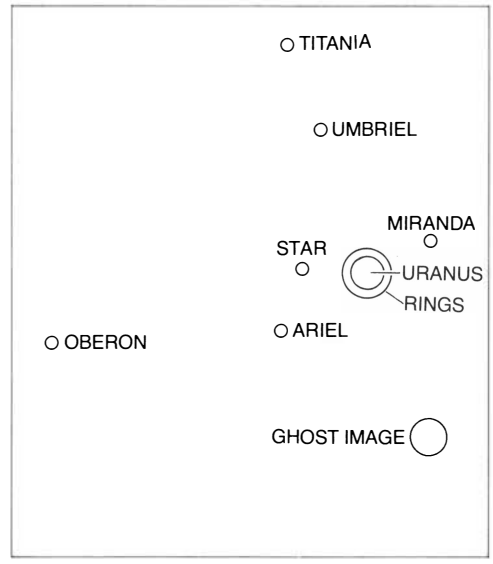
Just over 200 years ago an observation by an amateur astronomer doubled the size of the known solar system. Peering through a homemade six-inch telescope, William Herschel discovered Uranus, the seventh planet. Sixty-five years later Neptune was recognized; like Uranus, it is a giant planet roughly four times the size of the earth. Pluto, no larger than the earth's moon and twice as far as Uranus from the sun, did not emerge from the black until 1930. By that time, thanks to enterprising astronomers and rapidly improving astronomical instruments, it had become apparent that the outer three planets do not travel alone. Uranus was eventually found to have at least five moons and Neptune at least two. In 1978 Pluto's first satellite was detected on a grainy photograph of the planet.

Until recently little was known about the satellites of Uranus, Neptune and Pluto other than the parameters of their orbits. In contrast, the Pioneer and Voyager spacecraft had sent back stunning images and a wealth of data on the satellites of Jupiter and Saturn, transforming those remote specks into worlds whose history could be discerned, if only in outline. A similar era of discovery is about to begin for Uranus and Neptune. *Voyager 2,* launched in August, 1977, will fly by Uranus next January, passing within 29,000 kilometers of the planet's innermost moon, Miranda. The spacecraft will then encounter Neptune in August, 1989, on its last rendezvous before leaving the solar system. Unconstrained by the need to steer the probe toward another destination, the mission planners are taking full advantage of the final encounter. *Voyager 2* will fly within 10,000 kilometers of Neptune's inner satellite, Triton. Images of Triton made by a high-resolution camera are expected to reveal surface features as small as a few hundred meters in diameter.

In preparation for the Voyager flybys the pace of ground-based investigations of the moons of Uranus and Neptune has quickened considerably. The same is true of Pluto, which in some ways is more like a satellite than a planet, and of its moon, Charon. The increased interest, together with a vast improvement in telescopic detector technology, has begun to yield an understanding of the physical properties of these bodies. What follows, then, is a preview—a somewhat indistinct first look at objects that promise to be among the most interesting in the solar system.

With ground-based telescopes one can learn a number of important things about remote planetary satellites. First, one can determine the shape, size and period of their orbits and thereby make accurate predictions of their positions. More important, knowledge of a satellite's orbital parameters enables workers to calculate the mass of the planet and the strength of its gravitational field at various locations. The mass of the satellite itself can often be inferred by observing the perturbations it induces in the orbits of other satellites. If its diameter is known, one can then compute its mean density, which is the most important clue to bulk composition.

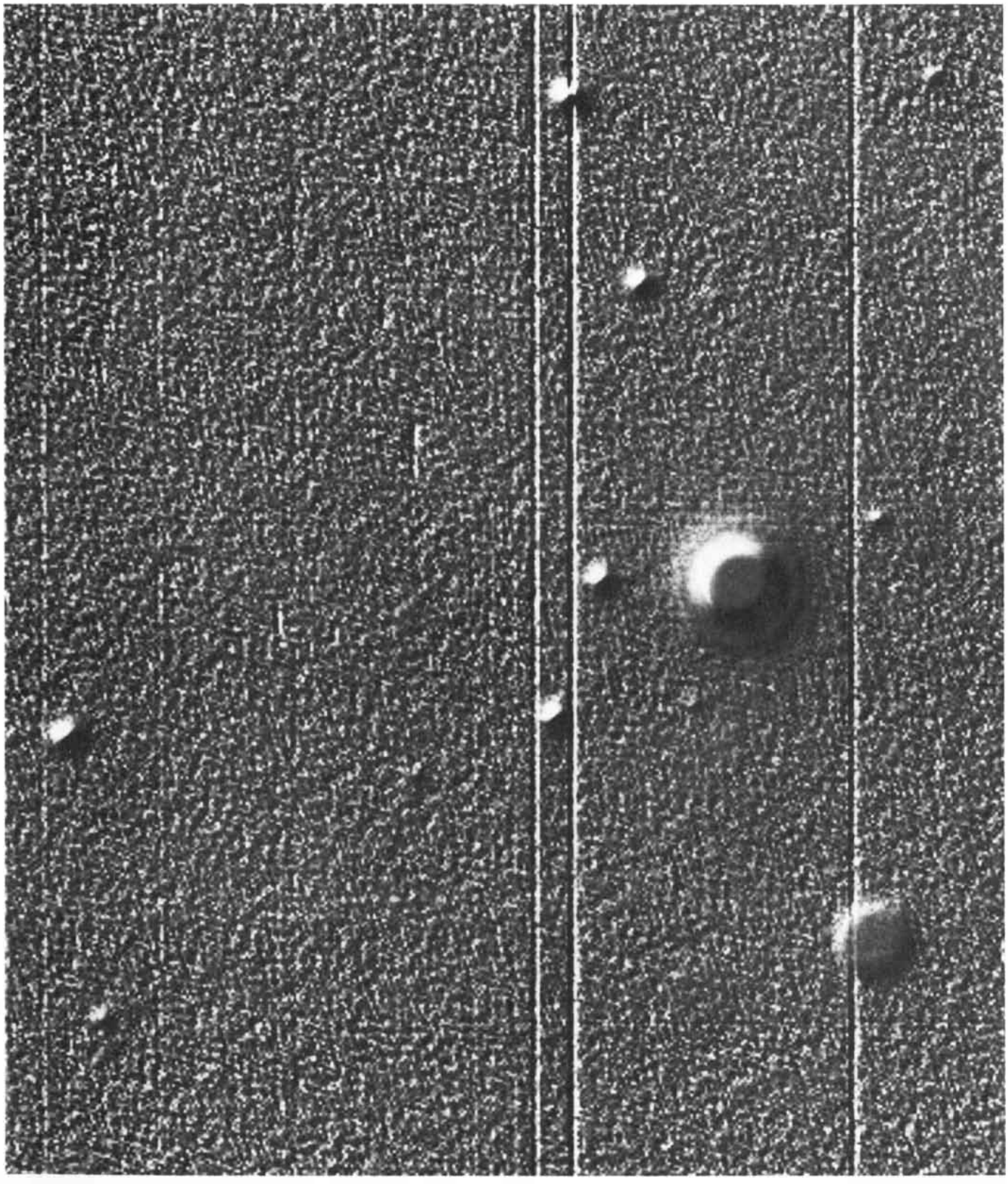
Information on surface composition can be gleaned more directly from absorption bands in the spectrum of sunlight reflected off the satellite. By comparing the observed spectrum with the laboratory spectra of various chemical elements or compounds one can identify some of the satellite's surface constituents. This work is best done in the near-infrared region (at wavelengths of one micrometer to five micrometers), because most of the molecular substances common in the solar system have strong absorption signatures in that range. Even when it is observed through the largest telescopes, however, the radiation from small, distant satellites is extraordinarily faint. Not until the late 1970's were infrared detectors capable of resolving it into

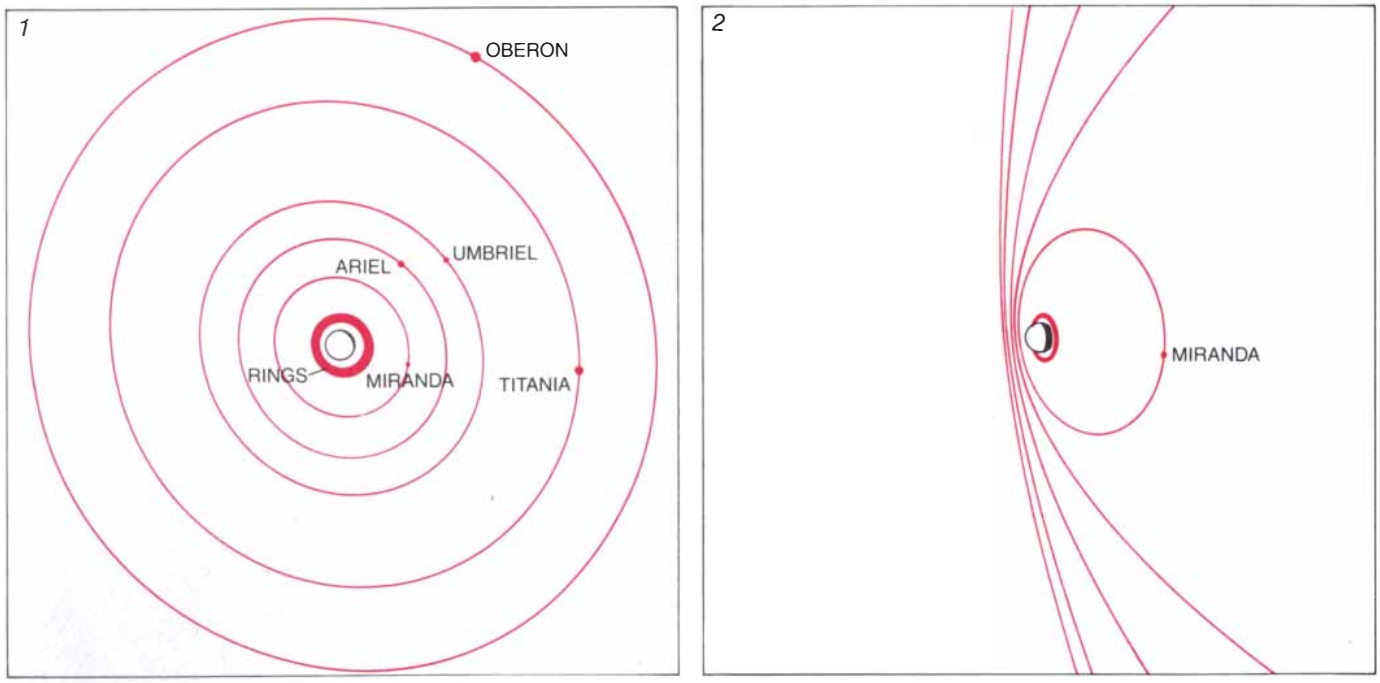


FIVE MOONS OF URANUS are seen on an electronic image made by Bradford A. Smith of the University of Arizona and Richard J. Terrile of the Jet Propulsion Laboratory. The visible-light image, produced with a charged-coupled device attached to the Carnegie Institution of Washington's 2.5-meter telescope in Chile, also shows the planet's thin rings of dark particles. Uranus' spin axis lies almost in its orbital plane; at present its south pole is pointed toward the earth and the sun. The orbit of Miranda, the innermost moon, is inclined by several degrees, but the other moons orbit Uranus in its equatorial plane. The vertical lines are caused by defects in the electronic detector.

spectra. (The new detectors use indium antimonide rather than lead sulfide as the light-sensitive compound; infrared radiation produces free electrons in the indium antimonide, generating an electric current whose intensity is proportional to that of the radiation.)

Theories of how the solar system evolved give some idea of what to expect in investigating the composition of bodies in its outer regions. The formation of both satellites and planets is thought to have begun with the accretion of frozen dust grains that condensed out of the primitive solar nebula as it cooled. The composition of the grains and hence of the larger bodies depends on their distance from the sun. In the hot region near the sun only relatively refractory substances could condense. As a result the inner planets—





PASSAGE OF *VOYAGER 2* through the Uranian system will offer a close look at the planet and at Miranda. The drawings, adapted from computer simulations prepared by mission planners at the Jet Propulsion Laboratory, show views through the spacecraft's cameras from several points on its trajectory. *Voyager 2* is approaching the sunlit south pole of Uranus at a speed of about 53,000 kilometers per hour. Ten days before its closest approach to the planet on January 24, 1986, the probe's wide-angle camera will still

Mercury, Venus, the earth and Mars—are rocky bodies consisting primarily of metals, metal oxides and silicates. At the distance of Jupiter the solar nebula cooled to temperatures low enough for ices to form; water ice seems to make up a significant part of the Jovian satellites and the bulk of Saturn's moons. Still farther from the sun, in the range of Uranus, Neptune and Pluto, one would expect to find water ice, but also ices of materials even more volatile than water: methane (CH_4) and ammonia (NH_3).

Uranus and its five moons—beginning with the outermost they are Oberon, Titania, Umbriel, Ariel and Miranda—make up the third-largest satellite system after the Jovian and Saturnian systems. The smallest of the group, Miranda, is six times fainter than the faintest of the others. It was discovered in 1948, and since then no new Uranian satellites have been observed in spite of concerted efforts with the largest telescopes and the most sensitive detectors. *Voyager 2* will undertake a special search. If experience with Jupiter and Saturn is a guide, the search may well reveal tiny moons that have escaped detection from the earth.

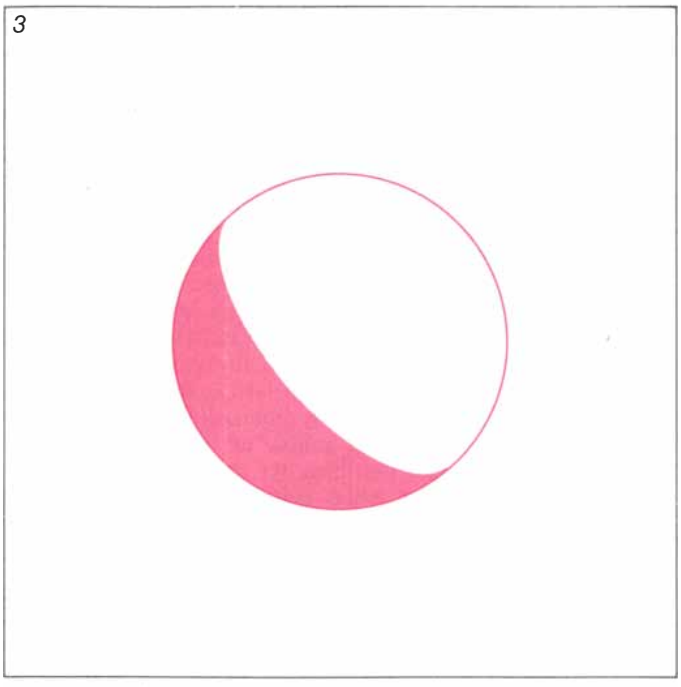
The known Uranian satellites travel in virtually circular orbits, and except for Miranda their orbital planes coincide with the planet's equatorial plane to within a few tenths of a degree. Although the satellite orbits are quite regular, the orientation of the entire system is unusual: it is tilted on its side, so that the rotation axes of both the planet and its moons lie nearly in the planet's orbital plane. Some workers attribute Uranus' odd orientation—as well as the less extreme inclinations of other planetary axes—to the impact of a planetesimal early in the planet's history. It has been suggested that such a catastrophe would have affected the evolution of the Uranian satellites, but so far there is no strong evidence to support or refute this hypothesis.

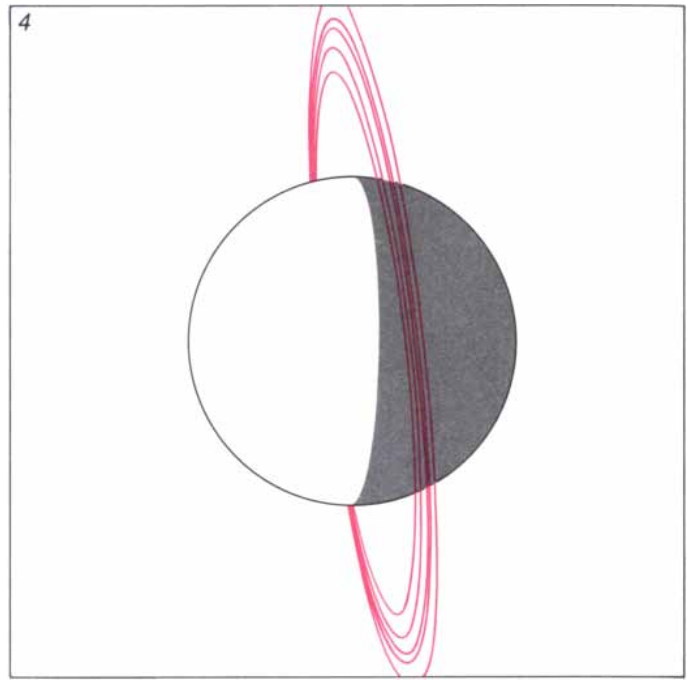
In 1979, even before the size and mass of the Uranian moons could be determined with confidence, we established the presence of water ice on the surface of the four outer ones. The near-infrared spectra we measured with the four-meter telescope at the Kitt Peak National Observatory all displayed the characteristic signature of water ice: broad, strong absorption bands at wavelengths of 1.5 and 2.0 micrometers. In 1983 one of us (Brown) and Roger N. Clark of the U.S. Geological Survey resolved the spectrum of Miranda for the first time; the data are relatively crude, but they show the 2.0-micrometer absorption band fairly clearly.

By then B. Thomas Soifer and his colleagues at the California Institute of Technology had confirmed our earlier results. Furthermore, their spectra suggested the presence of an additional compound on the surface of the Uranian satellites. Although the investigators were not able to identify the new component, they concluded it was significantly darker than water ice.

In the course of measuring the diameter of the Uranian moons we verified that their reflectivities are indeed lower than would be expected if they were composed of pure, pulverized water ice. The connection between diameter and reflectivity is worth explaining in detail. Because the Uranian satellites are so distant, the diameter of the disks they present to observers on the earth cannot be measured directly. At the same time the measurable intensity of their reflected light does not offer unambiguous information on their size: a satellite of a given brightness might be small and highly reflective or large and comparatively dark.

The ambiguity can be resolved by measuring the satellite's thermal flux, which also depends on its size and reflectivity. Sunlight that is not reflected heats the surface of the satellite and is reradiated as thermal energy at longer, infrared wavelengths (between 20 and 100 micrometers in the case of the Uranian moons). By combining radiometric measurements of the infrared heat flux with photometric measurements of brightness, one can in principle construct and solve a system of two equations in which the two unknowns





take in the entire satellite system (*1*). Two hours before the closest approach (*2*) the spacecraft will be nearing Uranus' equatorial plane between the orbits of Miranda and Ariel. As it passes within 29,000 kilometers of Miranda an hour later, its high-resolution camera will make images (*3*) that may reveal evidence of volcanic activity, such as geologically recent ice flows. About 13 minutes after that, as *Voyager 2* crosses the equatorial plane, high-resolution images will show the dark side of Uranus and the edge of its rings (*4*).

are the satellite's diameter and reflectivity. (In practice one must also make certain assumptions about the satellite's radiative properties. Fortunately these turn out to be quite simple for a surface consisting of fine ice grains loosely aggregated in a vacuum, because such a surface does not store or conduct much heat.)

Measuring the diameters of the Uranian satellites by means of the photometric-radiometric technique has nonetheless proved difficult. The surface temperature of the satellites is only 80 degrees Kelvin (degrees Celsius above absolute zero), and Uranus is never closer than about 2.7 billion kilometers from the earth. Furthermore, water vapor in the earth's atmosphere absorbs much of the incident infrared radiation. As a result the thermal flux from the Uranian moons is barely detectable even with the best infrared telescopes; indeed, Miranda's is not detectable at all.

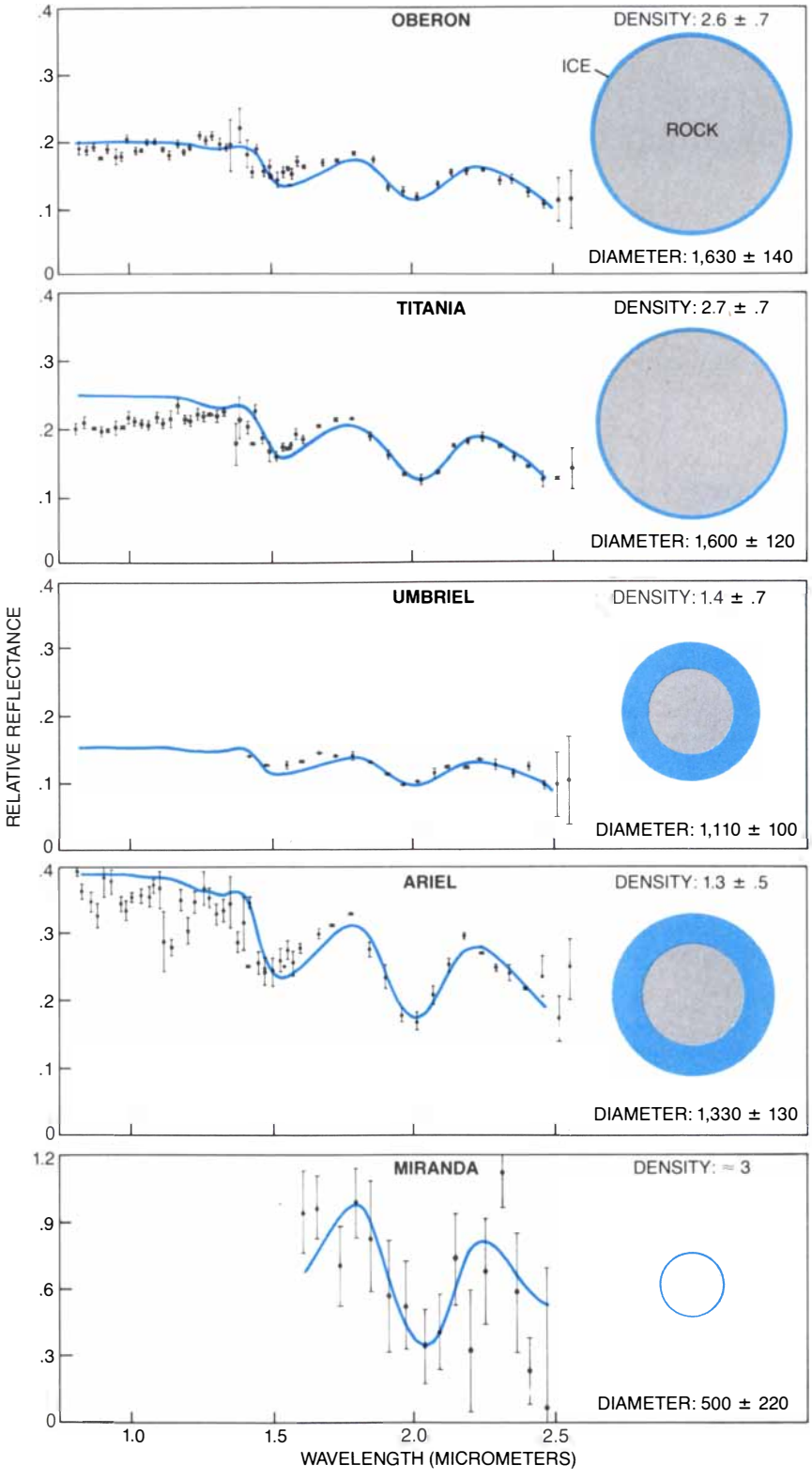
One of the best infrared telescopes is the National Aeronautics and Space Administration's facility on the Hawaiian peak Mauna Kea, which at an altitude of 4,200 meters benefits from a relatively dry atmosphere. Working there in 1981 with David Morrison of the University of Hawaii at Manoa, we measured the thermal flux and calculated the diameter and reflectivity of the outer four Uranian moons. We found they are much larger and darker than had been thought when it was assumed that their surfaces were pure water ice. Although they are considerably smaller than the four Galilean moons of Jupiter, Saturn's Titan and the earth's moon, they are among the largest satellites in the solar system.

What, then, is the substance that darkens their surface? At first we were puzzled by the absence, even on more recent high-resolution spectra, of absorption bands other than those attributable to water ice. Soon a possible explanation suggested itself: the dark substance might be spectrally neutral. In other words, it might be a colorless material that reflects very little light and absorbs all wavelengths more or less equally.

In laboratory experiments we found we could approximately match the satellite spectra with the spectra of fine-grained water ice, or frost, covered with isolated, uniformly distributed patches of charcoal. The satellites differ in reflectivity, and correspondingly the surface models that best reproduce the individual spectra differ in their proportions of water ice and charcoal. Charcoal is merely a convenient laboratory material; since it is derived from living organisms, it is probably not found on the Uranian moons. Like charcoal, however, the dark substance on the satellite surfaces is likely to be spectrally neutral and to reflect between 5 and 15 percent of incident light. Materials fitting this description include a magnetic oxide of iron called magnetite, several silicate minerals and a dark organic polymer found in primitive carbonaceous meteorites.

Yet there is reason to believe such rocky materials might not account for the dark patches on the Uranian moons. Theoretical work by David J. Stevenson of Caltech suggests the moons are large enough to have undergone physical differentiation. During this process heat from the decay of radioactive elements in the rocks would have melted the interior of the moons, allowing rocky material to settle to the core. (Most smaller satellites, it is generally thought, do not undergo differentiation because heat escapes from their interior before it can induce melting.) Moreover, according to Stevenson, the melting might have been accompanied by extensive volcanic activity. Water ejected in volcanic eruptions would have covered the surface with a layer of nearly pure, fine-grained ice, burying any remaining exposed rocks.

Steven W. Squyres, now at NASA's Ames Research Center, and Carl Sagan of Cornell University have proposed an explanation for the origin of the dark material that may be consistent with both differentiation and volcanic resurfacing. According to their hypothesis, ultraviolet light from the



COMPOSITION OF URANIAN MOONS can be inferred from their near-infrared reflectance spectra and from their mean densities. The spectra of the four outer moons all have deep absorption bands at wavelengths of 1.5 and 2.0 micrometers, both of which are characteristic of water ice. The data for Miranda are crude (the vertical lines show the error range) but a strong absorption band is evident at 2.0 micrometers. The satellite spectra (*dots*) are matched closely by the spectra of two-component laboratory models (*lines*) consisting of water frost and charcoal. All the moons are thought to have an icy surface and a rocky core, but the differences in mean density (measured in grams per cubic centimeter) suggest that Ariel and Umbriel have thicker ice layers than Titania and Oberon. Miranda's density is too uncertain to allow a conclusion about its internal structure. The diameters are given in kilometers. In comparison, the earth's moon has a diameter of 3,480 kilometers.

sun decomposes methane trapped in water-ice crystals at the surface of the satellites. The resulting atomic carbon and hydrogen recombine with methane to form complex hydrocarbon polymers whose color is a dark red. Laboratory experiments by a number of workers have confirmed that methane ice or methane-contaminated water ice does indeed form dark organic polymers—similar to those found in carbonaceous meteorites—when it is subjected to energetic radiation. The radiation need not be ultraviolet light: the ice can also be darkened by gamma rays like those emanating from the center of the galaxy or by energetic charged particles like those found in the Van Allen belts formed by the earth's magnetic field.

If much of the methane on the surface of the Uranian moons has been converted into dark matter, that would explain why methane has not yet been observed there even though most models of solar-system formation predict its presence. It is also possible that the strong absorption bands of water ice mask the spectral signatures of methane as well as of ammonia. Unfortunately *Voyager 2* is not carrying any instruments designed to gather detailed data on the surface composition of the satellites. Images transmitted by the probe will yield some new information, but conclusive evidence of methane and ammonia, and of the identity of the dark surface constituent, will have to be obtained through improved ground-based observations or by future spacecraft. Whatever the dark substance is, small particles of it may also make up Uranus' nine narrow rings: the rings too are spectrally neutral, and their reflectivity is even lower than that of the dark patches on the moons.

In trying to understand the origin and evolution of the Uranian satellites information on their bulk composition is even more important than knowledge of their surface properties. The most revealing indicator of an astronomical object's bulk composition is its mean density, which can be calculated if its mass and diameter are known. Christian Veillet of the Centre d'Études et de Recherches Géodynamiques et Astronomiques in Grasse, France, recently made the first reliable calculations of the masses of the Uranian moons by refining previous measurements of their orbits.

Using our measurements of the diameters of the outer four satellites, Veillet found that Ariel and Umbriel respectively have densities of approximately 1.3 and 1.4 grams per cubic centimeter, whereas the densities of

Titania and Oberon are roughly twice as large. (The comparable figure for the earth is 5.5 grams per cubic centimeter.) Ariel and Umbriel are about as dense as many of the icy satellites of Saturn, and so their bulk composition may be similar. A mixture consisting of 55 percent by weight of water ice, whose mean density is 0.9 grams per cubic centimeter, and 45 percent rocky material with a density of 3.0 yields the correct overall density value. By the same reasoning Titania and Oberon consist of a mixture of 95 percent rock and 5 percent ice. Assuming that melting has caused the rock to settle into the core, the outer two Uranian moons would be rocky spheres with thin skins of water ice; Ariel and Umbriel would have much thicker ice layers and smaller rocky cores.

This conclusion, if true, is somewhat surprising. If the moons formed at the same time as Uranus itself, then heat released by the accretion of interstellar grains and the gravitational contraction of the protoplanet should have resulted in a larger amount of heavy material condensing near the planet. Ariel and Umbriel should therefore have a higher proportion of rock and a lower proportion of volatiles than Titania and Oberon. Veillet's contrary finding could be taken as evidence for the hypothesis, advanced by some workers, that the Uranian satellites were actually formed later, by the same catastrophic impact that supposedly knocked the planet on its side. Current models of the satellites' internal structure must be treated skeptically, however, because the density calculations on which they are based contain large uncertainties.

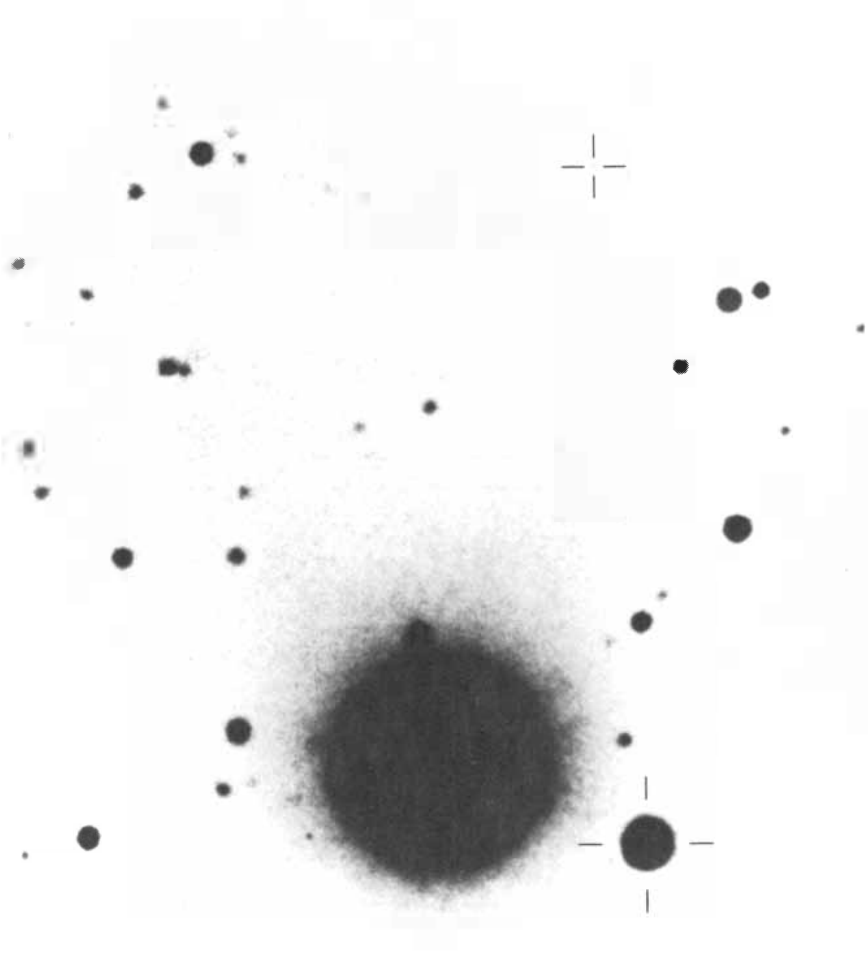
The greatest source of uncertainty is the diameter measurements, and so the situation should improve substantially when *Voyager 2* arrives at Uranus in January. Images transmitted by its high-resolution camera will enable workers to determine the diameter of all five satellites with high accuracy. Small deviations in the spacecraft's trajectory caused by the gravitational pull of Miranda will also make it possible to calculate that moon's mass with greater precision.

Just as important, the Voyager cameras will show what the satellites look like. The spacecraft will pass within 29,000 kilometers of Miranda and 127,000 kilometers of Ariel. The images of these two moons should reveal surface features as small as two kilometers wide. The angle of the probe's approach and its high velocity—about 10 times that of a rifle bullet—will allow only a small fraction of the satellite surfaces to be viewed in detail. Nevertheless, the images will probably show patches of light and dark material. They should also give an indication of the extent to which the Uranian satellites have been cratered by meteorites and resurfaced by fresh ice from below the surface.

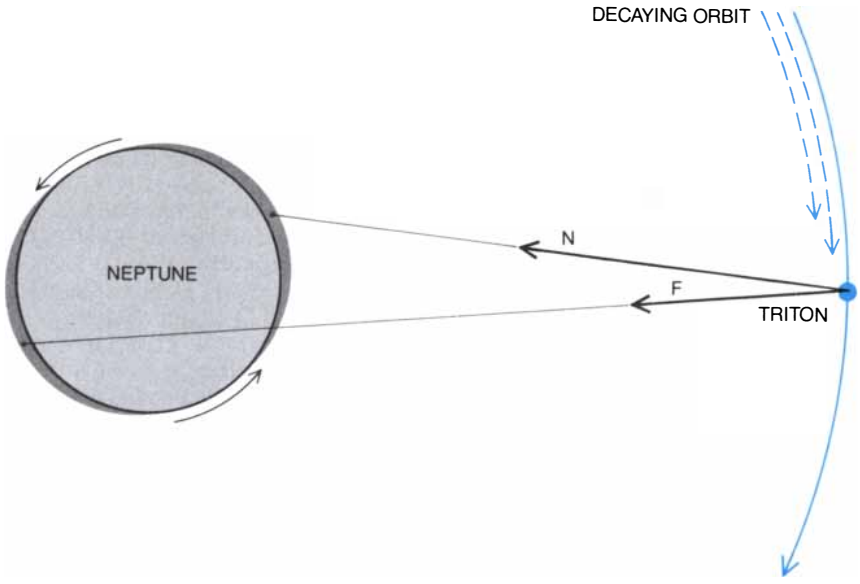
Theoretical studies suggest that volcanic activity and resurfacing may have occurred in the geologically recent past on Ariel and Miranda. Both are significantly more reflective than their companions, indicating they may be covered with extensive deposits of relatively fresh, undarkened ice. The Voyager images may reveal volcanic ice flows. If we are lucky, they may even record eruptions in progress, as they did on Jupiter's moon Io.

Three and a half years and more than a billion and a half kilometers later *Voyager 2* will encounter Neptune and its large satellite Triton. The spacecraft will not fly close to Nereid, Neptune's other moon, which is in a highly inclined elliptical orbit that takes it more than nine million kilometers from the planet. Almost nothing is known of Nereid's physical properties because it is very faint; from its luminosity it is judged to be between 150 and 525 kilometers in diameter, comparable to the largest asteroids.

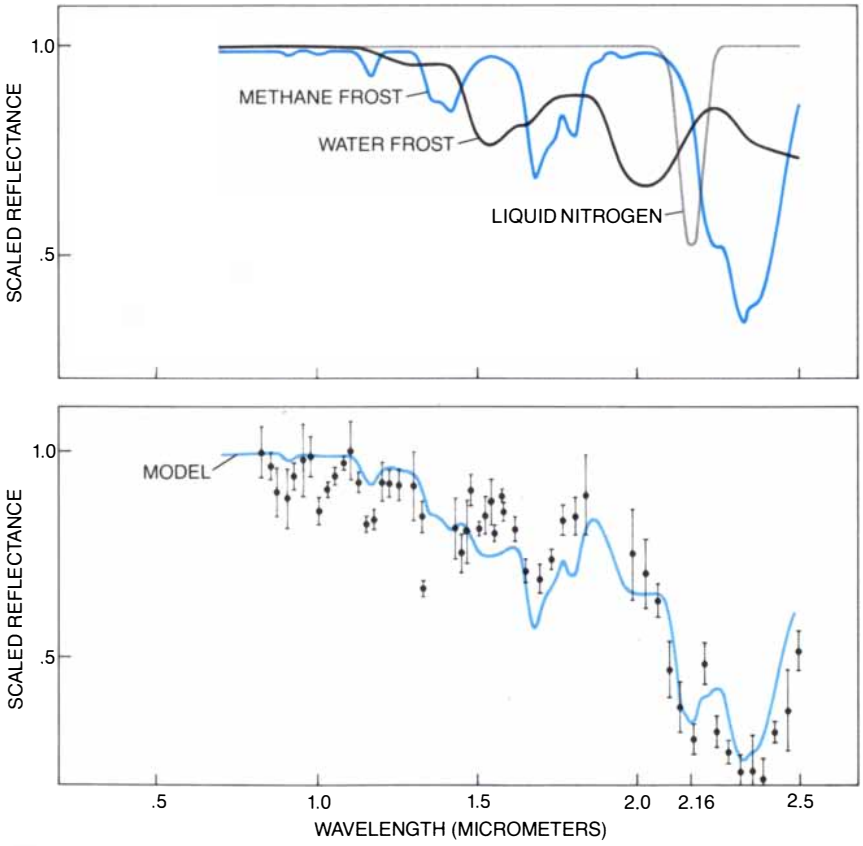
Triton, in a circular orbit roughly 350,000 kilometers from Neptune, is comparable in size to the earth's moon, which has a diameter of 3,480 kilometers. Triton's orbit is nearly coincident with the orbital plane of Neptune and is inclined by about 28 degrees to Neptune's equatorial plane. It is retrograde: the satellite revolves around Neptune in a direction opposite to that of the planet's rotation. A retrograde orbit slowly decays, because the tidal bulge raised on the planet by the satellite is carried in the opposite direction by the planet's rotation; the bulge exerts a gravitational pull on the satellite from behind, slowing it down [*see top illustration on next page*]. Early calculations predicted Triton's demise in less than 100 million years, but a more recent study by Alan W. Harris of the Jet Propulsion Laboratory indicates that the orbital decay is exceedingly



TWO MOONS OF NEPTUNE are Triton and Nereid (*bars*). Triton, the large inner satellite, is about the size of the earth's moon. (In the photograph it appears far bigger than it really is.) Nereid is probably between 150 and 525 kilometers in diameter. Photograph was made by Christian Veillet with the Canada-France-Hawaii Telescope on Mauna Kea.



RETROGRADE ORBIT OF TRITON is decaying almost imperceptibly under the influence of tidal forces. As Triton moves in a circular orbit some 350,000 kilometers from Neptune, its gravitational field distorts the planet into a more ellipsoidal shape. (The two bodies are drawn to a common scale but the distance between them is disproportionately small; the size of the tidal bulges is greatly exaggerated.) Because the planet is not perfectly elastic and because it spins in a direction opposite to that of Triton's orbital motion, the near-side bulge is always slightly behind the satellite; its gravitation exerts a drag (*N*) on Triton, causing the moon to slowly lose altitude. The opposing force of the far-side bulge (*F*) is smaller.



REFLECTANCE SPECTRUM OF TRITON in the near-infrared (*bottom*) suggests that its surface consists primarily of methane ice and liquid nitrogen. Most of the absorption bands (reflectance troughs) match bands found in the spectrum of fine-grained methane frost. The band at 2.16 micrometers does not, but it is prominent in the nitrogen spectrum. The "best fit" model is a superposition of these two spectra and the spectrum of water frost.

slow, and that Triton will survive for 10 billion years or so.

Neptune and its moons lie half again as far from the sun as Uranus, and the region of the solar nebula in which they formed was probably colder. One would therefore expect Triton's chemistry to differ somewhat from that of the Uranian satellites. On the other hand, before data became available it was certainly expected that Triton's near-infrared spectrum would show the signature of very cold water ice. The absence of this signature was a striking feature of the first crude spectrum of Triton, obtained in 1978 by one of us (Cruikshank) and Peter M. Silvaggio, then at the Ames Research Center. Equally striking was the presence of a strong absorption band at a wavelength of 2.3 micrometers, which is characteristic of methane. The work showed that water may not be the dominant volatile material on the surface of objects at the edge of the solar system. The region of the solar nebula in which Neptune formed seems to have been profoundly different in temperature and composition from the region near Uranus.

The discovery of methane on Triton raised another exciting prospect: Triton must have an atmosphere. Methane is more volatile than water; in other words, the vapor pressure of methane ice at a given temperature—its tendency to sublimate into a gas—is greater than that of water ice. At the low temperatures prevalent near Uranus and Neptune virtually no water ice can sublimate, but methane ice sublimates readily. Although the absorption bands on Triton's spectrum could have been produced by either the frozen or the gaseous form of methane, both are probably present on the satellite; generally the presence of one implies that of the other.

We believe Triton may have an ocean as well. Collaborating with Clark in 1981, we obtained improved spectra that revealed another important feature: an absorption band at 2.16 micrometers that cannot be attributed to methane or to closely related hydrocarbons. We have tentatively attributed this band to molecular nitrogen. Ordinarily nitrogen does not absorb in the near-infrared, but the interaction of nitrogen molecules under high pressure does produce weak absorption bands. To produce the strong spectral feature we have observed, the nitrogen on Triton would have to be under very high pressure; given the satellite's low surface temperature, the nitrogen would have condensed to a liquid or a solid.

Our interpretation of the observations is that a large fraction of Triton is

covered by liquid nitrogen at least a few tens of centimeters deep and possibly much deeper. The abundance of nitrogen appears to be much greater than that of methane. A small amount of methane may be dissolved in the nitrogen, but the data indicate methane exists as blocks of ice elsewhere on the surface.

We want to emphasize that the identification of nitrogen on Triton is tentative. If the nitrogen exists, it would not have to be in liquid form: Jonathan I. Lunine of the University of Arizona and Stevenson have proposed, on the basis of theoretical work, that it may be frozen solid. Further ground-based studies and *Voyager 2* data may help to settle the question.

Because of the high vapor pressure of nitrogen its presence in any form on Triton would imply that the satellite has a substantial nitrogen atmosphere; the atmospheric pressure at the surface would be between 10 and 30 percent of the level found on the earth. The earth and Saturn's moon Titan are the only other planet-size objects in the solar system known to have an atmosphere composed primarily of nitrogen. (If the earth were moved to the location of Triton, the low temperature would cause its atmosphere to condense to a liquid sea some 15 meters deep.) A nitrogen atmosphere on Triton would therefore be of considerable interest to planetary scientists. In particular it would bear on the question of the origin of nitrogen in the solar system: whether it originated as a pure substance in the solar nebula or whether it is derived from the photodissociation of ammonia gas into nitrogen and hydrogen. Ammonia is assumed to have been abundant in the nebula because hydrogen was abundant.

Solid ice, an ocean and an atmosphere on Triton might be expected to interact on seasonal and daily cycles, as they do on the earth. As a result of the inclination of its rotational axis Triton is subject to pronounced seasons. At present its south pole lies in permanent darkness and its north pole receives constant sunlight; in 82 years, when Neptune is on the other side of the sun, the situation will be reversed. Laurence M. Trafton of the University of Texas at Austin has speculated that the atmospheric abundance of methane in a particular region of Triton varies according to the season, increasing during the summer as methane ice sublimates and decreasing during the winter as methane gas freezes. Similarly, liquid nitrogen would freeze solid in the dark polar region. If the nitrogen sea is very shallow, it may even freeze at lower latitudes on a day-to-night cycle, in synchronism with Triton's rotation period of 5.877 earth days.

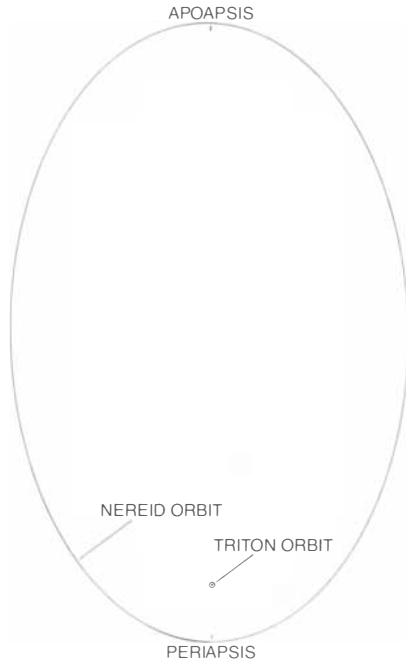
Another important attribute that Triton may share with the earth, as well as with Titan, is a complex organic chemistry. Both methane and nitrogen are colorless, yet Triton's color in the visible range is distinctly reddish. Clearly there must be an additional chemical component on Triton's surface. The red component may be produced by organic chemical processes like those proposed to account for the dark patches on the Uranian satellites. Sagan and Bishun N. Khare of Cornell have shown that ultraviolet irradiation of various mixtures of methane gas with ammonia, water vapor and other substances produces reddish hydrocarbon polymers. Similar compounds are formed, according to M. L. Delitsky of the Calgon Corporation, when methane dissolved in nitrogen is bombarded with charged particles. *Voyager 2* should determine whether Neptune is surrounded by Van Allen belts that could be the source of such particles.

The spacecraft will accomplish far more. Five hours after its closest approach to Neptune, on August 24, 1989, it will pass within 10,000 kilometers of Triton. It will transmit images of much of Triton's surface at a resolution of a few hundred meters—higher than that achieved during any previous flyby except the *Mariner 10* mission to Mercury and the Apollo missions to the moon. We may actually see methane ice blocks and the glint of sunlight reflected off the nitrogen sea; we may see the outlines of giant meteorite craters largely obliterated by the intense seasonal variations in Triton's weather. Moreover, *Voyager 2*'s infrared and ultraviolet spectrometers will determine the atmospheric abundances of methane, nitrogen and other gases, thereby testing crude models of the satellite's surface and atmosphere.

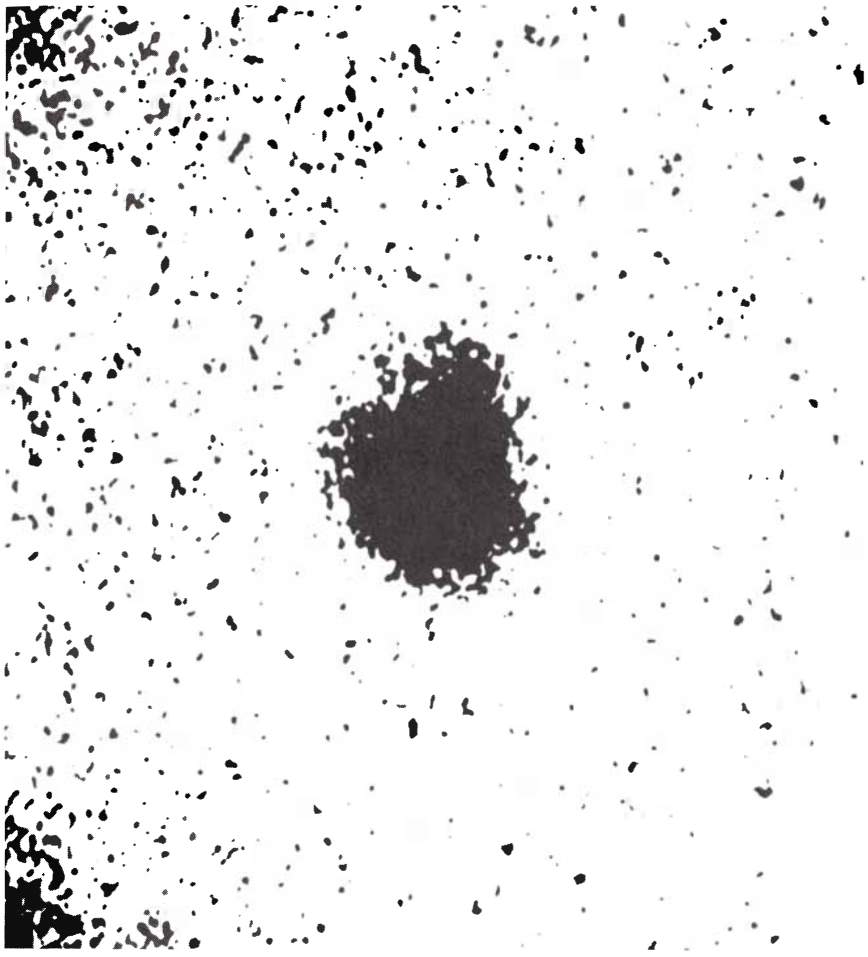
At the same time the spacecraft will add substance to studies of Triton's internal properties by providing the first reliable measurements of its mass and diameter. Recent radiometric measurements made by us and by others put the satellite's diameter at about 3,500 kilometers. This estimate could be wrong by as much as 40 percent, however, because it is based on the simplifying assumption that the solar energy absorbed by Triton is immediately reradiated as heat. A large, volatile ocean and an atmosphere would almost certainly store and redistribute a significant amount of heat, thereby violating the assumption.

Attempts to measure Triton's mass from the earth are hampered by the fact that Nereid, the only other satellite of Neptune, is too far away for its orbit to be measurably affected by Triton's gravitational field. Consequently workers have been forced to estimate Triton's mass from the almost imperceptible wobble it induces in the orbit of Neptune. Combined with the recent diameter measurements, such mass estimates yield a value of about eight grams per cubic centimeter for Triton's mean density, higher than that of any of the planets. This implausible result would imply that Triton is made mostly of iron or other metals, which is certainly not what one expects in the outer solar system. Precise mass and diameter figures derived from *Voyager 2* data will resolve the issue.

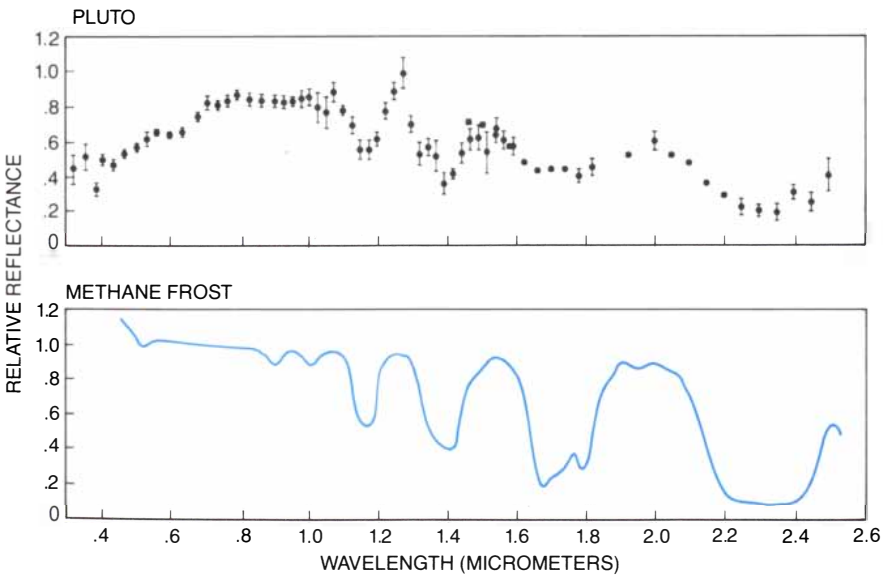
Finally, the spacecraft may settle the question of whether the Neptunian system has a third satellite or perhaps a ring. While monitoring the occultation of a distant star by Neptune in 1981, Harold J. Reitsema and his colleagues at the University of Arizona noted that the star dimmed several minutes before Neptune was to pass in front of it. The workers have suggested the star was occulted by a previously unknown satellite, one lying so close to Neptune that its image is lost in the planet's glare. In contrast William B. Hubbard, also of Arizona, has argued that Reitsema's occultation observations and those of others are best explained by positing a discontinuous ring around Neptune. If Hubbard is proved



ORBIT OF NEREID is a highly eccentric ellipse. At its closest approach (periapsis) the moon is 1.3 million kilometers from Neptune; at apoapsis it is 9.7 million kilometers away. Nereid's unusual orbit suggests it may have been captured by Neptune's gravity.



PLUTO'S MOON Charon is barely discernible as a bulge on the planet's image; here it is seen at the upper right. Charon is some 20,000 kilometers from Pluto, and it is probably between one-third and one-half the planet's size. The image was made by James W. Christy and Robert S. Harrington of the U.S. Naval Observatory, who discovered Charon in 1978.



PLUTO'S SPECTRUM in the near-infrared range shows at least four absorption bands characteristic of methane frost. The shape and strength of the bands, however, do not match those of the methane spectrum, indicating that Pluto has another surface component. A further discrepancy is found in the visible range (wavelengths between .3 and .7 micrometer), where Pluto has a reddish color because it reflects more red light than blue. The red component may be complex hydrocarbons derived from the photodissociation of methane.

correct, *Voyager 2*'s course may have to be changed: its present trajectory would take it right through the ring.

The spacecraft will not fly by Pluto, nor will any other probe in the foreseeable future. It is nonetheless appropriate to discuss Pluto here. Although it is a planet in a sun-centered orbit, accompanied by a satellite of its own, Pluto is comparable in size to Triton; at present it is also about the same distance as Triton from the sun, because its elliptical orbit has taken it temporarily inside the orbit of Neptune. The lessons learned from the Voyager encounters with Neptune and Uranus will expand our knowledge of Pluto as well as of Charon, its moon.

For nearly half a century after Pluto's discovery in 1930, the only thing known about it, other than its orbital parameters, was that it varied in brightness with a period of 6.4 days; the variation is attributable to the planet's rotation. Then in 1976 one of us (Cruikshank), collaborating with David Morrison and Carl B. Pilcher of the University of Hawaii, found spectrophotometric evidence that much of Pluto's surface is covered by methane ice. The high reflectivity implied by this finding suggested the planet was somewhat smaller and less massive than had been thought.

When Charon was discovered in 1978 by James W. Christy and Robert S. Harrington of the U.S. Naval Observatory, it became possible to calculate Pluto's mass. The results supported the hypothesis that Pluto is no larger than the earth's moon and is composed predominantly of volatiles. Charon is difficult to observe separately from Pluto—it is a mere appendage on the planet's image—and so its size and mass have not yet been calculated. It is probably between one-third and one-fifth the size of Pluto. Its mass must be a substantial fraction of the planet's, because its orbital period is the same as the planet's rotational period: the two bodies always keep the same face toward each other. Such synchronism can only arise when the satellite's mass is more than about 5 percent of the planet's mass. (In comparison, the mass of our own moon is less than 2 percent of the earth's; it keeps the same face toward the earth, but the earth's rotation is not constrained.)

Charon has a circular orbit oriented in a north-south direction, nearly perpendicular to the orbital plane of Pluto. This suggests that Pluto's equatorial plane is also roughly perpendicular to its orbital plane, a property once thought to be unique to Uranus. Earlier this year Edward F. Tedesco of the Jet Propulsion Laboratory, Richard P.

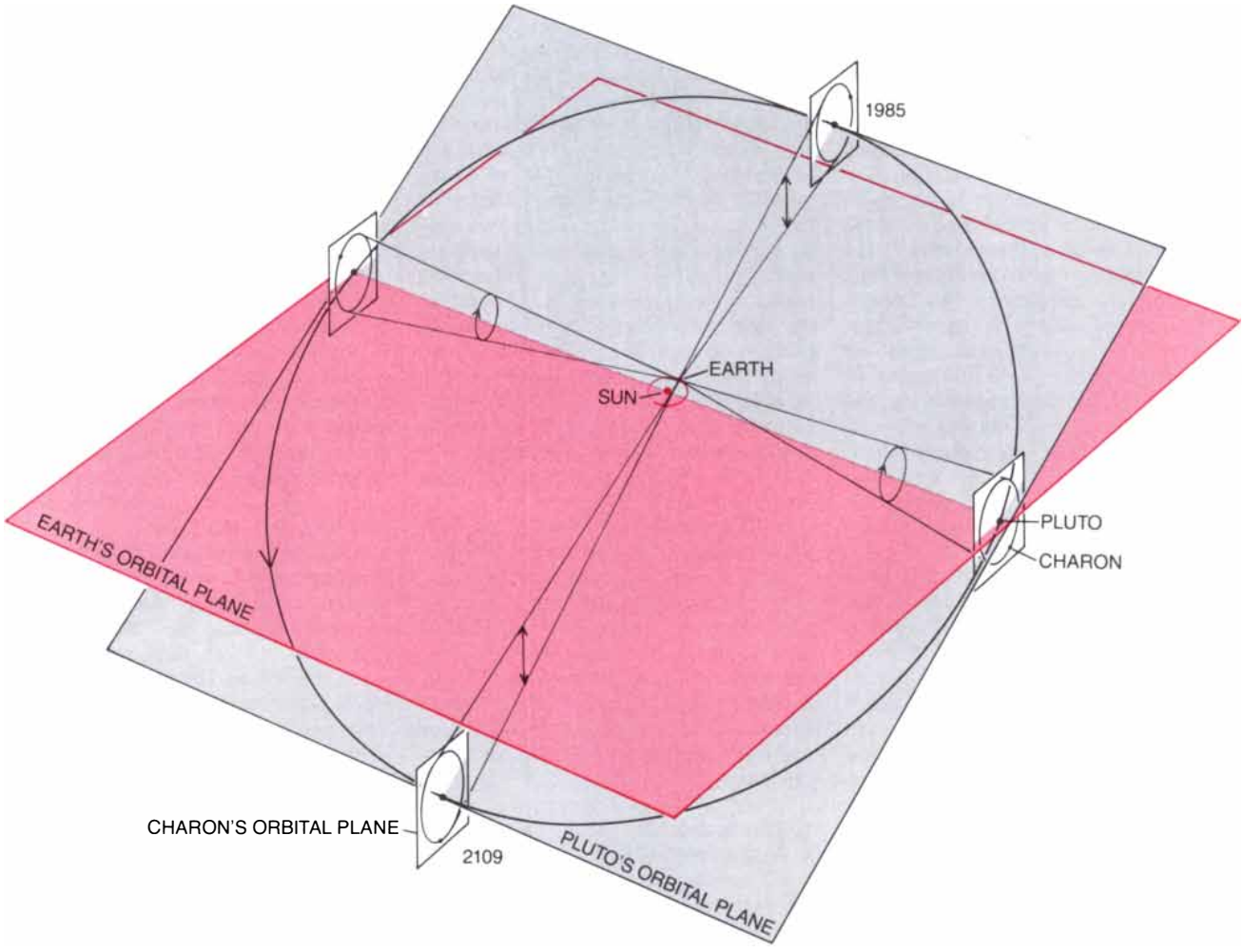
Binzel of the University of Texas at Austin and David J. Tholen of the University of Hawaii detected eclipses of Pluto (and occultations of Charon) for the first time. Eclipses can be observed only when the orbital plane of Charon is pointing toward the earth, which happens twice in Pluto's 248-year orbit. Observations of the current series of eclipses during the next few years should produce much better measurements of the dimensions of both bodies. It may be possible to determine whether they are similar in reflectivity and thus in surface composition.

The surface of Pluto is dominated by methane ice, but it must have another component as well. The strength of the methane absorption bands varies in different regions. Apparently the variation is associated with the variation in the planet's total brightness; bright regions seem to be those with the greatest coverage of methane ice. In dark regions a material of unknown composition covers the methane. On Pluto too the dark material may result from the photolysis of methane.

The presence of methane ice on Pluto implies, as it does on Triton, the existence of a tenuous methane atmosphere; Uwe Fink and Marc W. Buie of the University of Arizona have verified the presence of methane gas by means of spectral measurements of high resolution. Might Pluto also have a nitrogen sea? From the current data it is impossible to say, because even on high-resolution spectra the strong methane bands mask the region where a nitrogen band would appear. Nevertheless, in size and surface composition Pluto and Triton seem tantalizingly similar, and unlike the smaller, water-dominated Uranian moons.

The relation between Pluto and the Neptunian system is one of the enduring puzzles of the solar system. For a long time Pluto was considered an escaped satellite of Neptune, even though a mechanism that might have accounted for its transfer to a planetary orbit was never established. In fact, William B. McKinnon of Washington University has recently calculated that Pluto and Charon could not have survived an escape from Neptune without disintegrating. On the other hand, Triton's peculiar retrograde orbit suggests it may not be native to the Neptunian system. As McKinnon notes, the simplest hypothesis is that both Pluto and Triton condensed as planets in about the same region of the outer solar system—hence their chemical similarities—and that Triton was later captured by Neptune's relatively powerful gravitational field.

Of course, the simplest hypothesis may not be correct. The mysterious provenance of Pluto and Triton is another question that *Voyager 2,* together with the inventive use of telescopes on the earth, is likely to illuminate.



ECLIPSES OF PLUTO by Charon can be observed from the earth during two brief phases of Pluto's 248-year orbit; such observations were made for the first time earlier this year. Charon's orbit is nearly perpendicular to Pluto's orbital plane (*gray*) and to that of the earth (*color*). Near the points where Pluto is closest to and farthest from the sun, Charon's orbital plane is pointed toward the sun and the earth. Eclipses occur when the satellite passes in front of Pluto, occultations when it passes behind the planet. Both events produce a slight but measurable drop in the system's total brightness. The current series of events will continue for several years.

The Fundamental Physical Limits of Computation

What constraints govern the physical process of computing? Is a minimum amount of energy required, for example, per logic step? There seems to be no minimum, but some other questions are open

by Charles H. Bennett and Rolf Landauer

A computation, whether it is performed by electronic machinery, on an abacus or in a biological system such as the brain, is a physical process. It is subject to the same questions that apply to other physical processes: How much energy must be expended to perform a particular computation? How long must it take? How large must the computing device be? In other words, what are the physical limits of the process of computation?

So far it has been easier to ask these questions than to answer them. To the extent that we have found limits, they are terribly far away from the real limits of modern technology. We cannot profess, therefore, to be guiding the technologist or the engineer. What we are doing is really more fundamental. We are looking for general laws that must govern all information processing, no matter how it is accomplished. Any limits we find must be based solely on fundamental physical principles, not on whatever technology we may currently be using.

There are precedents for this kind of fundamental examination. In the 1940's Claude E. Shannon of the Bell Telephone Laboratories found there are limits on the amount of information that can be transmitted through a noisy channel; these limits apply no matter how the message is encoded into a signal. Shannon's work represents the birth of modern information science. Earlier, in the mid- and late 19th century, physicists attempting to determine the fundamental limits on the efficiency of steam engines had created the science of thermodynamics. In about 1960 one of us (Landauer) and John Swanson at IBM began attempting to apply the same type of analysis to the process of computing. Since the mid-1970's a growing number of other workers at other institutions have entered this field.

In our analysis of the physical limits of computation we use the term "information" in the technical sense of information theory. In this sense information is destroyed whenever two previously distinct situations become indistinguishable. In physical systems without friction, information can never be destroyed; whenever information is destroyed, some amount of energy must be dissipated (converted into heat). As an example, imagine two easily distinguishable physical situations, such as a rubber ball held either one meter or two meters off the ground. If the ball is dropped, it will bounce. If there is no friction and the ball is perfectly elastic, an observer will always be able to tell what state the ball started out in (that is, what its initial height was) because a ball dropped from two meters will bounce higher than a ball dropped from one meter.

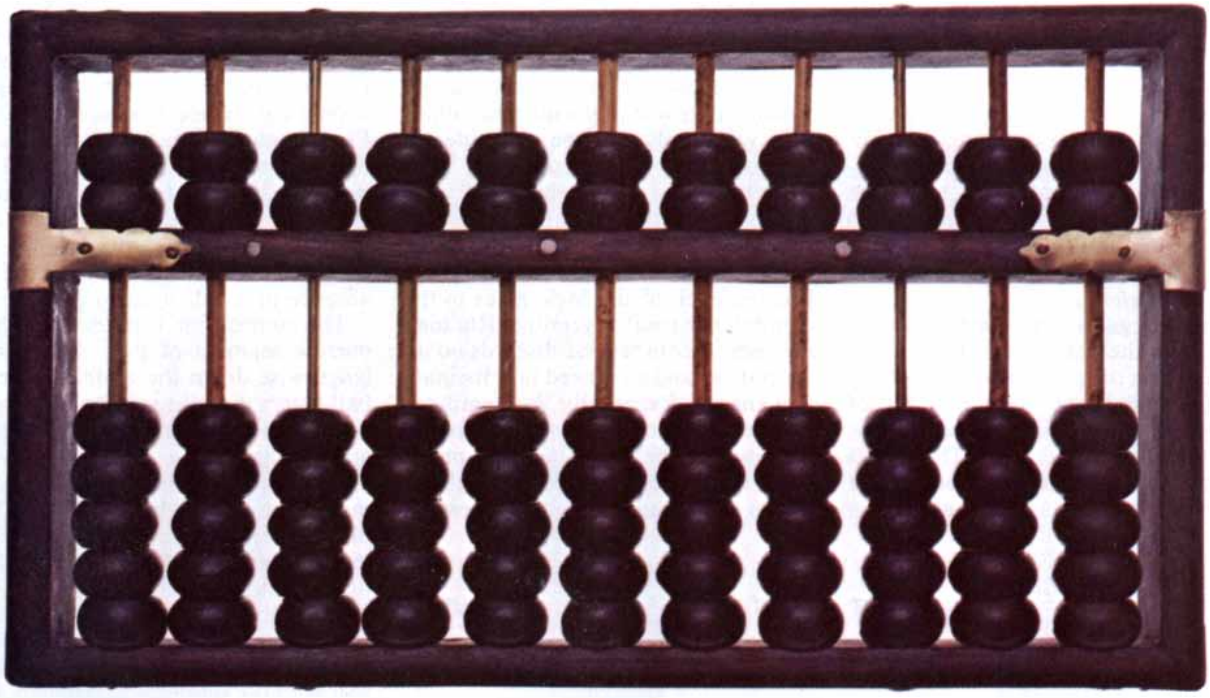
If there is friction, however, the ball will dissipate a small amount of energy with each bounce, until it eventually stops bouncing and comes to rest on the ground. It will then be impossible to determine what the ball's initial state was; a ball dropped from two meters will be identical with a ball dropped from one meter. Information will have been lost as a result of energy dissipation.

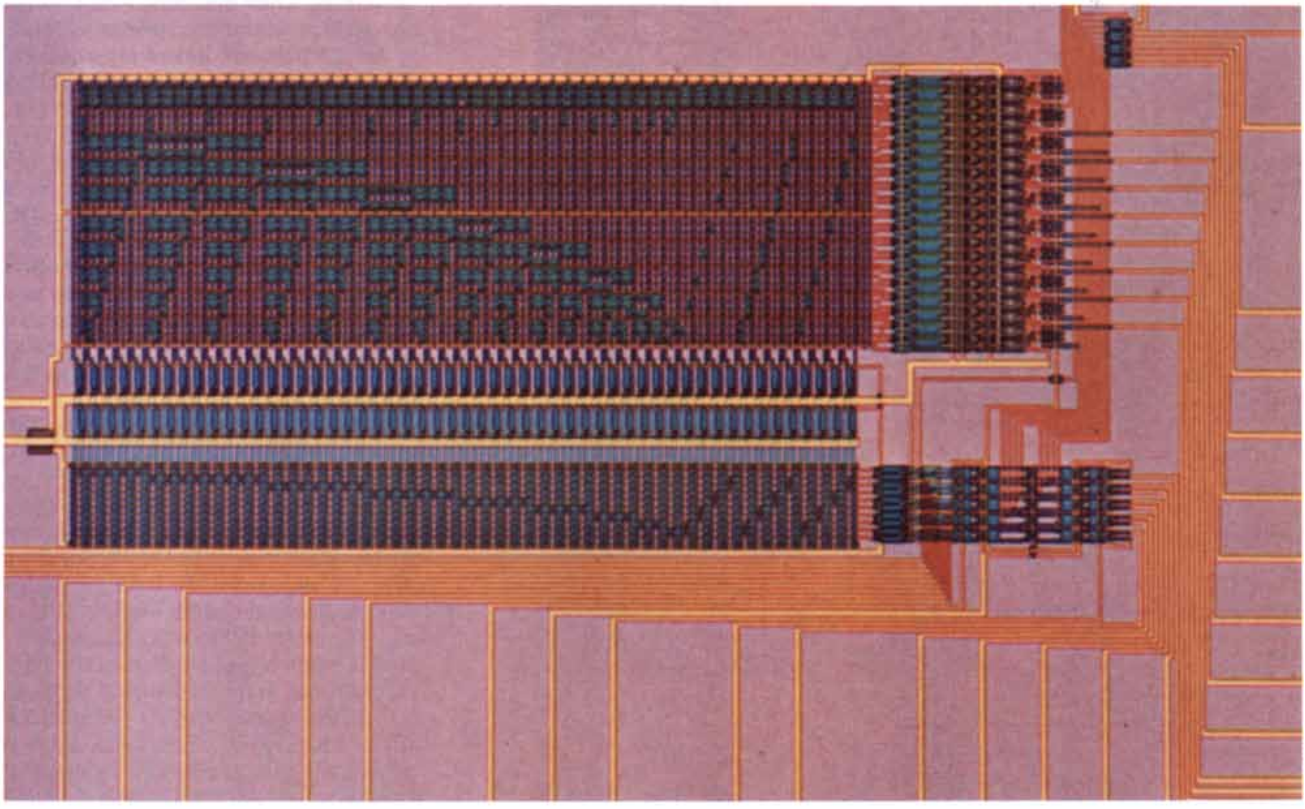
Here is another example of information destruction: the expression $2 + 2$ contains more information than the expression $=4$. If all we know is that we have added two numbers to yield 4, then we do not know whether we have added $1 + 3$, $2 + 2$, $0 + 4$ or some other pair of numbers. Since the output is implicit in the input, no computation ever generates information.

In fact, computation as it is currently carried out depends on many operations that destroy information. The so-called *and* gate is a device with two input lines, each of which may be set at 1 or 0, and one output, whose value depends on the value of the inputs. If both inputs are 1, the output will be 1. If one of the inputs is 0 or if both are 0, the output will also be 0. Any time the gate's output is a 0 we lose information, because we do not know which of three possible states the input lines were in (0 and 1, 1 and 0, or 0 and 0). In fact, any logic gate that has more input than output lines inevitably discards information, because we cannot deduce the input from the output. Whenever we use such a "logically irreversible" gate, we dissipate energy into the environment. Erasing a bit of memory, another operation that is frequently used in computing, is also fundamentally dissipative; when we erase a bit, we lose all information about that bit's previous state.

Are irreversible logic gates and erasures essential to computation? If they are, any computation we perform has to dissipate some minimum amount of energy.

As one of us (Bennett) showed in 1973, however, they are not essential. This conclusion has since been demonstrated in several models; the easiest of these to describe are based on so-called reversible logic elements such as the Fredkin gate, named for Edward Fredkin of the Massachusetts Institute of Technology. The Fredkin gate has three input lines and three outputs. The input on one line, which is called the control channel, is fed unchanged through the gate. If the control channel is set at 0, the input on the other two lines also passes through unchanged. If the control line is a 1, however, the





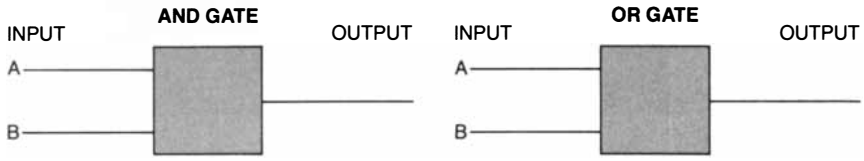
CONVENTIONAL COMPUTING DEVICES, the abacus and the logic chip, both dissipate energy when they are operated. The "logic gates" central to the design of a chip expend energy because they discard information. A chip consumes energy for a less fundamental reason as well: it employs circuits that draw power even when they merely hold information and do not process it. The abacus is dissipative because of friction between its beads and rods. It could not be built of frictionless components: if there were no static friction, the beads' positions would change under the influence of random thermal motion. Static friction exerts a certain minimum force no matter what the beads' velocity, and so there is some minium energy that the abacus requires no matter how slowly it is operated.

outputs of the other two lines are switched: the input of one line becomes the output of the other and vice versa. The Fredkin gate does not discard any information; the input can always be deduced from the output.

Fredkin has shown that any logic device required in a computer can be implemented by an appropriate arrangement of Fredkin gates. To make the computation work, certain input lines of some of the Fredkin gates must be preset at particular values [*see lower illustration below*].

Fredkin gates have more output lines than the gates they are made to simulate. In the process of computing, what seem to be "garbage bits," bits of information that have no apparent use, are therefore generated. These bits must somehow be cleared out of the computer if we are to use it again, but if we erase them, it will cost us all the energy dissipation we have been trying to avoid.

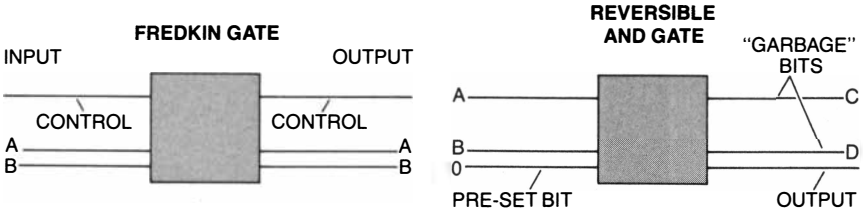
Actually these bits have a most important use. Once we have copied down the result of our computation, which will reside in the normal output bits, we simply run the computer in reverse. That is, we enter the "garbage bits" and output bits that were produced by the computer's normal operation as "input" into the "back end" of the computer. This is possible because each of the logic gates in the computer is itself reversible. Running the computer in reverse discards no information, and so it need not dissipate any energy. Eventually the computer will be left exactly as it was before the computation began. Hence it is possible to complete a "computing cycle"—to run a computer and then to return it to its original state—without dissipating any energy.



A	B	OUTPUT
1	1	1
1	0	0
0	1	0
0	0	0

A	B	OUTPUT
1	1	1
1	0	1
0	1	1
0	0	0

CONVENTIONAL LOGIC GATES dissipate energy because they discard information. For example, if the output of an *and* gate is 0, there is no way to deduce what the input was.



INPUT			OUTPUT		
CONTROL	A	B	CONTROL	A	B
1	1	1	1	1	1
1	1	0	1	0	1
1	0	1	1	1	0
1	0	0	1	0	0
0	1	1	0	1	1
0	1	0	0	1	0
0	0	1	0	0	1
0	0	0	0	0	0

			"GARBAGE"	
A	B	OUTPUT	C	D
1	1	1	1	0
1	0	0	1	0
0	1	0	0	1
0	0	0	0	0

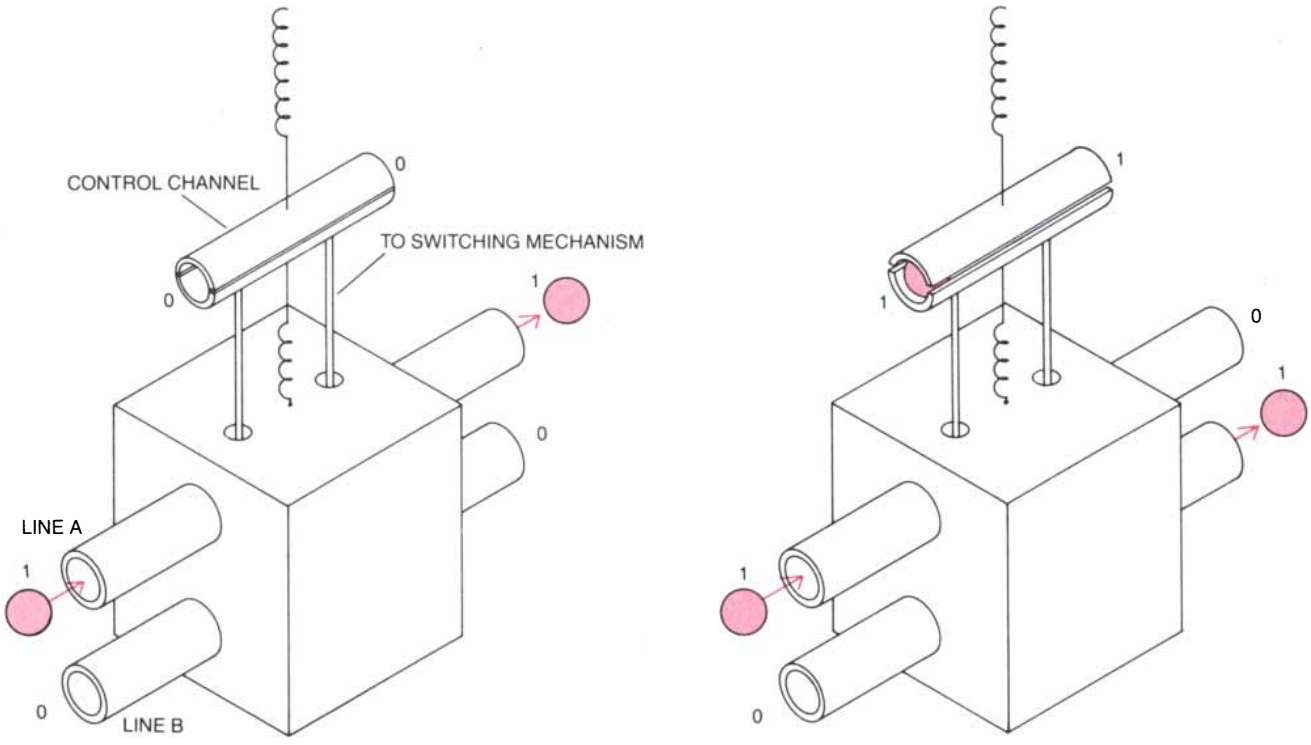
FREDKIN REVERSIBLE LOGIC GATE need not dissipate energy; the input can always be deduced from the output. The gate has a "control" line, the value of which is not changed by the gate. If the bit on the control line is a 0, the values of the other two lines are also untouched; if it is a *1*, however, the input of line *A* becomes the output of line *B* and vice versa. Reversible gates can be arranged to implement any function performed by an irreversible gate. To implement the *and* operation (*right*) one input is preset to equal 0, and two output bits, called garbage bits, are temporarily ignored. When the computation is complete, these bits are used to operate the gate in reverse, returning the computer to its original state.

So far we have discussed a set of logic operations, not a physical device. It is not hard, however, to imagine a physical device that operates as a Fredkin gate. In this device the information channels are represented by pipes. A bit of information is represented by the presence or absence of a ball in a particular section of pipe; the presence of a ball signifies a 1 and the absence of a ball signifies a 0.

The control line is represented by a narrow segment of pipe that is split lengthwise down the middle. When a ball enters the split segment of pipe, it pushes the two halves of the pipe apart, actuating a switching device. The switching device channels any input balls that may be in the other two pipes: when a ball is present in the control line, any ball that enters an input pipe is automatically redirected to the other pipe. To ensure that the switch is closed when no control ball is present, there are springs that hold the two halves of the split pipe together. A ball entering the split pipe must expend energy when it compresses the springs, but this energy is not lost; it can be recovered when the control ball leaves the split pipe and the springs expand.

All the balls are linked together and pushed forward by one mechanism, so that they move in synchrony; otherwise we could not ensure that the various input and controlling balls would arrive at a logic gate together. In a sense the forward progress of the computation is really motion along a single degree of freedom, like the motion of two wheels rigidly attached to one axle. Once the computation is done we push all the balls backward, undoing all the operations and returning the computer to its initial state.

If the entire assembly is immersed in an ideal viscous fluid, then the frictional forces that act on the balls will be proportional to their velocity; there will be no static friction. The frictional force will therefore be very weak if we are content to move the balls slowly. In any mechanical system the energy that must be expended to work against friction is equal to the product of the frictional force and the distance through which the system travels. (Hence the faster a swimmer travels between two points, the more energy he or she will expend, although the distance traveled is the same whether the swimmer is fast or slow.) If we move the balls through the Fredkin gates at a low speed, then the energy expended (the product of force and distance) will be very small, because the frictional force depends directly on the balls' speed. In



IDEALIZED PHYSICAL REALIZATION of a Fredkin gate substitutes pipes for wires and the presence or absence of a ball for a 1 or 0. A narrow, split segment of pipe represents the control channel. When a ball passes through it, the pipe spreads apart, operating a switching mechanism; the mechanism in turn switches any input ball from line *A* to line *B* and vice versa. A pair of springs keeps the control channel closed when no ball is in it. This gate does not need static friction in order to operate; it could be immersed in a viscous fluid, and the frictional forces could be made to depend only on the balls' velocity. Then the energy dissipation could be as small as the user wished: to lower the amount of energy dissipated, it would only be necessary to drive the balls through the device more slowly.

fact, we can expend as little energy as we wish, simply by taking a long time to carry out the operation. There is thus no minimum amount of energy that must be expended in order to perform any given computation.

The energy lost to friction in this model will be very small if the machine is operated very slowly. Is it possible to design a more idealized machine that could compute without any friction? Or is friction essential to the computing process? Fredkin, together with Tommaso Toffoli and others at M.I.T., has shown that it is not.

They demonstrated that it is possible to do computation by firing ideal, frictionless billiard balls at one another. In the billiard-ball model perfect reflecting "mirrors," surfaces that redirect the balls' motion, are arranged in such a way that the movement of the balls across a table emulates the movement of bits of information through logic gates [*see illustration on next page*]. As before, the presence of a ball in a particular part of the computer signifies a 1, whereas the absence of a ball signifies a 0. If two balls arrive simultaneously at a logic gate, they will collide and their paths will change; their new paths represent the output of the gate. Fredkin, Toffoli and others have described arrangements of mirrors that correspond to different types of logic gate, and they have shown that billiard-ball models can be built to simulate any logic element that is necessary for computing.

To start the computation we fire a billiard ball into the computer wherever we wish to input a 1. The balls must enter the machine simultaneously. Since they are perfectly elastic, they do not lose energy when they collide; they will emerge from the computer with the same amount of kinetic energy we gave them at the beginning.

In operation a billiard-ball computer produces "garbage bits," just as a computer built of Fredkin gates does. After the computer has reached an answer we reflect the billiard balls back into it, undoing the computation. They will come out of the machine exactly where we sent them in, and at the same speed. The mechanism that launched them into the computer can then be used to absorb their kinetic energy. Once again we will have performed a computation and returned the computer to its initial state without dissipating energy.

The billiard-ball computer has one major flaw: it is extremely sensitive to slight errors. If a ball is aimed slightly incorrectly or if a mirror is tilted at a slightly wrong angle, the balls' trajectories will go astray. One or more balls will deviate from their intended paths, and in due course errors will combine to invalidate the entire computation. Even if perfectly elastic and frictionless billiard balls could be manufactured, the small amount of random thermal motion in the molecules they are made of would be enough to cause errors after a few dozen collisions.

Of course we could install some kind of corrective device that would return any errant billiard ball to its desired path, but then we would be obliterating information about the ball's earlier history. For example, we might be discarding information about the extent to which a mirror is tilted incorrectly. Discarding information, even to correct an error, can be done only in a system in which there is friction and loss of energy. Any correctional device must therefore dissipate some energy.

Many of the difficulties inherent in the billiard-ball computer can be made less extreme if microscopic or submicroscopic particles, such as electrons, are used in place of billiard balls. As Wojciech H. Zurek, who is now at the Los Alamos National Laboratory, has pointed out, quantum laws, which can restrict particles to a few states of motion, could eliminate the possibility

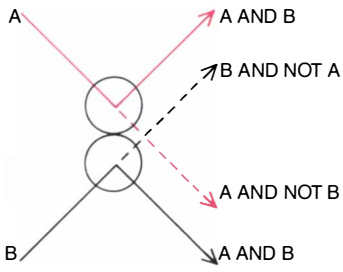
that a particle might go astray by a small amount.

Although the discussion so far has been based primarily on classical dynamics, several investigators have proposed other reversible computers that are based on quantum-mechanical principles. Such computers, first proposed by Paul Benioff of the Argonne National Laboratory and refined by others, notably Richard P. Feynman of the California Institute of Technology, have so far been described only in the most abstract terms. Essentially the particles in these computers would be arranged so that the quantum-mechanical rules governing their interaction would be precisely analogous to the rules describing the predicted outputs of various reversible logic gates. For example, suppose a particle's spin can have only two possible values: up (corresponding to a binary 1) and down (corresponding to a 0). The interactions between particle spins can be prescribed in such a way that the value of one particle's spin changes depending on the spin of nearby particles; the spin of the particle would correspond to one of the outputs of a logic gate.

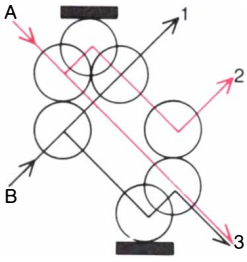
So far this discussion has concentrated on information processing. A computer must store information as well as process it. The interaction between storage and processing is best described in terms of a device called a Turing machine, for Alan M. Turing, who first proposed such a machine in 1936. A Turing machine can perform any computation that can be performed by a modern computer. One of us (Bennett) has shown that it is possible to build a reversible Turing machine: a Turing machine that does not discard information and can therefore be run with as small an expenditure of energy as the user wishes.

A Turing machine has several components. There is a tape, divided into discrete frames or segments, each of which is marked with a 0 or a 1; these bits represent the input. A "read/write head" moves along the tape. The head has several functions. It can read one bit of the tape at a time, it can print one bit onto the tape and it can shift its position by one segment to the left or right. In order to remember from one cycle to the next what it is doing, the head mechanism has a number of distinct "states"; each state constitutes a slightly different configuration of the head's internal parts.

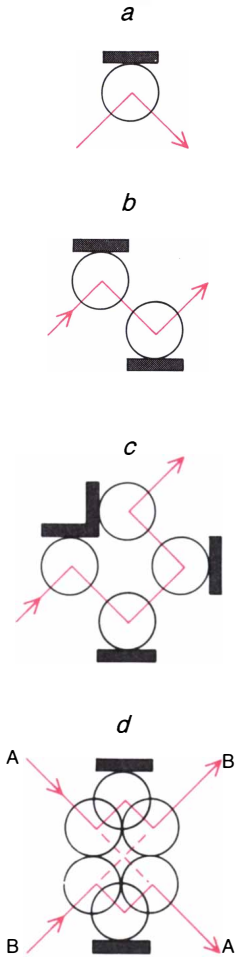
In each cycle the head reads the bit on the segment it currently occupies;

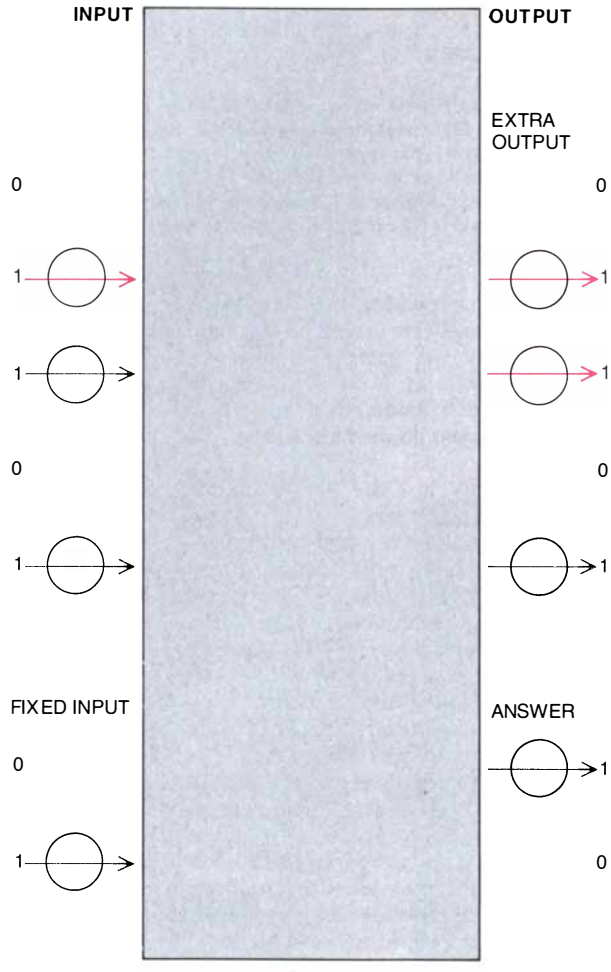


INPUT		OUTPUT			
A	B	1	2	3	4
1	1	1	0	0	1
1	0	0	0	1	0
0	1	0	1	0	0
0	0	0	0	0	0



INPUT		OUTPUT		
A	B	1	2	3
1	1	0	1	1
1	0	0	0	1
0	1	1	0	0
0	0	0	0	0



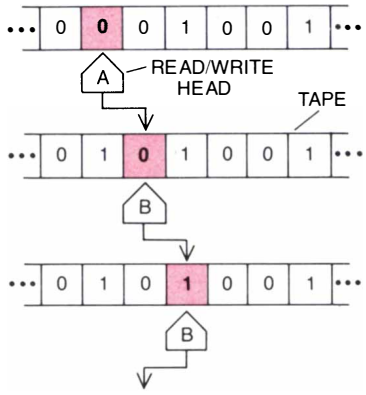


BILLIARD-BALL COMPUTER employs the movement of billiard balls on a table to simulate the movement of bits through logic gates. In billiard-ball logic gates (*left*) the balls' paths are redirected by collisions with one another or with reflecting "mirrors." In addition to their role in gates, mirrors can deflect a ball's path (*a*), shift the path sideways (*b*), delay the ball's motion without changing its final direction or position (*c*) or allow two lines to cross (*d*). It is possible to arrange mirrors so that the resulting "computer" implements the function of any logic chip. For example, a billiard-ball computer could be made to test whether a number is prime. One such computer (*right*) accepts as input any five-bit number (in this case 01101, or 13) and the fixed input sequence 01. Like a Fredkin gate, a billiard-ball computer typically returns more output bits than its user needs. In the case shown, the computer returns the original input number itself (which is the "extra" output), and an "answer" sequence: 10 if the input number is prime and 01 if it is composite.

then it prints a new bit onto the tape, changes its internal state and moves one segment to the left or right. The bit it prints, the state it changes into and the direction in which it moves are determined by a fixed set of transition rules. Each rule specifies a particular set of actions. Which rule the machine follows is determined by the state of the head and the value of the bit that it reads from the tape. For example, one rule might be: "If the head is in state *A* and is sitting on a segment of tape that is printed with a 0, it should change that bit to a 1, change its state to state *B* and move one segment to the right." It may happen that the transition rule instructs the machine not to change its internal state, not to print a new bit onto the tape or to halt its operation. Not all Turing machines are reversible, but a reversible Turing machine can be built to perform any possible computation.

The reversible Turing-machine models have an advantage over such machines as the frictionless billiard-ball computer. In the billiard-ball computer random thermal motion causes unavoidable errors. Reversible Turing-machine models actually exploit random thermal motion: they are constructed in such a way that thermal motion itself, with the assistance of a very weak driving force, moves the machine from one state to the next. The progress of the computation resembles the motion of an ion (a charged particle) suspended in a solution that is held in a weak electric field. The ion's motion, as seen over a short period of time, appears to be random; it is nearly as likely to move in one direction as in another. The applied force of the electric field, however, gives the net motion a preferred direction: the ion is a little likelier to move in one direction than in the other.

It may at first seem inconceivable that a purposeful sequence of operations, such as a computation, could be achieved in an apparatus whose direction of motion at any one time is nearly random. This style of operation is quite common, however, in the microscopic world of chemical reactions. There the trial-and-error action of Brownian motion, or random thermal motion, suffices to bring reactant molecules into contact, to orient and bend them into the specific conformation required for them to react, and to separate the product molecules after the reaction. All chemical reactions are in principle reversible: the same Brownian motion that accomplishes the forward reaction sometimes brings product molecules together and pushes them backward through the transition.



TRANSITION RULES

HEAD STATE	BIT READ	CHANGE BIT TO	CHANGE STATE TO	MOVE TO
A	1	0	A	LEFT
A	0	1	B	RIGHT
B	1	1	A	LEFT
B	0	0	B	RIGHT

TURING MACHINE can be constructed in such a way that it can perform any computation a computer can. An infinitely long tape is divided into discrete segments, each of which bears either a 0 or a 1. A "read/write head," which can be in any of several internal states (here there are only two states, *A* and *B*), moves along the tape. Each cycle begins as the head reads one bit from a segment of the tape. Then, in accordance with a fixed set of transition rules, it writes a bit onto that segment, changes its own internal state and moves one segment to the left or right. This Turing machine, because it has only two head states, can do only trivial computations; more complicated machines with more head states are capable of simulating any computer, even one much more complicated than themselves. To do so they store a representation of the larger machine's complete logical state on the unlimited tape and break down each complex cycle into a large number of simple steps. The machine shown is logically reversible: it is always possible to deduce the machine's previous configuration. Other Turing machines, with different transition rules, are not reversible.

In a state of equilibrium a backward reaction is just as likely to occur as a forward one.

In order to keep a reaction moving in the forward direction, we must supply reactant molecules and remove product molecules; in effect, we must provide a small driving force. When the driving force is very small, the reaction will take nearly as many backward steps as forward ones, but on the average it will move forward. In order to provide the driving force we must expend energy, but as in our ball-and-pipe realization of the Fredkin gate the total amount of energy can be as small as we wish; if we are willing to allow a long time for an operation, there is no minimum amount of energy that must be expended. The reason is that the total energy dissipated depends on the number of forward steps divided by the number of backward steps. (It is actually proportional to the logarithm of this ratio, but as the ratio increases or decreases so does its logarithm.) The slower the reaction moves forward, the smaller the ratio will be. (The analogy of the faster and slower swimmers is valid once again: it requires less total energy to go the same net number of reaction steps forward if the reaction moves slowly.)

We can see how a Brownian Turing machine might work by examining a Brownian tape-copying machine that already exists in nature: RNA polymerase, the enzyme that helps to construct RNA copies of the DNA constituting a gene. A single strand of DNA is much like the tape of a Turing machine. At each position along the strand there is one of four "bases": adenine, guanine, cytosine or thymine (abbreviated *A, G, C* and *T*). RNA is a similar chainlike molecule whose four bases, adenine, guanine, cytosine and uracil (*A, G, C* and *U*) bind to "complementary" DNA bases.

The RNA polymerase catalyzes this pairing reaction. The DNA helix is normally surrounded by a solution containing a large number of nucleoside triphosphate molecules, each consisting of an RNA base linked to a sugar and a tail of three phosphate groups. The RNA-polymerase enzyme selects from the solution a single RNA base that is complementary to the base about to be copied on the DNA strand. It then attaches the new base to the end of the growing RNA strand and releases two of the phosphates into the surrounding solution as a free pyrophosphate ion. Then the enzyme shifts forward one notch along the strand of DNA in preparation for attaching the next RNA base. The result is a strand of RNA that is complementary to the template strand of DNA. Without RNA polymerase this set of reactions would occur very slowly, and there would be little guarantee that the RNA and DNA molecules would be complementary.

The reactions are reversible: sometimes the enzyme takes up a free pyrophosphate ion, combines it with the last base on the RNA strand and re-

leases the resulting nucleoside triphosphate molecule into the surrounding solution, meanwhile backing up one notch along the DNA strand. At equilibrium, forward and backward steps would occur with equal frequency; normally other metabolic processes drive the reaction forward by removing pyrophosphate and supplying the four kinds of nucleoside triphosphate. In the laboratory the speed with which RNA polymerase acts can be varied by adjusting the concentrations of the reactants (as Judith Levin and Michael J. Chamberlin of the University of California at Berkeley have shown). As the concentrations are brought closer to equilibrium the enzyme works more slowly and dissipates less energy to copy a given section of DNA, because the ratio of forward to backward steps is smaller.

Although RNA polymerase merely copies information without processing it, it is relatively easy to imagine how a hypothetical chemical Turing machine might work. The tape is a single long backbone molecule to which two types of base, representing the binary 0 and 1, attach at periodic sites. A small additional molecule is attached to the 0 or 1 group at one site along the chain. The position of this additional molecule represents the position of the Turing machine's head. There are several different types of "head molecule," each type representing a different machine state.

The machine's transition rules are represented by enzymes. Each enzyme is capable of catalyzing one particular reaction. The way these enzymes work is best demonstrated by an example.

Suppose the head molecule is type *A* (indicating that the machine is in state *A*) and is attached to a 0 base. Also suppose the following transition rule applies: "When the head is in state *A* and reads a 0, change the 0 to a 1, change state to *B* and move to the right." A molecule of the enzyme representing this rule has a site that fits a type-*A* head molecule attached to a 1 base. It also has one site that fits a 0 base and one site that fits a *B* head [*see illustration on opposite page*].

To achieve the transition, the enzyme molecule first approaches the tape molecule at a location just to the right of the base on which the *A* head resides. Then it detaches from the tape both the head molecule and the 0 base to which the head was attached, putting in their place a 1 base. Next it attaches a *B* head to the base that is to the right of the 1 base it has just added to the tape. At this point the transition is complete. The head's original site is changed from a 0 to a 1, the head molecule is now a type *B*, and it is attached to the base that is one notch to the right of the previous head position.

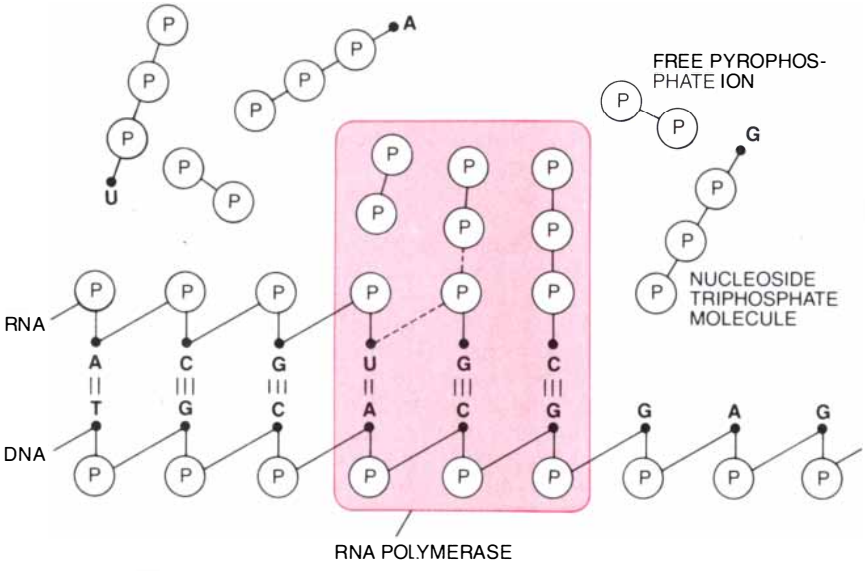
During the operation of a Brownian Turing machine the tape would have to be immersed in a solution containing many enzyme molecules, as well as extra 0's, 1's, *A*'s and *B*'s. To drive the reaction forward there would have to be some other reaction that cleaned the enzyme molecules of detached heads and bases. The concentrations of the reactants that clean the enzyme molecules represent the force that drives the Turing machine forward. Again we can expend as little energy as we wish simply by driving the machine forward very slowly.

The enzymatic Turing machine would not be error-free. Occasionally a reaction that is not catalyzed by any enzyme might occur; for example, a 0 base could spontaneously detach itself from the backbone molecule and a 1 base could be attached in its place. Similar errors do indeed occur during RNA synthesis.

In principle it would be possible to eliminate errors by building a Brownian Turing machine out of rigid, frictionless clockwork. The clockwork Turing machine involves less idealization than the billiard-ball computer but more than the enzymatic Turing machine. On the one hand, its parts need not be manufactured to perfect tolerances, as the billiard balls would have to be; the parts fit loosely together, and the machine can operate even in the presence of a large amount of thermal noise. Still, its parts must be perfectly rigid and free of static friction, properties not found in any macroscopic body.

Because the machine's parts fit together loosely, they are held in place not by friction but by grooves or notches in neighboring parts [*see illustration on page 56*]. Although each part of the machine is free to jiggle a little, like the pieces of a well-worn wood puzzle, the machine as a whole can only follow one "computational path." That is, the machine's parts interlock in such a way that at any time the machine can make only two kinds of large-scale motion: the motion corresponding to a forward computational step and that corresponding to a backward step.

The computer makes such transitions only as the accidental result of the random thermal motion of its parts biased by the weak external force. It is nearly as likely to proceed backward along the computational path, undoing the most recent transition, as it is to proceed forward. A small force, provided externally, drives the computation forward. This force can again be



RNA POLYMERASE, an enzyme, acts as a reversible tape-copying machine; it catalyzes the reaction that constructs RNA copies of segments of DNA. As the enzyme moves along a strand of DNA, it selects from the surrounding solution a nucleoside triphosphate molecule (an RNA base bound to a sugar and a "tail" of three phosphate groups) that is complementary to the DNA base about to be copied. It then attaches the new base to the end of the RNA strand and releases a free pyrophosphate ion consisting of two phosphates. The reaction is reversible: sometimes the enzyme takes up the last link of RNA, attaches it to a pyrophosphate ion and returns the resulting molecule to the solution, backing up a notch on the DNA strand. When the reaction is close to chemical equilibrium, the enzyme takes almost as many backward as forward steps and the total energy needed to copy any segment of DNA is very small. The reaction can be made less dissipative by being run more slowly; there is no minimum amount of energy that must be expended to copy a segment of DNA.

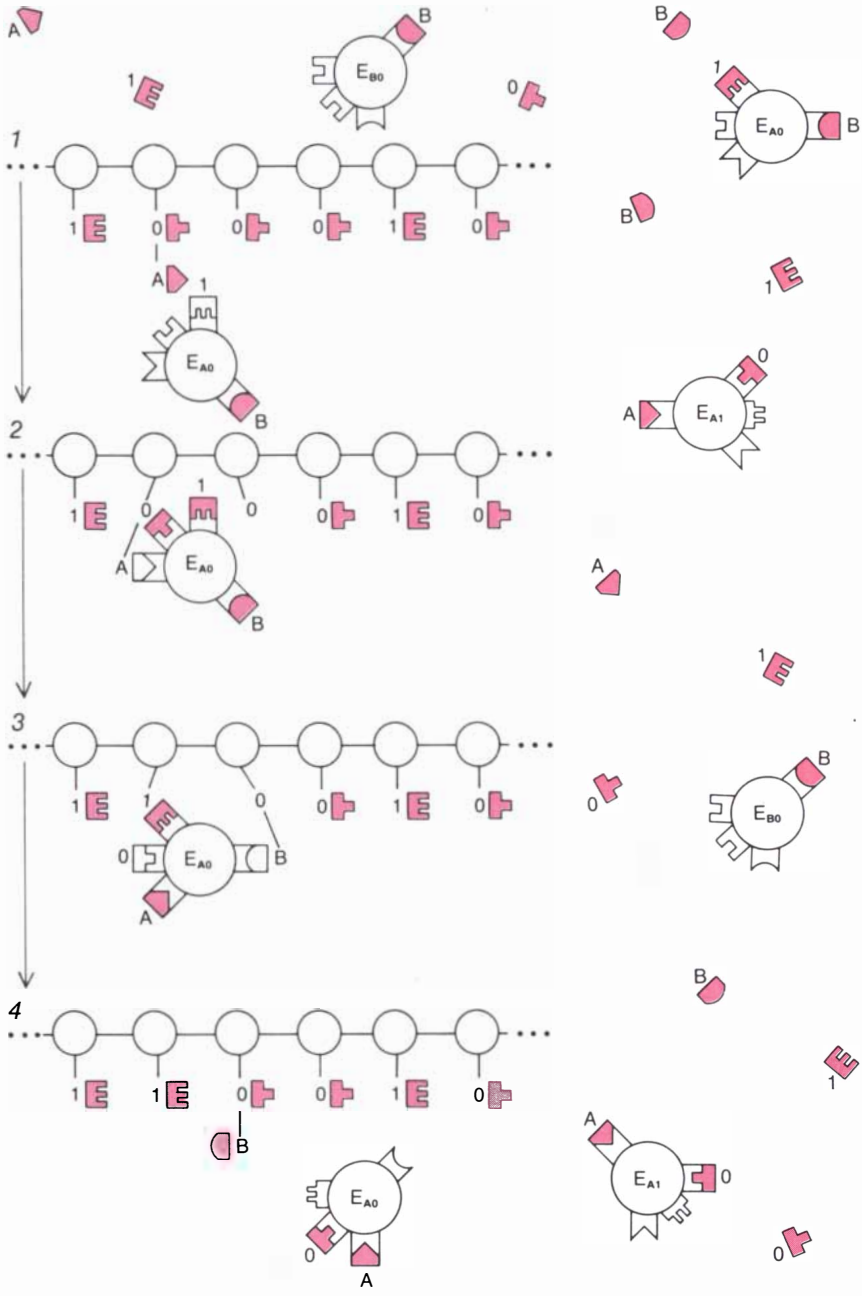
as small as we wish, and so there is no minimum amount of energy that must be expended in order to run a Brownian clockwork Turing machine.

According to classical thermodynamics, then, there is no minimum amount of energy required in order to perform a computation. Is the classical thermodynamical analysis in conflict with quantum theory? After all, the quantum-mechanical uncertainty principle states there must be an inverse relation between our uncertainty about how long a process takes and our uncertainty about how much energy the process involves. Some investigators have suggested that any switching process occurring in a short period of time must involve a minimum expenditure of energy.

In fact the uncertainty principle does not require any minimum energy expenditure for a fast switching event. The uncertainty principle would be applicable only if we attempted to measure the precise time at which the event took place. Even in quantum mechanics extremely fast events can take place without any loss of energy. Our confidence that quantum mechanics allows computing without any minimum expenditure is bolstered when we remember that Benioff and others have developed models of reversible quantum-mechanical computers, which dissipate no energy and obey the laws of quantum mechanics.

Thus the uncertainty principle does not seem to place a fundamental limit on the process of computation; neither does classical thermodynamics. Does this mean there are no physical limitations to computing? Far from it. The real limitations are associated with questions that are much harder to answer than those we have asked in this article. For example, do elementary logic operations require some minimum amount of time? What is the smallest possible gadgetry that could accomplish such operations? Because scales of size and time are connected by the velocity of light, it is likely that these two questions have related answers. We may not be able to find these answers, however, until it is determined whether or not there is some ultimate graininess in the universal scales of time and length.

At the other extreme, how large can we make a computer memory? How many particles in the universe can we bring and keep together for that purpose? The maximum possible size of a computer memory limits the precision with which we can calculate. It will limit, for example, the number of decimal places to which we can calculate pi. The inevitable deterioration proc-



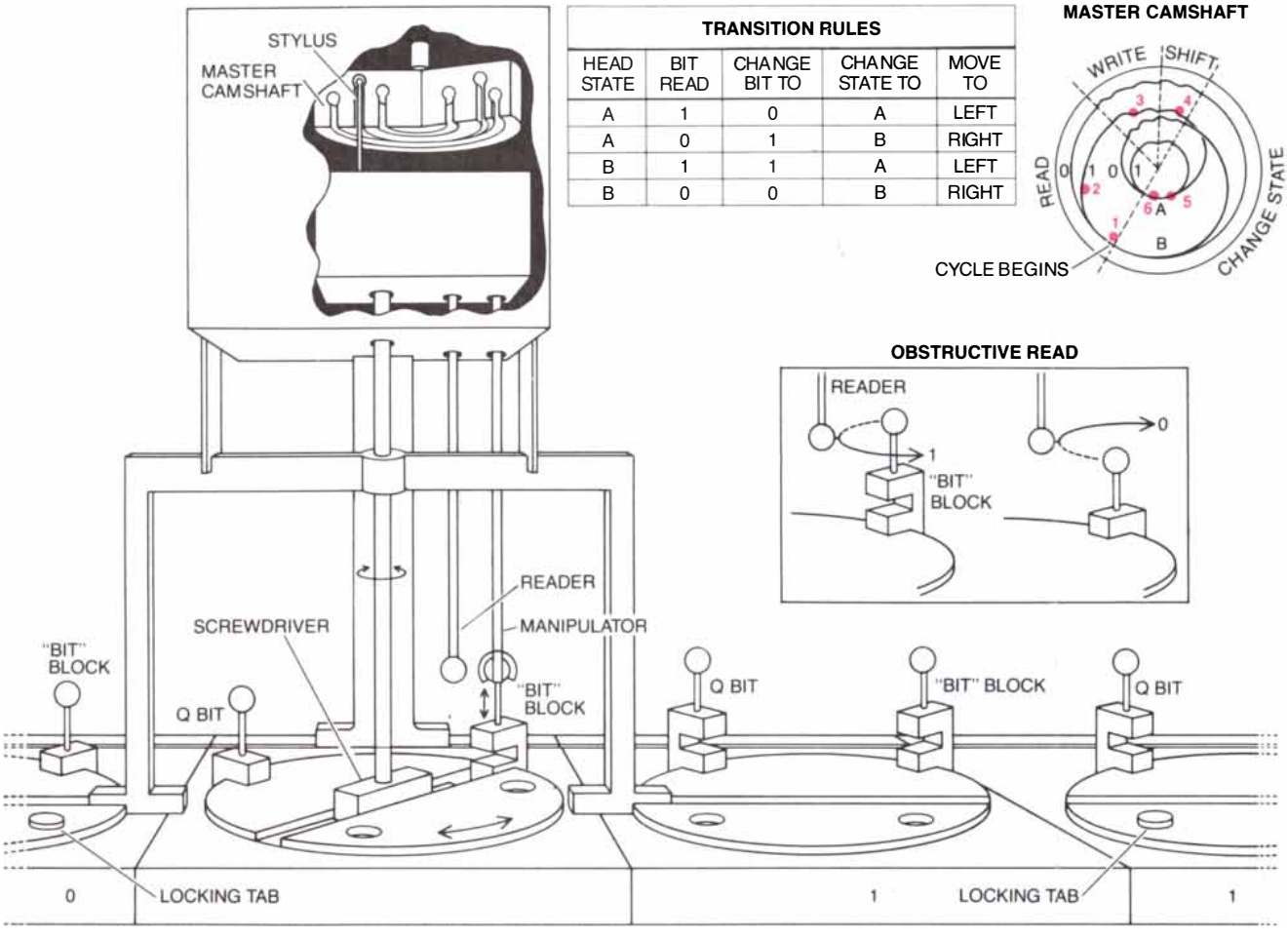
TRANSITION RULES

HEAD STATE	BIT READ	CHANGE BIT TO	CHANGE STATE TO	MOVE TO
A	1	0	A	LEFT
A	0	1	B	RIGHT
B	1	1	A	LEFT
B	0	0	B	RIGHT

HYPOTHETICAL ENZYMATIC TURING MACHINE could perform a computation with no minimum expenditure of energy. Molecules representing 0 and 1 bits are attached at periodic intervals to a backbone molecule. A small additional molecule, representing the Turing machine's head, is attached to the 0 or 1 group at one site on the chain (*1*). There are several types of head molecule, each type representing a different internal machine state. Transition rules are represented by enzymes. In each cycle an enzyme attaches itself to the head molecule and the bit molecule to which the head is attached (*2*); then it detaches them from the chain, putting in their place the appropriate bit molecule (*3*). As it does so it rotates, so that it attaches the appropriate head molecule to the bit that occupies the site one notch to the right or left of the bit it has just changed. Now the cycle is complete (*4*): the value of a bit has been changed, and the head has changed state and shifted its position. Each kind of enzyme is able to catalyze one such set of reactions. As in the case of RNA synthesis, these reactions can be made to dissipate an arbitrarily small amount of energy.

esses that occur in real computers pose another, perhaps related, question: Can deterioration, at least in principle, be reduced to any desired degree, or does it impose a limit on the maximum length of time we shall be able to devote to any one calculation? That is, are there certain calculations that cannot be completed before the computer's hardware decays into uselessness?

Such questions really concern limitations on the physical execution of mathematical operations. Physical laws, on which the answers must ultimately be based, are themselves expressed in terms of such mathematical operations. Thus we are asking about the ultimate form in which the laws of physics can be applied, given the constraints imposed by the universe that the laws themselves describe.



TRANSITION RULES

HEAD STATE	BIT READ	CHANGE BIT TO	CHANGE STATE TO	MOVE TO
A	1	0	A	LEFT
A	0	1	B	RIGHT
B	1	1	A	LEFT
B	0	0	B	RIGHT

BROWNIAN CLOCKWORK TURING MACHINE, made out of rigid, frictionless parts, relies on random jiggling of its loosely fitted parts to change from state to state. When a part is held in place, it is not by friction but by grooves or notches in neighboring parts. Parts interlock in such a way that they can follow only one "computational path"; although they are free to jiggle a little, the only large-scale motions they can make correspond to forward or backward computational steps. The operation of the machine is driven slowly forward by a very weak force; at any instant the machine is almost as likely to move backward as forward. On the average, however, the machine will move forward and the computation will eventually end. The machine can be made to dissipate as small an amount of energy as the user wishes, simply by employing a force of the correct weakness. Segments of tape are represented by grooved disks; bits are represented by E-shaped blocks, which are locked onto the disks in either the up (1) or the down (0) position. The head consists of a rigid framework and a complicated mechanism (most of which is not shown) from which are suspended a reader, a manipulator and a screwdriver-shaped rod. The machine's operation is controlled by a grooved "master camshaft," which resembles a phonograph record (*top left and far right*); different grooves correspond to different head states. At the beginning of a cycle the head is positioned above one of the disks and a "stylus" is in the "read" segment of the groove in the master camshaft that corresponds to the machine's current head state. During the "read" part of the cycle (*1*) the reader determines whether the disk's "bit" block is up or down by a process called an obstructive read (*center right*). In an obstructive read the reader moves past the block, following a high or a low path; one of the paths will be obstructed by the knob on the end of the block, and so there will be only one path for the reader to follow. At the point on the master camshaft that corresponds to this "decision" the grooves branch; each groove splits into two, and the stylus is guided into the groove that corresponds to the bit's value (*2*). Then the master camshaft turns until the stylus is in the "write" segment (*3*). In this segment each groove contains a different set of "instructions" for the machine to follow; the instructions are transmitted by a complex linkage between the stylus and the rest of the mechanim. If the instructions call for the bit's value to change, the manipulator moves over and grasps the knob; then the screwdriver rotates the disk until the block is free to move, the manipulator moves the block up or down and the screwdriver rotates the disk again to hold the block in place. After the stylus passes through the "write" segment of the master camshaft it enters the "shift" segment (*4*). Each groove in this segment contains instructions to move the head one segment to the left or right. Then the stylus enters the "change state" segment of the camshaft (*5*), where grooves merge in such a way that the stylus falls into the groove representing the next head state. The cycle is now complete (*6*). Disks adjacent to the one being read are held in place by the head's framework. Disks that are farther away are held by "locking tabs." The locking tab on each disk is coupled to a special bit, called the Q bit, on an adjacent disk. The linkages between Q bits and locking tabs are constructed so that the disk currently being read is free to move, while disks far to the right or left are held still.

BREAKTHROUGH:
SPEED PATIENT RECOVERY WITH A COMPUTERIZED GET WELL CARD.

When infection threatens, identifying it–and the antibiotic to fight it–takes time. It's time a patient may not have.

We found a way to pinpoint an infection and a treatment in just hours instead of days.

Our Vitek diagnostic system uses cards containing "wells" of dehydrated nutrients. When a patient sample is introduced, the infection feeds and grows. Light beamed through the wells to a computer identifies bacteria which are growing. Antibiotics in wells of a second card stop the growth and help the physician choose the best treatment. The time saved is days. The ache saved is appreciated. The lives saved are priceless.

We're creating breakthroughs not only in healthcare but also in helicopters, information systems and spacecraft.

We're McDonnell Douglas.

MCDONNELL DOUGLAS

SCIENCE AND THE CITIZEN

Birthday Blowout

More than two years ago President Reagan called in his "Star Wars" speech for the development of a space-based defense that would render nuclear missiles "impotent and obsolete." In recent months workers engaged in implementing the Administration's Strategic Defense Initiative (SDI) have made advances in solving two of the many technical challenges that may need to be met if the president's goal is to be realized. They have increased the brightness of an X-ray laser powered by a nuclear explosion and they have guided electron beams through low-pressure gases by means of ultraviolet laser beams.

The top-secret development of the X-ray laser began nearly five years ago at the Lawrence Livermore National Laboratory in California. The device, called Excalibur, consists of a nuclear explosive surrounded by a cylindrical array of thin metal fibers. X rays released during the nuclear explosion stimulate the emission of a beam of X rays from the fibers in the microsecond that elapses before the device immolates itself. Since X rays penetrate most matter rather than being reflected or bent, workers found it difficult to focus the beam. In a second-generation version, named Super Excalibur, the investigators exploited the index of refraction of free electrons to bend the path of the X rays, thereby focusing the beam and increasing its brightness by six orders of magnitude. The new device was tested in an underground explosion in Nevada on March 23, the second anniversary of the president's speech.

Workers at Livermore and at the Sandia National Laboratories in New Mexico have independently developed methods to improve the range of another potential weapon, a high-current electron beam. The distance such a beam can travel is limited by instabilities arising from the repulsive force between electrons. To counteract the forces, investigators use an ultraviolet laser that forms a cylindrical channel in a gas by stripping its constituent atoms of electrons. The ionized channel can guide a high-energy electron beam just as a metallic pipe would. Employing one variant of the basic technique, Sandia investigators have propelled 1.5-million-volt beams, whose current strength ranged between 10,000 and 20,000 amperes, through nitrogen gas maintained at a pressure of one torr (1/760 atmosphere) over a distance of 1.3 meters. Livermore investigators, employing another variant of the technique, have propelled 50-million-volt beams, whose current strength was 10,000 amperes, through benzene gas maintained at a pressure of 1/10,000 torr, over distances exceeding 100 meters.

The two successes have not mollified critics of the SDI. Hans A. Bethe, who attended a pretest briefing on Super Excalibur at Livermore, still characterizes the overall program as "nonsense." For their part developers of the X-ray laser and the high-energy electron beam are heartened but observe that they must overcome many technical hurdles before they can actually build weapons. Posttest enthusiasm has not caused the Administration's spokesmen to revert to the high optimism expressed by President Reagan two years ago. At a symposium on lasers and particle beams held at the University of Rochester a few weeks after the successful test Louis C. Marquet, then director of the Department of Defense's directed-energy office, stated: "The goal of the program is not to build a perfect defense system. The goal of the current Strategic Defense Initiative is to bring to the table the technical issues."

Ailing Arrivals

Some 700,000 refugees from the war and turmoil in Southeast Asia now live in the U.S., and the influx continues at the rate of 50,000 a year. Language, housing and jobs are the major challenges Southeast Asians face in this country; the status of their health constitutes a challenge to the medical-care system.

Among the arriving refugees there is a high incidence of anemia, intestinal parasites, tuberculosis and hepatitis. A study published in 1982 revealed that of the refugees in the San Diego area who were examined less than two months after settling in, 61 percent were infected with intestinal parasites, 55 percent had positive skin tests for tuberculosis and 14 percent showed signs of hepatitis B.

As refugee programs and health facilities have adapted to meeting such pressing health needs less obvious and more intractable problems have gained prominence. One is sudden, unexplained nocturnal death (SUND), whose victims almost invariably are young, apparently healthy male refugees; in perhaps half of the cases they are members of the Hmong tribe from Laos. Among Southeast Asian men from 25 through 44 years old SUND occurs at an annual rate approaching one case per 1,000, making it one of the leading causes of death in that age group. The Centers for Disease Control (CDC) in Atlanta has received reports of more than 80 cases that have occurred since July, 1977.

Recent reports suggest the immediate cause of such deaths is ventricular fibrillation: rapid, uncoordinated contraction of the heart ventricles. The underlying cause remains mysterious. It has been proposed that SUND has a genetic basis, that it results from emotional stress, that it is associated with sleep apnea (a pause in breathing during sleep) or that it is caused by certain toxic folk remedies used in Hmong, Cambodian and Laotian cultures. For want of evidence all four hypotheses remain viable, according to Mark S. Eberhardt, a CDC epidemiologist.

The CDC has been pursuing a possible link between some folk remedies and another health problem that has emerged among Southeast Asian refugees: lead and arsenic poisoning. Cases of arsenic poisoning of Hmong adults were reported in St. Paul in 1984. The source of the arsenic could not be positively identified, but folk medicine used by the Hmong community contained between 60 and 80 percent arsenic. In Hmong children in St. Paul, symptoms of lead poisoning were commoner than they were among other children in that city; again suspicion has focused on folk remedies.

A less acute problem, but one that may grow in importance, is the prevalence of variant hemoglobin genes among some Southeast Asian groups. Such genes usually do little harm when they occur singly, but they result in debilitating or fatal blood disorders such as alpha and beta thalassemia and hemoglobin *E* when an individual carries two of them. In Southeast Asia between 20 and 40 percent of the individuals in some populations have been found to carry the genes.

No reliable studies of the genes' frequency have been done among Southeast Asians in this country, but the issue is already attracting the attention of the Division of Maternal and Child Health of the Department of Health and Human Services. If the variant genes are common among refugees, the incidence of thalassemia and hemoglobin *E* may rise as refugees marry and have children.

Richard F. Mollica, a psychiatrist at

Room. That was the idea behind the Camry. Legroom. Headroom. Family of five room. But along the way, something happened. Advanced technology made the Camry more than a roomy, comfortable sedan. It made it exciting, innovative; a whole new car was born. And the critics loved it. They loved its gripping front-wheel drive. They applauded its Electronic Fuel Injection and its sophisticated, electronically controlled 4-speed automatic overdrive transmission. They declared the Camry, every gorgeous inch of it, a star. The Camry's space and weight saving design helps make it fuel efficient, too. With all that responsiveness, you can still expect 29 City MPG, 34 Highway MPG!* A performance like *that* deserves reviews like this...

CAR AND DRIVER
"Toyota has fielded the most elaborate electronically controlled transmission to date." (And it's roomy. With almost 39" of headroom, even a 6'4" basketball player feels at home.)

CONSUMER'S DIGEST
"Toyota Camry has been elected by our readers to the Consumer's Digest Hall of Fame, 'for the consistent high quality of its entire model line, its high owner satisfaction and commitment to safer design.'" (And it's roomy. The Camry seats five with room to spare.)

THE CHRISTIAN SCIENCE MONITOR
"The Toyota Camry is a landmark vehicle...the Camry is fun to drive and think of all the gasoline you save to the volume of space inside the car. If Toyota were to get a letter grade, it would have to be an A." (And it's roomy. Camry's got 93 cubic feet of passenger space...that's real stretch-out room for a family of five.)

MOTOR TREND
"...even with an electric sunroof installed, the five-passenger Camry will accommodate six-footers in all locations." (See, we told you it was roomy!)

*Remember: Compare these estimates to the "EPA Estimated MPG" of other cars. You may get different mileage, depending on how fast you drive, weather conditions and trip length.

OH WHAT A FEELING!
TOYOTA

BUCKLE UP—IT'S A GOOD FEELING!

THE 1985 FAMILY CAMRY. IT'S DEVELOPED QUITE A ROOMY REPUTATION.

the Harvard Medical School, believes the psychiatric needs of the refugees also constitute a major public-health problem. Mollica, who with James Lavelle, a social worker, directs the Indochinese Psychiatry Clinic (IPC) at the Brighton Marine Public Health Center at St. Elizabeth's Hospital in Boston, thinks many refugees bear deep psychological scars. No formal studies of the rate of psychiatric disorder among refugees have been done, but trauma is abundant in their histories. Many of the refugees lost family members; some were imprisoned, tortured or subjected to other forms of severe physical deprivation and degradation.

At the IPC, which is funded by grants from the Federal Office of Refugee Resettlement, the Massachusetts Department of Mental Health and the Boston Foundation, the most disturbed among the refugees from Boston and surrounding areas receive treatment. Since its founding in 1982 the clinic, the first of its kind, has seen more than 400 patients, most of them suffering from severe depression rooted in their experiences in Southeast Asia.

Mollica hopes clinics patterned on the IPC will be established elsewhere in the U.S. The immediate concerns of most refugees are the practical difficulties of life in a new country. For the minority whose experiences in Southeast Asia remain an unmanageable burden, however, Mollica thinks psychiatric care is essential if they are to adjust to their new life.

Prion Gene

Scrapie, an infectious neurological disorder in sheep, has for two reasons excited uncommon interest. The first is the similarity of its pathological symptoms to those of certain dementias in human beings, including Alzheimer's disease and Creutzfeldt-Jakob disease. The second reason is the strangeness of the infectious agent that transmits scrapie: unlike a virus, it seems to contain little or no genetic material. Purified preparations of the agent yield a single, characteristic protein, but not the nucleic acid that codes for it. How does the scrapie agent—often called a prion—replicate?

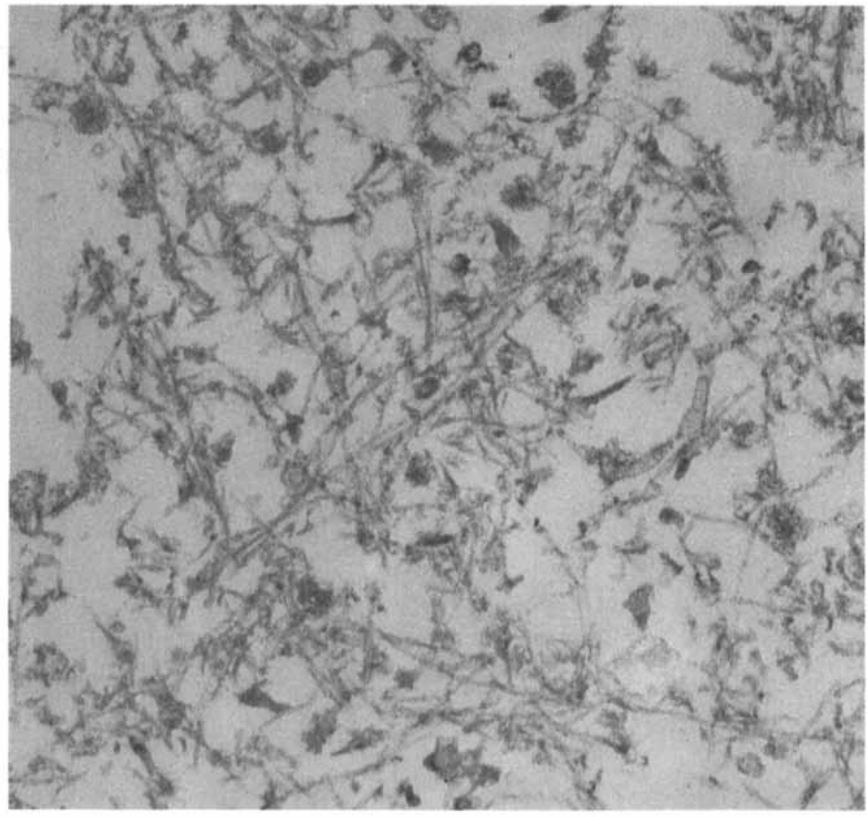
Part of the answer has now been reported in *Cell* by investigators at the University of California at San Francisco School of Medicine, the California Institute of Technology and the University of Zurich. They have identified the gene for prion protein, or PrP, in the brain cells of hamsters as well as in mouse and human tissue. Apparently the prion subverts a protein that is present in normal, uninfected animals.

The first step, taken by Leroy E. Hood and his colleagues at Caltech, was to determine the amino acid sequence of a small fraction of the prion protein and to derive from that sequence a bit of single-strand DNA. Charles Weissmann and his co-workers in Zurich then used the synthesized DNA as a probe to pick the PrP gene out of a "library" of DNA culled from the brains of scrapie-infected hamsters; the radioactively labeled probe hybridized with the one DNA molecule whose nucleotide sequence matched its own. By cloning and sequencing the entire gene, the workers verified that it codes correctly for the known parts of PrP.

With the cloned PrP DNA now available as a probe, the San Francisco group, led by Stanley B. Prusiner, showed that the PrP gene is not present in purified preparations of the scrapie agent. On the other hand, the gene was found in the chromosomes not only of infected hamsters but also of healthy ones. More surprising, it is not a "silent" gene that is "switched on" by infection; PrP-associated messenger RNA (which initiates protein synthesis by transcribing DNA) is present in healthy animals, indicating that the protein itself is too. The normal function of PrP is a mystery. Its gene does not correspond to that of any known protein.

Unlike normal PrP, the PrP in diseased animals aggregates into "rods" and filaments, and it resists breakdown by the enzyme proteinase *K*. It is still not clear how the infectious version of PrP is reproduced, but the finding of a cellular gene for PrP has eliminated the most radical hypothesis: that the prion protein somehow codes for itself. Some investigators believe the scrapie agent may yet emerge as a conventional virus bearing its own genes, but if it does, the viral protein would have to be something other than PrP. That hypothesis is contradicted by strong evidence linking scrapie to the presence of PrP.

Two possibilities remain. The prion could still be a pure protein that invades a cell and modifies normal PrP. The infectious protein might alter the PrP gene, or it might attack PrP directly, rendering the normal molecule infectious by, for example, cleaving it at a particular place. Alternatively, the change in the gene or the protein might be produced by a tiny bit of DNA or RNA that has escaped detection. The infectious nucleic acid would not code for PrP but would merely modify the normal protein, "recruiting" it as, say, a protective capsule. Prusiner and his colleagues are now trying to decide between these two hypotheses by synthesizing PrP from the gene isolated from hamsters with scrapie. If the protein



Prion filaments in a scrapie-infected hamster brain

SCIENCE/SCOPE

A new radar can map military targets with high resolution equal to that of infrared devices, even in rain and other bad weather. The Advanced Synthetic Aperture Radar System (ASARS-2), designed to complement electro-optic sensors, is flown on a U.S. Air Force TR-1 reconnaissance aircraft and provides real-time radar imagery to a ground station. ASARS-2 operates in all weather at ranges far exceeding the capabilities of infrared and other electro-optic devices, thanks to new state-of-the-art signal processing and other advances. The Air Force gave the system an excellent rating after it underwent strict operational performance tests as part of a "fly-before-buy" program. Hughes Aircraft Company is producing the system under a development and production contract. Eventually ASARS-2 is expected to be adapted for tactical aircraft and mobile tactical stations.

Jam-resistant communications have been introduced into NATO by a new terminal for AWACS early-warning aircraft. The Joint Tactical Information Distribution System (JTIDS) Class 1 terminal is designed to combat the formidable and growing electronic countermeasures threat to tactical communications. JTIDS uses principles of time division multiple access to provide secure, high-capacity communications for AWACS radar planes and ground stations. The system relays a wide variety of information, such as command and control, surveillance, intelligence, force status, target assignments, warnings and alerts, weather, and logistics. Software filtering lets each participant select data pertinent to his own needs. Hughes is producing the JTIDS terminal for use with NATO's Airborne Early Warning/Ground Environment Integration Segment (AEGIS).

Lasers soon will be inspecting solder joints of fighter aircraft radars, thanks to new manufacturing technology being set in place at Hughes. Solder joints will be examined by a computerized technique using lasers and fiber optics, the glass threads that carry laser light transmissions. The process will free manufacturing personnel from tedious and time-consuming inspections of more than 36 million solder joints created in a single year's production. The project is part of an Industrial Modernization Incentive Program (IMIP) awarded by the U.S. Navy and Air Force. IMIP is a share-the-savings concept that will reduce costs of the F-14, F-15, and F/A-18 radar programs by more than $10 million, while improving the quality and reliability of the systems.

Two high-power direct-broadcast satellites will carry family-oriented and religious programming for Dominion Video Satellite, Inc. The Hughes satellites will incorporate a novel design, combining existing technologies of spin-stabilized satellites and body-stabilized satellites. Each will carry large, winglike solar arrays similar to those on body-stabilized satellites. The wings will extend 110 feet, generating 5,000 watts of electrical power for eight channels of communications. A central spinning section will provide gyroscopic stability and additional electrical power for basic housekeeping functions. The satellites will be equipped with batteries to provide full power when the spacecraft pass through Earth's shadow and the sun is blocked from the solar cells.

Hughes' Santa Barbara Research Center is seeking experienced engineers and scientists to further develop advanced IR focal plane technologies. We need custom integrated circuit designers, nuclear effects engineers, material scientists, semiconductor device scientists and process engineers, and IR system analysts. To learn how to become involved in this innovative technology, contact the Santa Barbara Research Center, Professional Employment, Dept. S2, 75 Coromar Drive, Goleta, CA 93117. Equal opportunity employer. U.S. citizenship required.

For more information write to: P.O. Box 11205, Dept. 70-12, Marina del Rey, CA 90295

HUGHES
AIRCRAFT COMPANY

alone is found to transmit scrapie, the workers will have determined what a prion is by making one.

Past Present

Analyses of certain astrophysical objects suggest that astronomers can now observe the early evolution of a galaxy. One of the most intriguing of these objects is a galaxy labeled II Zwicky 23, which was initially catalogued by the late Fritz Zwicky of the California Institute of Technology. II Zwicky 23 is a so-called compact blue galaxy that lies about 500 million light-years from the earth. Its predominant blue-light emissions indicate the presence of a large population of hot, young, blue stars.

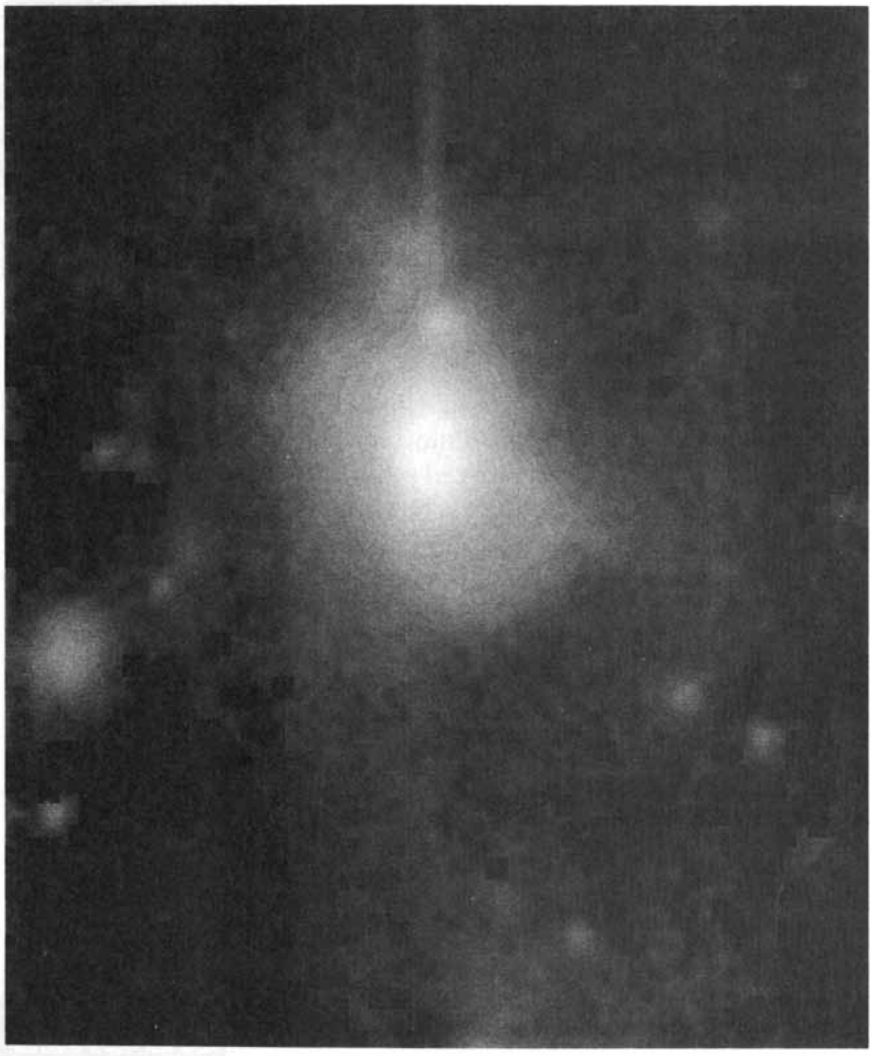
According to calculations made by William C. Keel of the Kitt Peak National Observatory, II Zwicky 23 converts into stars each year an amount of gas equivalent to about 40 solar masses. Keel estimates the rate of formation from the intensity of lines in the galaxy's hydrogen spectrum. Only hot, new stars are energetic enough to ionize gas in their vicinity, and the recombination of ionized gas generates a characteristic spectrum. A standard assumption about the number of new stars in any generation hot enough to ionize hydrogen gas yields the rate of formation.

In the high-contrast image made by the National Optical Astronomy Observatories (*see illustration below*) several faint filaments can be seen projecting from the bright, central core of the galaxy. The filaments are made up of infalling gas that condenses as it rushes toward the core to feed the star-making process. Keel calculates that the rate at which stars are forming is roughly equal to the rate at which the filaments appear to be plunging into the galactic core.

The prodigious rate of star formation in II Zwicky 23 must be a relatively recent phenomenon. If that rate had been roughly constant since the birth of the Milky Way, the hydrogen gas that envelops II Zwicky 23 would have long since been used up and the galaxy would be much brighter than it is today. The rapid star formation is not the sudden burst of activity that has been observed in some other galaxies. If the current flare-up had begun only a few million years ago, there would be no spectral signature from stars older than 100 million years. The presence of such a signature suggests to Keel that most of the stars now in the galaxy were formed at a roughly constant rate during a protracted period of about 200 million years.

A second line of observational data on galactic birth may come from quasars, whose enormous red shifts indicate that many of them are on the order of 10 billion light-years from the earth. If they are this distant, we see them at an early epoch in the evolution of the universe when most galaxies were forming. Yet quasars seem to be missing from nearby regions of space. What has become of them?

To answer this question Alexei Filippenko of the University of California at Berkeley and Wallace L. W. Sargent of Caltech subtracted from the spectrums of 75 nearby galaxies the light that stars would be expected to contribute. In about a third of these galaxies the remaining spectrums resemble closely those of distant quasars. Presumably the net spectrums consist of the faint emissions of mildly active galactic nuclei, which are usually overwhelmed by the brighter emissions of stars. The finding is the strongest evidence to date for a hypothesis that has frequently been made but never proved, namely that the remnants of quasars are indeed nearby. They have not been recognized as such because they have evolved in the past 10 billion years to become nuclei dominated by normal stars. Hence when astronomers peer at a quasar, what they see may well be an embryonic version of the Milky Way.



Infalling gas streams toward the center of a young galaxy

Water Power

The flashes of lightning in thunderstorms often seem to be followed within a few moments by notable gushes of rain. The sequence suggests that the lightning somehow brings on the downpour. Writing in *Journal of Geophysical Research,* E. R. Jayaratne and C. P. R. Saunders of the University of Manchester Institute of Science and Technology suggest that the truth is the other way around: the downpour brings on the lightning.

Jayaratne and Saunders simulated a cloud of ice crystals and supercooled water droplets and measured the electric charge imparted to simulated hail-

Scanning Tunneling Microscopy

IBM

To: Robin
From: Roger
Subject: IBM Technology

I've been reviewing some of our past and present technological achievements, and it occurred to me that the scientific, engineering, and academic communities might like to know more about them. Will you select a topic from the following list or suggest another one? Thanks.

Vacuum tube digital multiplier
IBM 603/604 calculators
Selective Sequence Electronic Calculator (SSEC)
Tape drive vacuum column
Naval Ordnance Research Calculator (NORC)
Input/output channel
IBM 608 transistor calculator
FORTRAN
RAMAC and disks
First automated transistor production
Chain and train printers
Input/Output Control System (IOCS)
STRETCH computer
"Selectric" typewriter
SABRE airline reservation system
Removable disk pack
Virtual machine concept
Hypertape
System/360 compatible family
Operating System/360
Solid Logic Technology
System/360 Model 67/Time-Sharing System
One-transistor memory cell
Cache memory
Relational data base
First all-monolithic main memory
Thin-film recording head
Floppy disk
Tape group code recording
Systems Network Architecture
Federal cryptographic standard
Laser/electrophotographic printer
First 64K-bit chip mass production
First E-beam direct-write chip production
Thermal Conduction Module
288K-bit memory chip
Robotic control language
Masterslice and the Engineering Design System

Roger -
IBM's researchers have developed a powerful new technique for studying surfaces at the atomic level: let's tell this story!
Robin

Figure 1. *This three-dimensional representation of a silicon surface was obtained by scanning tunneling microscopy, developed by IBM. The individual hills or bumps indicate actual atoms separated by as little as six angstroms. (One angstrom is one ten-billionth of a meter.)*

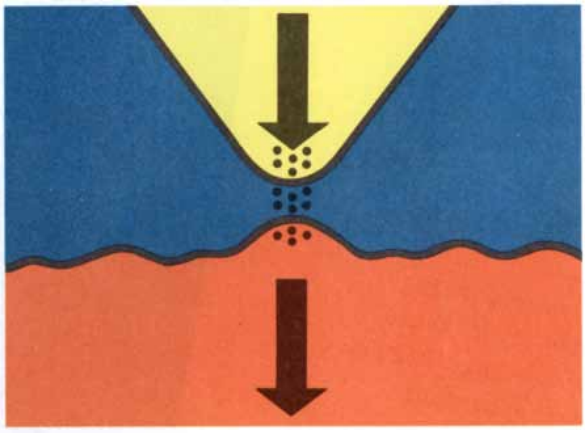


Figure 2. *IBM's new microscopy technique makes use of a phenomenon called vacuum tunneling, which involves the passage, or tunneling, of electrons between two conducting or semiconducting solids that are narrowly separated by a vacuum. Tunneling occurs because electrons have wavelike properties as well as particle properties. This means, according to quantum theory, that electrons appear as electron clouds that spill out slightly beyond the surfaces of the solids in which they originate. As a result, there is a finite probability that electrons will tunnel through the vacuum.*

Miniaturization is the driving force behind the computer revolution. As computer chips continue to evolve, their structural details are becoming so small that it is vital to understand them at the atomic level.

Recently, IBM researchers have succeeded in examining structures at the atomic level by developing an absolutely new kind of microscopy technique—scanning tunneling microscopy, or STM. Specifically, they have produced three-dimensional images of the surface topography of solids that show vertical position differences as small as 0.1 angstroms (one angstrom is one ten-billionth of a meter) and horizontal position differences as small as six angstroms. Such simultaneous resolution is unprecedented.

The new microscopy technique makes use of a quantum-mechanical phenomenon called vacuum tunneling, which involves the passage, or tunneling, of electrons between two conducting or semiconducting solids that are narrowly separated by an insulator or a vacuum.

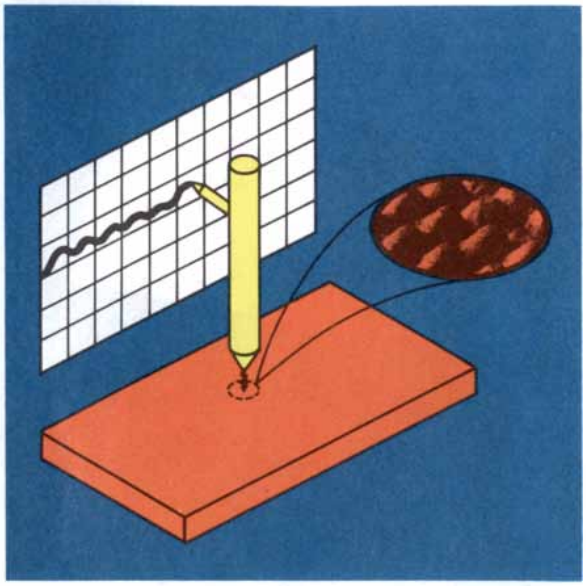


Figure 3. *The principle of the scanning tunneling microscope is quite simple. As a probe tip is scanned across a surface's microscopic hills and valleys, its vertical position is adjusted to maintain a constant tip-to-surface distance (by keeping tunnel current constant). The probe consequently follows the surface contour as it moves, so that monitoring its vertical position can be used to yield a two-dimensional representation of the surface contour for each scan. The full three-dimensional image is obtained by assembling an entire sequence of scans.*

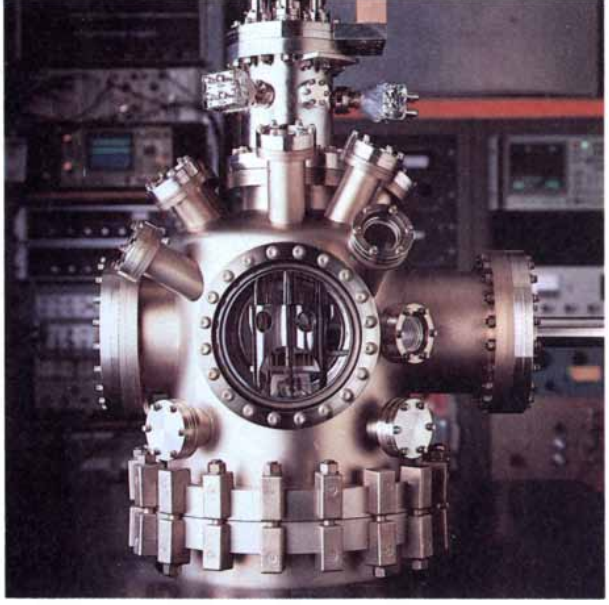


Figure 4. *The scanning tunneling microscope is contained inside this chamber. The electronics (background) collect and process the measurements and then display the results on a screen or plotter. An absolute necessity for making measurements is a vibration-free suspension, which also had to be specially developed by the IBM researchers.*

Tunneling through solid insulating barriers was first demonstrated in 1957; it was only early in 1982 that controlled vacuum tunneling was demonstrated by IBM in an experimental configuration suitable for microscopy.

In principle, the scanning tunneling microscope takes advantage of the strong dependence of the tunnel current on the separation between two solids. One solid has its surface under investigation; the other, a metal tip, is a probe electrode. As the probe moves laterally across the surface (while separated from it by about ten angstroms), the tunnel current will vary in accordance with changes in the tip-to-surface distance. The tunnel-current variation in effect is a measure of the surface topography.

In practice, the vertical position of the probe is changed to keep the tunnel current, and thus the tip-to-surface distance, constant for all points. In that way, monitoring the position of the tip while scanning yields a topographic picture of the surface. The technique is so sensitive that a change in tip-to-surface distance by the diameter of a single atom produces a tunnel-current change by a factor of 1,000.

By providing a more detailed view of surface structures, STM has already significantly advanced the understanding of important materials such as silicon. However, STM is more than a surface structural tool with atomic resolution: it also images surface parameters (such as composition and oxidation state) and can determine electronic properties. This opens fascinating possibilities in many areas of science and technology.

STM can be performed at ambient pressure and can see surfaces covered by nonconducting liquids. The ability to operate under such conditions makes STM attractive in many different fields, from engineering to biology.

Scientists at the IBM Zurich Research Laboratory developed the world's first scanning tunneling microscope. Their contributions are only part of IBM's continuing commitment to research, development, and engineering.

IBM®

For free additional information on STM, please write: IBM Corporation, Dept. 1006, P.O. Box 5089, Clifton, NJ 07015

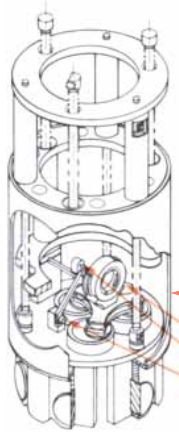


Figure 5. *Photograph and illustration of the scanning tunneling microscope. Note that the microscope is lowered into the chamber (shown in Figure 4) when in use.*

stones by collisions of the hailstones with the crystals and the droplets. If the temperature in the cloud was colder than some −20 degrees Celsius, the hailstones acquired a negative charge; at higher temperatures the acquired charge was positive. The presence of sodium chloride dissolved in the droplets (a further attribute of the interior of a cloud) increased the temperature at which the charge reversal happens.

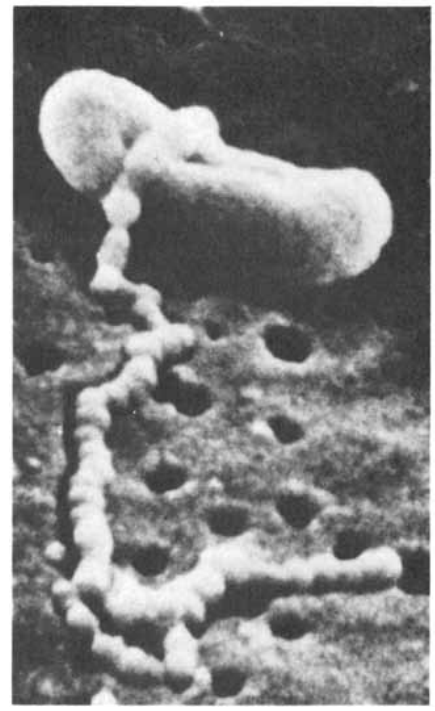
In an actual thunderstorm the crystals and droplets would be supported by updrafts. Meanwhile the hailstones would fall. The resulting collisions would impart negative charge to the stones, and the charge would continue to accumulate until it attained a maximum negative value at the charge-reversal level: the level in the cloud where the temperature is about −20 degrees. (The actual temperature depends on the impurities in the cloud.) There the acquisition of positive charge would begin to neutralize the negative. The charge-reversal level is typically some distance above the base of the cloud.

Suppose, then, a sudden spurt of rainfall develops in the cloud below the level where the temperature is −20 degrees. The rainfall, being below the level of charge reversal, should accumulate positive charge, and its attraction for negative charge should facilitate the discharge (that is, the passage to the earth) of the negative charge above it. The result would be a bolt of lightning. Some moments later the rainfall that initiated the bolt would itself arrive at the earth.

Industrious Bacteria

Bacteria have been an indispensable part of industrial enterprise for thousands of years. They have been employed in the manufacture of cheese, the fermentation industry exploits them to make such products as vinegar and lactic acid and biotechnologists have learned how to induce them to synthesize proteins specified by genetically engineered DNA. Yet the possible industrial applications of bacteria seem far from being exhausted. Investigators at several institutions have found that bacteria can break down a variety of troublesome toxic substances and can even manufacture commercial quantities of a biodegradable plastic.

William A. Chantry, a doctoral candidate at the University of Wisconsin at Madison, has selectively cultured bacteria, including species of *Pseudomonas* and *Nocardia,* that can break down the recalcitrant chemical compounds known as PCB's (polychlorinated biphenyls). Since PCB's are



Alcaligenes eutrophus *secreting polyester*

good insulators and are nonflammable, manufacturers have incorporated them in transformers and other electrical equipment. In the process PCB's have been discharged into rivers and streams and have been disposed of in open landfills.

In large amounts PCB's have been linked to reproductive damage in laboratory animals and are suspected of causing birth defects in human beings. The toxicity of small amounts of the compound is in dispute, however, and so some groups have challenged the cost-effectiveness of expensive cleanup efforts.

The potential cost of ridding the Madison Metropolitan Sewerage District's wastewater treatment plant of PCB's, for example, has been estimated at $45 million if the PCB's are incinerated and at $40 million if they are removed to an approved chemical-waste landfill. Based on preliminary work with sewage sludge, Chantry thinks the cost of employing bacteria to remove the compound at the plant would be between $4.5 and $9 million. The bacteria should reduce the concentration of PCB's by from 60 to 65 percent for every 20 days during which the sludge is exposed to them. He foresees that further studies of the interaction between the bacteria and the PCB's should be applicable to other toxic hydrophobic compounds (substances that do not dissolve in pure water) such as DDT and dioxin.

A less familiar but potentially dangerous compound, benz(a)anthracene,

SAVE $49.00

SALE!

One of our best sellers—Was $149.95

Magnificent, Hand-Crafted Wood

Genuine OAK File Cabinet

Why pay the same price for ordinary metal cabinets!

2 Drawer Size

Now Only $99.95

Turn-of-the-century design... Elegant hand-finished OAK with SOLID BRASS HARDWARE

SIZES AVAILABLE:

- **TWO DRAWER (16x17x28″)** ..... **$ 99.95**
- **FOUR DRAWER (16x17x53″)** **189.95**
- **TWO DRAWER, LEGAL SIZE (32x28x22″)** **149.95**
- **FOUR DRAWER, LEGAL SIZE (54x28x22″)** **289.95**

(All with top-drawer lock & 2 keys)

MIGHTY OAK returns to the American business scene . . . Fine 'antique' filing cabinets, at less than you'd plan to pay for common steel cabinets. Sturdy and functional, with an added *plus:* the warmth and elegance of fine wood furniture.

AN AMERICAN CLASSIC, faithfully reproduced. In order to locate an original oak cabinet, you'd probably have to spend months searching the antique shops. On the front are raised panels of solid oak, rabbeted in for maximum strength. The top and sides are a routed oak veneer. Handles, top-drawer lock, and nameplates are solid brass. Drawers are rated to hold 100 lbs., and have full-extension ball bearings. Made to accommodate manila folders or suspension files.

EXCELLENT FOR THE HOME, TOO

. . . Stores a wealth of household or home-business information. So attractive, it looks good in any room. Easy assembly; comes with 2 keys. *May be tax deductible*—Check with your tax advisor. (Legal size shipped by truck)

MONEY-BACK GUARANTEE

Returnable within 30 days for *full refund* of purchase price (except ship. & hdlg.). Our policy is to process all orders promptly. We charge your credit card order only when order is shipped. Delays notified promptly. Shipment *guaranteed within 60 days.*

SYNCHRONICS,
Hanover, PA 17333

CALL TOLL FREE (1-800)-621-5800 Or Use Handy Coupon

SATISFACTION GUARANTEED—MAIL TODAY!

SYNCHRONICS, Dept. AZ-4032
Bldg. #43, Hanover, PA 17333

Yes, please rush my *Oak File Cabinet*(s) as indicated below, on full money-back guarantee.

_____TWO DRAWER (Z748947), only $99.95, add $19.95 hvywt. ship. & hdlg.
_____FOUR DRAWER (Z736652D), each $189.95, add $39.95 hvywt. ship. & hdlg.
_____TWO DRAWER, LEGAL SIZE (Z737288D), $149.95, add $24.95 hvywt. ship. & hdlg.
_____FOUR DRAWER, LEGAL SIZE (Z737296D), each $289.95, add $49.95 hvywt. ship. & hdlg.

Encl. is $_____(PA & NV res. add sales tax)

CHARGE TO MY: ☐ Diners Club ☐ VISA ☐ American Express ☐ MasterCard

Acct. No. _____ Exp. Date _____

PRINT NAME _____

ADDRESS _____

CITY _____

STATE _____ ZIP _____

appears to be susceptible to a soil bacterium. The discovery was made by David T. Gibson and William R. Mahaffey of the University of Texas at Austin and Carl E. Cerniglia of the National Center for Toxicological Research. Benz(a)anthracene is one of a group of hydrocarbons present in automobile exhaust, cigarette smoke, barbecued meats and crude oils. The substances have been shown to induce cancer in laboratory animals. The bacterium, tentatively identified as a member of the genus *Beijerinckia,* degrades benz(a)anthracene to carbon dioxide and water-soluble products. The investigators think it should be possible to exploit a genetically engineered variant of the bacterium to remove benz(a)anthracene and perhaps other carcinogenic agents from the environment.

With other colleagues Gibson has also engineered a bacterium capable of producing large amounts of an enzyme that breaks down the toxic pesticide parathion. Gibson, Cuneyt M. Serdar and Douglas M. Munnecke worked with the soil bacterium *Pseudomonas diminuta,* an organism that naturally synthesizes small amounts of the enzyme parathion hydrolase. By means of recombinant-DNA techniques they significantly increased the amount of the enzyme produced by each bacterial cell. Serdar told a meeting of the American Society of Microbiology that *P. diminuta* should be effective in disposing of parathion spills and accumulations in wastewater discharged by production plants.

Bacteria can make compounds as well as break them down. When *Alcaligenes eutrophus* and some other bacteria are fed sugars, they synthesize beads of polyester. Until recently only a few hundred grams of the plastic had been made by the bacteria. Now, under the direction of John Adsetts, Imperial Chemical Industries in England has manufactured tons of the polyester. The bacterial product, named Biopol, is indistinguishable from the polyester derived from oil, with one exception: it is biodegradable.

Death, Be Not Proud

Does death make class distinctions? A team of investigators at the London School of Hygiene and Tropical Medicine reports that the mortality rate for civil servants in higher employment grades is substantially less than the rate for workers in lower grades. The team attributes part of the finding to wide disparities among the grades in such individual characteristics as height, weight and blood pressure and in such behavioral factors as smoking and amount of leisure-time physical activity.

The conclusions are based on a 10-year prospective study of 17,530 civil servants working in London. Michael G. Marmot (now at University College London), Martin J. Shipley and Geoffrey Rose write in *The Lancet* that they classified the civil servants into four employment grades: administrative, professional and executive, clerical, and unskilled manual labor. All the subjects were men and all worked in offices. The initial examinations, made over a two-year period beginning in 1967, included a health history, an electrocardiogram and measurements of height, weight, blood pressure and plasma cholesterol. In the course of the next 10 years the investigators then kept track of the causes of death among members of the group.

They found a steep inverse relation between employment grade and mortality. The overall rate among unskilled laborers was three times as high as the rate for administrators; rates for intermediate-grade workers fell in between. The three-to-one ratio applied specifically in the case of coronary heart disease, which accounted for 43 percent of the deaths. Similar ratios were observed for a wide range of other causes including lung cancer, cerebrovascular disease and bronchitis.

What accounts for the differences in mortality rate? Although the nature of the subjects' work itself may play a role, disparities in individual characteristics and personal habits account for at least part of the inverse relation. The investigators found, as they moved from the highest to the lowest grade, a steady increase in the ratio of body weight to height, in average blood pressure and in the number of cigarettes smoked, and a decrease in height and in the amount of leisure-time physical activity.

The findings confirm the threat to health known to be posed by certain personal characteristics and behavior. No single factor, however, could account for the observed differences in mortality rate. Smoking, for instance, was strongly associated with death from coronary heart disease, but even among nonsmokers coronary-disease mortality was highest at the bottom of the grade structure. The decrease in average height from the highest to the lowest grade suggests to the investigators that what they call "early life factors" such as childhood nutrition and family income may have an impact on adult death rates. Since a large part of the grade differences in mortality remains unexplained, the investigators will examine in another study the impact of exercise, diet and psychosocial factors. Women will be among the subjects in future studies.

Tit for Tat

Why do animals cooperate? One explanation is kinship: an animal advances its fitness by helping relatives. Would the same advantage encourage cooperative behavior among individuals who are not related? Michael P. Lombardo of Rutgers University has attempted to answer the question by doing experiments with tree swallows (*Tachycineta bicolor*) at a wildlife refuge on Long Island. Lombardo describes the work in *Science*.

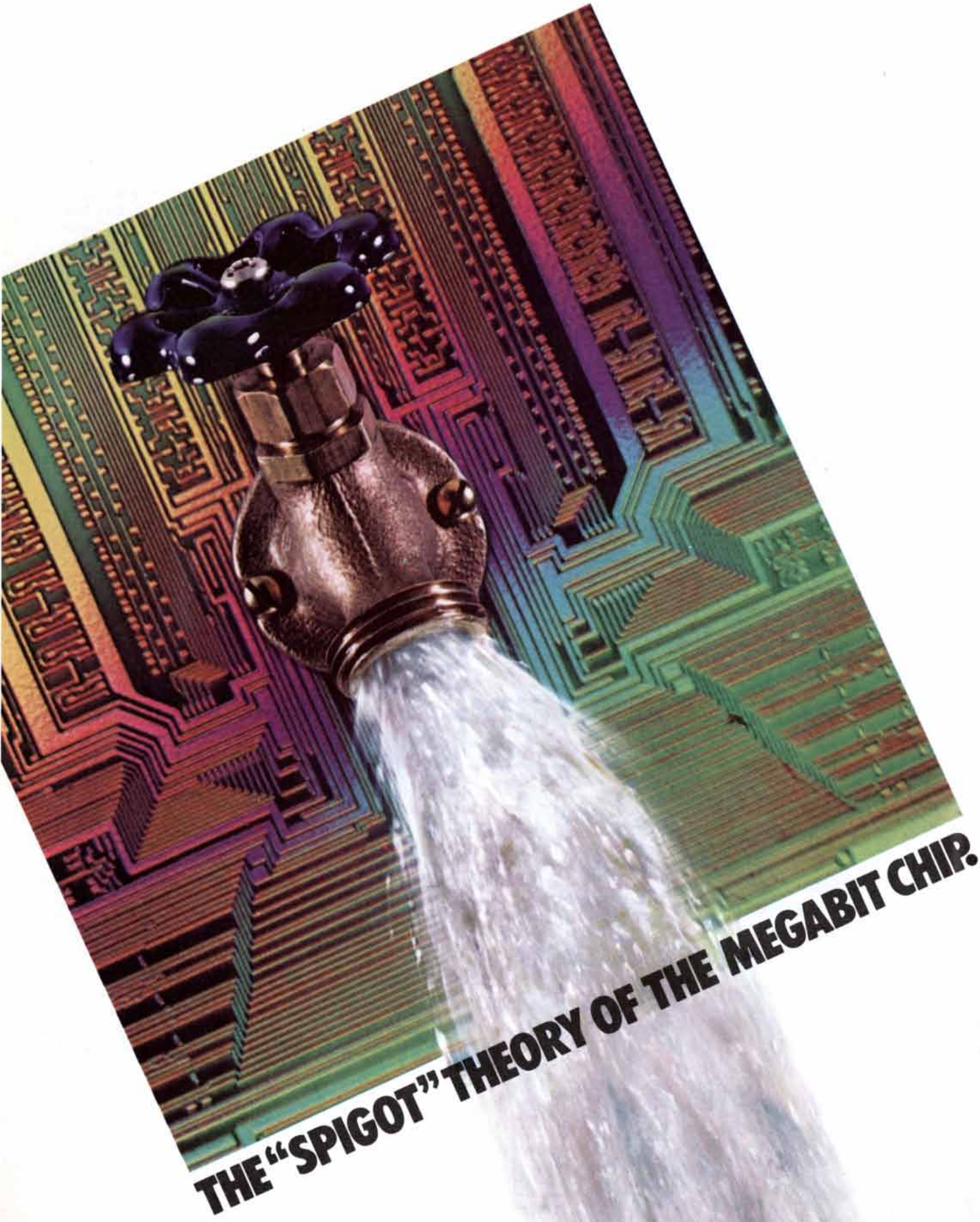
The experiments were modeled on the game called Prisoner's Dilemma, a two-player encounter in which the participants have a potential conflict of interest. In any round a player has an option: to cooperate (refrain from taking a selfish action) or to defect (take a selfish action). The payoff is higher for defection than it is for cooperation unless both players defect; then they both do less well than they would if they had cooperated.

For repeated interactions the optimal strategy (called tit for tat) requires a player to start by cooperating but to retaliate if his competitor defects. The strategy leads to restraint of conflict before a defection, retaliation after a defection and then a resumption of cooperation. From an evolutionary point of view this pattern might establish a stable situation in which cooperation would be likely to develop and persist.

Lombardo began to test the assumption by identifying breeding and nonbreeding swallows in the wildlife refuge. He ascertained, by banding more than 500 birds, that nonbreeders were not close relatives of breeders.

Because nonbreeders were in search of nest sites, Lombardo says, they had a potential conflict of interest with breeders. According to the tit-for-tat model, a pair of breeders would show cooperative behavior by allowing their nest to be visited by nonbreeders; vigorous nest defense would be an act of defection. Nonbreeders would show cooperative behavior by making a benign visit or helping to drive off a predator; an attack on a nest or nestlings would constitute defection.

To test the tit-for-tat model, Lombardo made it appear as though nonbreeders had killed nestlings. He found that 12 out of 17 pairs of parents behaved in a way consistent with the tit-for-tat prediction: they attacked both nonbreeders and the stuffed birds Lombardo put near the ravaged nests. The investigator found this act of retaliation was followed by a return to mutual restraint.



There are megabit chips. And there are megabit chips. All of them can access more than 1 million bits of information. But not all of them can access these 1 million bits of information AT&T fast. And fast to AT&T is 20 million bits per second.

That means we can pour out 1,048,576 bits much faster than you can say "byte." And we can do it because only the AT&T megabit chip has a fast column access mode.

The "Trickle-Down" Theory Vs. The "Spigot" Theory

Most other megabit chips use a "page mode" to access their memory cells. When a chip is in this mode, its fastest data rate is about 10 million bits per second.

But the fast column cycle, developed by AT&T Bell Laboratories, pours out data twice as fast. And not only does data flow out faster, getting data out is easier with the fast column, because access timing requirements are less demanding. Valid data are always available during the entire access cycle.

And speed, ease of use and timing tolerance will make it easier for AT&T to design new and improved products and high-speed systems around the megabit chip.

The Megabit Special

Because this megabit chip is designed for manufacturability, you might want to know what, besides its fast column, makes it special. It's made using an advanced CMOS—Complimentary Metal-Oxide Semiconductor—process that makes it possible to provide high performance at reduced power consumption.

As a matter of fact, it uses 1/8 the power per bit of most 256K DRAMs. The lower power requirement is a plus for any system using the megabit—allowing for lower operating costs and reduced cooling requirements.

The AT&T megabit chip is smaller than a dime, yet has more than 2 million elements on it. Among those 2 million-plus elements are more than 20,000 spare memory cells for

Actual size

safety's sake. If for some reason there's a bad cell in a chip, a computer-controlled test targets the cell—and its whole row or column—for replacement. A laser redundancy technique, pioneered by AT&T, then disconnects the offenders and automatically replaces the entire memory row or column from the spares.

Cleanliness Is Next To Goodliness

A speck of dust on a megabit chip is like a boulder on a railroad track.

To keep our chip clean, it has to be made in a room that's C-L-E-A-N. And an AT&T class-10 clean room is one of the cleanest rooms in the world—where the air contains fewer than 10 particles of dust in a cubic foot of air. And the largest particle must be smaller than 1/150 the diameter of a human hair. That makes our room 10,000 times cleaner than a hospital operating room.

To keep our air clean, we have to keep it moving. Ours is constantly circulated and filtered by fans that could inflate the Goodyear blimp in only 30 seconds.

You Ain't Seen Nothing Yet

Looking toward the year 2000, today's technology predicts submicron design widths, with lines only 400 atoms wide. Chips containing 100 million components—50 times the current amount—could have tiny regions with capabilities that more than match the megabit chip.

Now we're talking computer POWER, because these new chips with micro-parts could work at picospeeds (trillionths of a second). Current estimates suggest 10-picosecond transistors.

AT&T will not let the chips fall where they may. We'll put them where they'll do the most good—in the powerful and reliable digital systems creating the general-purpose information technology we call Universal Information Services. Our continuing flow of advanced technology is one reason why AT&T is the right choice.

AT&T

The right choice.

The Evolution of Darwinism

Recent developments in molecular biology and new interpretations of the fossil record are gradually altering and adding to the synthetic theory, for 40 years the standard view of the process of evolution

by G. Ledyard Stebbins and Francisco J. Ayala

When biologists refer to the theory of evolution nowadays, they are usually thinking not of Charles Darwin's original statement of the theory but of a modified and expanded version of Darwinism that took shape in the 1930's and 1940's. First known as the Neo-Darwinian theory and later as the synthetic theory, it affirmed the fundamental tenets of Darwinism. It held that evolution proceeds through the natural selection of heritable differences arising at random in each generation. Differences that render their carriers better adapted to the environment are multiplied and those that are harmful are eliminated. Like Darwinism, the synthetic theory stressed that evolution through natural selection is opportunistic, in that variations arise by chance and are selected in accordance with the demands of the environment, and that it takes place steadily.

To this Darwinian foundation the architects of the synthetic theory, who included the geneticist Theodosius Dobzhansky, the biogeographer and systematist Ernst Mayr, the paleontologist George Gaylord Simpson, the biologist Julian Huxley and one of us (Stebbins), added new elements. The science of genetics made it possible to identify the determinants of the traits on which natural selection acts as genes: heritable units of information governing structure, development and function. Variant traits were held to result from mutations, or lasting alterations that arise at random in individual genes. From population biology came a second new feature of the synthetic theory: an emphasis on the importance of population structure and distribution in the development of new species. The synthetic theory also incorporated the biological concept of species, which leans more heavily on reproductive isolation (an inability to interbreed) than on visible differences in distinguishing species. Although some biologists initially resisted the synthetic theory, for four decades most evolutionists have considered it the best explanation of evolutionary processes, and it has taken a central place in biology.

In the 1970's and 1980's new developments confronted the synthetic theory. An explosion of investigation into the structure of DNA, the carrier of genetic information, has enabled biologists to study the mechanisms of evolution at the molecular level. The new work has thereby amplified the synthetic theory much as the discovery of genes amplified Darwinism.

The molecular studies have also led to two direct challenges to the synthetic theory. One is a proposal that a kind of molecular determinism, rather than pure chance, impels the development of variations in DNA. The other is a contrasting claim, known as the neutral theory, that chance governs not only the initial appearance of genetic variations but also their subsequent establishment in a population. A different kind of challenge, based on new interpretations of the fossil record, has emerged from paleontology. Known as punctuated equilibrium, it holds that evolution proceeds not at a steady pace but irregularly, in fits and starts.

At the outset it must be said that—unlike attacks by creationists and other nonscientists—none of these challenges denies that evolutionary change occurs, that current species have descended from common ancestors or that Darwinian natural selection plays an important part in the process. The disputes are conflicts of degree and emphasis within a shared evolutionary outlook. We believe, moreover, that with modifications both to the traditional views and to the competing theories most of the challenges can be accommodated within the encompassing vision of the synthetic theory.

Molecular Evolution

The most dramatic changes in thinking about evolution stem from new knowledge about genetic processes at a molecular level, and yet many of the implications of that knowledge for evolutionary theory remain obscure. The molecular pathways that lead from genes to visible characteristics are long, complex and as yet largely unexplored. Until the relation of genes to development is better understood at a molecular level the full impact of molecular biology on evolutionary theory cannot be assessed.

Certain consequences of the recent findings are already evident, however. They have shown genetic variation to be far more complex than was thought, involving changes in the number and configuration of genes as well as mutations in individual genes. Mutations and structural changes are now known to affect not only genes but also sequences of DNA that cannot be called genes because they do not code for proteins. Even as the new molecular biology has complicated the traditional picture of genetic variation, it can also be said to have buttressed the synthetic theory by supplying mechanisms underlying many processes the theory invokes. The meaning of evolu-

MOSAIC EVOLUTION is exemplifed by a fossil of *Archeopteryx*, a species that lived some 150 million years ago and is considered to be a transitional form between the dinosaurs and the birds. In spite of the intermediate position of *Archeopteryx* as a species its individual traits are not transitional in character: it unites a skeleton that remains dinosaur-like with feathers that closely resemble those of modern birds. The concept of mosaic evolution, which holds that different characteristics tend to evolve at different rates because they face different selective constraints, has recently enriched the synthetic theory of evolution.

tion at a molecular level is beginning to come clear. It is now possible, for example, to give incipient answers to the question: How do new genes arise?

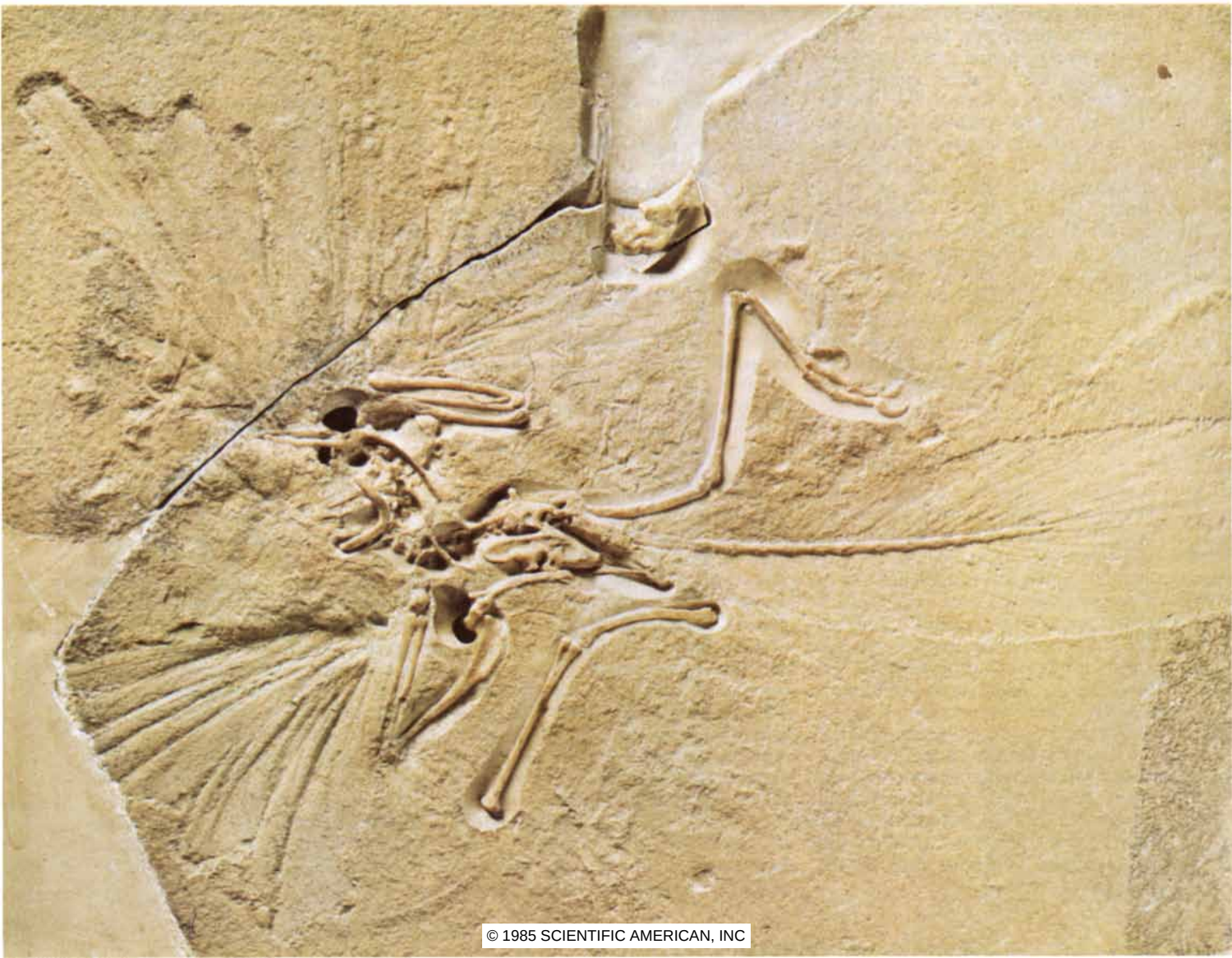
The importance of the development of new genes (and not simply the modification of old ones) in the process of evolution is reflected in the relation between the amount of genetic material and the complexity of organisms. The genome, or total genetic complement, of a virus amounts to between 1,300 and 20,000 nucleotide pairs (np). (Nucleotides are the subunits of DNA, which are strung together in pairs to form the two strands of the double helix.) The genome of bacteria includes, on the average, about four million np. Among eukaryotes (organisms with distinct cell nuclei) fungi have between 10 and 20 million np per cell; most animals and plants have several billion np per cell.

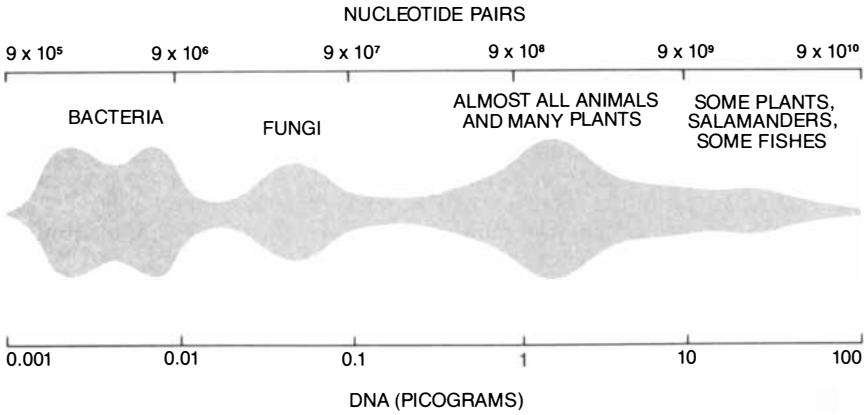
The relation between the amount of DNA and the complexity of the organism or its developmental pattern is not precise. A few groups of higher plants, salamanders and some primitive fishes have 10^{10} np per cell, much more than most mammals. The largest amounts of DNA, some 10^{12} np per cell, are found in eukaryotes that are relatively simple in structure and development, such as certain species of amoeba and the *Psilopsida,* primitive relatives of the ferns. Such species can have thousands of copies of some genes as well as long stretches of DNA that do not code for proteins and are not considered to be genes at all.

Indeed, segments of DNA with no known function have been found in surprising numbers. In the human genome a sequence called *Alu* that is about 300 np long is present in some 300,000 copies, corresponding to 3 percent of the total human DNA; another short segment of about 100 np recurs nearly a million times in the mouse genome. The discovery of these seemingly meaningless repetitions led to the speculation that some molecular evolution is deterministic, proceeding in a particular direction that is independent of chance and natural selection. Such sequences are held to have multiplied not through a series of random events but because a kind of "molecular drive" impelled each sequence to reproduce itself within the genome. Mathematical models have indicated, however, that traditional concepts of evolution suffice to explain the proliferation of such segments.

In spite of the exceptions we have noted and in spite of the fraction of total DNA made up of meaningless sequences, the amount of genetic information carried in each cell does in general increase steadily—from bacteria to molds to higher plants and animals. A typical gene consists of a thousand or more nucleotides arrayed in an order as crucial as the sequence of letters in a sentence. A random sequence of a thousand letters is not likely to make sense; they must be organized in order to convey information. How have meaningful sequences of DNA accumulated in the course of evolution?

One way in which an organism's





AMOUNT OF DNA PER CELL increases in rough proportion to the complexity of the organism. The chart expresses the amount of DNA, the genetic material, in picograms (trillionths of a gram) and in number of nucleotide pairs (the chemical units of the two-strand DNA molecule). The thickness of the band at each point corresponds roughly to the number of species; the shape of the bulges shows that within certain groups of organisms most species have about the same amount of DNA per cell. The right-hand end of the diagram is accounted for by a class of exceptions to the general relation of DNA and complexity: less complex organisms with extremely large amounts of DNA, among them certain seed plants, salamanders and some primitive fishes such as the sturgeon, the gar and the coelacanth.

complement of genes can increase abruptly is through polyploidy: a doubling of the number of chromosomes from one generation to the next, producing offspring that usually are reproductively isolated from the parent generation and in effect constitute a new species [see "Cataclysmic Evolution," by G. Ledyard Stebbins, Jr.; SCIENTIFIC AMERICAN, April, 1951]. Some organisms with very large amounts of DNA, such as *Psilopsida,* are polyploid. Indeed, about 47 percent of all flowering plants are polyploid, including cultivated plants such as the potato, the strawberry and wheat. Nevertheless, many complex plants and animals are not polyploid, and polyploidy does not explain the origin of genes. It is simply a mechanism by means of which existing genes are multiplied.

New Genes

A process that does create entirely new genes is the tandem multiplication, over evolutionary history, of a short sequence of nucleotides. A tandem multiplication is the repetition of a sequence in adjacent sites along a chromosome. The multiplication can occur as matching chromosomes pair during the process of meiosis, the cell divisions that give rise to gametes, or sex cells. As the chromosomes line up they sometimes exchange segments; if the chromosomes have paired out of alignment, the exchange can yield one chromosome in which a particular sequence is repeated. Because it is carried by a sex cell, the repetition will be preserved in the genome of the next generation.

In some cases, such as the $\alpha 2$(I) collagen gene in the chicken, the origin of a gene through tandem multiplication is evident in its structure. Collagen is the main structural protein of bone, cartilage, connective tissue and skin in vertebrates. In chickens the $\alpha 2$(I) gene consists of more than 50 exons, which are discrete sequences of DNA that code for proteins and are separated by noncoding intervening sequences, or introns; the gene has a total length of about 38,000 np. The exons consist of a repeated sequence nine nucleotides long. The repetitions differ somewhat in their component nucleotides, but their common origin can be recognized because the triplet of amino acids for which each sequence codes invariably begins with the amino acid glycine; proline often occupies either or both of the two subsequent positions in the triplet.

From the gene's structure its evolutionary history can be reconstructed. A series of five tandem repetitions of the ancestral 9-np sequence yielded an exon 54 np long. Next the basic exon was multiplied about 50 times. The repetitions have yielded some exons with nucleotide numbers other than 54, but they are always a multiple of nine. Because of the recurrence of the 9-np sequence within each exon, misaligned pairing of chromosomes during meiosis sometimes left corresponding exons overlapping by some multiple of nine. Changed nucleotide numbers resulted when the chromosomes exchanged segments.

Susumu Ohno of the City of Hope Medical Center and his collaborators have traced a similar origin for some of the genes coding for immunoglobulins, or antibody molecules, in mice. Immunoglobulins are proteins consisting of two large ("heavy") and two smaller ("light") polypeptide chains. In each heavy and light chain a region whose nucleotide sequence is nearly invariant is joined to an extremely variable region; possible combinations of the variable regions yield the enormous diversity of antibodies that serve to protect mammals from foreign substances.

Ohno and his colleagues have found that the genes, each about 600 np long, coding for the variable region of the immunoglobin heavy chain evolved from an ancestral gene that in turn was the result of 12 tandem repetitions of a primordial sequence of 48 np. Further analysis showed that the 48-np building block represents an association of three segments, 14, 21 and 15 np long. Although they have been modified by point mutations over the course of evolutionary history, the three segments retain enough similarities to suggest that they represent a tandem triplication of an even smaller building block.

The gene encoding the constant region of an immunoglobulin heavy chain in mice evinces a different evolutionary history, one that may also be common. The heavy chain's constant region includes three protein domains with distinct funtions. One domain interacts with cell surfaces, another activates complement (a group of proteins that destroy foreign cells) and a third forms the attachment point for the light chain. In addition the heavy chain has a fourth segment, a hinge region that separates two portions of the molecule. Each of the four exons in the gene codes for one of the four components of the chain, suggesting each represents a small primordial gene that once coded for a separate polypeptide with a function ancestral to that of the modern protein component.

Gene Duplication

In such cases diverse genes have united to form a single gene. In other instances a single complex gene has given rise to a number of separate genes through duplication. Like the tandem multiplication of a short DNA sequence, the process probably occurs as matching chromosomes exchange segments in the course of meiosis. The original gene and the duplicate may retain the same function; alternatively, one of the genes may preserve the original function while the other evolves a different, albeit related, function. Even before recombinant-DNA techniques were introduced the origin of related genes through the duplication in toto of an ancestral gene was well known;

the amino acid sequences of groups of related proteins gave incontrovertible evidence of the common ancestry of the genes coding for them. The globin genes, which code for the four polypeptides that make up human hemoglobin molecules, are an example of a group of genes for which such an origin is apparent.

The globin genes have maintained varying degrees of similarity since the duplication events that gave rise to them. In another evolutionary pattern, a duplicated gene may diverge freely from the original gene because it is superfluous and therefore not subject to the constraints of natural selection. Examples came to light when the length of DNA containing the globin genes was sequenced. Known as pseudogenes, they are sequences homologous to functional genes, from which they arose through duplication, but they contain mutations that prevent them from making a functional polypeptide. The mutations were able to accumulate unhindered by selective pressures because the organism retained a duplicate gene capable of performing the original function. Pseudogenes are now thought to be common in vertebrates and perhaps in other organisms as well.

When new genes arise through duplication, the original and the duplicate gene are usually transmitted together to the descendants of the organism in which the duplication occurred. An apparent twist in the process has been discovered, however—another of the seemingly limitless ways in which evolution proceeds at the genetic level. Occasionally the original gene is found in one species and the duplicate in a totally unrelated organism. The phenomenon is called horizontal DNA transfer, because the DNA is passed from one species to another coexisting species rather than vertically, from parents to their descendants within a single species.

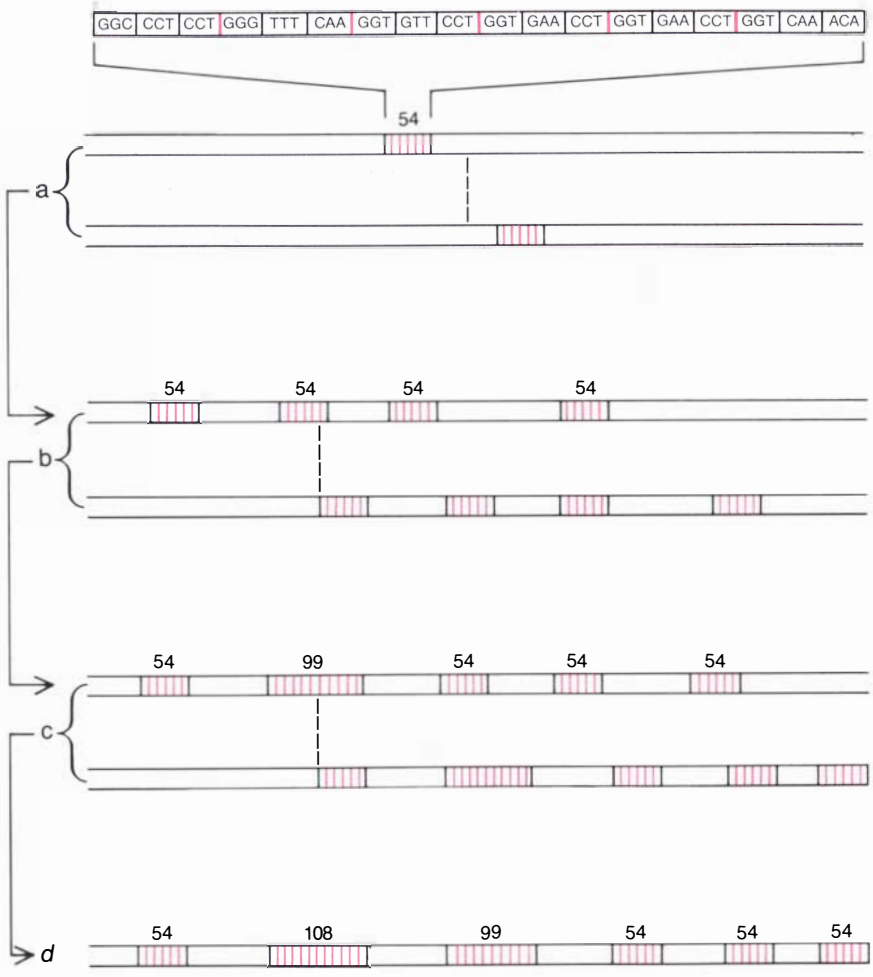
The genetic material of certain sea urchins gives evidence of the process. Among different species of sea urchins the genes encoding proteins of the family known as histones differ in their precise nucleotide sequence by an amount roughly proportional to the time since the species separated in the course of evolution. In the species *Psammechinus miliaris,* however, the genes coding for two histones, *H*3 and *H*4, seem to have diverged much more slowly from corresponding genes in related species than the other histone genes have.

A possible explanation is that unknown selective constraints reduce the rate of evolution of the *H*3 and *H*4 genes in *Psammechinus* to a fraction of their rate in other species. An alternative hypothesis is that a cluster of genes coding for *H*3 and *H*4 was transferred to *Psammechinus* from another species, *Strongylocentrotus drobachensis,* less than a million years ago; the original and the duplicate gene cluster have evolved since then at the usual rate. Because the last common ancestor of the two urchin species lived some 65 million years ago, the histone genes not affected by gene transfer have diverged much further.

The actual mechanisms of horizontal gene transfer are not known. Perhaps the agents are the small circular bits of DNA known as plasmids, which can carry genetic material from one cell to another. Whatever its mechanism, horizontal gene transfer cannot be common. The genes of an organism are coadapted: new or altered genes are favored by natural selection not only because of the function they perform in isolation but also because they complement other genes. Genes are like the players in an orchestra: however virtuosic they may be as soloists, they must also play the same piece or the effect will be cacophony. The coadaptation of the genome reduces the incorporation of functional foreign genes, if it occurs at all, to the realm of rare evolutionary events.

Variation

Molecular biology has yielded insights into not only genetic change over time but also the precondition



EVOLUTION OF A GENE coding for collagen in the chicken took place as an exon (a gene segment coding for amino acids) was multiplied over many generations. The sequence of 54 nucleotides making up the exon is listed at the top; *A* is for the nucleotide base adenine, *C* is for cytosine, *G* is for guanine and *T* is for thymine. A roughly similar series of nine nucleotides recurs six times within the exon. Incorrect alignment of matching chromosomes during meiosis (*a*), followed by an exchange of chromosome segments at the site indicated by the broken line, resulted in a new chromosome in which the exon was duplicated; over evolutionary history the process was repeated, producing the chromosomes in *b*. Because of the repetitive structure of the exon, misaligned pairing and exchange sometimes took place within exons, yielding new exons with nucleotide numbers other than 54. The process is shown giving rise to exons of 99 nucleotides (*b, c*) and 108 nucleotides (*c, d*). Only a part of the modern gene is illustrated (*d*), but the same processes operated throughout the gene.

for such evolution: genetic variation. Evolutionists have shown that among populations in identical or similar environments, the rate of evolution is proportional to the amount of genetic variation within each population [see "The Mechanisms of Evolution," by Francisco J. Ayala; SCIENTIFIC AMERICAN, September, 1978]. The techniques of molecular biology allow the genetic variation within a population to be gauged much more readily than was possible when the synthetic theory was formulated.

Significant measures of the genetic variation within populations became available in the late 1960's. Gel electrophoresis (in which proteins are embedded in a gel and their mobility in an electric field is compared) and other simple techniques made it possible to determine how many forms of a protein exist in a given species, and in what frequencies. The variation can be expressed as the degree of heterozygosity: the proportion of gene loci in an average individual where the two members of a pair of genes, one gene from each parent, encode different proteins. In the several hundred species studied by electrophoresis and other methods the proportion of heterozygous loci detected ranges from 5 to 20 percent.

The amount of variation implied by such studies turns out to be much larger than expected. One reason is that as the various alleles (the distinct forms of a variable gene) are assorted into gametes in the course of meiosis, an individual heterozygous at n loci can give rise to 2^n different kinds of gametes. If an organism with 10,000 gene pairs is heterozygous for 10 percent of them, it could produce $2^{1,000}$ different gametes, a number much larger than the number of atoms in the universe.

Moreover, the amount of genetic variation in a population is usually larger than the results of electrophoresis suggest. Because the technique relies on the differential mobility of proteins in an electric field, differences in the amino acid sequence of a protein that do not change its net electric charge may not be detected. Furthermore, not all nucleotide differences in the DNA sequence result in different protein sequences. The reasons are that different triplets of nucleotides can code for the same amino acid and that some nucleotide sequences do not code for amino acids at all.

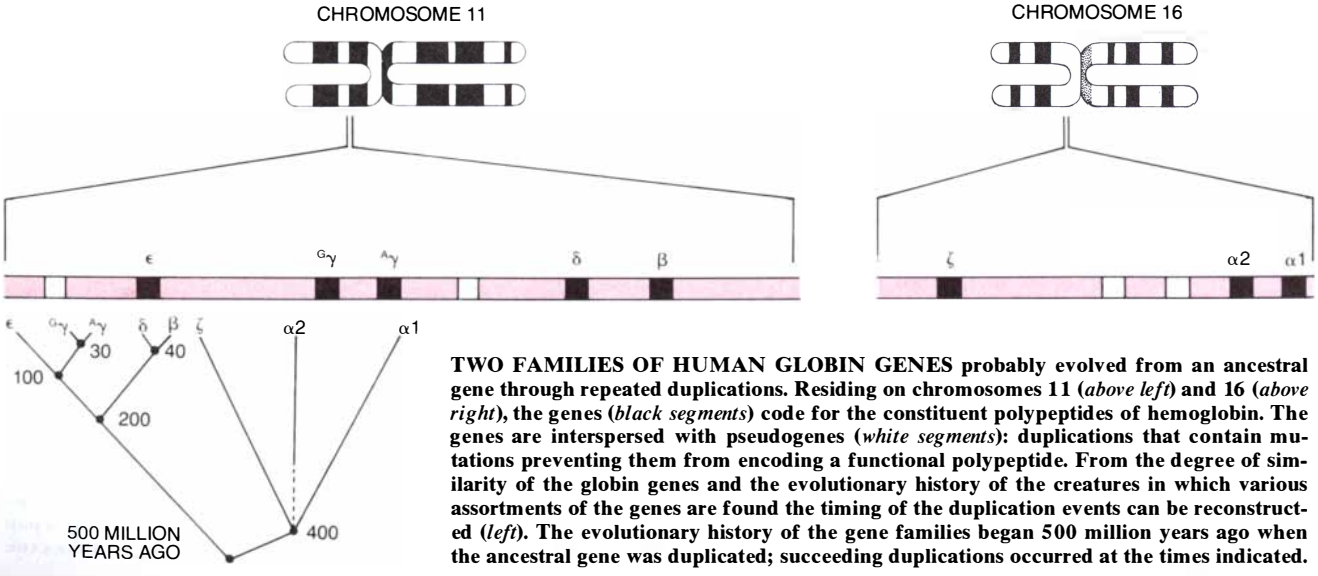
One way to detect some of the genetic variation thus hidden from ordinary electrophoresis is to digest each polypeptide chain of a protein into its component peptides (sequences of a few amino acids) and then compare the individual peptides by means of electrophoresis or chromatography. Differences in electric charge and other properties that were concealed when the peptides were incorporated in the longer polypeptide chain are thereby revealed. One of us (Ayala) applied the technique to two enzymes, alcohol dehydrogenase and superoxide dismutase, from the fruit fly *Drosophila*. In both cases about one in 10 polypeptides that had appeared to be identical when they were tested by electrophoresis turned out to be different in amino acid sequence when they were digested into peptides.

The most direct way to determine genetic variability is to obtain the nucleotide sequence of the same gene in different individuals of a species. This has been done for only a few genes. In 1980 two human genes coding for the same hemoglobin protein were sequenced in the laboratory of Oliver Smithies at the University of Wisconsin at Madison. Although the genes encoded the same product, they differed in their nucleotide sequences by .8 percent if only actual substitutions of one nucleotide for another were counted and by 2.4 percent if nucleotides present in one gene but not in the other were also included. Other genes that have been sequenced, from human beings, mice and fruit flies, have shown comparable degrees of variation. Such results suggest that at the level of the DNA sequence, organisms may be heterozygous at almost all gene loci.

The Neutral Theory

The extensive variation such studies have revealed is one of the bases of the neutral theory, the second of the challenges to the synthetic theory that have emerged from molecular biology. Its chief exponent is Motoo Kimura of the National Institute of Genetics in Japan [see "The Neutral Theory of Molecular Evolution," by Motoo Kimura; SCIENTIFIC AMERICAN, November, 1979]. Neutralists contend that if most genetic differences are under the control of natural selection (as the synthetic theory implies), the observed degree of variation should be low because adaptively beneficial differences should spread through the population and harmful ones should be eliminated. The finding of considerable variation suggests to neutralists that most of the genetic differences neither foster nor hinder an organism's survival and that their persistence or elimination within a population is a matter of chance.

Yet the magnitude of the variation also makes it possible to reconcile the assertion that most differences among alleles are adaptively neutral with the synthetic view, which affirms the importance of Darwinian natural selec-



TWO FAMILIES OF HUMAN GLOBIN GENES probably evolved from an ancestral gene through repeated duplications. Residing on chromosomes 11 (*above left*) and 16 (*above right*), the genes (*black segments*) code for the constituent polypeptides of hemoglobin. The genes are interspersed with pseudogenes (*white segments*): duplications that contain mutations preventing them from encoding a functional polypeptide. From the degree of similarity of the globin genes and the evolutionary history of the creatures in which various assortments of the genes are found the timing of the duplication events can be reconstructed (*left*). The evolutionary history of the gene families began 500 million years ago when the ancestral gene was duplicated; succeeding duplications occurred at the times indicated.

tion. There is so much variation within each generation that even if the majority is neutral, the differences that do have an adaptive effect would supply abundant raw material for the creative force of natural selection.

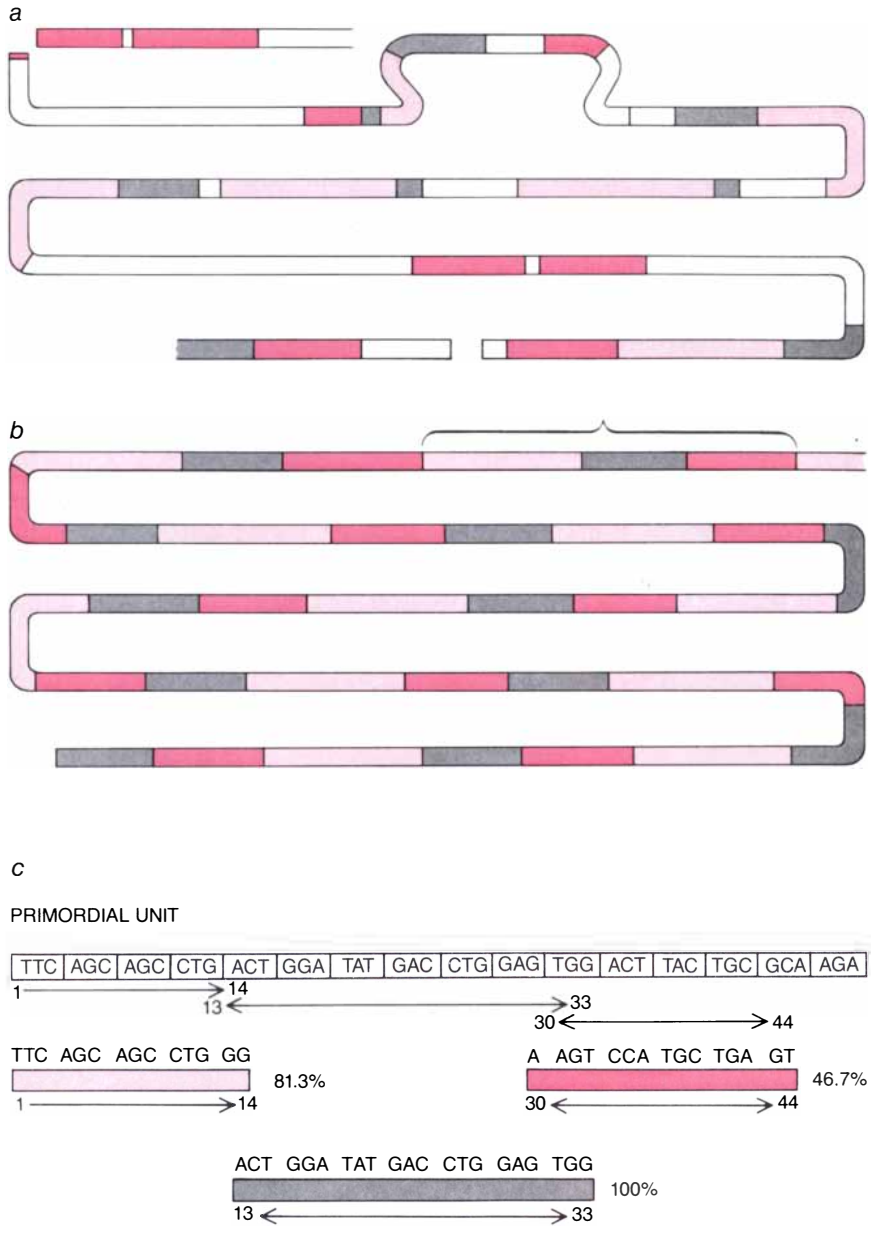
The question, then, is not which is correct, neutrality or Darwinian natural selection. It is instead how much of the genetic variation persists through chance and how much persists because it confers an adaptive advantage. For example, a mutation of the third nucleotide in a codon (the triplet of nucleotides coding for a particular amino acid) often produces a new codon that is synonymous in translation: both the original and the altered triplet code for the same amino acid. Such "silent" mutations, which do not affect the protein encoded by the gene, might indeed be adaptively neutral, and their frequencies in populations may be due largely to chance. On the other hand, the frequency of the mutation causing the single amino acid difference between the hemoglobin of sickle-cell anemia and normal hemoglobin is clearly under the control of natural selection. When an individual has two copies of the sickle-cell allele, the mutation is fatal, but heterozygosity (possession of one mutant and one normal allele) allows the carrier to be reasonably healthy and confers an adaptive advantage: resistance to malaria.

The Molecular Clock

The neutral theory is based not only on the amount of variation prevailing within a given generation but also on discoveries about the rate of genetic change across generations. Studies relating evolutionary history to the number of molecular differences in DNA sequences or proteins common to several species suggest that a given gene or protein can be regarded as a molecular clock. Its rate of evolution is fairly constant over long periods, and it evolves at much the same rate in different species.

This apparent constancy, the neutralists argue, is incompatible with the notion that molecular change reflects the activity of natural selection. They maintain that the synthetic theory implies variable rates of molecular evolution because selective pressures should vary in intensity over time and from species to species. The chance incorporation of variations into the gene pool of a species is a better way to account for the molecular clock, say the neutralists, because the process would take place at a more or less constant rate.

The synthetic theory does not, however, require the rate of molecular evolution to be as irregular as the critics assume. As long as the function of a gene or a protein is the same in different evolutionary lineages it is not surprising that over millions of years it would appear to evolve at the same rate, since the functional constraints to which it is subject will be much the same. Histone proteins evolve very slowly because they act as a structural support for DNA, and extensive amino acid changes would impair their function; fibrinopeptides, substances involved in blood clotting, evolve much faster because they are not subject to such tight constraints and because the clotting process is often modified in response to environmental



STRUCTURE OF A GENE suggests its evolutionary history. In the mouse IgV_H gene, which encodes a component of antibody molecules, segments in the same color represent sequences of DNA sharing structural similarities (*a*). Breaks near each end of the gene indicate junctions with sequences that do not code for amino acids; the loop represents an intron, or noncoding sequence within the gene. Through statistical analysis Susumu Ohno of the City of Hope Medical Center and his colleagues determined the structure of the ancestral sequence of DNA (*b*) from which the modern gene evolved. The ancestral sequence consisted of 12 repetitions of a unit of DNA (*bracketed*) that in turn was made up of three subunits 14, 21 and 15 nucleotides long. Similarities among the subunits as they exist in the current gene suggest they represent a threefold repetition of an even more ancient segment of DNA. The nucleotide sequence of the primordial unit (*c*) was inferred from the fragments persisting in the current gene. Three subunits from the modern gene are shown next to the corresponding sections of the primordial unit, and the degree of similarity between each modern subunit and the matching part of the reconstructed ancestral unit is indicated.

changes. The constraints on the evolution of most molecules are intermediate between these two extremes.

The constancy over time of rates of molecular evolution can also be reconciled with the synthetic theory. The number of nucleotide or amino acid substitutions may indeed be greater during a given interval as a result of natural selection, when an evolutionary lineage is becoming adapted to a new environment that requires simultaneous changes in many functions. Chance itself can result in abrupt increases in the rate of genetic change. When a species passes through a population bottleneck in which the number of individuals is drastically reduced, virtually neutral variations present in the genes of the survivors may become dominant in the genetic makeup of the population as a whole.

If the fundamental function of a gene or a protein does not change in the course of evolution, however, there is no reason to expect that fluctuations in the rate at which it evolves will be frequent or extended. The enormous lengths of time over which rates of molecular evolution are calculated cause the fluctuations that do occur to average out, yielding the apparent constancy of evolutionary rates. John Gillespie of the University of California at Davis has constructed mathematical models showing the existence of a molecular clock is compatible with the assumption that molecular evolution is driven by natural selection.

Punctuated Equilibrium

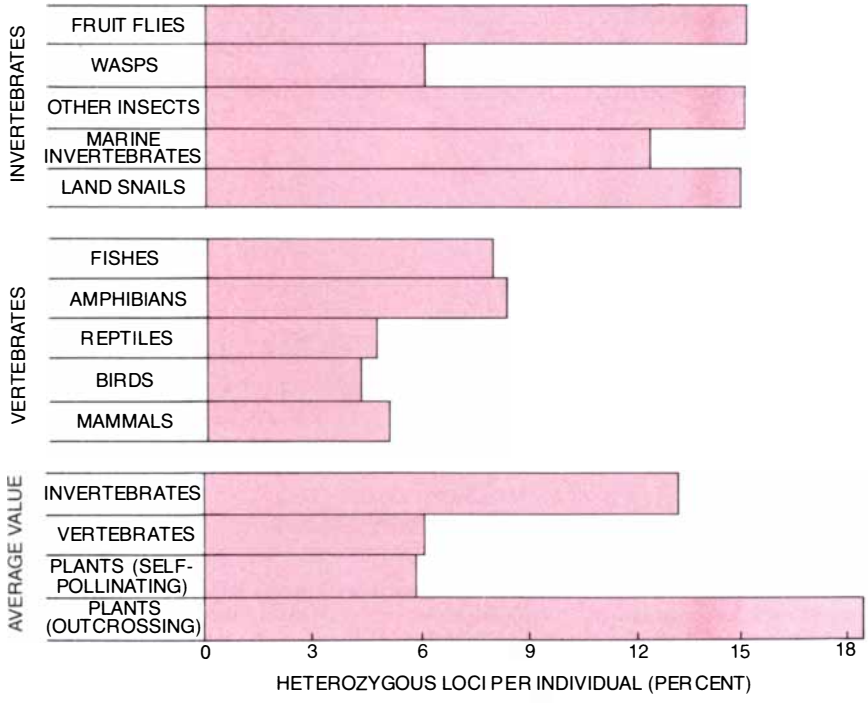
A different kind of question about evolutionary rate is central to the conflict between the synthetic theory and the theory of punctuated equilibrium, advanced by Niles Eldredge of the American Museum of Natural History and Stephen Jay Gould of Harvard University. In this case the argument centers on morphological evolution, in which visible characteristics are changed, rather than on evolution at the molecular level, and in this dispute the positions are reversed. Whereas the neutralists contend that the rate of molecular evolution is too regular for the synthetic theory to hold, the punctualists argue—from fossil evidence—that the rate of morphological evolution is less regular than the synthetic theory requires.

Eldredge and Gould reject the explanation Simpson and others had offered for the rarity or absence in the fossil record of specimens that are intermediate in morphology between successive fossil forms: that the record is incomplete. Instead they take the record at face value and maintain that the abrupt appearance of new fossil species reflects their development in bursts of evolution, after which the species may have changed little over millions of years. In their view the notion of gradualism must be replaced by a picture in which spurts of change alternate with long periods of stasis.

The dispute with the punctualists loses some of its focus when one recognizes that it is partly an artifact of a radical difference in time scales: the time scale of the paleontologists who propose the theory of punctuated equilibrium and that of the geneticists who were instrumental in formulating the synthetic theory. Since successive layers in geologic strata may have been laid down tens of thousands of years apart, morphological changes that developed over thousands of generations may make an abrupt appearance in the fossil record. In contrast, geneticists refer to changes that require 200 generations or more as gradual, since they exceed the time span of all experiments except those on microorganisms. In speaking on the one hand of sudden change and on the other of gradual evolution, the punctualists and the gradualists are in many cases talking about the same thing.

The apparent episodes of abrupt change in the fossil record, then, do not necessarily weigh against the synthetic theory and its emphasis on gradual processes. Can the same be said for the long periods of stasis the punctualists point to, in which the visible characteristics of a species change either little or not at all? Taking a closer look at the fossil record, many paleontologists outside the punctualist camp agree that the stability of visible characteristics over millions of years is much greater than the architects of the mid-century synthetic theory thought at the time. Some of the apparent constancy in the record, however, may reflect a phenomenon known as mosaic evolution, first described by the British evolutionist Sir Gavin de Beer.

The term refers to the fact that different parts of an organism do not change at a uniform rate in the course of evolution. Just as genes and proteins do not evolve in lockstep, so the visible characteristics they govern do not evolve at the same rate either. A fossil that lies halfway along a temporal sequence from one organism to the next is not intermediate between the two with respect to every trait. Instead the creature can be compared to a mosaic, resembling the ancestral organism in some characteristics but the descendant in others. *Archeopteryx,* the fossil species intermediate between the rep-



AVERAGE AMOUNTS OF VARIATION detected in the gene pools of species from different groups of organisms are compared. The measure of variation is the percentage of gene loci in an average individual that are heterozygous: those in which the two members of the gene pair differ in their nucleotide sequence. In general more variation is found among invertebrates than among vertebrates. Among plants, outcrossing species (in which individuals exchange pollen and thereby mix genes) show greater variation than plants that pollinate themselves. The true percentages of heterozygous loci may be somewhat higher than what is shown; the technique used here, gel electrophoresis, fails to detect some variation.

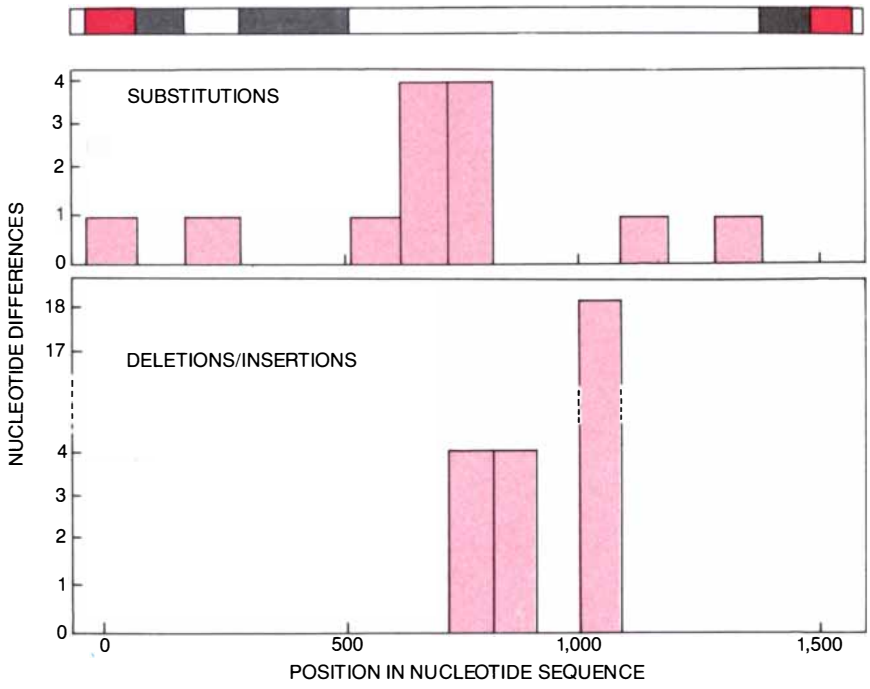
tiles and the birds, has a skeleton considered to be fully reptilian, but its feathers are birdlike, even to the differentiation between fluffy feathers on the body and longer, stiffer wing feathers. The australopithecines (creatures of two to four million years ago that are ancestral or very nearly ancestral to human beings) resemble humans in that their hipbones are shaped for walking erect, but their cranial capacity is more simian than human.

Some of the stasis apparent in the fossil record may be a product of this evolutionary pattern. When the evolution of a single salient trait is followed in the fossil record, long periods of little change may be evident. While one trait is becalmed by evolution, however, other characteristics that may not be recorded in the fossils can continue to evolve. The punctualist argument may be more convincing with respect to single traits than it is with respect to entire organisms.

Morphology and Speciation

Along with a distinctive claim about the rate of evolution the theory of punctuated equilibrium also makes a specific argument about the mode of evolution. Eldredge and Gould believe anagenesis (changes in morphology along a lineage) and cladogenesis (the splitting of a single species into two species) are causally associated. They maintain that there is a brief burst of morphological change precisely when a small population diverges from an original species and forms a new species; little change then ensues until the species either goes extinct or gives rise to new ones. The opposing view, which the punctualists associate with the synthetic theory, is that gradual morphological change occurs within a species, dividing it into races and subspecies, long before new species can be said to have formed.

Actually evolutionary change follows both patterns (and many others). Numerous studies have shown that morphological change and the development of reproductive isolation (the condition that defines a new species) are genetically distinct phenomena: they can occur either together or separately. Arne Müntzing of the University of Lund showed many years ago that the genes coding for morphological differences among species of flowering plants are not linked with those that interfere with the fertility of hybrids and thereby keep species reproductively distinct. He crossed two species of hemp nettle. The resulting hybrids were partially fertile, enabling Müntzing to produce a second generation of hybrids.



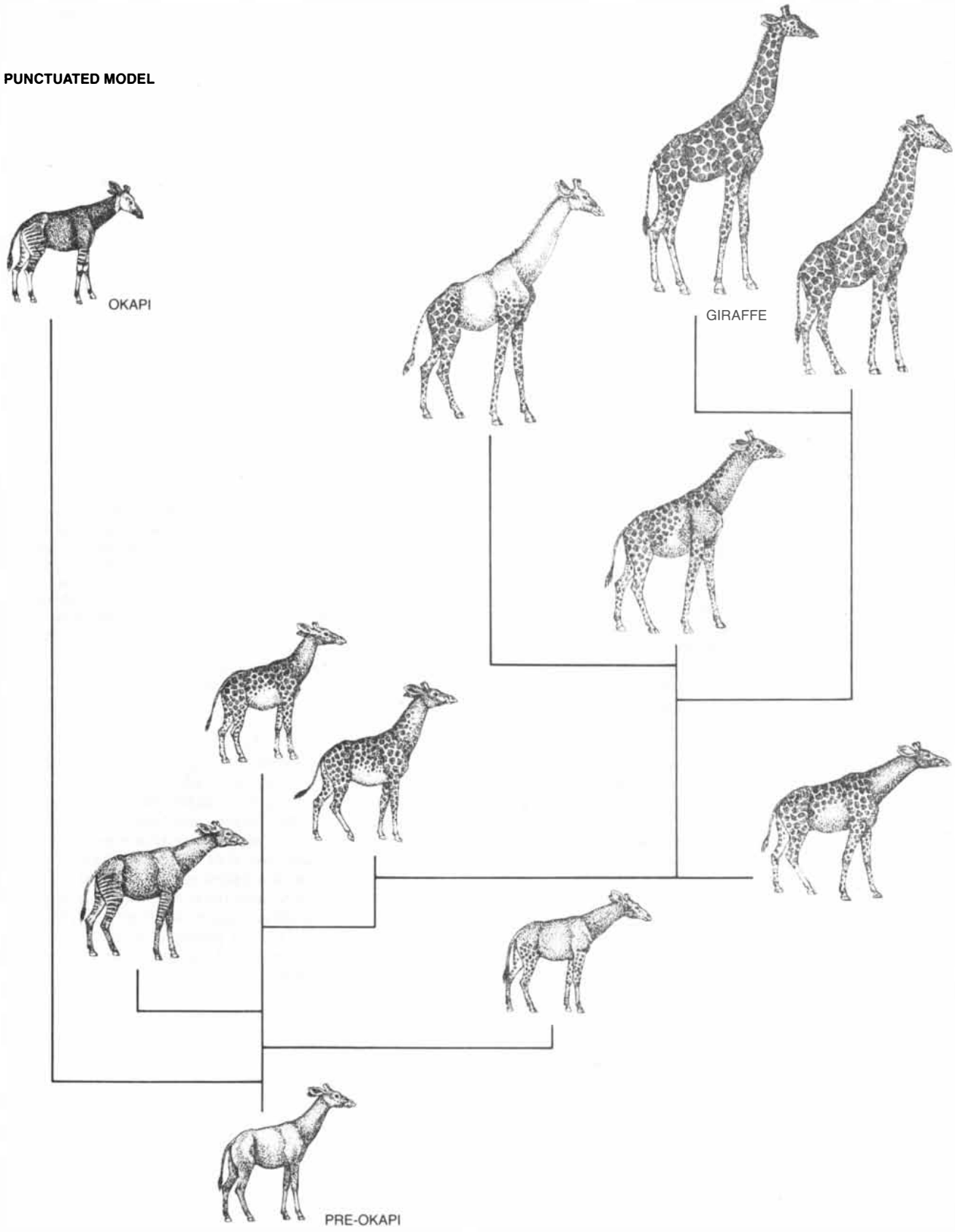
NUCLEOTIDE DIFFERENCES were revealed in a comparison of the DNA sequence of the human $^{A}\gamma$ globin gene from two different individuals. At the top the components of the gene are represented: flanking sequences of DNA (*color*), which do not code for amino acids but contain instructions governing the transcription of the gene, exons (*gray*) and introns (*white*). The bar graphs indicate the kind, number and location within the gene of the nucleotide differences. In a substitution different nucleotides occupy a particular site in the DNA sequence; in a deletion or an insertion a nucleotide present in one variant of the gene is absent from the other. The largest number of nucleotide differences fall within the introns, where mutational changes do not affect the amino acid sequence encoded by the gene and therefore can accumulate largely unconstrained by natural selection. The exons show no nucleotide differences, and so the gene is functionally identical in both individuals.

Some of the second-generation hybrids were morphologically intermediate between the parent species, suggesting they had inherited genes governing morphology from both species. In these intermediate species, however, the genes that maintain reproductive isolation appeared not to have been inherited with the genes for morphology, because a few of these second-generation hybrids were completely fertile. Other hybrids, which were not intermediate morphologically but instead resembled one of the parent species, were sterile. Since Müntzing's work similar results have been obtained in other plants.

Morphological differentiation and the reproductive isolation that distinguishes species develop independently in animals as well. In the continental U.S. many sibling species of *Drosophila* are morphologically alike but are reproductively isolated. In the Hawaiian archipelago, where several hundred species of *Drosophila* have arisen, the pattern is quite different. There are conspicuous morphological differences between species that are little differentiated genetically and even between geographically separated, interfertile populations of the same species. Many groups of mammals also exhibit this disjunction between morphological evolution and chromosomal differentiation or other modes of reproductive isolation.

These and numerous other examples indicate that the conditions leading to innovative morphological adaptations and those that favor speciation need not occur together. Morphological change proceeds in response to environmental challenges, such as exposure to new predators or the opening of a new habitat, whereas the splitting of a population into new species occurs as a consequence of population structure: patchy distribution in spatially restricted habitats, for example.

Both kinds of circumstance do often occur at once or in close succession, thereby bringing about the linkage of speciation with morphological change that is central to the theory of punctuated equilibrium. The fossil snails Gould studied from Pleistocene (Ice Age) strata in Bermuda may fit the pattern. They seem to define a branching phylogeny in which each branch represents a new species as well as a recognizable change, although not a profound one, in shell morphology. One must recognize the possibility of circu-



CONTRASTING MODELS of evolutionary processes give rise to two scenarios for the differentiation of the okapi and the giraffe from a common ancestor. In the punctuated model (*left*) morphological change occurs in spurts that accompany the formation of new, reproductively separate species. Thus the evolutionary route from the ancestral pre-okapi to the modern giraffe forks repeatedly as small, isolated populations rapidly evolve into distinct species, each representing a new morphological variant. The okapi developed from the ancestral species in similar but more direct fashion: in a single burst of evolution an isolated population developed okapi morphology and then ceased to change. In the gradualistic model (*right*) slow morphological change takes place continuously with-

lar reasoning, however. A new species is identified each time morphological change is apparent along a temporal sequence of fossils. Hence speciation and morphological evolution are necessarily associated. Actually speciation may have taken place, unrecorded in the fossils, between the recognizable changes, and some of the visible differentiation may define populations that were interfertile and so do not qualify as separate species.

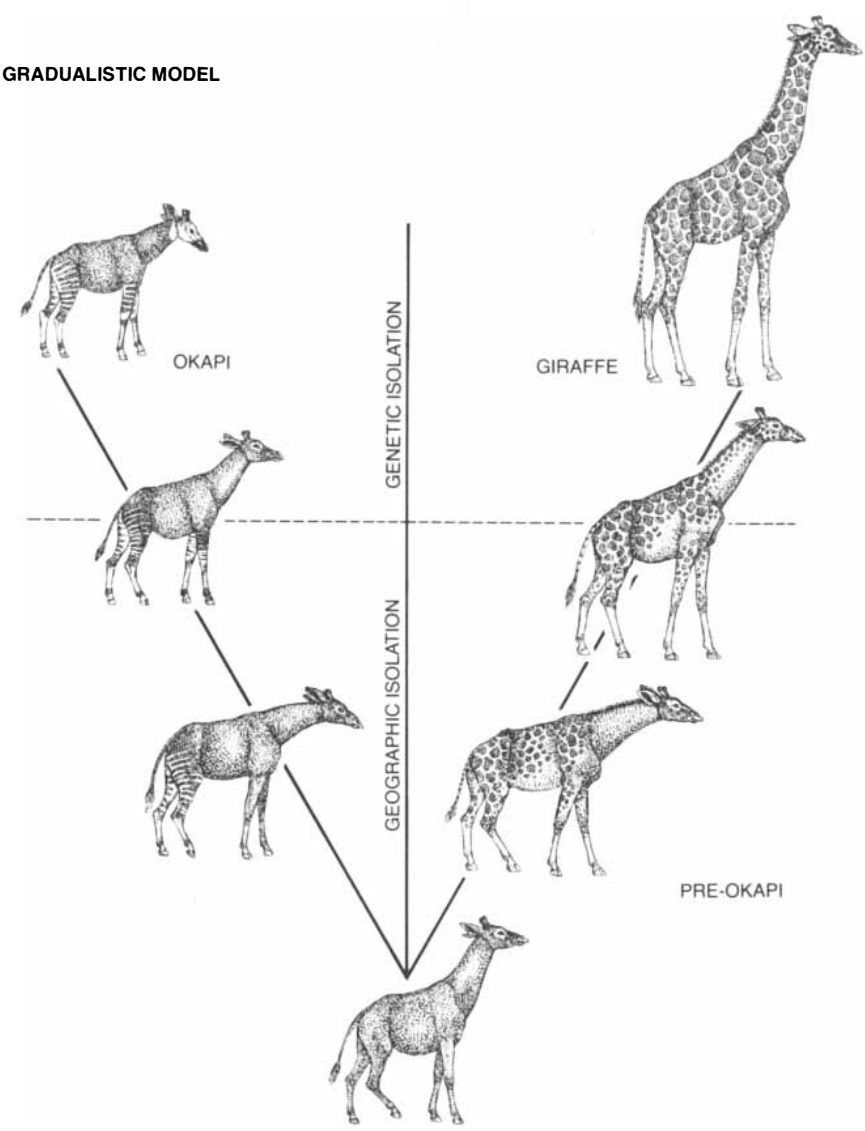
Species Selection

In keeping with their contention that evolution within a species is largely irrelevant to evolutionary trends, the punctualists have embraced a concept known as species selection. Species selection expands the province of natural selection (which is ordinarily held to act on individuals or, more precisely, on genes) to include entire species. The concept assumes that a species with a certain attribute will be more likely to persist and give rise to new species than one lacking the attribute. In a given group of organisms, for example, species with large individuals might be more likely to "bud off" new species, or less likely to become extinct, than species with small individuals. Over evolutionary time the average size of the organisms within the group of species would increase. The increase would result not from evolution toward a larger size within any given lineage but from an increase in the number of species that have large individuals.

There is a fundamental difference, however, between the comparative fates of species and the spread of genes and gene combinations within populations. Since the alleles of a gene contain alternative bits of information, only one of which can be present at a chromosomal locus, an increase in the number of individuals carrying one allele of a gene necessarily entails a reduction in the frequency of the other alleles. Alleles can therefore be said to compete, through their adaptive effects on their carriers, for prevalence in the gene pool of a population.

For species the situation is different. Related species do not usually compete directly. Morphologically similar sibling species of *Drosophila,* for example, exist side by side, exploiting the environment in slightly different ways. The origin and spread of a species does not necessarily cause other species in the same lineage to decline or become extinct. When a species does become extinct, it is usually not because of competition with related species but because its gene pool fails to become adapted to a changing environment. Environmental challenges may include competition with other species, to be sure, but those species are often entirely unrelated to the one that becomes extinct. Insofar as it implies a parallel to the natural selection of genes, then, the notion of species selection is of dubious worth.

It is true, as Simpson recognized some 40 years ago, that evolutionary trends in morphology are the result not only of genetic evolution within species but also of the differential survival and multiplication of species. The former process is the fundamental one, however. The fate of species depends on the ability of the individuals making up the species to cope with the enviroment, and such an ability can only result from the natural selection of genes.



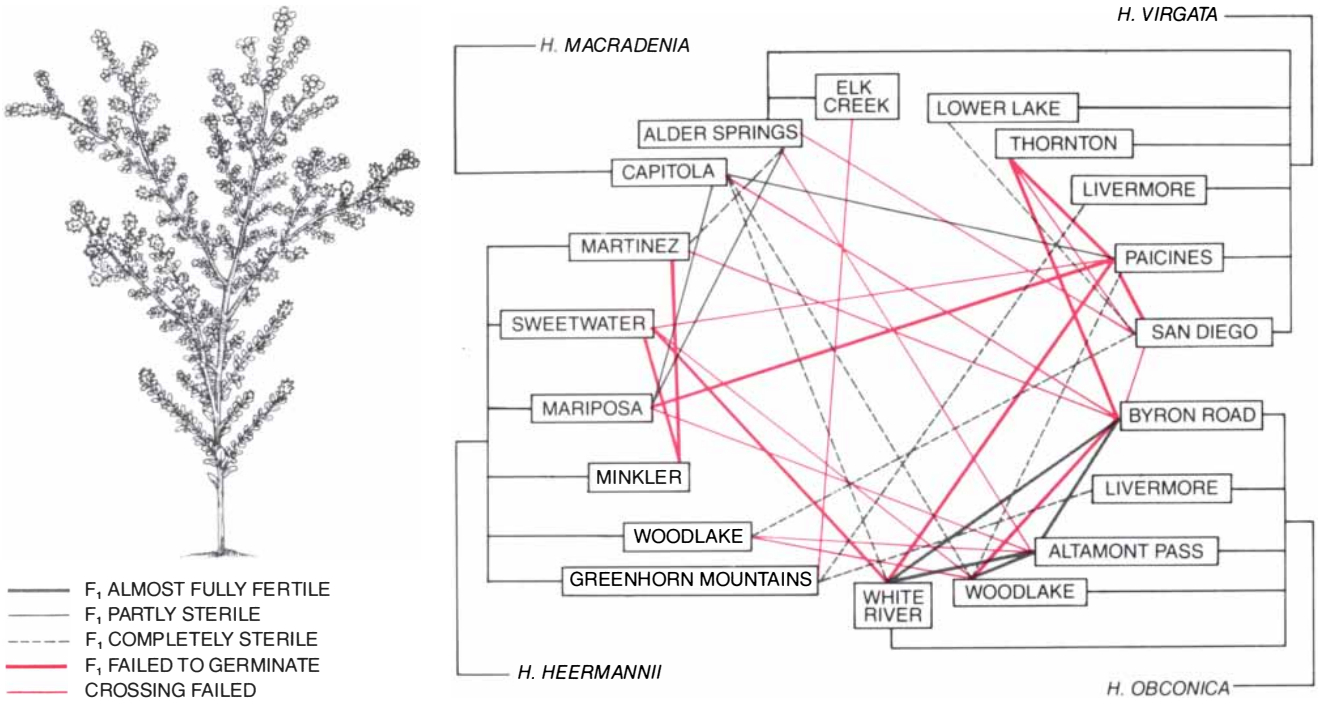
in single species. Accordingly the morphological divergence of the okapi and the giraffe began long before they developed into separate species and continued long after. It was initiated when the pre-okapi species was divided into two populations by a geographic barrier; eventually enough genetic differences accumulated to render the two groups incapable of interbreeding, thereby qualifying them as separate species. Actually evolution has followed both patterns, and both probably figured in the development of the okapi and the giraffe.

A New Synthesis

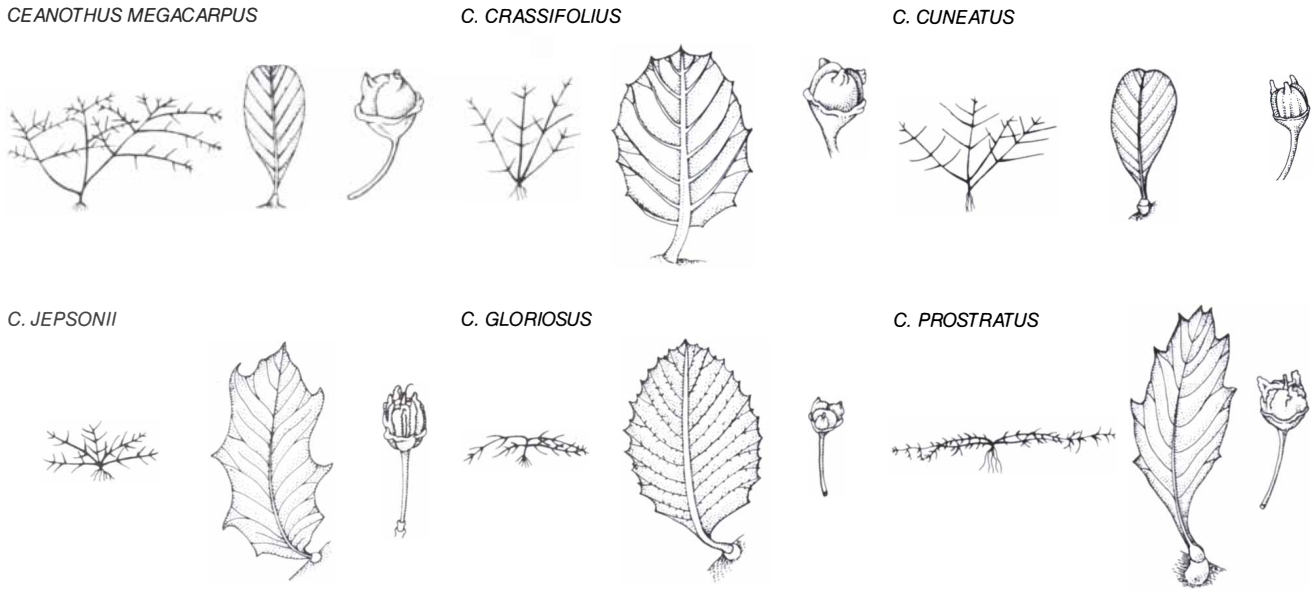
How has the mid-century synthesis weathered recent developments in research and theory? The new molecular biology, by showing that the evolutionary process at the level of DNA is far more complex than had been thought, casts doubt on some old certainties. It has also, however, provided the beginnings of answers to a fundamental matter the architects of the synthetic theory could not address: how genetic information accumulates over

evolutionary history. Neutralists and selectionists, strange bedfellows at first glance, can both retain their basic postulates within the harmony of a more comprehensive theory, one that allows chance a greater role in genetic change. Some of the tenets of punctuated equilibrium can be refuted; others are compatible with a modified synthetic theory that encompasses the notions of species stasis and mosaic evolution.

Whatever new consensus emerges from ongoing research and controversy, it is not likely to require rejection of the basic tenets of Darwinism and the mid-century theory. The synthetic theory of the 21st century will differ considerably from the one developed a few decades ago, but the process by which it emerges will be one of evolution rather than upheaval.



FAILURE TO INTERBREED in spite of close morphological resemblances was observed in several species of tarweed, genus *Holocarpha* (*left*). The crossing polygon (*right*) indicates the outcome of crosses among 19 populations of four species from various locations in California: in many cases no hybrids could be produced, and in other instances the first-generation hybrids (F_1) were sterile. In contrast to the punctuated model of evolution, which holds that distinctions in morphology and the reproductive barriers that define separate species are closely associated and evolve simultaneously, the tarweed has evolved into separate species with few visible changes.



MORPHOLOGICAL DIFFERENCES among six species of buckbrush, genus *Ceanothus*, are exemplified by growth habits, leaves and seed capsules. Like the tarweed (*above*), the buckbrush shows that reproductive barriers and morphological differences need not evolve together: the six visibly different species can be crossed successfully, yielding fertile hybrids. Although an inability to interbreed is the usual criterion for distinguishing species, the species of buckbrush are considered to be separate because of their considerable differences in morphology, their adaptation to various climates and soils and the fact that they do not interbreed in nature.

Introducing The Single Solution To Many Problems That Can Tie Up Your Personal Computer.

You need to use your computer for a critical job, but can't—because it's tied up running your printer or modem. Or you're rushing to compose and print a report, when you realize the wrong printer is hooked up, and you have to stop and swap cables.

Fortunately, there's a simple solution to these problems—and many more. It's a remarkable new device called Transet 1000, from Hayes.

Transet 1000 relieves your computer from operating your printer or modem, so you and your computer can go on to another task. As fast as your computer can send it, Transet 1000 receives and stores more than 75 pages of text. Then it feeds that data to your printer or modem, at the speed at which the peripheral can accept it. At the same time, you can be using your computer for an entirely different job.

While performing other functions, Transet 1000 can receive data through a modem. In fact, it operates as an unattended electronic mailbox, receiving messages 24 hours a day. Even if the computer's turned off.

Transet 1000 automatically routes information among peripherals, so you don't have to physically change connecting cables. It also allows two computers to share an expensive peripheral, like a laser printer.

And Transet 1000 works with a wide range of systems and configurations.

But these are just a few of the problems Transet 1000 can solve for you. For more information, see your authorized Hayes dealer. Soon.

Because you should be working with your computer. Not waiting for it.

Hayes Transet 1000™

Innovative products for enterprising people

Hayes Microcomputer Products, Inc.
P.O. Box 105203, Atlanta, GA 30348
404-441-1617

Tomorrow's buggy whips: **With the sophisticated tools shown here a communications network could be designed in a few weeks—and it could be obsolete before it got off the drawing board. Contel Information Systems' MIND* system now harnesses computer**

power to get the same job done in minutes. Minutes vs. weeks—and done better, too! No wonder more than 800 network pros have become MIND converts. They get easy-to-update network designs that precisely balance high performance and minimum costs. Continental Telecom and its subsidiaries lead in this new era of network services with the equipment and design tools to make them work with increased efficiency. From telephony to satellites. Architects of telecommunication.

FOR A FREE BROCHURE ON ALL OF CONTEL'S CAPABILITIES WRITE TO:
CONTEL, DEPT. 509, 245 PERIMETER CENTER PARKWAY, ATLANTA GA 30346 © 1984 CONTINENTAL TELECOM INC.

CONTEL

*MIND℠—MODULAR INTERACTIVE NETWORK DESIGNER

The Chemosensory Recognition of Genetic Individuality

Genes regulating immunologic function impart to individual mice a characteristic scent. A mouse can thus discriminate genetic differences among its potential mates by smell alone

by Gary K. Beauchamp, Kunio Yamazaki and Edward A. Boyse

For human beings, a species whose dominant senses are sight and hearing, the sense of smell seems to be little more than a luxury. It is hard to imagine how important it may be for members of other species to recognize one another by scent. Our study of such recognition among rodents shows that the individual odor of an animal is in part determined by its genes, notably by genes in a chromosomal region that in all mammals plays a major part in immunologic recognition and thus largely decides the fate of transplanted organs and tissues.

The work began at the Memorial Sloan-Kettering Cancer Center in New York as a result of incidental observations on the social behavior of mice making up pedigreed breeding colonies. Specifically, it began with the observation that mice differing genetically only at a chromosomal region occupied by a string of genes with immunologic functions, known as the Major Histocompatibility Complex, or MHC (the mouse MHC is known as *H*-2 and the human MHC as HLA), can distinguish one another by scent. At the same time Lewis Thomas, who was president of the center but did not yet know of the observations, proposed that the histocompatibility genes might impart to each individual a characteristic scent. The fortunate coincidence of observation and theory was the starting point for a series of studies, now conducted chiefly at the Monell Chemical Senses Center of the University of Pennsylvania, on the olfactory recognition of genetic identity.

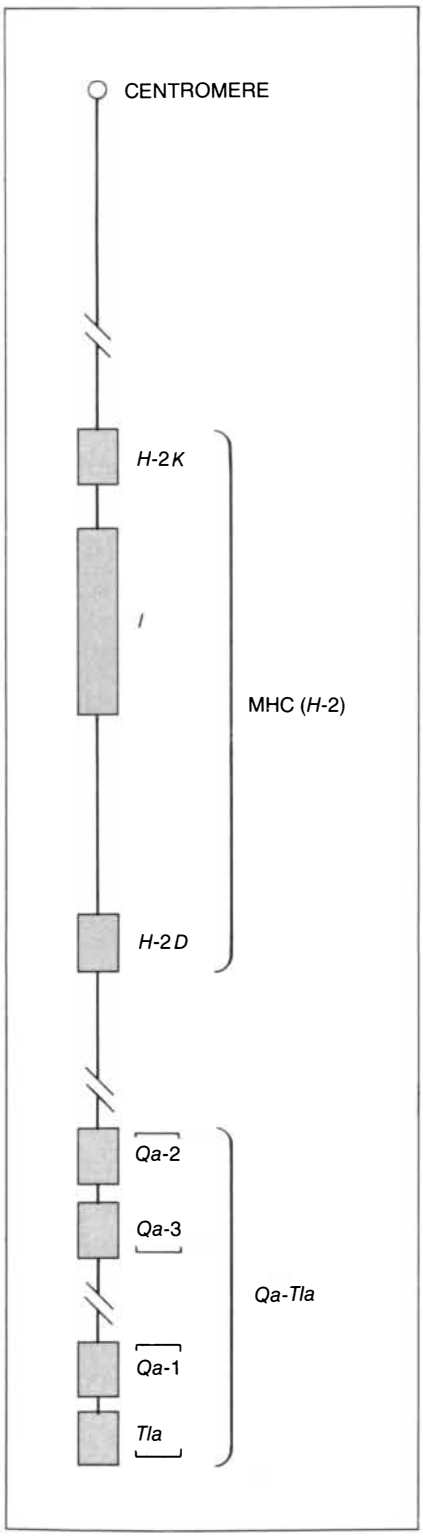
The importance of the MHC group of genes can be gauged from the fact that a similar set of genes is thought to operate in all vertebrates. In man the complex is best known for its influence on transplanted kidneys and other organs and tissues: the donor and the recipient of the transplant should have similar HLA types, or the transplant will be recognized as foreign and therefore attacked and rejected by the recipient's immune cells, called lymphocytes. (The immune cells can respond to a vast range of chemical information from the environment, providing protection against all kinds of infection. Olfaction too involves the recognition of a multitude of chemical signals, by cells of the nervous system.)

Some of the genes composing the MHC, notably the genes called *H-2K* and *H-2D,* are extremely mutable. They are, so to speak, genetic hot spots: places in the genome where mutations occur with remarkable frequency. Hence there are many alleles, or alternative forms, of these genes. As a result two unrelated individual animals are extremely unlikely to have identical MHC types. That is the main reason it is difficult to find MHC-matched human organ donors and recipients, except within families. The products synthesized by the cells of different individuals from instructions encoded by MHC genes are correspondingly diverse. These products are proteins, which, with the addition of carbohydrate chemical groups, become glycoproteins and are mostly inserted into the outer membrane that forms the surface of a cell. Thus the variability among MHC genes establishes in large measure the biochemical individuality of the cell surface and what P. B. Medawar called "the uniqueness of the individual."

As in many other fields of biomedical research, inbred strains of mice have been of inestimable value in the work described in this article. Inbred strains are derived by serial brother-to-sister matings over many generations, to the point of genetic uniformity [*see illustration on page 89*]. Congenic inbred strains represent a further refinement. They are derived by first crossing one inbred strain to another. The principle is to mate the progeny through many succeeding generations back to mice of the first inbred strain, selecting progeny in each backcross generation for a defined genetic difference such as MHC (*H*-2) type [*see illustration on page 90*]. The result is two strains of mice that differ in only one segment of the genetic material. In the case of the MHC, for example, the congenic strains would be genetically identical except for the region of chromosome 17 (the mouse has 20 pairs of chromosomes) that bears the MHC (*H*-2) group of genes.

Behavioral consequences of the sensory recognition of MHC types were first noticed in 1974. In the mouse-

TRAINING OF A RAT to distinguish between the odors of two congenic, or closely related, strains of mice was photographed at the Monell Chemical Senses Center in Philadelphia. The rat began each trial in the training procedure when it positioned itself to receive a measured amount of air scented with the urine of a mouse of one of two strains (*top*). If the scent was the one for which the rat is rewarded, the rat was required (within four seconds) to press a key near its head (*middle*). The reward was a drop of water (*bottom*). The results show that the sense of smell enables rodents of one genus (the rat) to recognize in another genus (the mouse) genetic differences at the Major Histocompatibility Complex (MHC), a group of genes with vital functions in immunologic recognition and response, including the acceptance or rejection of tissue or organ transplants. The training apparatus was designed and constructed at Monell by Heather Duncan, Burton Slotnick and Charles J. Wysocki.



***H*-2 REGION on chromosome 17 of the mouse contains MHC genes. These include the *H-2K* gene, the *H-2D* gene and the intervening *I* genes, all of which specify the composition of glycoproteins of the cell surface. They in turn contribute to immunologic recognition and response. The nearby *Qa* and *Tla* genes are closely related to *H*-2 genes, both in their DNA composition and in the structure of the cell-surface glycoproteins they specify, but their functions are not known. Brackets in the drawing indicate uncertainties in the precise order of genes.**

breeding rooms at the Sloan-Kettering Center some inbred male mice had each been housed with a pair of *H*-2 congenic females. In that situation each male encountered a genetically identical female and a female of a different *H*-2 type but otherwise genetically identical. It was observed that males and females of dissimilar *H*-2 type were more disposed to nest and associate with each other than males and females of the same *H*-2 type.

These incidental observations were systematically investigated. Single males presented with two females in estrus were watched until one of the females was inseminated. The females were *H*-2 congenic (differing genetically only for *H*-2), one being of the same inbred strain as the male. With several different *H*-2 combinations there was a substantial bias in favor of *H*-2-dissimilar matings.

To verify the genetic basis of this sensory distinction we turned next to a conventional F_2 linkage test [*see upper illustration on page 91*]. First a pair of *H*-2 congenic strains (genetically identical except for *H*-2) is crossed to produce the F_1 (first filial) generation. The F_1 population consists of genetically uniform mice that are *H*-2 heterozygotes: every F_1 mouse carries both the paternal and the maternal *H*-2 alleles. The F_1 *H*-2 heterozygotes are then crossed to give the F_2 generation. The F_2 generation consists of two groups of *H*-2 homozygotes, each genetically identical with the respective inbred grandparent strain, and a group of *H*-2 heterozygotes genetically identical with the F_1 mice. The three groups are identified by reacting their cells with antiserum to the cell-surface glycoproteins specified by particular alleles of *H*-2 genes. (In the same way human organ donors are matched with organ recipients by HLA typing.)

Although the F_2 homozygotes of such a cross are genetically identical with the respective grandparents, they have experienced a different environment with respect to *H*-2. The inbred grandparents were never exposed, before or after birth, to any type of *H*-2 other than their own. In contrast, the F_2 homozygotes were born of *H*-2 heterozygous F_1 parents and so were exposed, before and after their birth, to an *H*-2 type different from their own.

We found that the *H*-2 mating preferences of F_2 homozygotes were not in all cases identical with the preferences of genetically identical mice of the grandparental strains. Our tentative conclusion is that the MHC-related mating preferences of mice are at least in part the result of familial chemosensory imprinting.

By 1978 our initial work was complete: we had confirmed that mice can sense one another's *H*-2 types and that this is the basis of a mating bias favoring partners that differ from themselves at the *H*-2 locus. The bias, if it operates in the wild, should favor outbreeding and *H*-2 heterozygosity, promote the diversity of *H*-2 genes and help to propagate their mutations. The phenomenon of "hybrid vigor" (better fitness of overall genetic heterozygosity) is well known. In the case of MHC heterozygosity and diversity, the advantages may include the capacity to respond to a wide range of antigens (foreign material of any kind that excites an immune response) and improved capacity to adapt the immune system to new threats, such as infectious organisms that can subtly adapt to the immunologic armament of their hosts. Inbreeding, which must tend to reduce the diversity of the gene pool, would seem to be a particular hazard to mice because in nature large seasonal populations derive from a few mating pairs or even a single pair—hence an evident value of a bias toward *H*-2-dissimilar matings.

Observation of mating preferences was far too cumbersome a method for inquiry into the nature and mechanism of genetically determined individuality of odor. We therefore built a Y maze [*see illustration on page 92*]. Differentially scented currents of air are drawn from odor chambers and conducted through the two arms of the Y. Mice are trained to distinguish between pairs of odors. The trainee mouse is deprived of water for a period beforehand. A drop of water is supplied each time the thirsty animal enters the arm scented by one of the odors. The mouse first learns to distinguish the strong scents of juniper and cinnamon. It is rewarded, for example, whenever it enters the juniper-scented arm. Then comes a new task: distinguishing between the odors of mice of two unrelated inbred strains (occupying the odor chambers) that differ from each other throughout the genome. The final task is to distinguish between the odors of *H*-2 congenic strains differing only with respect to the *H*-2 gene complex. A performance score of 80 percent correct responses or better is the criterion for the completion of each phase of training. (Chance responses would give a score of 50 percent.)

The initial Y-maze studies established that males and females could readily be trained to distinguish *H*-2 congenic mice, male or female. It proved to be immaterial whether either or neither of the *H*-2 types pre-

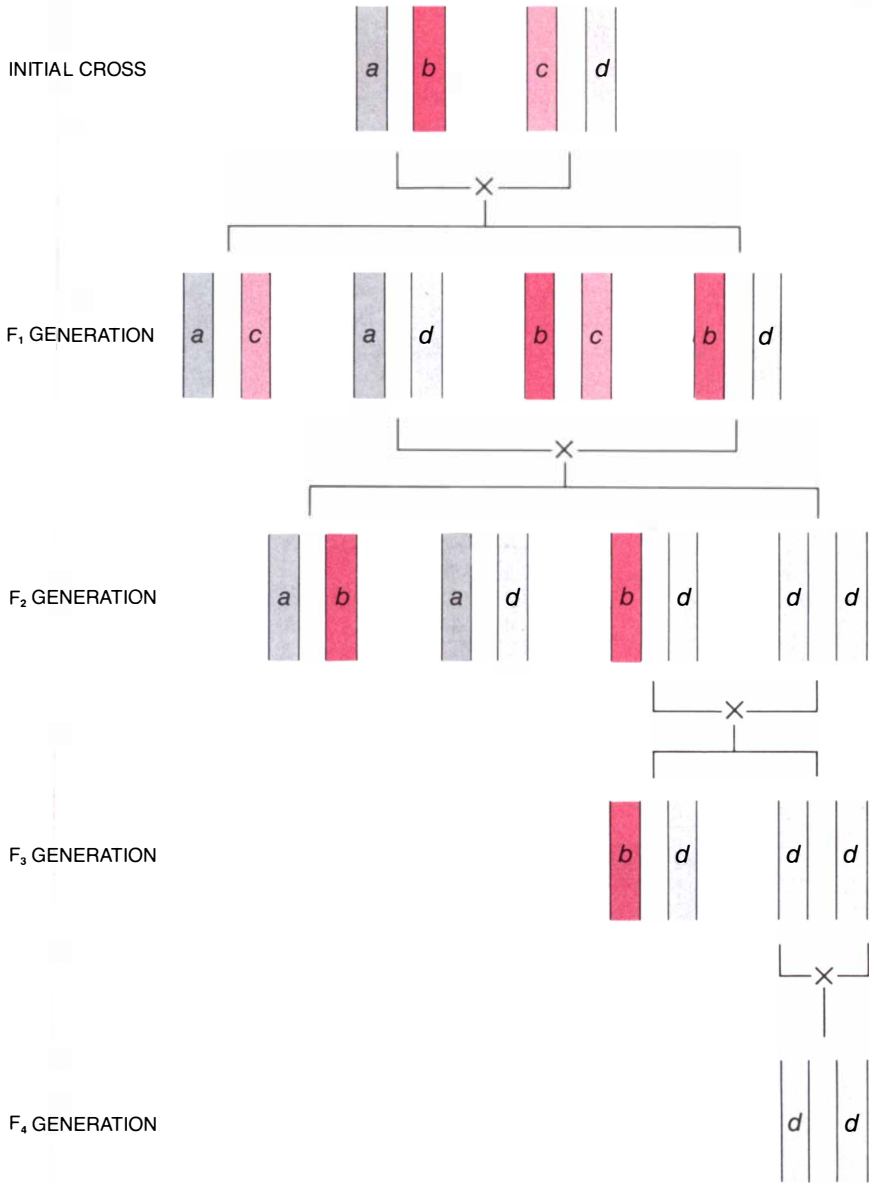
sented to the trained mouse was the same as that of the trained mouse itself. It was further established that the urine of *H*-2 congenic mice could be substituted in the odor chambers for the mice themselves, and urine was used as the test material thereafter. Olfaction, then, is a sufficient basis for the sensory discrimination of *H*-2 types. Other sensory channels are not excluded. For example, ultrasound (by which mice communicate with one another) might conceivably be influenced by *H*-2 genetic differences.

The Y-maze studies took into account individual variations in body and urine scent due to circumstances such as health, diet, age, stress and state of estrus. In order to "tune out" this nongenetic background, we assembled panels of mice to act as urine donors. The mice in any given panel of mice to be compared were matched for age and sex and were maintained under uniform conditions. Their urines were used in rotation, and a particular combination of donors was never used twice in a single series of trials. The placement of alternative urine samples in the odor chambers of the maze was governed by random numbers. Under these conditions mice were readily trained to distinguish between paired panels of *H*-2 congenic mice but could not be trained to distinguish between paired panels of genetically identical mice.

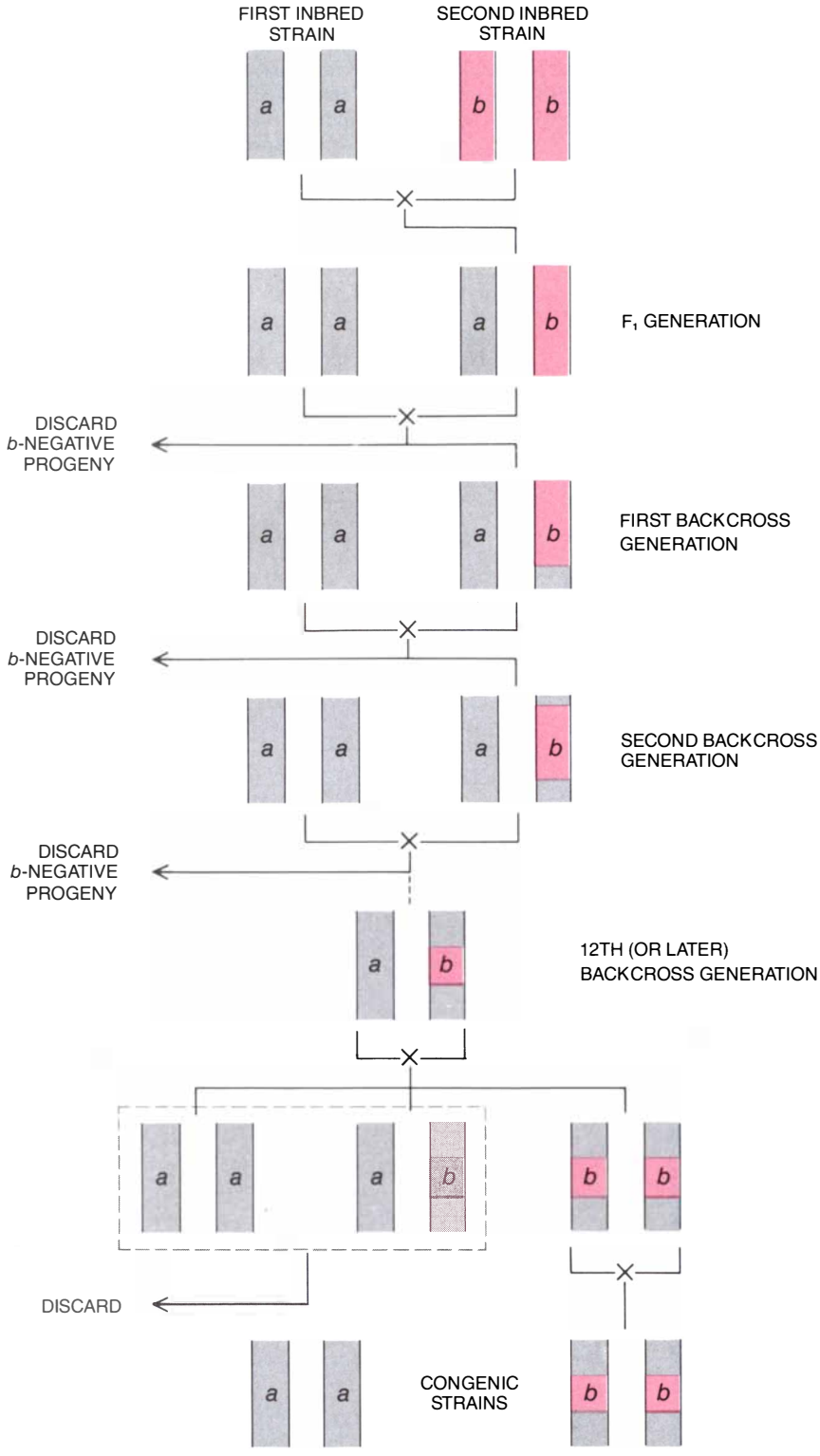
To confirm the genetic basis of the sensory discrimination of *H*-2 types in the Y maze we resorted again to an F_2 linkage test. Here testing proceeded by transfer of training, a protocol in which the mouse can apply only what it has already learned in order to deal with the new stimuli presented, in this case the urine samples of homozygous F_2 mice. In the transfer of training technique a mouse trained to discriminate between odors of *H*-2 congenic mice is first made accustomed to getting no reward on a certain proportion of a trial series, even if the correct arm of the maze is chosen. Then coded urine samples from the genetically corresponding F_2 homozygote panels (urine samples the trained mouse has not yet encountered) are substituted in the same proportion of unrewarded trials. A person conducting the trials therefore cannot know what the mouse would have to do to fulfill the predictions underlying the experiment. Since the crucial trials are conducted without reward and new learning cannot occur without reward, and since no one doing the experiment knows which sample is which until the tests are completed, a successful transfer of training series eliminates the possibility of response to unintentional prompting or to any cues other than those arising from an MHC genetic difference. The transfer of training was altogether successful, confirming the MHC genetic origin of the sensory cues in question.

In addition to congenic strains that differ throughout the *H*-2 complex there are congenic mouse strains that differ in only a part of the complex. The successful training of mice to distinguish between such strains tells us that at least three subregions within this stretch of genes can each impart a distinctive odor. Hence each complete *H*-2 odor "profile" reflects the action of several genes and is likely to be compounded of multiple odorants, or individual odor elements.

We asked ourselves whether the *H*-2



INBRED STRAIN of mice consists of animals that are genetically uniform and also homozygous; that is, each mouse has the same alleles, or versions of genes, on its chromosome pairs. The strain results from some 20 to 30 generations of brother-to-sister matings, with a single brother and sister producing each generation. First two ordinary mice are crossed. The letters *a, b, c* and *d* represent alleles of the MHC (*H*-2) gene complex or any other chromosomal region. The rest of the chromosomes are not shown. (The mouse has 20 pairs.) The F_1 (first filial) generation consists of four groups of progeny: *ac, ad, bc* and *bd* (all heterozygotes). A mating pair is chosen at random; here it is $ad \times bd$, so that allele *c* is lost to the F_2 generation. Again a pair is chosen at random; here it is $bd \times dd$, so that *a* is lost to the F_3 generation. By chance the cross engendering the F_4 generation is $dd \times dd$. Except for mutants, the mice thereafter will be *dd* homozygotes. Probability theory dictates that by the 20th to 30th generation uniformity and homozygosity will emerge for virtually all genes.



CONGENIC STRAINS of mice are homozygous and genetically uniform except for a difference confined to a small genetic region. First two inbred strains are mated. In this diagram the strains have different *H*-2 types (*a, b*). The mice of the F_1 generation are crossed back to the first inbred strain. Then the progeny (the first backcross generation) are typed for *H*-2 by reacting their cells with *H*-2(*b*) antibody; *b*-positive mice are crossed back to the first inbred strain. The procedure is repeated for at least 12 backcross generations. By that time the size of the chromosome segment bearing the *H*-2 genes from the second inbred strain has been pared down by crossing-over: the exchange of genetic material between the chromosome pairs in an animal's cells. Next the *ab* heterozygotes of the 12th (or later) backcross generation are crossed among themselves. The progeny are typed so that *bb* homozygotes can be crossed among themselves, yielding the desired congenic strain, which matches the first inbred strain except for the substitution of *b H*-2 alleles for *a H*-2 alleles.

odor profile is constituted by known genes within the MHC or by unknown genes within the complex: discrete ones that might have a primary function in sensory recognition. The dilemma, common in biology, amounts to asking whether a certain trait known to be inherited along with a particular gene is determined by that gene or by an unknown gene in the vicinity. Classically the problem is approached by determining whether the trait in question is altered by mutations of the known gene. We have examined a suitable mutation, one that affects the *H-2K* gene, altering the amino acid sequence of the *H-2K* protein. The results indicate that an *H-2K* mutant strain can be distinguished, by trained mice in the Y maze, from the nonmutant companion strain. It is the strongest suggestion so far that one may need to look no further than the known genes of the Major Histocompatibility Complex to account for the individuality of odor profiles.

Here, then, is a plausible explanation of how the genetically determined odor of individual animals is constituted with respect to the MHC. The MHC exhibits a very high degree of genetic diversity, and the genes of the MHC are intimately involved in immunologic recognition and response, whether to infection or to foreign materials such as transplants. In addition MHC genetic diversity is known to be associated with a corresponding diversity of numerous anatomical, physiological and developmental variations of the kind that make up the normal range of structural and biochemical variations among members of any species. This suggests that MHC genes have primordial functions in the development of an animal from the fertilized ovum, functions that may depend on chemical recognition but that extend much further back in evolutionary time than the highly evolved functions of the immunologic defense system.

One is not surprised that identical twins look alike. They are genetically identical; the element of genetic variation affecting appearance is lacking. By the same token, it seems probable that inbred genetically identical animals smell alike; the element of genetic variation hypothetically affecting odorous by-products voided in urine and secretions is lacking. It is well to remember that readily distinguishable compound odors can be generated simply by varying the proportions of the chemical constituents of a given mixture of odorants. Such proportional variation caused by variation in the output of odorous metabolites

arising from MHC genetic variation might alone account for individual odors related to MHC types.

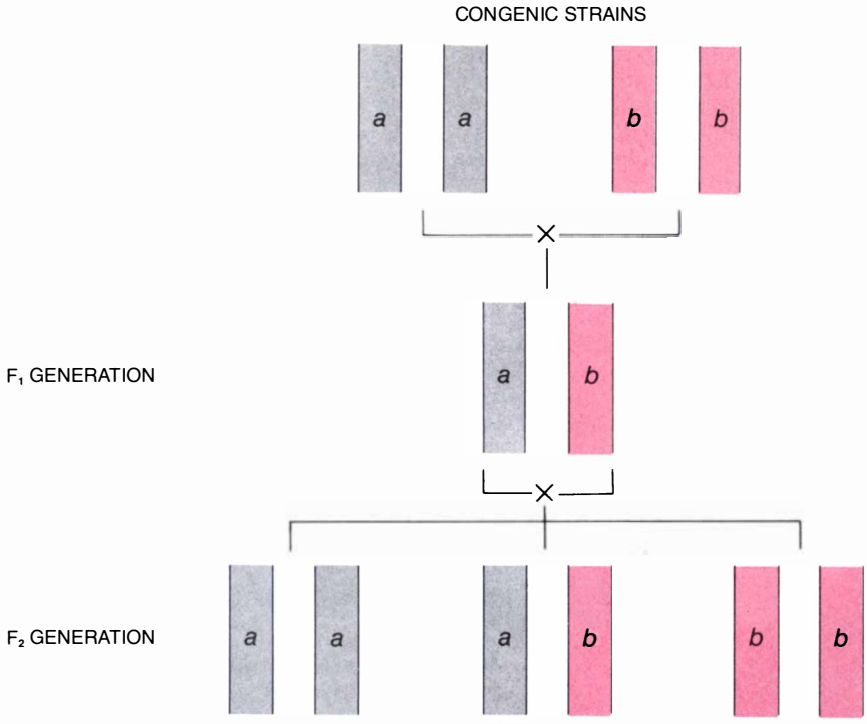
Some corroborating insight into the subtleties of the olfactory world of the mouse emerges from a further series of experiments. During the early phase of pregnancy in the mouse, up to four days after conception, the growing embryo is not yet attached to the wall of the uterus; the uterine wall is being prepared for the implantation of the embryo by a change in the balance of hormones. If during that time a pregnant mouse is separated from her mate and exposed to the presence or the odor of a strange male, from a strain genetically different from her mate's, the hormonal balance of the female is disturbed and the risk of abortion due to a failure of implantation increases.

Realizing that a difference in *H*-2 types enables a pregnant female to distinguish the scent of a strange male from the scent of her mate, we compared the incidence of abortion in isolated pregnant females experiencing the scent of males genetically identical with the mate with the incidence of abortion in similar females experiencing the scent of congenic males whose *H*-2 differed from that of the mate. The incidence of abortion proved to be considerably greater when the new male, or scent of urine, came from the congenic strain. A lesser but significant increase in the incidence of abortion was produced by exposure to congenic females whose *H*-2 type differed from that of the mate. Clearly the scent of an unfamiliar *H*-2 type (even if unaccompanied by the stimulus of the pheromone that occurs in male urine only) can initiate the neuroendocrinological events that produce a hormonal imbalance inimical to continued pregnancy.

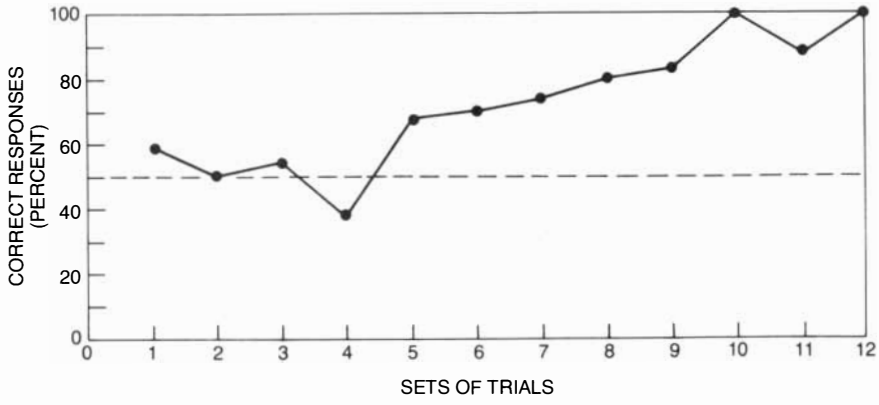
Chemical sensory information indicative of genetic constitution is communicated from one mouse to another. Still, we have discussed only one set of genes, the Major Histocompatibility Complex, and only one species, the mouse. Moreover, we have neglected one side of the coin. The two elements of a communication system are first the production and transmission of the signal, and second its reception and the ensuing response. Clearly, genetic variation acting at sites of odor reception, and thereafter in the animal's neural and endocrine systems, might be another agent of differential recognition and response in addition to the genetic variation of the signal, which has been our concern so far. One asks, therefore, whether two genetically dissimilar mice might perceive the odor of one mouse differently, as opposed to the question of whether one mouse can distinguish two genetically dissimilar mice.

The genetic tuning of reception and response to odor is suspected from the occurrence in human beings of specific hyposmia, a variation in human sensitivity to odor. Certain people either fail to sense a particular odorant or sense it only at a concentration substantially greater than other people require. Moreover, the quality of the odor such people describe is sometimes different from the one that other people describe. No systematic study, however, has yet been directed exclusively to genetic variation of reception response, as opposed to variation of the signals (odors), in the context of *H*-2 sensory recognition.

There is also the matter of how the



CONGENIC F_2 MICE identical genetically except for their *H*-2 genes are produced by crossing a pair of *H*-2 congenic strains. The F_1 generation consists of identical heterozygotes (*ab*). They in turn yield an F_2 generation of homozygotes and heterozygotes. The F_2 homozygotes (*aa* and *bb*) are identical with their grandparents, but their familial experience is different: they were born of heterozygous parents, whereas their grandparents (being inbred) had homozygous parents. Congenic F_2 mice aid studies of the inheritability of a trait.



TRAINING RECORD for a mouse displays the animal's progress in learning to discriminate the odors of the urine of congenic mice that are identical genetically except in the *H*-2 region. The mouse to be trained was deprived of water for a time and then was placed in a Y maze; the arms of the maze were differentially scented with urine samples from congenic mice of different *H*-2 types. The mouse was rewarded with a drop of water for entering the arm scented by the urine of a particular strain of congenic mice. Trials came in sets of 24, with two sets per day. In the initial four sets the choices made by the mouse were no better than random. By the 12th set, however, the mouse's performance was virtually faultless.

work we describe can be extended to other mouse genes and to other species. Let us consider the other genes first. The selection of mates based on *H*-2-related odors would directly affect the genetic composition of mouse populations only with respect to chromosome 17, where the MHC (*H*-2) genes are situated. If genes elsewhere in the mouse genome contribute to the identification of individuals by scent, reproductive selection could operate on other chromosomes also. We have found that unrelated inbred strains that happen to share the same *H*-2 type can be distinguished in the Y maze, showing that the mouse genome en masse can participate in odor specification.

Still, it is unlikely that any chromosomal region rivals *H*-2 in its potential for labeling individual mice by scent. For one thing, no other chromosomal region is known to exhibit such great constitutional diversity from one mouse to another. Also, we have studied, in the Y maze, a few congenic strains that differ at a single region of a chromosome other than chromosome 17. These strains are distinguishable by mice in the maze only with difficulty, if at all.

Finally, there is the extension of our investigations to species other than mice. For the work described in this article laboratory mice were indispensable: they offer the inbred, congenic and mutant strains essential to such studies. But now we can begin to go further. The first step we have taken is to investigate the ability of rats, a second species within the order Rodentia, to distinguish congenic mice on the basis of *H*-2 types. The rat is more able and trainable than the mouse, so that we can employ an automated olfactometer, a device we have constructed that greatly simplifies testing. Under the control of a microcomputer the olfactometer shunts a flow of air to any of seven odor samples and then administers a measured amount of the scented air to a rat (primed by previous limitation of water intake) positioned in a test chamber. The rat initiates each test: sensors above and below the animal signal when its nose is in a position to receive the stimulus. If the odor presented is the one for which a response is rewarded, the rat must touch a bar near its head to receive the reward, a drop of water. In this way we have successfully trained rats to distinguish the *H*-2 types of congenic mice.

It is a considerable step from the recognition of genetic individuality among rodents to such a perennial topic as the identification and tracking of individual human subjects by dogs. The logical starting point will be to test the ability of dogs to distinguish the *H*-2 types of congenic mice, thereby postponing the complications of overall genetic variation that will have to be faced in testing the dogs' ability to discriminate HLA types in man (a prediction Lewis Thomas made more than a decade ago). The relatedness of the mammals, a class of vertebrates that evolved less than 100 million years ago, leads us to think that what applies to the mouse and the rat will apply at least to other mammalian species with an adequate sense of smell. It may even be that human subjects placed under suitable conditions can perceive the HLA types of other individuals, and physiological responses are not beyond the bounds of possibility.

Y MAZE is employed at the Monell Center; it is the apparatus in which mice are trained to discriminate olfactory cues related to the MHC. Air is drawn by a fan through left and right odor boxes. Each box has a hinged lid for admitting an odor source, such as a mouse-urine sample. The currents of air then pass to the left and right arms of the maze, each of which includes a water tube guarded by a fence. If a test mouse placed at the base of the Y enters the correct arm, the fence is raised to give the mouse access to the water. The left-right placement of odor sources is governed by a series of random numbers. Typically a trained mouse takes two or three seconds to choose between the arms; it chooses either without pause or after sniffing the entrance to the arms, or sometimes with a brief retracing.

A dark sky and threatening forecast is enough to humble the greatest drivers among us. At times like these you need an automobile like Eagle.

With the simple flick of a switch, Eagle gives you the confidence and security of 4-wheel drive. You can even "shift-on-the-fly," which is a luxury you shouldn't be without when faced with a sudden shower, blizzard or stretch of bad road.

Of course, security isn't the only luxury you'll enjoy in an Eagle, with features available like power windows and door locks, cruise control, 6-way power seats and plush carpeting.

The American Eagle. It's the best thing going for you when you're going in bad weather.

EAGLE THE 2-WHEEL/4-WHEEL DRIVE SECURITY SYSTEM

AVAILABLE FOR PURCHASE OR LEASE AT PARTICIPATING AMC DE

SAFETY BELTS SAVE LIVES

Minoan Palaces

At the center of Crete's rich Bronze Age culture was the palace, which had economic, political and religious functions. Studying those functions makes it possible to reconstruct Minoan society

by Peter M. Warren

How is it possible to understand the culture and social structure of a civilization that came to an end some 3,500 years ago? The most direct method would be to read and interpret the writings left behind by that civilization. In some instances, however, written information provides little help. For example, the writing of the Minoan culture of Bronze Age Crete (the script known as Linear A) has not yet been deciphered. Moreover, it is probable that the Minoan writings consist mainly of economic accounts. Even if Linear A were to be deciphered, such documents would not completely illuminate Minoan culture. Of course nonlinguistic artifacts also yield considerable information. One of the best sources of information about the social structure and culture of Bronze Age Crete is not linguistic but architectural: the Minoan palaces.

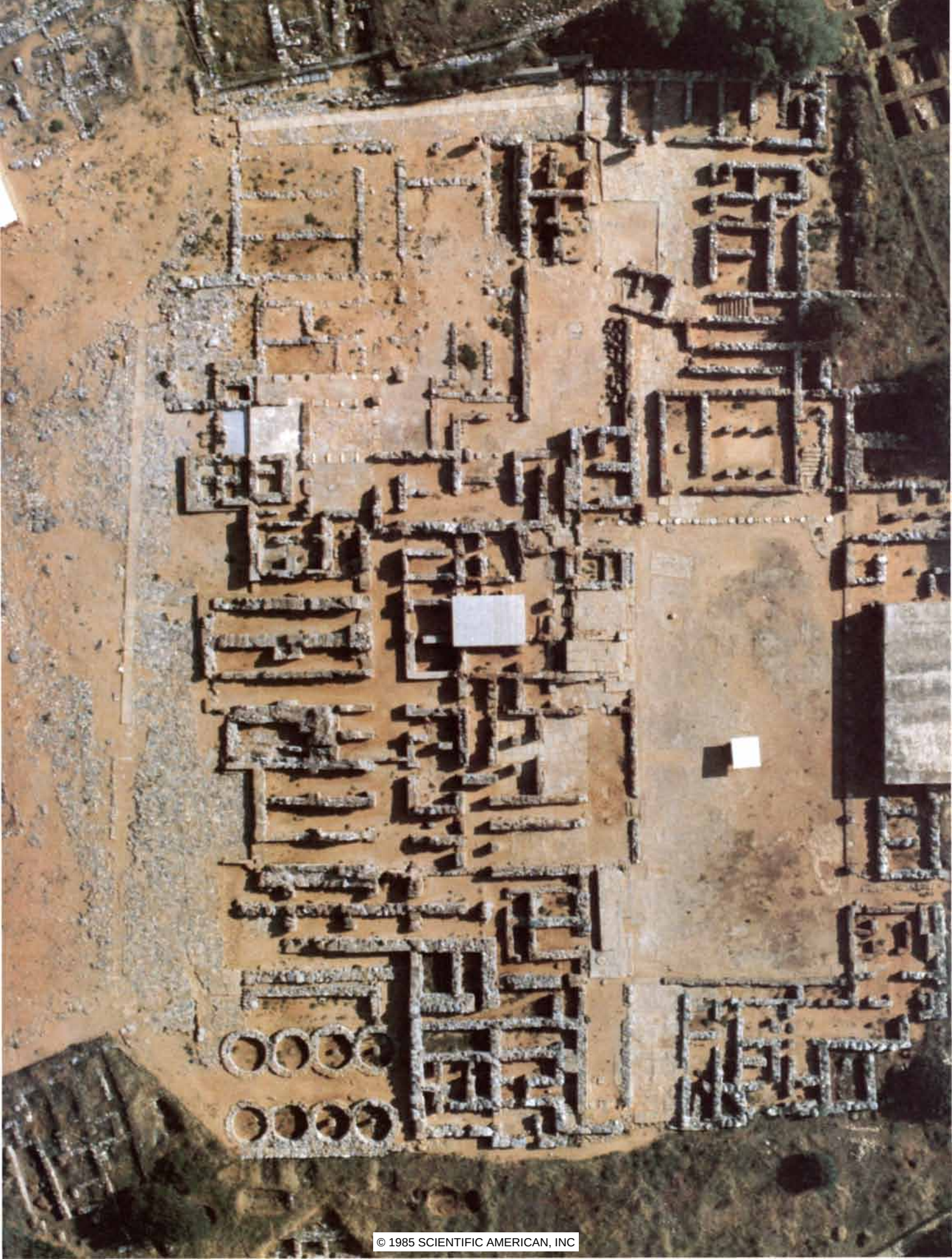
Four Minoan palaces have been excavated so far, beginning with the discovery of the greatest of them at Knossos in 1900 by the British archaeologist Sir Arthur Evans. The four known examples share features that could serve as a definition of the palace at its height. Each palace was the main building at its site, which was generally a major Minoan town. Each palace included a central court; generally there were other courts as well. Around the central court were grouped storage facilities, production areas, archives of inscribed tablets, rooms for ritual activity and rooms for state functions. Great power was concentrated in these elaborate structures: both the secular and the religious authorities of Minoan societies lived in them. Recent scholarly work has concentrated on the economic functions of the palace. It has been shown that the palace rulers probably had considerable control over agriculture in the region around the palace and also over Crete's rich foreign trade.

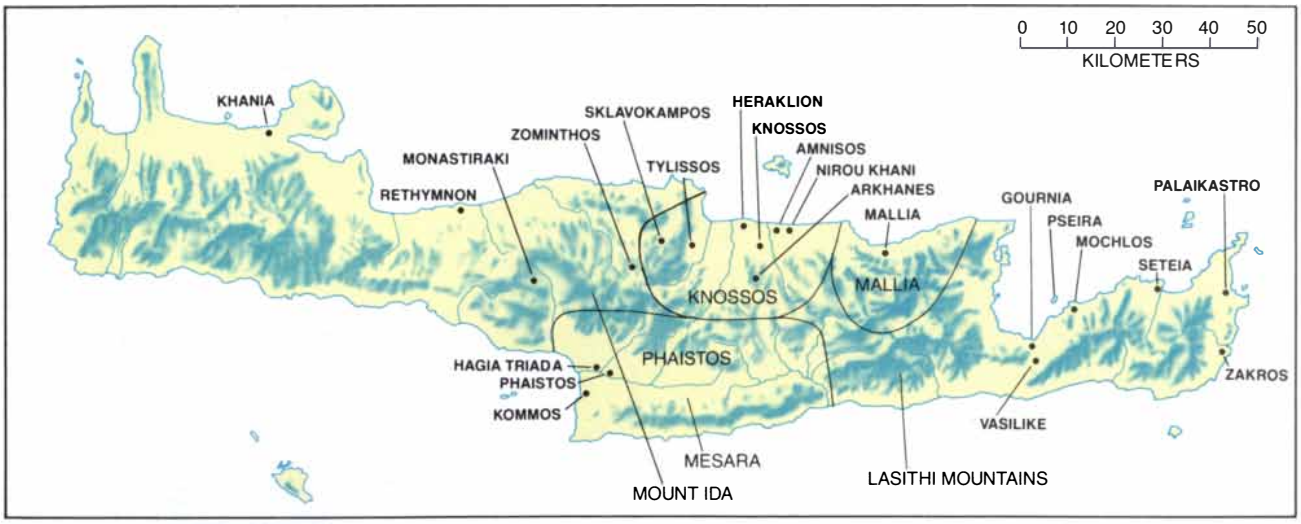
Since the palace had functions that extended into every area of life, a study of how the palace operated can form the basis for a reconstruction of the entire Minoan society. Bronze Age society on Crete was hierarchical but not divisive. Under the authority of the palace the various social groups appear to have lived in relative harmony. Furthermore, the art found in the palaces reflects a vitality and humanism that distinguishes Crete from contemporaneous societies such as Egypt and Assyria. The origin of the European tradition of humanism and individualism is generally attributed to the Greeks, but there is a sense in which Minoan culture can be considered the first example of a distinctively European tradition. In the absence of written documents some questions will always remain, but a study of the palaces can contribute greatly to understanding the successes of Minoan society.

The sites of all four known palaces share a significant common feature: they are on or near the coast. Knossos, which lies five kilometers inland from the northern coast, commands the fertile valleys of northern central Crete. In the south Phaistos dominates the Mesara, Crete's only large plain. The palace at Phaistos was also discovered in 1900 and was excavated by workers from the Italian School of Archaeology under the direction of Luigi Pernier. Mallia, east of Knossos on the northern coast, may have included the Lasithi Mountains in its territory. The Greek archaeologist Joseph Hatzidhakis began digging at Mallia in 1915; in the 1920's a team from the French School of Archaeology took over the site and continued the work. Zakros, the site of the fourth palace, is tucked in the southeast corner of the island. Zakros is difficult to reach by land and the Minoan site makes sense only as a major port for traffic from the east and southeast. The palace there is the most recently excavated, having been uncovered in the 1960's by a Greek team under the leadership of Nicholas Platon.

There may well be Minoan palaces in addition to the four known examples. At Khania in the west on the northern coast there was probably a fifth palace. The palace at Khania has not yet been excavated but its existence is suggested by the discovery there of Linear A tablets, which probably form part of an archive. Several other regions of Crete may yield palatial towns, notably the Rethymnon area of western central Crete and the Seteia area in the east. In addition large structures dominated the center of other towns, including Gournia and Monastiraki. Although not palaces in the full architectural sense, they are large and important Minoan buildings that share some features with the palaces. The fact that all these unfortified sites are in the coastal zone raises the possibility that Crete's location was a significant factor in the development of

PALACE AT MALLIA, like the three other known Minoan palaces, consisted of many small buildings grouped around a central court (*lower right*). The roof in the central court covers an altar. The open space at the left is the west court. Between the courts lies the west wing, which included living quarters, storage magazines and rooms for religious ceremonies. Facing the west wing across the central court is the smaller east wing, where olive oil and perhaps wine were made. The circular forms at the lower left were probably granaries. At its height the palace may have covered some 9,800 square meters, or 2.4 acres. The photograph is from the forthcoming *Aerial Atlas of Ancient Crete,* by J. Wilson Myers, Eleanor Emlen Myers and Gerald Cadogan. An unmanned balloon served as the camera platform.





BRONZE AGE SITES on Crete are concentrated in the coastal zone. The palaces that have been excavated are at Knossos, Mallia, Phaistos and Zakros. Knossos, Mallia and Phaistos appear to have dominated large regions (*solid lines*). Zakros seems to have functioned mainly as a port. The finding of inscribed tablets at Khania suggests a fifth palace stood there, but it has not been excavated.

PALACE AT KNOSSOS, discovered by Sir Arthur Evans in 1900, was the grandest of the four known Minoan palaces. It may have covered 17,400 square meters, or 4.3 acres. The view in the photograph is to the south. The main court is at the center; the west court is at the right. The west wing between the courts had two floors. Storage magazines and rooms for religious ceremonies occupied the ground floor; archives of inscribed tablets, a shrine treasury (containing stone vessels) and state reception rooms occupied the upper floor. The east wing (*far left*) held the living quarters. It had at least four floors and was reached by the Grand Staircase, which is covered by the smaller roof flanking the central court. The east wing included a room for bathing and a toilet supplied with running water. Raymond V. Schoder of Loyola University, using a hand-held camera, took the photograph through the open cargo door of a DC-3.

civilization there. The island was far enough from mainland powers such as the Egyptians, the Hittites and the Mycenaean Greeks to discourage frequent attacks but not far enough to prevent sustained trade.

The Minoan palaces, which represent the highest achievement of Bronze Age culture in Crete, were the culmination of a long period of social development on the island. The Bronze Age, which lasted from about 3000 to 1000 B.C., was preceded by 3,000 years in which the Neolithic inhabitants of Crete successfully established agriculture and animal husbandry. Neolithic achievements in farming contributed to the material basis of Minoan culture: a rich agricultural economy of cereals, olives, vines and herds. Following the Neolithic was the prepalatial Bronze Age, or Early Bronze Age, lasting from about 3000 to 1930 B.C.

In the prepalatial period three trends helped to create the social conditions needed for the establishment of the palaces. The first trend was the establishment of closely knit agricultural settlements. The excavation of several prepalatial settlements shows that early Bronze Age villages were quite dense. Many of the closely packed houses shared walls, giving the village an overall form reminiscent of a honeycomb. In such villages the early Bronze Age Cretans lived a communal life in which ties were close; there is little evidence of social ranking in the villages. The second trend was the probable development within these tightly knit communities of nuclear or extended families, which formed subunits of the village.

The third trend was the establishment of territories defined as being for the use of a particular settlement. Such territories were small regions that evolved through continuous agricultural exploitation. The farmers of the Early Bronze Age lived close together and walked to the fields to work, much as their modern counterparts on Crete do. As a result a village's defined territory may have been limited by the distance a farmer could travel each day to the fields and back on foot. In addition to the geographic definition the agricultural settlements were defined socially by customary ways of living.

Over the period from about 2200 to about 1900 B.C. some prepalatial centers grew in size and importance. One aspect of the increase in importance is that some communities appear to have commanded larger territories. Each of the larger territories, or provinces, may have been dominated by a powerful family that built and occupied what eventually became the palace. The process by which the provinces were defined and authority was established over them is not well understood. The unthreatening physical setting of the palaces, however, suggests they were built with the collective acceptance of the rest of the community. None of the palaces is in a physically commanding position and all four of them are readily accessible from the surrounding towns.

The Age of Palaces lasted from 1930 to 1450 B.C., ending with a massive episode of destruction whose nature is still being debated. During those five centuries continuous progress was made in architectural design. Among scholars of Minoan Crete a division is often made between the First (or Old) Palace Period (1930 to 1700 B.C.) and the Second (or New) Palace Period (1700 to 1450 B.C.); a major episode of destruction, probably caused by an earthquake, divided the two epochs. The architectural history of the palaces is actually much more complex than such a division suggests. Each palace was probably built and rebuilt many times. In all this development no trace exists of foreign invaders; what is observed is the heightening and elaboration of the native styles.

The ultimate result of this prolonged architectural refinement was the design of the palace at its zenith in about 1450 B.C. As noted above, that design included many small buildings fringing a central court. Some elements of the design had Minoan precedents. The plan of the early palace at Phaistos clearly shows a cellular arrangement of small buildings linked by passages that resembles the honeycomb arrangement of the Early Bronze Age settlements. Other design elements do not seem to have Cretan antecedents and may have been imported. The monumentality of the palaces and the arrangement of buildings around a central court can be seen in contemporaneous Near Eastern palaces such as the one at Mari, a Mesopotamian city on the Euphrates. It is known that by 1700 B.C. the Minoans had trading links with Mari, and architectural ideas may have been brought back to Crete along with trade goods.

Although the palaces shared a basic architectural plan, they varied considerably in grandeur. Knossos was the most splendid. If it is assumed the palaces were roughly rectangular and if allowance is made for areas that have been lost, it is found that the buildings and central court at Knossos covered 17,400 square meters, or 4.3 acres. The corresponding area at Mallia was 9,800 square meters, at Phaistos 8,300 and at Zakros 4,250. At Knossos there were at least four upper stories in the east wing; all the palaces had at least one upper floor. More work has been done at Knossos than at any of the other palaces, but in spite of its preeminence Knossos has a significant disadvantage for a study of the pure Minoan culture. Although it was destroyed in 1450 B.C. along with the other palaces, Knossos was subsequently reoccupied by Mycenaean Greeks. The Greek occupiers may well have learned to write from the Minoans: Mycenaean writing, the transliteration of Greek known as Linear B, employs Minoan symbols. Thus the largest and richest Cretan palace, and the one with the greatest number of economic records, had a final occupation controlled not by Minoans but by Greek immigrants.

The four known palaces differed not only in grandeur but also in the detailed arrangement of functional areas. There was generally a west wing extending from the central court to a western court. The ground floor of the west wing was divided into two sections: storage magazines filled with great jars called *pithoi* and rooms for religious ceremonies. At Mallia the ceremonial rooms were aligned with an altar in the central court. The functions of the upper floor of the west wing varied from one palace to another. At Knossos the upper floor contained a treasury where stone vessels for a shrine were stored, an archive of tablets and a group of reception rooms for state functions. At Zakros, on the other hand, the shrine treasury and archives were on the ground floor of the west wing; the upper floor included workshops and stores of vessels for burning incense.

The other wings also varied in their details and function. At Knossos and Zakros the east wing held what are generally thought to be the living quarters of the rulers. At Knossos the residential area was reached by the Grand Staircase, a marvel of architectural engineering. The living quarters there included a room for bathing and a toilet supplied with running water. At Zakros there was a fine circular swimming pool. In contrast at Mallia the main component of the eastern part of the palace was industrial: a carefully constructed system of jar stands and channels for processing oil and perhaps also wine. At Knossos and Zakros industrial installations appear in other parts of the palace. In the northeastern section at Knossos were a stone carv-

er's workshop and possibly a potter's workshop.

The shrine rooms opening onto the central court at several palaces and the altar found in the court at Mallia strongly suggest that the central open area served as the site for religious ceremonies. Noting the existence of protective grills in the main court at Mallia and Phaistos, some workers have contended that the central court was also the site of the famous Minoan bull sports. The unprotected cult rooms facing the main court in several palaces and the fragile altar in the court at Mallia, however, make it seem improbable that bulls cavorted in the open area. At Knossos and at Mallia bull sports may have been conducted northwest of the palace. Buildings there could have served as grandstands from which the inhabitants of the palatial town watched young athletes being tossed by the bulls.

When their economic, religious and political functions are taken into account, the palaces must be considered the chief material expression of the structure of Minoan society. A useful way to elucidate that structure is

PALACE AT PHAISTOS has the same overall plan as the other Minoan palaces but differs from them in its details. The central court is at the bottom, the west court at the upper left. As in the other palaces, the storage magazines were between the courts in the west wing. Several features of the palace at Phaistos, however, are unique: they include the large staircase ascending from the west court and the colonnaded residential rooms in the north wing (*under roofs*). The south and east wings of the palace have been lost to erosion. Including its lost parts, the palace at Phaistos may have covered 8,300 square meters, or 2.1 acres. At the left are houses and circular storage pits from the early palace period. The photograph, like the one on page 95, is from the forthcoming *Aerial Atlas of Ancient Crete*. The diagonal line extending from the upper left to the lower right is the tether for the unmanned balloon that held the camera.

to examine the palace in three progressively wider contexts: the town, the province and the world. Each palace stood at the heart of a town. The palatial towns were large and highly developed. At Knossos, the largest palatial town, intensive settlement covered some 75 hectares (185 acres) during the time of the palaces. In contrast Palaikastro in eastern Crete, one of the largest excavated towns that do not have a palace, probably covered between 25 and 35 hectares.

The palatial towns often included large and impressive private houses. The houses at Knossos were constructed of the finest dressed and neatly jointed ashlar masonry, just as the palace was. At Mallia early in the Age of Palaces there stood a group of important buildings not connected with the palace that had cult rooms and family-size workshops attached to the exterior. It has been argued by Jean-Claude Poursat of the French School of Archaeology at Athens that this group of buildings housed the religious authorities of the community. If that is so, there was a division between sacred and secular authority in the early part of the Age of Palaces, since the secular rulers undoubtedly lived in the palace. It is known that in the later part of the Age of Palaces, after the Middle Minoan buildings in question had gone out of use, both kinds of authority were concentrated in the palace. It is possible that during palace times the authority structure of Minoan society was consolidated, with the result that the authority of the palace was increased still further.

The fundamental question with respect to the towns is that of their relations with the palace. The spectrum of possibilities extends from absolute palace control to complete town independence. The post-Minoan, Mycenaean relationship, which has been adduced from Linear B tablets, entailed a strong degree of palatial control, including a uniform system of taxation in kind throughout the kingdom. In the absence of documents containing explicit social information the relations between town and palace in the Minoan period cannot be precisely characterized. Spatial and architectural evidence, however, suggests that the Minoan system was not as centralized as the Mycenaean one. For example, at Knossos a series of large mansions stood in close proximity to the palace. The residents of the mansions must have had much contact with the rulers, including ready access to the palace. Furthermore, at Zakros there is no sharp physical or architectural boundary between the palace and the houses: the outbuildings of the palace extend into the town without a break. Such arrangements imply that the Minoan social hierarchy was accompanied by little social division or conflict. The combination of hierarchy and harmonious relations is a key to understanding the success of Minoan civilization.

The second zone of economic and social organization is the dependent, or controlled, territory beyond the town. It cannot be proved that there was a formal territory for each palace, but the locations of three palaces suggest natural territories for them: northern central Crete for Knossos, southern central Crete and the Mesara plain for Phaistos and the bay of Mallia and perhaps the Lasithi Mountains for Mallia. In an agricultural society such as that of the Minoans territorial control would undoubtedly have entailed appropriation of a significant fraction of the region's agricultural surplus. Analysis of the palatial storage capacity suggests the rulers may have received the surplus from a fairly large surrounding territory.

J. W. Graham of the University of Toronto showed that the west magazines of the palace at Knossos could have held 420 *pithoi,* each with a maximum capacity of 586 liters. Therefore the maximum storage capacity of the west magazines was about 246,000 liters. Suppose a third of that capacity were allotted to olive oil, which may well be too small a fraction. Assuming modern yields, from 16,000 to 32,000 olive trees would have been needed to keep the jars filled. Assuming also a modern distribution of trees in the olive groves (from 10 to 20 trees per 1,000 square meters), the palace magazines could have held the yield of some 320 hectares. If the remaining two-thirds of the storage capacity was taken up by cereals and wine, the land area under cultivation would have been considerably greater than 320 hectares. Major crops such as coriander and saffron would have covered additional land. Furthermore, some of the private houses in Knossos had their own storage facilities, hence the palatial magazines received only part of the produce of the region. Thus to the palatial town might have gone much of the produce from a surrounding region of about 1,000 hectares (2,500 acres). That region, which lies within the much larger natural territory of the palace, includes the Knossos valley and land to the north on the outskirts of the modern town of Heraklion.

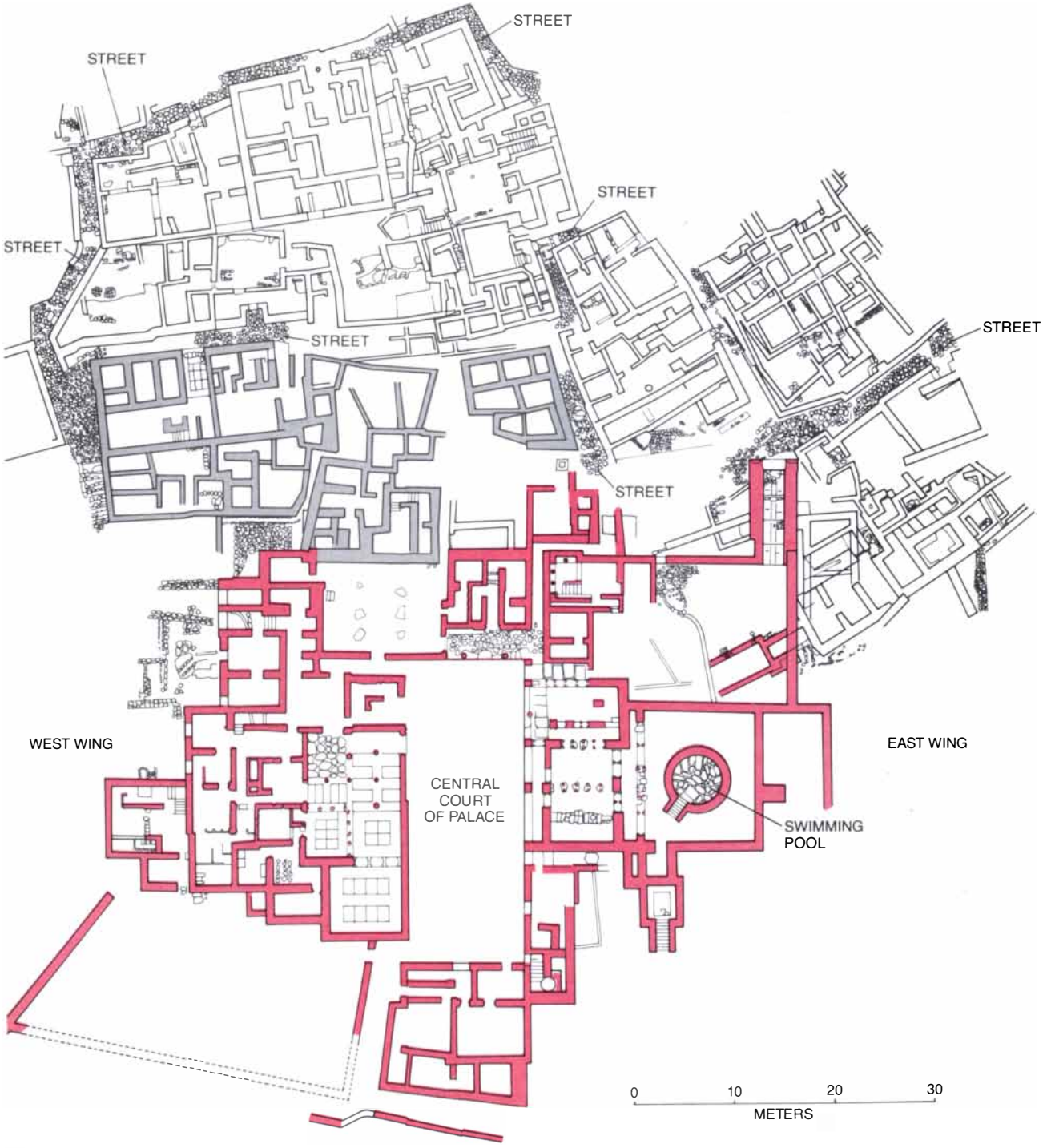
A similar, although somewhat less precise, computation can be made for Phaistos. It is possible that in palatial times the rulers of Phaistos controlled the entire Mesara. The area of the Mesara (below an altitude of 200 meters) is about 18,000 hectares, or 70 square miles. Wheat yields in Crete in the 1930's were about 140 kilograms per hectare per year. Assuming that the yield was the same in Minoan times and also that the entire plain was under wheat, the Mesara could have produced about 2.5 million kilograms of wheat annually. Among the records from Knossos are tablets that may indicate the yield from the Mesara. If the tablets have been interpreted correctly, they show a maximum wheat yield from the Mesara of about 775,000 kilograms. That figure is somewhat less than the plain's capacity, but there are several unknown quantities in the computation. In addition it is unlikely that no crops other than wheat were planted on the Mesara. In spite of such uncertainties, the computations indicate within an order of magnitude how much agricultural land the palace controlled.

If the occupants of the palace were able to appropriate the agricultural surplus of a fairly large region, they probably had representatives in the countryside. Set outside the palatial towns are sites that have been described variously as country mansions, villas, farms and estates. In a functional sense the villas resembled miniature palaces without the great central court. The villas included storage magazines, oil- and wine-making plants, weaving areas, shrines and other, more imposing rooms. It cannot be assumed that the villas were palace satellites, helping to organize the collection of produce. They may have been completely independent of the palace, and the stored produce may have been their own. A third possibility is that the villas were neither completely independent of nor completely dependent on the palace. Further work will be needed to decide among the alternatives.

Some mansions clearly had functions unrelated to collecting produce, and some of them may have been out of the orbit of the palace altogether. The villas at Nirou Khani, Sklavokampos and the newly discovered structures at Zominthos, which straddles a critical junction of roads on Mount Ida, may have been stations on economically significant routes. At Tylissos and Mochlos stood clusters of mansions that can be interpreted as communities commanding land or sea territories. The relation of the mansion clusters to the palace is not understood, but some of them, including Tylissos, are far enough from the pal-

ace to appear independent. In addition to mansion clusters the region around the palace included towns. The towns varied considerably in size and in the richness of the artifacts found there. The size, form and geographic setting of some towns suggest that they were independent of the palaces. Notable among the towns in this group are Gournia, Pseira and others on Mirabella Bay.

Two towns in particular raise interesting questions about the balance of independence and dependence: Arkhanes and Hagia Triada. Both sites were close to palaces but both seem to have been more than satellite towns. Arkhanes is 10 kilometers south of Knossos. Rich cemeteries found there suggest the town had a long and independent history. Moreover, during the Second Palace Period, Arkhanes had fine buildings with palatial masonry and architectural elements; in one place Linear A tablets were also found. Such findings imply that Arkhanes was far more than a satellite of Knossos.

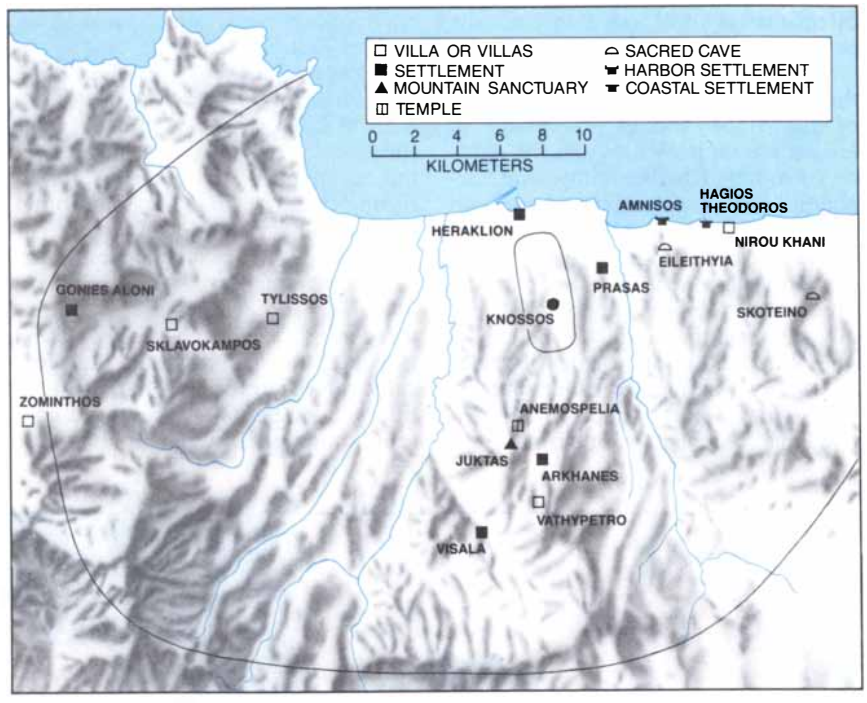


PALACE AND TOWN in Minoan society were closely related, as is suggested by the plan of Zakros. The palace buildings are shown in color, the town buildings in white and the buildings that cannot be assigned definitely to either category in gray. The absence of fortifications or even a clear physical boundary between the palace and the town implies that palace authority was not imposed by force. The living quarters at Zakros were in the east wing of the palace; the circular structure was a swimming pool. The southern part of the palace included workshops where craftsmen made perfumes as well as goods of bronze, stone, crystal, ivory and earthenware.

The location of the town and the stylistic resemblance of artifacts found there to artifacts found at Knossos, however, suggest some degree of dependence on the palace.

Hagia Triada also combines aspects of dependence and independence. The town, three kilometers west of Phaistos, is a particularly rich site where some of the finest Minoan ritual vessels known were found along with a large store of copper ingots. These finds come from a sizable structure that some observers have maintained was a palace. Recent work, however, suggests the "palace" was actually two joined mansions. If that is so, Hagia Triada might have resembled Tylissos as an independent mansion cluster. The proximity to Phaistos, on the other hand, strongly suggests some dependence on the palace there. It is possible that Hagia Triada had several functions, serving as a religious center and also organizing agriculture for Phaistos in the western Mesara.

The model of a capital manifesting control of its territory by acquiring and storing agricultural surplus fits what is known of Knossos, Mallia and Phaistos fairly well. The same cannot be said of Zakros. Set in a bowl of mountains on the southeastern coast of the island, Zakros does not have a natural territory in the sense that the other palatial towns do. That Zakros was a port town is clear not only from its location but also from the stores of imported ivory, copper and semiprecious stones found in the palace. It seems likely that Zakros was an independent entity with its own ruler and administrative archives. At least two factors, however, suggest a special connection between Zakros and Knossos. First, a large group of clay-seal impressions found in one of the houses has connections with impressions found in Knossos. Second, both the palace and the town have yielded an unusual quantity of fine pottery that almost certainly came from Knossian workshops. Perhaps Zakros was the eastern trading port for Knossos: the first landfall for ships sailing from the southeastern Mediterranean. If Zakros also served as the first landfall for ships traversing the southern coast, then Phaistos may have had some say in its affairs, but the connection with Knossos seems stronger.



PALACE AND PROVINCE probably formed a political and economic unit. Solid lines indicate the natural territory of Knossos and the smaller region within that natural territory from which the palace received agricultural produce. The palace magazines at Knossos were large enough to accommodate much of the agricultural surplus from the smaller region, which includes about 1,000 hectares, or 2,500 acres. Within the territory of Knossos were several types of sites. "Villas" were large, freestanding houses. Some may have housed those who organized agriculture for the palace. Others, including the ones at Nirou Khani, Sklavokampos and Zominthos, may have been stations on important routes. Most of the settlements were dominated by Knossos, but two may have possessed some independence: Arkhanes, a rich town, and Tylissos, which may have commanded its own small territory.

Zakros' status as a port has a direct bearing on the third and widest context in which the palace functioned: the world beyond Crete. Minoan society had extensive contacts with the mainland, chiefly in the form of peaceful trade relations. Ports were the point of entry into the web of commercial contacts. Mallia, on the northern coast, served as its own port. The harbor towns of Amnisos and Kommos stand close enough to the palaces of Knossos and Phaistos respectively to suggest they were the ports for those palaces. From these ports Minoan products were carried as far as Delphi on the Greek mainland, Troy, Cyprus, the Levant coast and Egypt. Closer to Crete were regions of much more intense economic and social contact, in particular the Cycladic islands and the northeastern Peloponnese. Indeed, at some sites in the Aegean there appear to have been communities of Cretans among the local population; such overseas communities were a natural outgrowth of economic and demographic expansion on Crete.

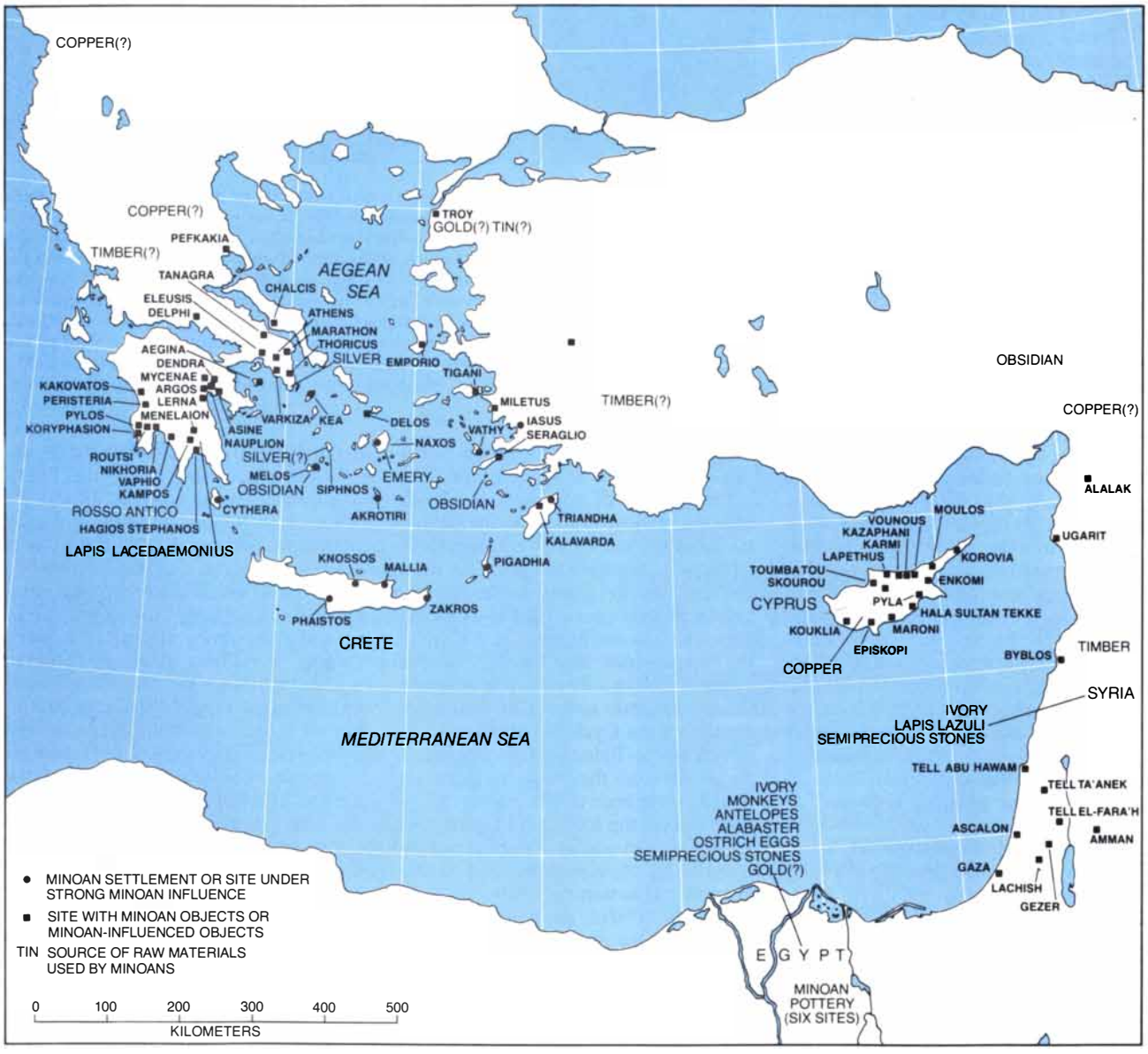
The successful, expanding Minoan economy also required imports. The island's chief needs were metals: copper, gold, lead, silver and tin. Other desirable goods included ivory and precious stones for carving into vessels and seals. The major question in relation to trade is that of control. In contemporaneous Egypt trade was controlled by the rulers. In Mesopotamia and Levantine cities such as Ugarit, on the other hand, there was a semi-independent merchant class. Did the Minoan system resemble either the Egyptian or the Mesopotamian? In the absence of written archives the question cannot be answered definitively. It is notable, however, that imported goods found in Minoan towns are not confined to the palaces: bulk imports, ivory and fine stones appear in houses. Copper ingots have also been found in houses, although the largest nonpalatial hoard, at Hagia Triada, is an ambiguous find. It would appear that a Mesopotamian-Levantine model of economic structure is more appropriate for Minoan Crete than the Egyptian one. The Mesopotamian model attributes overall control of trade to the palace but allows considerable latitude for a semi-independent merchant class.

The traverse of town, province and world forms the basis for a sketch of the groups that made up Minoan society. The sketch begins with the occupants of the palace itself. Who ruled Minoan Crete is not known with certainty. It has been argued above that the construction of the palaces was part of a process of unifying large territories. This process may have been carried out by powerful families who

established dynasties as they built the palaces. Yet little can be said of the individual Minoan rulers; not even their sex is known. The three members of the Minoan dynasty celebrated in legend are male: Minos and his shadowy brothers Rhadamanthys and Sarpedon. Strong arguments based on iconographic evidence from the palace at Knossos, however, suggest that a chief priestess was the ruler. (Mallia, Phaistos and Zakros are archaeologically silent on the matter of the ruler's sex.) Whether or not the ruler of Minoan society was female, other findings imply that some groups of women had equality with men, or perhaps preeminence, a remarkable status in the Bronze Age.

Below the rulers of the palace in the social hierarchy there must have been high officials who organized agriculture. Such officials may have been the occupants of the mansions spread through the provincial hinterland. They may have stemmed from the ruling family or from independent families. A prominent role was also played by priests and priestesses. In addition there must have been a group who organized and carried out the critical overseas trade. As noted above, the traders may have constituted a semiautonomous merchant class within the society. At all levels of Minoan society there must also have been craftsmen, ranging from palace dependents to those who maintained workshops in the towns. In an agricultural society such as that of Bronze Age Crete, however, the majority of the population would have been farmers, agricultural laborers or herdsmen.

The hierarchical but peaceful society made up of these groups suffered a catastrophic destruction in about 1450 B.C. The immediate agent of destruction was fire. The source of the fire, however, remains in dispute. Internal strains leading perhaps to civ-



PALACE AND WORLD BEYOND CRETE were linked by the island's rich foreign trade. Minoan Crete had an export economy. The main imports were raw materials: metals, ivory and fine stones to be carved into ceremonial vessels and seals for making impressions in clay. Exports consisted of finished goods, notably pottery. The influence of the Minoans extended throughout the eastern Mediterranean. The most intense contact was with the Greek mainland and the nearby islands, where there were Cretan communities. It is probable that trade in Minoan society was carried out by a semi-independent merchant class under the authority of the palace.

il war or rebellion, a Mycenaean invasion and a major volcanic eruption have all been proposed. There is little trace of internal strain in the material record, and it seems unlikely that the Mycenaean newcomers present after the destruction would have ruined economically everything they had just occupied. The third possibility is perhaps the likeliest: earthquake and other effects generated by the cataclysmic eruption of the volcano on the nearby island of Thera (Santorin). It is known that the volcano erupted early in the late Bronze Age, but opinion has been divided between a date of about 1500 B.C. and one of about 1450. Recent work in Greenland on ice cores containing acidity that can be dated by its depth in the ice supports the later date, as does paleomagnetic information from Thera and Crete. The scale of destruction on Crete is compatible with the effects of a massive volcanic eruption. It is quite possible that this major debate will be settled in the foreseeable future.

Even when the end of Minoan civilization is understood, much will be left to learn. Current work on the functioning of the Minoan palaces is focused on several socioeconomic questions. Did the palaces emerge as the result of extended, incremental development or did they arise rather suddenly after a gathering of social momentum? What was the precise form of the social hierarchy in the Bronze Age settlements? What was the status of women? What organization lay behind the successful foreign trade?

Significant as these questions may be, none of them touches on the most fundamental problem: What was the source of the relative harmony of Minoan society? At the height of the Age of Palaces in 1450 B.C. no Minoan settlement was fortified. All of them were in easily accessible, scarcely defensible positions. Crete's location may account for its peaceable relations with the mainland powers but it cannot account for the apparent absence of armed conflict within Cretan society during the palace period. It has been suggested that the rich ritual practices of Minoan society could have served to redirect aggressive impulses. Such an observation, however, merely conceals the deeper question. At the heart of what we call civilization is the sublimation of aggressive impulses: their redirection toward higher and more abstract purposes than killing other human beings. Why that capacity appeared in a highly developed form among the seafaring inhabitants of a small, sun-baked island in about 2000 B.C. remains a profound and intriguing mystery.

Fossil Water under the Sinai-Negev Peninsula

By studying the water issuing from springs and wells scattered across the Sinai and the Negev deserts, hydrogeologists have identified a great aquifer formed during the last glacial age

by Arie Issar

The large springs called 'Ayun Musa at the northwestern edge of the Sinai desert seem inexhaustible. Water from them floods an extensive marshy area partially covered by thick clusters of reeds, tamarisk trees and date palms. As the spring water surfaces its carbonates and sulfates precipitate and form mounds of porous tufa. The mounds have grown to heights of several meters, indicating that the springs have existed for many thousands, if not tens of thousands, of years. Biblical passages confirm the antiquity of the springs: "So Moses brought Israel from the Sea of Reeds and they went out into the wilderness of Shur and they went three days in the wilderness and found no water. And when they came to Marah, they could not drink the waters for they were bitter. Therefore the name of the place was called 'Marah'" (Exod. 15:22–23). "Mar" translates as "bitter" in Hebrew and 'Ayun Musa as "Springs of Moses" in Arabic.

What is the source of this rich flow of water at the edge of the desert? Hydrogeologists have determined that the 'Ayun Musa springs provide an outlet for a deep underground aquifer, or reservoir, storing a huge quantity of water. The water is tens of thousands of years old and is therefore called paleowater, or fossil water. Artesian pressure forces the water to the surface. By tapping this aquifer Egypt and Israel could develop large areas of the deserts of the Sinai-Negev peninsula into agricultural regions.

Four distinct lines of evidence support the conclusion that such an aquifer underlies the peninsula. Geological surveys of the peninsula show the setting is appropriate for an aquifer. Carbon-14 dating and chemical and isotopic analyses indicate that the waters of springs hundreds of kilometers apart have such strikingly similar characteristics that they must have one common source. The chemical and isotopic studies in conjunction with archaeological evidence suggest that the aquifer holds rainwater that was trapped during the most recent ice age.

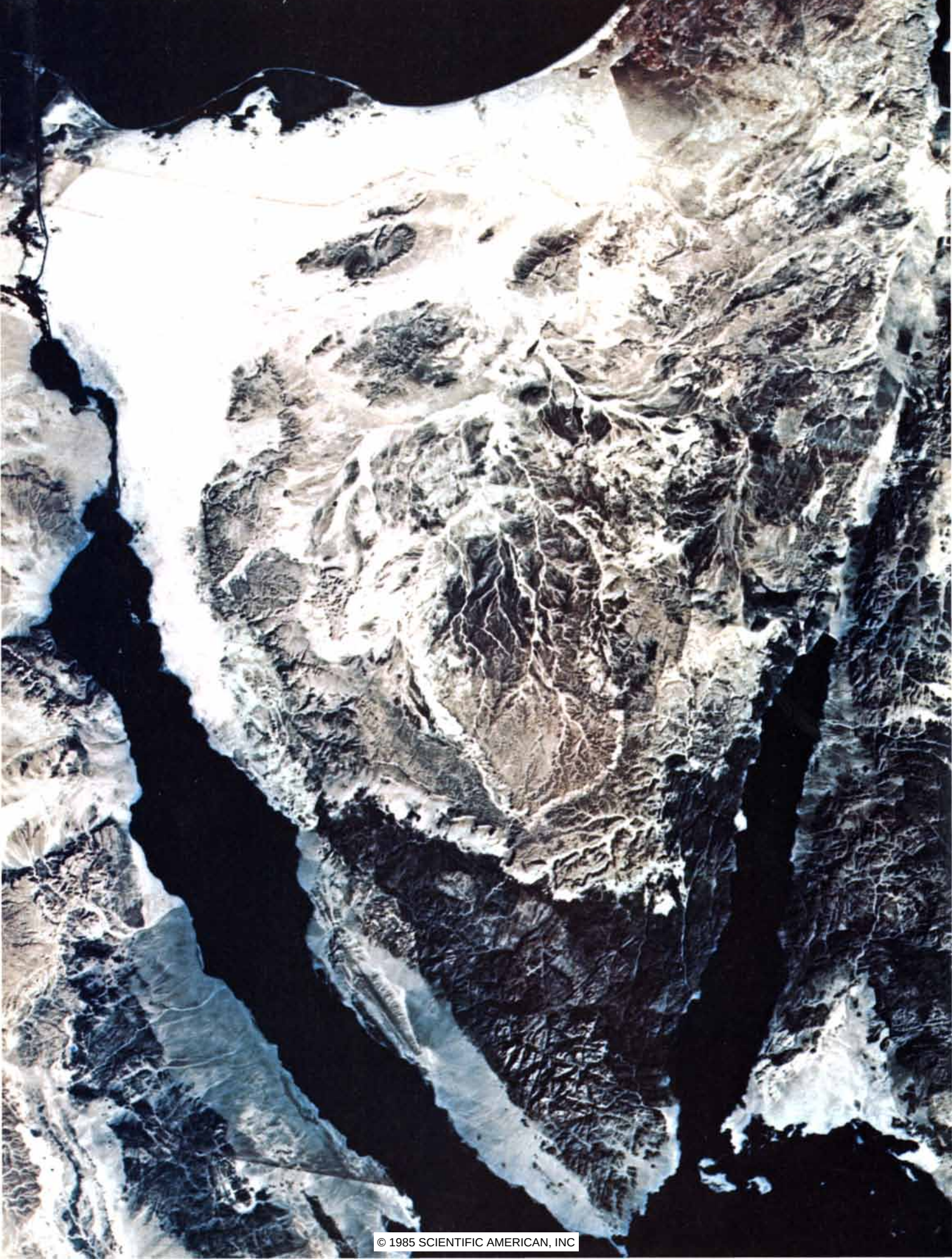
Spacecraft pictures of the region reveal surface characteristics consistent with subterranean geology that could support an aquifer. The triangular shape of the peninsula of Sinai (Egypt) and the Negev (Israel) that forms a transition zone between Asia and Africa is particularly striking. From the vantage of space the two regions divided by the political border between Egypt and Israel are geologically and geographically indistinguishable from each other. In fact, the Sinai and the Negev have existed as a single geological unit throughout most of history. Furthermore, this unit has features in common with the land areas bordering it: the Sinai peninsula resembles the deserts of Egypt, and the Negev resembles southern Jordan.

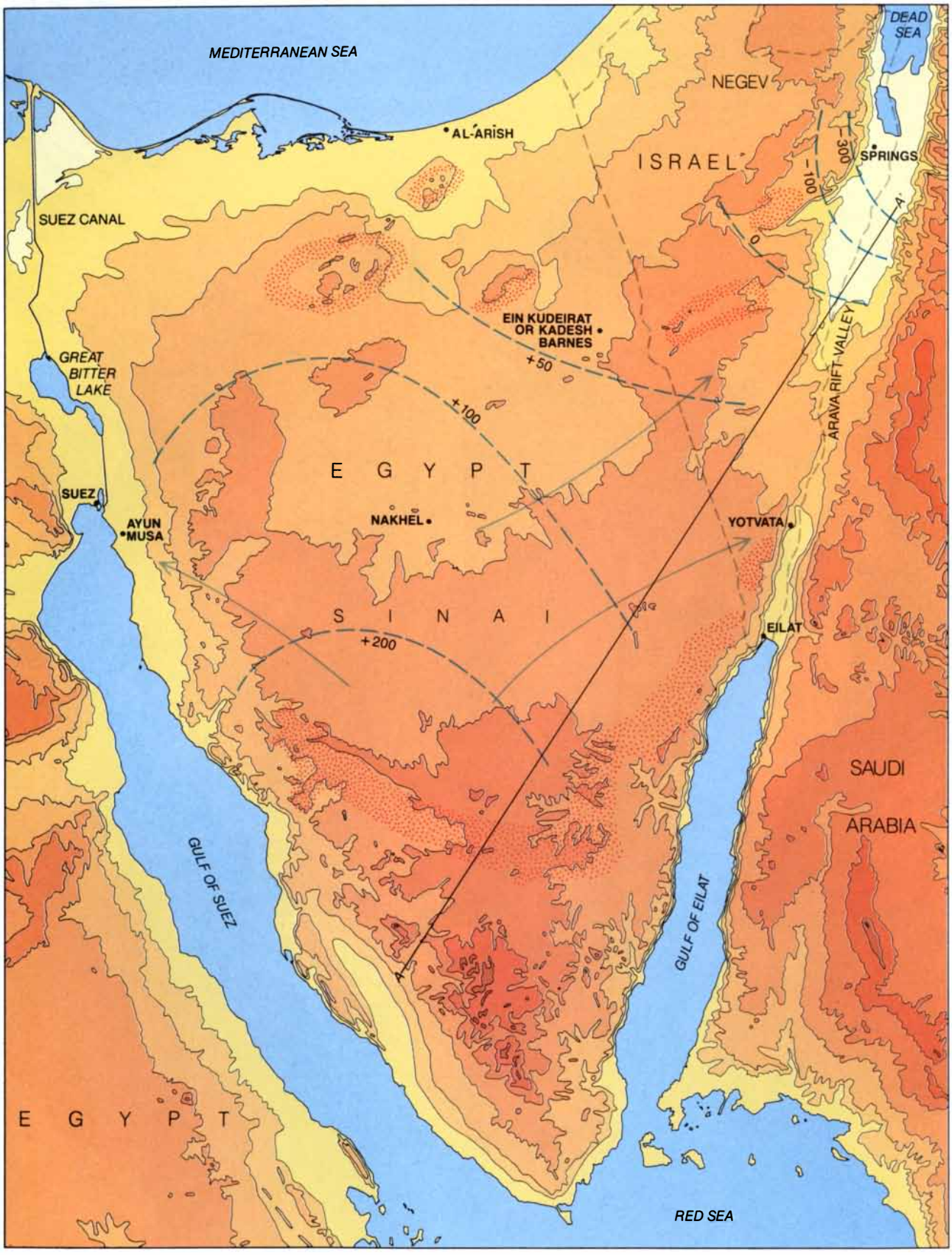
Indeed, the geological, structural and stratigraphic similarities between the Sinai-Negev peninsula and the African and Arabian plates allow the peninsula to be regarded as a miniature plate. The resemblance of the Sinai-Negev peninsula to its environs bears witness to its recent (geologically speaking) separation from Asia on the east and Africa on the west. The Gulf of Eilat, the Arava rift valley and the Dead Sea rift now divide the Negev from Asia, whereas the Gulf of Suez rift divides the Sinai from Africa. The separation of the Sinai and the Negev from Arabia and Africa, which took place at the dawn of the Quaternary period some two million years ago, was the culmination of plate movements during the Tertiary period.

The shearing and the down-faulting movements that began at the end of the Tertiary and continued through the lower and middle Quaternary gave the rift valleys their present morphological features and caused diversion of the drainage patterns from the Mediterranean Sea to the rifts. The rifting also profoundly affected the subsurface flow of groundwater. The regional faults along the rift valleys' borders fractured the rocks along the faces of the faults and displaced vertically impermeable layers, thereby opening flow routes between the deep subsurface and the surface. Water still flows up from the depths to the surface following these routes, forming springs and oases.

Not all desert oases, however, are fed by deep underlying water. For instance, Al-'Arīsh and other oases along the ancient routes near the Mediterranean that connect Egypt to the Fertile Crescent are fed by shallow groundwater tables under the sand dunes. The scanty rains (between 50 and 100 millimeters, or some two to four inches, per year) falling on the dunes seep quickly into the sand and so do not evaporate. The water accumulates on the impermeable underly-

SINAI-NEGEV PENINSULA shares geological, structural and stratigraphic similarities with both Asia and Africa, from which it separated about two million years ago. The similarities between the peninsula and northern Africa, where a vast quantity of water lies under the Sahara, suggest that a huge aquifer also lies under the Sinai-Negev deserts. The illustration is a composite of false-color Landsat images made with infrared radiation.





NUBIAN SANDSTONE OUTCROPS

WATER TABLE IN AQUIFER

DIRECTION OF FLOW

ARTESIAN SPRINGS AND WELLS situated hundreds of kilometers apart offer evidence for the existence of the aquifer. Water from the Nakhel well in the central Sinai is about 20,000 years old and water from the 'Ayun Musa springs on the western edge of the Sinai and from wells near the Dead Sea is about 30,000 years old. The waters from the springs and wells are remarkably similar in chemical and isotopic composition; most notably, the high sulfate content gives them a bitter taste. The profile along line *A*—*A'* is shown on page 109.

ing rock layers and forms a shallow groundwater table with a mild gradient toward the sea. The groundwater table is close to the surface of the land near the coast, so that inhabitants and travelers could dig shallow wells to reach the water.

This formation meant that a well-guided traveler in ancient times could depend on a reliable supply of water for his party and their pack animals over most of the northern route from Egypt to the Fertile Crescent. He would have to carry only enough water to last for some of the longer stretches between oases. Crossing the Sinai desert, however, becomes more difficult for even small groups, not to mention large ones, if they must choose a more southerly course. Water resources are scarcer along this route than they are along the northern route. Only two widely separated clusters of large springs are to be found. One is the 'Ein Kudeirat and Qadis spring cluster (the biblical Qadesh Barnea) near the Israeli border and the other is the 'Ayun Musa near the Gulf of Suez.

It has been clear for some time that the springs of 'Ayun Musa could not be sustained by the negligible local rainfall and flooding. As early as the 1940's Leo Picard of the Hebrew University of Jerusalem suggested that the source of these springs might instead be a huge aquifer under the Sinai-Negev peninsula. He formulated the hypothesis on the basis of the geological similarities between the region and northern Africa, where a large aquifer exists under the Sahara desert [see "Water under the Sahara," by Robert P. Ambroggi; SCIENTIFIC AMERICAN, May, 1966].

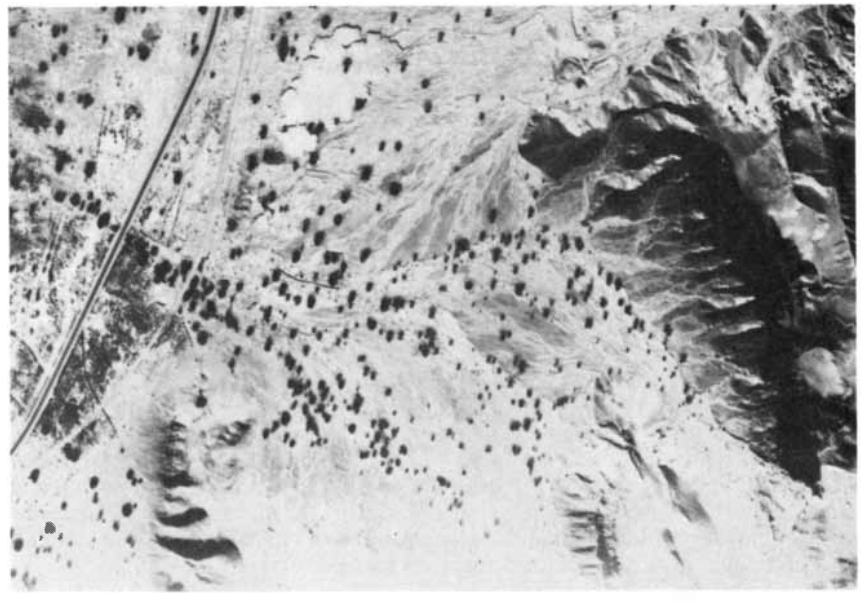
During the 1950's Zeev Shiftan of the Geological Survey of Israel observed that the springs emerging along the fault line near the Dead Sea on the eastern margin of the Negev desert yield a steady flow of water. Unlike most springs in Israel, these do not fluctuate with the season. Shiftan suggested the springs might be an outlet of a vast aquifer in the Nubian sandstone layer underlying the Negev. An evaluation of the quantities of water stored under the deserts of the Sinai and the Negev was not carried out, however, because of the scarcity of the data available for these regions. The springs of 'Ayun Musa and those along the shore of the Dead Sea, which taken together turn out to be strong evidence for the existence of a large aquifer, are hundreds of miles apart, and for some time their common source was not apparent.

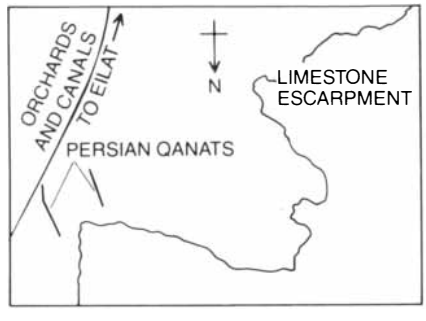
In the late 1950's a team of geophysicists from the University of Heidelberg produced the first scientific data supporting Picard's suggestion. By means of carbon-14 dating the investigators established that the water in the 'Ayun Musa is about 30,000 years old. Later studies indicate that water from artesian wells in the Nubian sandstone layer near the Dead Sea is also 30,000 years old and that water from an abandoned oil well dug in the Nubian sandstone layer in Nakhel in the central Sinai is 20,000 years old. Incidentally, geologists drilling deep exploration wells for oil in the central Sinai and near the Dead Sea shore found that the sandstone layer contains brackish water under artesian pressure.

The ages determined from carbon-14 dating agree with results of hydrologic flow models. Such models determine the time span required for water infiltrating the subsurface of the outcrop areas to flow to desert oases hundreds or even thousands of kilometers away. These ages also agree with those established for the water in the Sahara. They coincide with the latter part of the most recent glacial age, which dominated the globe from approximately 80,000 to 15,000 B.P. (before the present).

My own hydrogeologic investigations along the Arava rift valley fit nicely into the emerging pattern of evidence. I discovered a striking similarity in the chemical composition of the groundwater flowing along the fault lines that run the length of the western margin of the valley about 300 kilometers from 'Ayun Musa. The similarity is quantitatively expressed by comparing the sulfate content and the ratio of dissolved salts in water samples from different locations. In all the samples I found a high content of gypsum ($CaSO_4$) and epsomite (or English salt, $MgSO_4$). Other groups have found the same characteristics in water from the Dead Sea rift valley borders and from the 'Ayun Musa in the Suez rift valley. These sulfates, known for their bitter taste, explain the adverse digestive reactions of inhabitants and travelers, from the time of Moses, who drank the water from the fault-line springs of the Arava rift valley and from the 'Ayun Musa springs.

In spite of all the evidence pointing to the existence of the aquifer, the early 1970's found the community of hydrogeologists in Israel split in two. One group, convinced that the aquifer existed, argued that it could be exploited to change profoundly the prospects of development in these regions. The other group contended the data were too





CHAINS OF WELLS (called qanats), visible in the aerial photograph at the top, were planned by Persian engineers and dug in the Arava rift valley some 2,000 years ago to tap the underground water flow from the Sinai-Negev aquifer. Qanats are subsurface galleries that drain the groundwater table and bring the water to the surface. Shafts were dug for ventilation and to facilitate the removal of excavated material. Some of this gravel and soil was used to make ringlike mounds around the shafts, thereby revealing their locations to the overhead observer.

limited and scattered to support the conclusion that such an aquifer exists.

In light of this debate among contemporary hydrogeologists, it is interesting to note that ancient water engineers, presumably Persians or students of the Persian school of water finders, were aware some 2,000 years ago of the underground water flow from the Sinai. They dug a chain of wells in a direction perpendicular to the cliff of the Arava rift valley in order to tap the flow. Since no stream or alluvial fans extend in this direction, the ancients' discovery of the underground water is particularly impressive.

These chains of wells, called qanats [see "The Qanats of Iran," by H. E. Wulff; SCIENTIFIC AMERICAN, April, 1968], are subsurface galleries that form a drain under the groundwater table and bring the water to the surface. Along the galleries, which run parallel to the groundwater gradient, shafts were excavated to facilitate the inflow of air and the removal of excavated material. These shafts, each surrounded by a ringlike mound of excavated gravel and soil, are visible in aerial photographs.

Modern hydrogeologists did not resolve the debate over the existence of the aquifer until the recently developed method of isotope analysis was applied. This method measures the amounts in a water sample of heavy hydrogen (hydrogen 2, or deuterium) and heavy oxygen (oxygen 18). Water as it appears in nature consists predominantly of two normal hydrogen atoms (hydrogen 1) and one normal oxygen atom (oxygen 16). The relative amounts of deuterium and oxygen 18 in a particular water sample serve as its "fingerprint." Water samples from the 'Ayun Musa, from the abandoned oil-exploration well dug into the Nubian sandstone layer in Nakhel in the central Sinai and from the artesian wells in the Nubian sandstone layer near the Dead Sea all have the same relative amounts of deuterium and oxygen 18. Furthermore, the amounts differ from those of the contemporary rains and the groundwater of the Sinai and Israel. The results also demonstrate that the aquifer's outlets extend along the regional fault lines of the Suez and the Dead Sea rift valleys.

To study the isotopic composition of water in the peninsula, my colleagues and I developed a special sampling device. We were motivated by the difficult terrain, the scarcity of settlements, the randomness of rainstorms and flood events and the need to prevent evaporation of water samples. The sampler consists of a funnel whose diameter was designed to yield 200 centiliters of water, the amount needed for complete chemical and isotopic analyses. A magnetic device in each sampler seals the container hermetically when it is full. The waters are then diverted to the next container in a series of such samplers. We placed these samplers in observation stations all over the desert. To collect floodwater that infiltrates the gravel beds of desert streams, we dug special trenches in the stream beds and covered the samplers with gravel.

In addition to contributing to the resolution of the aquifer question our isotopic and chemical analyses suggest where the water in this great resource came from. We are indebted to the pioneering work of Joel R. Gat and Willfrid Dansgaard, who did their research at the Weizmann Institute of Science. Gat and Dansgaard found that the relative abundance of heavy isotopes in the rainwater and groundwater from the Mediterranean differs from the abundance in the rainwater and groundwater from the oceans. In general rainwater has a lighter isotopic composition than seawater. The phenomenon has a simple explanation. Since water molecules that contain any combination of deuterium and oxygen 18 are heavier than normal water molecules, they evaporate less readily than the normal molecules. The relative abundance of heavy isotopes in the water that evaporates from the seas or oceans is therefore less than that of ordinary ocean water. (The heavy-isotope concentration of ocean water is known as the standard mean ocean water.) As a result the relative abundance of heavy isotopes in clouds and rainwater is also less than the standard mean ocean water.

Rainwater reaching the surface of the land again evaporates, leading to a further concentration of heavy water. Any rainwater that seeps immediately to the subsurface to form groundwater escapes evaporation, however, and therefore retains its original isotopic composition. In other words, the isotopic composition of groundwater is generally the same as the isotopic composition of the rainwater from which it is formed.

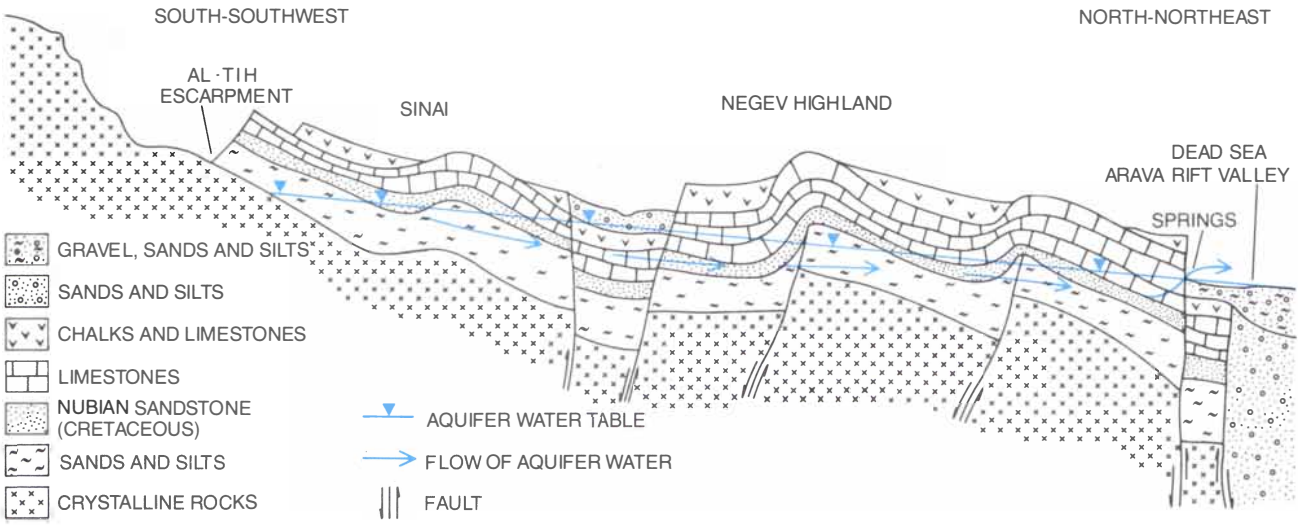
By analyzing the groundwater and rainwater of the Sinai-Negev, my colleagues and I have reconstructed the paleoweather patterns that dominated the region when the aquifer was filled. In the course of the reconstruction we also discovered why the aquifer water is bitter.

To follow the explanation it is important to grasp another aspect of the technique of isotopic analysis. For thermodynamic reasons air has eight times as many molecules incorporating deuterium as it does molecules incorporating oxygen 18. The relative abundances of these heavy isotopes are further affected by local conditions of evaporation. The effect of these conditions is reflected by a measurable quantity called the deuterium excess.

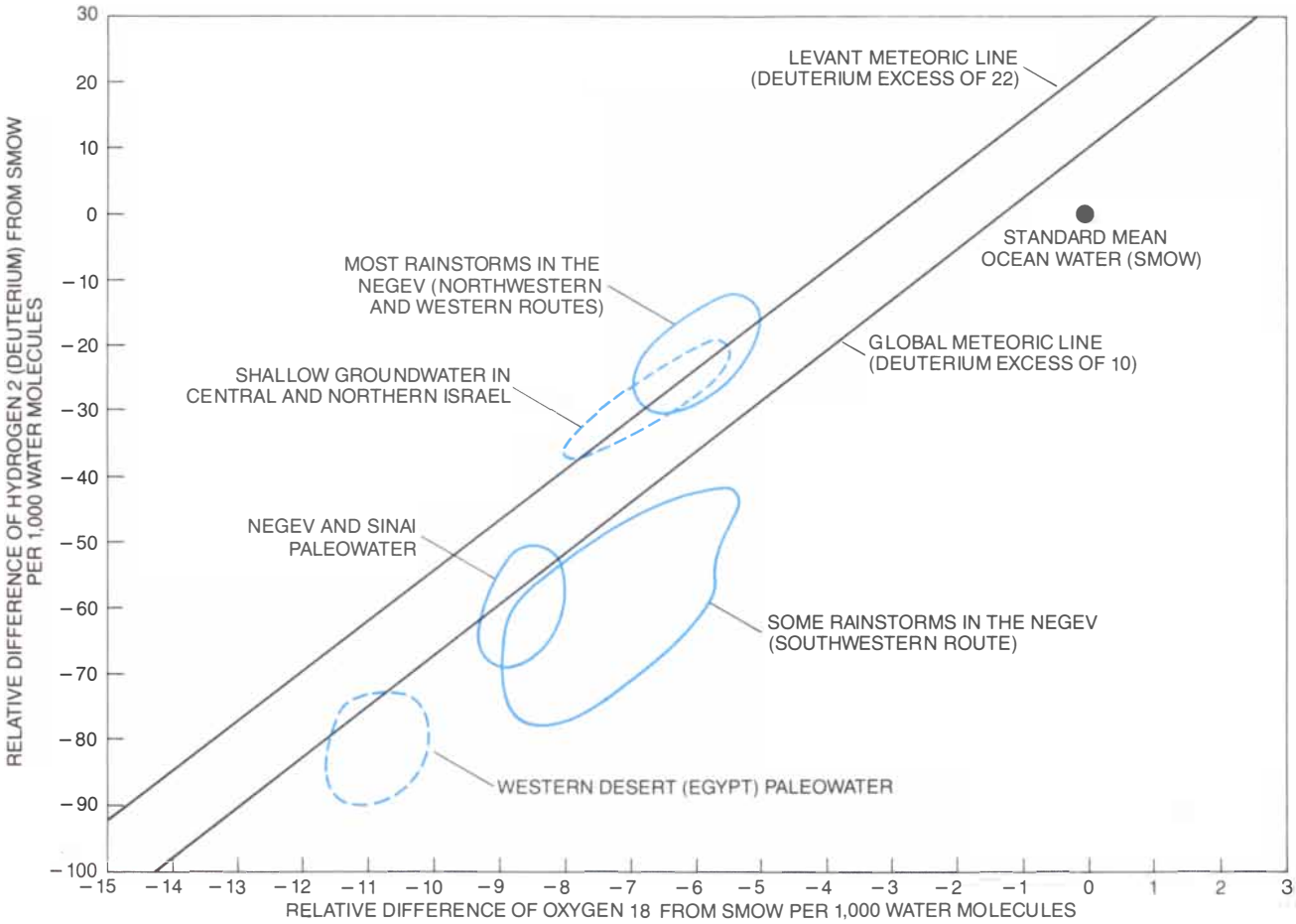
A network of rain-sampling stations distributed all over the world by the International Atomic Energy Agency has established that on the average the world's rainfall has a deuterium excess of about 10. In the Mediterranean area, on the other hand, the deuterium excess of most contemporary rains is about 22. Thus the heavy-isotope ratios of the rains in countries near the oceans fall on one line, known as the global meteoric line, whereas the heavy-isotope ratios of most of the rains in countries near the Mediterranean fall on another line, known as the Levant meteoric line [*see bottom illustration on opposite page*].

After a rainstorm and subsequent flood events in the Negev my colleagues and I would analyze the water collected by our sampling devices for its chemical and isotopic composition. The analysis yielded interesting results regarding the chemistry of the rains and floods. One finding was that the amount of airborne salts brought in by the rainstorms was larger than the amount exported by the floods. The finding explains why desert soils become salty. A comparison between the path of a rainstorm and the rain's isotopic composition showed a statistical correlation between the isotopic composition and the direction from which the storm came. Most of the rainstorms over the Negev come from the northwest and north and are similar in isotopic composition to the rains of northern Israel. They have a deuterium excess of about 22, as does the shallow groundwater of the Sinai and Israel. A few storms come from the west and southwest. They are depleted in deuterium and oxygen 18 and are characterized by a deuterium excess of 10. The paleowater from the Sinai-Negev aquifer also has a deuterium excess of 10.

The isotopic composition of the water from rainstorms coming from the south and southwest is due to the thermodynamic conditions prevailing over the Mediterranean, from which these storms receive most of their moisture, and the route they follow over the deserts of Libya and Egypt. The conformity in isotopic composition between the storms and the paleowater of the Nubian sandstone gives a clue to the nature of the climatic conditions that prevailed in this part of the world during the most recent glacial period



GEOLOGIC CROSS SECTION shows the flow of water in the Sinai-Negev aquifer from the mountains of the central Sinai to the springs along the fault line near the Dead Sea. The water is stored in Nubian sandstone, a layer that was formed in the Cretaceous period some 100 million years ago. Artesian pressure forces the water to the surface, where it flows steadily regardless of the season.



HEAVY-ISOTOPE ANALYSIS of water collected by samplers led to an explanation of how the Sinai-Negev aquifer was formed. The relative amounts of heavy hydrogen (hydrogen 2, or deuterium) and heavy oxygen (oxygen 18) in ocean water is called the standard mean ocean water (SMOW) and is assigned the value of 0. (Most naturally occurring water consists of two normal hydrogen atoms, hydrogen 1, and one normal oxygen atom, oxygen 16.) For thermodynamic reasons air has eight times as many water molecules comprising deuterium as it does molecules comprising oxygen 18. A plot of the relative amount of deuterium as opposed to the relative amount of oxygen 18 in rain samples therefore determines a line having a slope of 8. (The relative amounts of heavy isotopes are negative because they are less than SMOW.) Local conditions of evaporation shift this line up by an amount called the deuterium excess. Both the contemporary rains reaching the Negev from the north and northwest and the shallow groundwater in the peninsula lie on the Levant meteoric line, whose deuterium excess is 22. The isotopic composition of the contemporary rains from the southwest and that of the aquifer water both lie on the global meteoric line. The deuterium excess is 10. These results are consistent with the supposition that the rains filling the aquifer some 30,000 years ago followed a southwesterly route over the southern Mediterranean, northern Africa and the Gulf of Suez. The storms originated over the North Atlantic and were driven south by a cold front.

(Würm), when the Nubian sandstone aquifer was filled.

The Würm ice age began in the uppermost Pleistocene, about 80,000 years ago. During this period the seas retreated as the glaciers captured water. Bathymetric charts of the Gulf of Suez and the shore of the Mediterranean indicate that the water in these places is very shallow, not exceeding a few tens of meters. Therefore it can be concluded that during the most recent glacial period the Gulf of Suez and the upper shelf area of the Mediterranean shore along the Sinai dried up, causing the formation of salt flats, marshes and lagoons. The salt flats consisted mainly of calcium and magnesium sulfate salts (gypsum and epsomite).

During this period a cyclonic cold high-pressure area over Europe deflected in a southerly direction many rainstorms coming from the North Atlantic and passing over the Mediterranean. Propelled by the high-pressure area, the storms entered the African continent over the Libyan coast and drove before them huge clouds of dust. When the storms reached the dried salt flats, they picked up salt and gypsum, which became mixed with the dust. Consequently the rains that followed the dust storms were laden with mud, and they caused floods that reached the outcrops of the Nubian sandstone and filled its pores with brackish water.

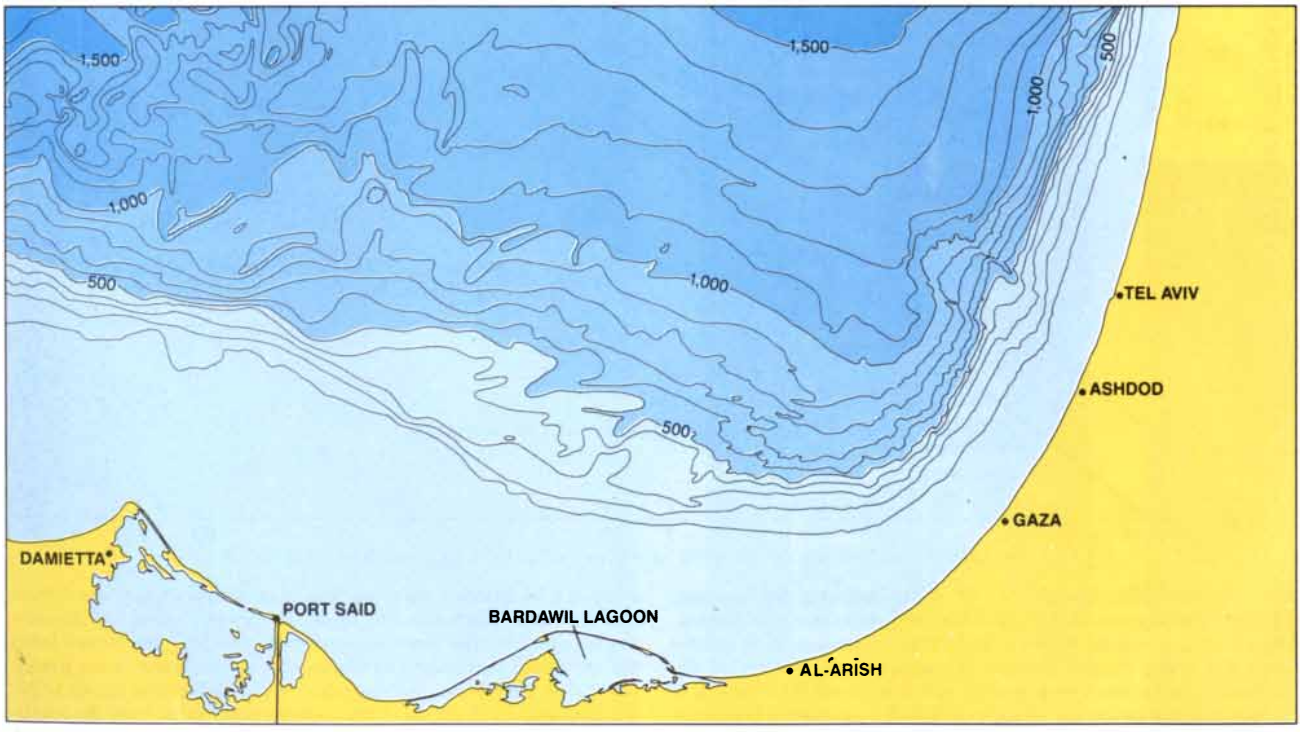
One corollary of our reconstruction of the prevailing paleoweather patterns during the filling of the aquifer is that the climate of the Würm ice age must have been drastically different from the climate of today. Indeed, Gat and I concluded that the climate prevailing in the regions south and east of the Mediterranean in the most recent glacial age was a semiarid one with summer rains.

The results of archaeological exploration confirm the supposition. Numerous prehistoric sites and tools from Paleolithic to Neolithic times are scattered all over the Sahara. Many rock engravings of such savanna animals as giraffes, elephants and ostriches, as well as their fossilized bones, also attest to the sultry paleoclimate of the region.

Tens of thousands of years later the tribes of Israel escaped serfdom in Egypt and traveled eastward. Led by Moses, the Jews crossed the valley between the northern edge of the Gulf of Suez and the Mediterranean. Then, after escaping from the pursuing Egyptian army, they turned away from the coastal route, where Egyptian garrisons were stationed. Instead they took a more southerly route. Traveling southeast, they reached the springs of 'Ayun Musa. Since they were used to the fresh Nile water, the taste of the water—particularly the bitterness of the sulfate—was not to their liking. They protested the taste and thereby produced one of history's earliest hydrochemical records.

Hydrogeologists now calculate that the Nubian sandstone aquifer under the Sinai and the Negev holds 200 billion cubic meters of water, 70 billion cubic meters of which is under the Negev. Agricultural settlements in the Negev demonstrate that the water is low enough in salt content to be suitable for irrigation. Since the water, like that found under the Sahara and Arabia, cannot be replenished, its extraction is analogous to the extraction of oil. A long-range plan must be devised for managing the resource.

At present 25 million cubic meters per year of the Nubian sandstone water is being used for industrial and agricultural purposes by Israel. Israeli engineers and hydrogeologists are investigating the feasibility of driving galleries into the sandstone layers in order to pump 300 million cubic meters of water per year in the coming century. They dream of turning the Arava rift valley into a densely populated agricultural region. Preliminary studies give encouraging results and research continues in that direction.



SHALLOW WATER along the southern coast of the Mediterranean provides the final clue to the story of the formation of the Sinai-Negev aquifer. (Depths in this bathymetric chart are indicated in meters below sea level.) Throughout the last ice age, which lasted from about 80,000 to 15,000 B.P. (before the present), the capture of water in glaciers caused the shallow regions (depths down to 100 meters) of the Mediterranean and the Gulf of Suez to dry up, forming salt flats in their wake. The southerly rains swept the salts from the flats. The resulting mud-laden floods seeped into the outcrops of the Nubian sandstone and filled the aquifer with bitter water.

COME ALIVE, COME AND DRIVE

NISSAN

MAJOR MOTION

MOVE OVER!

MAKE WAY FOR THE NISSAN 300 ZX.

FUEL-INJECTED • TURBO-CHARGED • V-6 • 200 HORSEPOWER • 3-WAY ADJUSTABLE SHOCKS.

AWESOME!

AT DATSUN DEALERS

NISSAN NISSAN

BELT YOURSELF

Milkweeds and Their Visitors

The insects and other animals that frequent milkweed form a model community for the study of interactions among species. The animals come to forage, but a few of them also serve the needs of the plant

by Douglass H. Morse

The common milkweed is host to an annual gathering of insects and other animals. Activity at the plants reaches a peak in early summer, when the milkweed's flowers attract bees, butterflies, moths and various smaller insects. These species are drawn to the sugary nectar secreted by the flowers, but several of the insects also serve the needs of the plant by carrying pollen, often over considerable distances. Other visitors to the milkweed feed on the plant itself: aphids suck the sap, for example, and caterpillars of the monarch butterfly eat the leaves. Later in the summer milkweed bugs attack the seeds. The presence of all this animal life brings in turn parasites and predators, including certain wasps and spiders, and scavengers that lay claim to whatever the predators leave behind.

The activity at milkweed plants attracts still another occasional visitor: the ecologist, who finds in the milkweed community a convenient microcosm for the study of interactions among species. The community is rich and varied enough to be of interest, and yet it is not so complex that the number of links between species becomes unmanageably large. Furthermore, the community is reasonably self-contained: at any given time the animals found at milkweed plants have few interactions with other plant species, and some of the animals spend their entire lives in association with milkweed. For these reasons milkweed provides a useful model for ecological field studies.

Milkweed takes its name from the milky white sap exuded when the plant is cut or broken. The family name, Asclepiadaceae, is from Asclepios, the Greek god of healing, and also refers to properties attributed to the sap: at one time or another preparations of milkweed were recommended for many medicinal uses, from the treatment of heart disease to contraception. Whether or not any of these medications were effective, it is true that some milkweed plants contain potent alkaloids (plant substances active in higher animals).

The milkweed I have been studying is *Asclepias syriaca,* the common milkweed, which grows throughout much of eastern and central North America. It is a colonizer of open ground and is often found in fields and pastures, along roadsides and even on vacant urban land. Like other invaders of clearings, milkweed tends to grow quickly but has little tolerance for shade. Over a period of a few years dense stands of dozens or even hundreds of plants sometimes develop. The stands are created by the underground growth of rhizomes, or root shoots, rather than by sexual reproduction. Thus all the plants in such a stand are genetically identical; they constitute a clone.

A typical plant is about three feet tall with oval leaves arrayed in pairs on an upright stem. The plants bloom in June and July, then the fertilized flowers grow into large conical seedpods. In fall the pods open, releasing seeds attached to silky tufts, which are scattered by the wind.

One interaction between the milkweed plant and an insect is a classic of the ecological literature. Larvae of the monarch butterfly, *Danaus plexippus,* feed on milkweed foliage and store the plant's alkaloids in their tissues. The alkaloids make the adult butterflies unpalatable to birds, which therefore learn to avoid them. Other butterflies that mimic the monarch's appearance benefit from its chemical defenses even though they do not feed on milkweed and do not contain the noxious alkaloids.

Although the evolutionary biology of the monarch and its mimics is full of fascinating intricacies, the monarch's relation to the milkweed itself is that of a simple herbivore. Other insects have more complex interactions with the plant. The species that bring about pollination raise some of the most interesting issues.

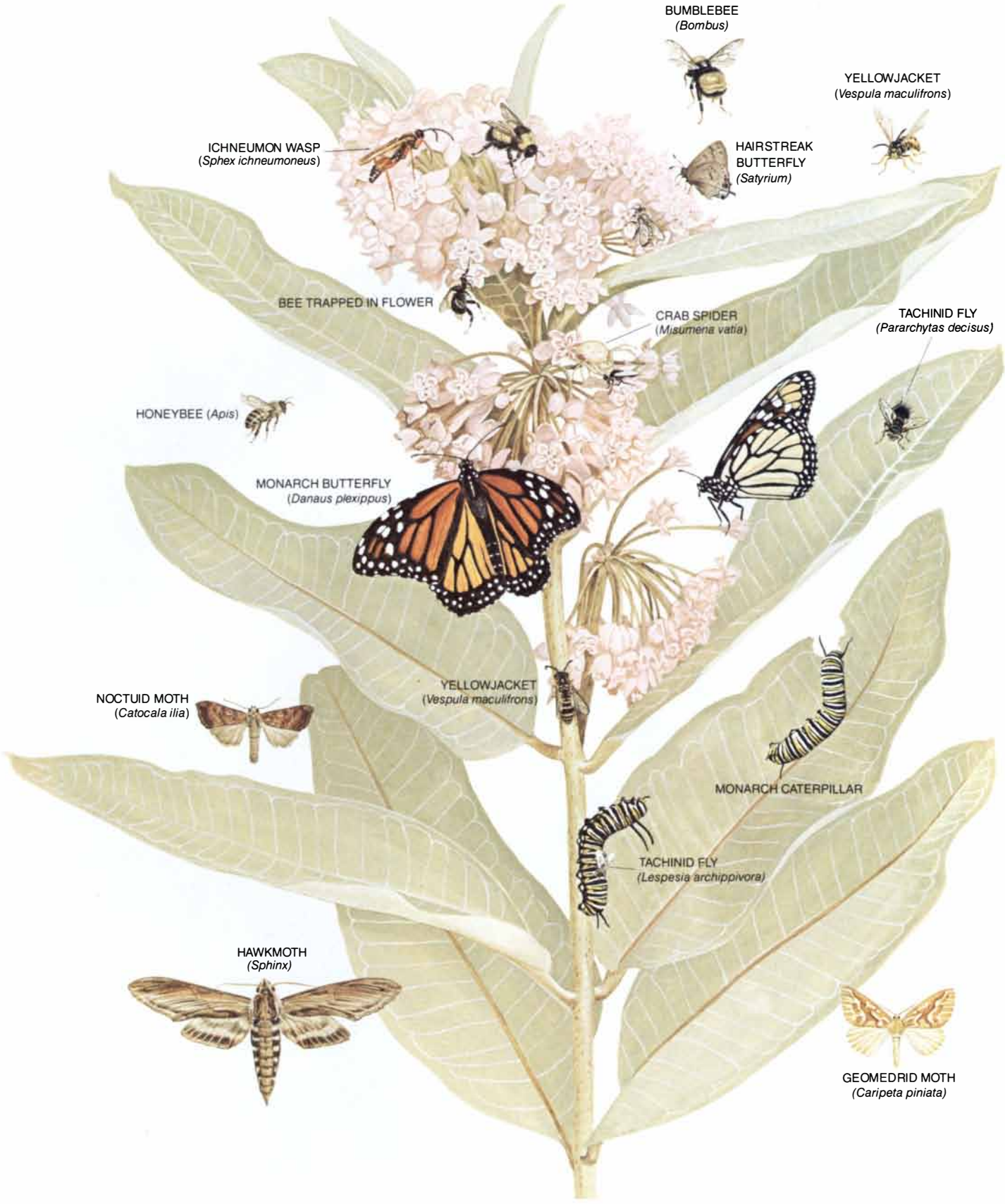
In the majority of flowering plants pollen grows on stalklike anthers; the ovules are at the base of an elongated pistil. The pollen is shed in single grains or in loose clusters of a few grains; when a grain reaches the sticky surface of the stigma at the tip of the pistil, a pollen tube grows downward to the ovules, where fertilization takes place. The architecture of milkweed flowers is different in several respects. The pistil is very short, whereas the stigma is greatly enlarged and has the form of a deep slit. Moreover, the pollen is not shed in single grains but is packaged in structures called pollinia, in which many grains (enough to fertilize the seeds of a single pod) are held together by a waxy coating.

A milkweed flower consists of five parts; each part has a stigma and a pair of pollinia lodged in sacs on each side of the stigma. The pollinia of each pair are connected by thin bands called translator arms to a clip, the corpusculum, at the top of the stigmatic slit. The entire pollination unit (two pollinia, two translator arms and a corpusculum) has the general shape of a wishbone and is called a pollinarium.

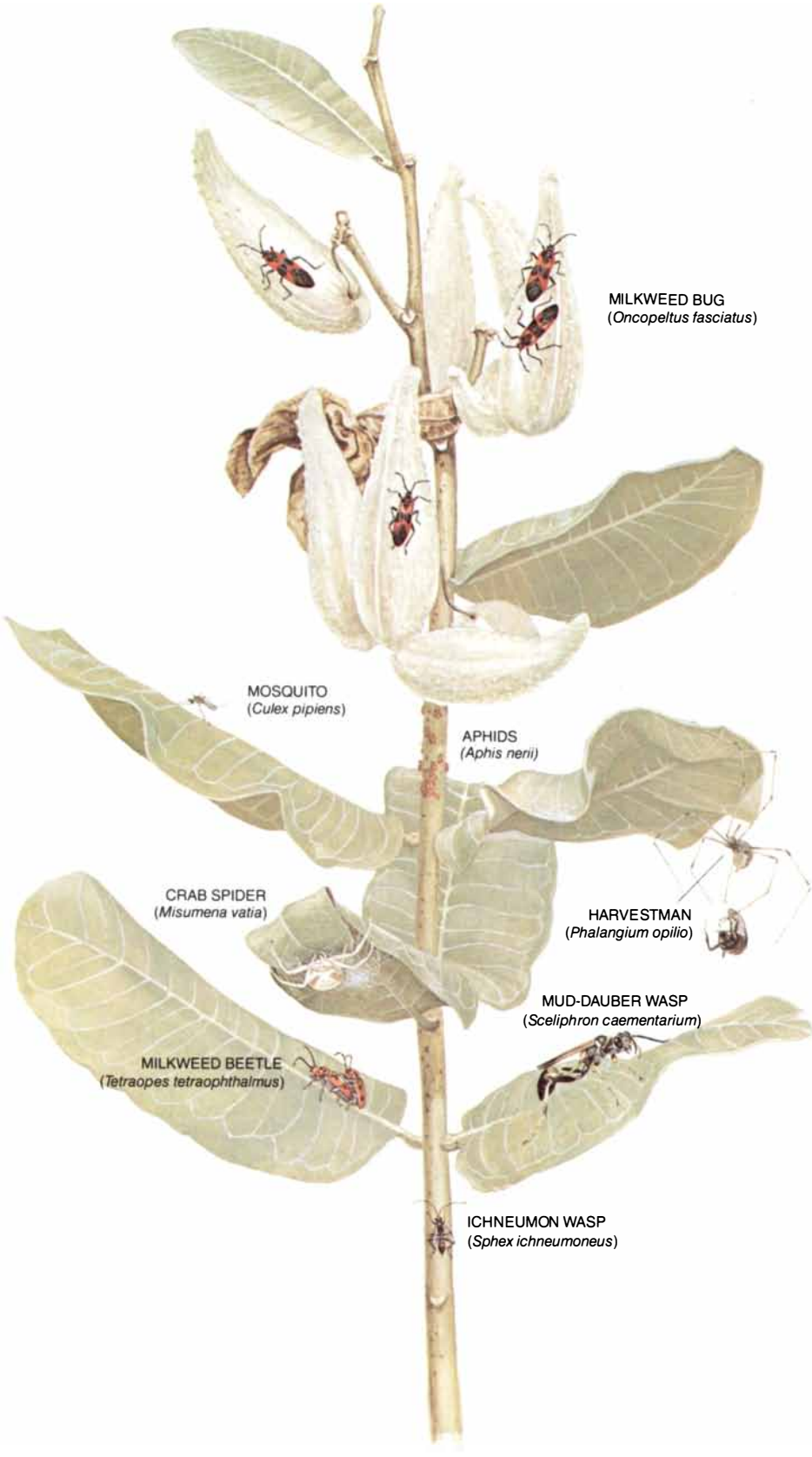
On the exposed surface of the corpusculum is a groove that narrows from bottom to top. As an insect walks over the flower, its legs or mouthparts can become wedged in the groove. When the insect then leaves, the pollinia are pulled from their sacs and the entire pollinarium is carried away. A few minutes later a twisting of the translator arms rotates the pollinia by 90 degrees. Before the rotation the pollinia are arranged rather like feet with the toes pointing forward; afterward the toes point outward.

If an insect carrying a pollinarium subsequently visits another milkweed flower, one of the pollinia may enter

BURSTING SEEDPOD of a milkweed plant releases between 100 and 200 seeds borne on airy tufts of silk. Although wind is the agent of seed dispersal, at an earlier stage in the reproductive cycle the plant is dependent on the services of animals: cross-pollination, which is essential to reproductive success, can be carried out only by bees, moths and other large insects. Milkweed is an early colonizer of open ground. Because suitable sites may be widely scattered, both seeds and pollen need to be capable of covering great distances.



SUMMER VISITORS to milkweed include a variety of herbivores, predators and parasites, but the nectar-feeders that forage on the flowers are the most important to the welfare of the plant. Bees attracted by the nectar are the chief pollinators of milkweed, although some nocturnal moths (including certain noctuid and geometrid species) may also contribute. Other insects feed at the flowers but seldom carry pollen. Prominent among the herbivores is the monarch butterfly, which feeds on foliage as a larva and returns to consume nectar as an adult. Crab spiders prey on visiting insects. Tachinid flies are parasites of other animals, including monarch larvae.



MILKWEED COMMUNITY IN FALL has a different composition. Nectar-seekers leave. Aphids, which suck the sap, infest the plants intermittently. Milkweed bugs suck on the developing seeds and milkweed beetles eat the foliage. The crab spider continues its predation, and scavengers such as the harvestman, or daddy longlegs, recover the remains of what the predators have killed. The ichneumon and mud-dauber wasps are parasites.

the enlarged lower end of the stigmatic slit. As the insect moves, the pollinium is pulled farther into the slit until, at the upper end, it bumps against the corpusculum of the "resident" pollinarium (assuming that structure has not already been removed). At this point the translator arm often breaks, leaving the pollinium in the stigmatic chamber. The individual pollen grains within the pollinium then put out pollen tubes that grow toward the ovules at the base of the flower. Meanwhile the resident pollinarium may well have been carried off by the insect. In many cases the corpusculum of one pollinarium becomes attached to the broken translator arm of another, so that long strings of pollinaria build up on an insect's appendages.

In analyzing the milkweed's reproductive strategy, it is essential to know that the plant is an obligate or nearly obligate outbreeder. Self-pollinated flowers are rarely fertile. Susan R. Kephart of Willamette University and others have shown that some populations of milkweed do have a limited capacity for self-fertilization, but even in those cases the pods formed by self-pollination usually abort early in development. The requirement of cross-pollination is particularly important in view of the plant's tendency to grow in extensive clones. Insects may forage only at a single clone and so bring about only self-pollinations.

Certain features of milkweed's pollination mechanism may favor cross-pollination. For example, the rotation of the pollinia a few minutes after they are pulled out of their sacs gives them an orientation in which they readily fit into a stigmatic slit. Before the rotation they are much less likely to be inserted. The delay could allow time for an insect to move from one clone to another. The means by which the translator arm breaks, leaving the pollinium in the stigma, may also promote outbreeding, although the analysis of this factor is complicated. The translator arm is more likely to break if the resident pollinarium has not already been removed. Hence, at least during the early stages of the bloom, those flowers that have not given up their pollen are the likeliest to be fertilized. They may be the ones that most often intercept pollen from other clones.

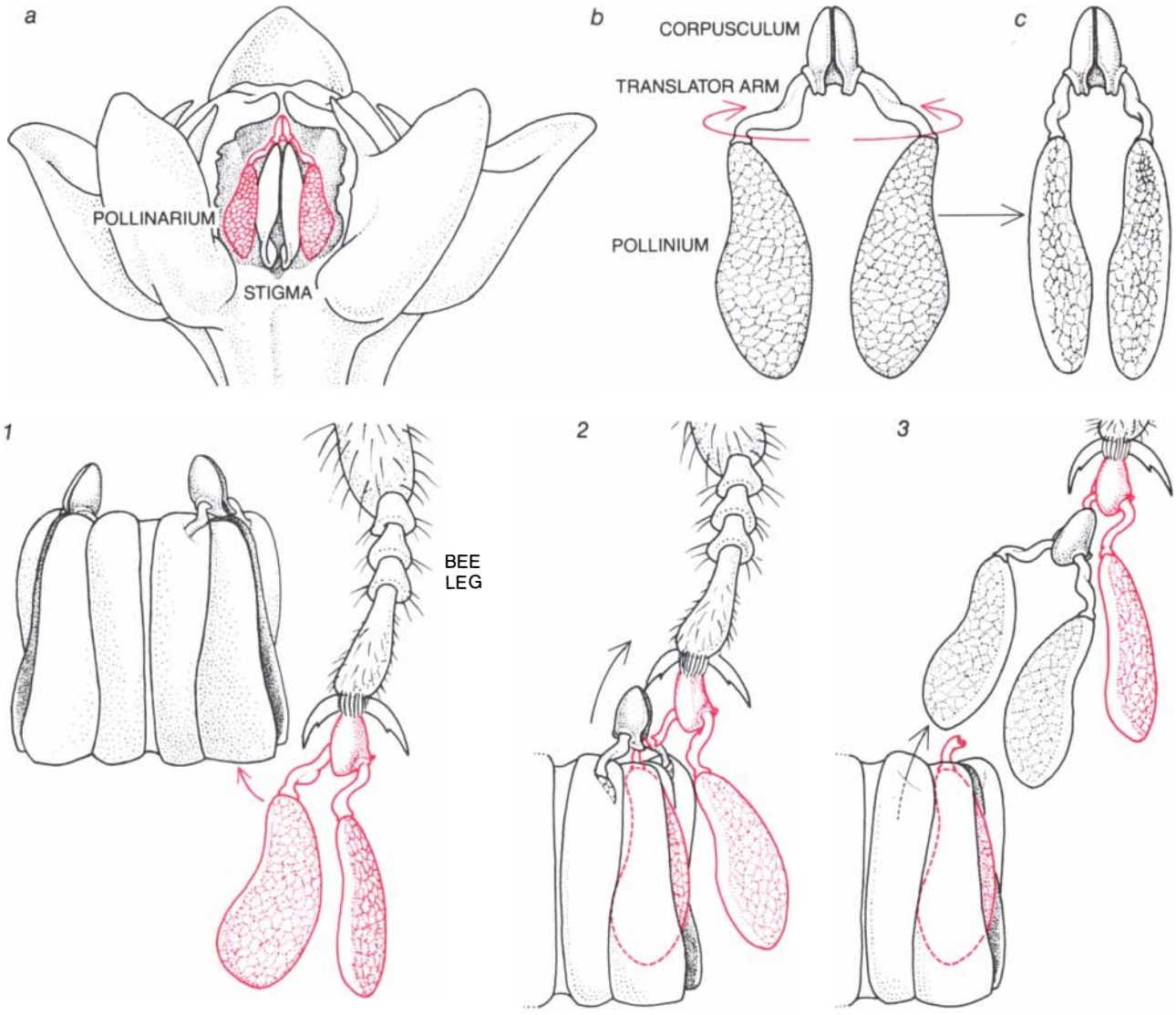
Milkweed flowers have two ovaries, so that in principle each flower could give rise to two seedpods; in most cases, however, no more than one pod develops to maturity. In many areas the factor limiting the number of pods is probably the supply of soil nutrients, as Mary F. Willson and her colleagues

at the University of Illinois at Urbana-Champaign have shown. Under these conditions a form of sexual selection may operate. In birds and mammals sexual selection is a straightforward process: typically a female chooses her mate according to some conspicuous trait, such as brilliance of plumage, which therefore tends to be accentuated in the population. In milkweed the process is somewhat different. A flower cannot choose which pollinia it receives, but if multiple pollinia are available and only one flower is allowed to develop a mature seedpod, there is a possible means of control.

Susan S. Bookman of Washington State University has tested the concept in experiments with the showy milkweed (*Asclepias speciosa*), a species closely related to the common milkweed. By hand-pollinating flowers with pollinia from various clones she has demonstrated that certain clones are favored as male parents. Milkweed plants evidently select the "best" fathers from those available, although the criteria for judging fitness for fatherhood are not known.

My own studies of *A. syriaca* growing in a coastal region of Maine lead to a different conclusion. In this area it seems pod production is limited not by nutrient resources but by the availability of pollen. With hand pollination the number of pods reaching maturity can be increased substantially. Given a shortage of genetically suitable pollen, it seems unlikely that sexual selection has much of a role in the reproduction of these clones. The study of milkweed pollination therefore offers an opportunity to evaluate the importance of sexual selection under differing natural conditions.

Because milkweed typically occupies early successional areas, such as transitory forest clearings, the nearest clones available for cross-pollination may be at a distance of many meters or even kilometers. This sparse distribution puts a premium on securing the services of large insects capable



POLLINATION OF MILKWEED relies on the services of large, strong insects. The milkweed flower (*a*) has an unusual anatomy, shown here in a cutaway view. Pollen is packaged in waxy structures called pollinia; the stigma, which receives the pollen, has the form of a slit. For each stigma there is a pair of pollinia connected by thin bands called translator arms to a corpusculum: a clip with a central groove that narrows from bottom to top. The structure of pollinia, translator arms and corpusculum is called a pollinarium (*b*). After a pollinarium is removed from a flower, a twisting of the translator arms rotates the pollinia by 90 degrees (*c*), changing their orientation so that they fit into the stigma of another flower. The pollinium enters the lower end of the stigma (*1*). It is drawn farther in until it bumps against the corpusculum of the resident pollinarium (*2*). The translator arm breaks, leaving the pollinium implanted in the stigma. The resident pollinarium may become attached to the stub of the broken translator arm and be carried away (*3*).

of flying great distances. Such insects can be attracted only by offering large rewards, which milkweed supplies in the form of copiously secreted nectar.

In my study area the commonest large visitors to milkweed flowers during the daylight hours are bees, primarily bumblebees. Anatomically they are well suited to the task of pollination, and indeed they are found to bear pollinaria on their legs and mouthparts. On the other hand, most of the bees are workers and must travel back and forth between the flowers and the hive. An individual bee can therefore be expected to visit comparatively few clones and to carry pollen shorter distances than an insect that does not have to return regularly to a fixed base. Foragers such as butterflies, moths and drone bees, for example, are free to wander over large areas.

Butterflies are frequent visitors to milkweed and are not constrained to radial flights from a central base. The monarch, in particular, visits the clones in my study area in large numbers and seems to be a good candidate for dispersing pollinia over long distances. The findings so far, however, indicate that butterflies do not have an important role in pollinating common milkweed. Bumblebees usually outnumber them at milkweed plants by a ratio of 10 or 20 to one, and the bees carry pollinia four times as often. Furthermore, it turns out the two groups move between clones with similar frequency. In the case of the monarch butterfly the evidence is particularly clear: in 150 captures of 75 marked individuals I have not found a single pollinium of the common milkweed.

Thus in spite of being tethered to a central hive bees appear to be the most important daytime pollinators. In my study area two bumblebee species, *Bombus terricola* and *B. vagans,* probably account for between 75 and 90 percent of the diurnal pollinations.

There is more to the story, however. At night large numbers of moths visit milkweed, and some of them carry pollinia on their feet and mouthparts. Brown moths of the family Noctuidae, about the size of a small bumblebee, are the commonest nocturnal visitors. In order to measure the contributions of the diurnal and the nocturnal pollinators I carried out a series of experiments. I covered sets of flowering stems with nylon netting to exclude flying insects either by day or by night and counted the pods produced in each group of plants. The results were consistent from clone to clone and from year to year: most of the plants were pollinated during the day, but not all of them. Contributions of the moths ranged from 5 to 25 percent.

MONARCH BUTTERFLY, ***Danaus plexippus,*** **is one of the species most closely associated with milkweed. Here monarch caterpillars are seen feeding on the leaves. Alkaloids present in some milkweed plants make the adult butterflies unpalatable to birds, giving both the monarch and species that mimic its appearance a measure of protection from predators.**

In view of this result it is surprising to find that milkweeds secrete considerably more nectar at night than they do during the day. How can one explain this bias, which seems contrary to the plants' own interest? One hypothesis is that the nectar is secreted not when it is needed but when the cost to the plant, in terms of water loss, is lowest. The idea would be plausible if nectar tended to accumulate at night, but it does not. Nocturnal secretions are exploited by moths or other night-flying insects; little remains in the morning to draw daytime pollinators.

Milkweed does have some traits ordinarily associated with nocturnally pollinated plants. It is notably fragrant at night, at least to human senses, and its whitish flowers should aid recognition. Moths are known for their response to vanishingly small amounts of female scent, but they can also orient to other odors. N. B. M. Brantjes of the University of Nijmegen in the Netherlands has shown that both noctuid and sphingid moths can find point sources of floral odors with an accuracy of a few meters; at that range milkweed flowers might serve as effective visual cues, even in the dark.

The presence of attractants for both diurnal and nocturnal insects suggests that milkweeds may have evolved under conditions in which there were periodic shortages of pollinators. Being able to exploit either population would then be advantageous. If the original primary habitat of milkweed was made up of scattered forest clearings, then insects that routinely move through the intervening forest may have been essential to cross-pollina-

tion. Moths are likely candidates for the role. I have begun to test this hypothesis in two ways: by monitoring the pollinia loads of moths captured in the forest and by marking moths found at various clones and following their subsequent dispersal. Preliminary results indicate that pollinia-bearing moths of several species course widely through the forest, hundreds of meters from the nearest milkweed clone, and that few moths stay for long periods at a single clone. It remains to be determined whether the movements of the moths are rapid enough for them to carry viable pollen over the long distances between clones.

It is not known with any precision how long milkweed pollen remains viable. Through hand-pollination experiments I have found that pollinia removed from three-day-old flowers are still viable, and even pollinia from five-day-old flowers can sometimes produce pods if they are inserted into the stigmas of newly opened flowers. The question, however, is whether pollen that has been removed from the flower and is being carried on the legs of an insect retains its viability equally long. The waxy coating of milkweed pollinia may well extend their lifetime. So far I have demonstrated that pollinia carried by insects for up to three days can still produce pods, but the studies need to be extended. (A century ago Charles Robertson, an Illinois naturalist, reported that pollinia of *Asclepias sullivantii* were viable for two weeks after their removal from flowers.)

From a pollinator's point of view, foraging at milkweed flowers can be somewhat risky. Only large insects are strong enough to remove the pollinaria; smaller insects sometimes become caught in the corpuscular grooves and die in the flowers. The trapped insects are typically moths and flies with a mass of from 15 to 50 milligrams (a housefly weighs about 20 milligrams), but even small bumblebees and honeybees weighing as much as 75 to 100 milligrams occasionally become ensnared. Larger bees are able to extricate themselves, but they may lose claws or even segments of a leg.

In addition to the actual and potential pollinators, many other insects feed at milkweed flowers without pollinating them. Among flying insects the commonest are small moths, flies and mosquitoes weighing 10 milligrams or less. Although they remove only a small proportion of the nectar produced, it is probably a major source of nutrition for them. The most important nectar parasites, however, are not flying insects but ants. Robert S. Fritz of Vassar College and I have shown that nectar does not accumulate in milkweed flowers attended by the ants *Lasius neoniger* and *Tapinoma sessile* even in the absence of the flowers' flying visitors. Ants seldom carry off pollinaria; in any case, the ants' range of action is so limited that they could not contribute to cross-pollination. Fritz and I have found that ants significantly lower the pod production of milkweed. By depleting the flowers of nectar they probably reduce the number of visits from pollinating insects.

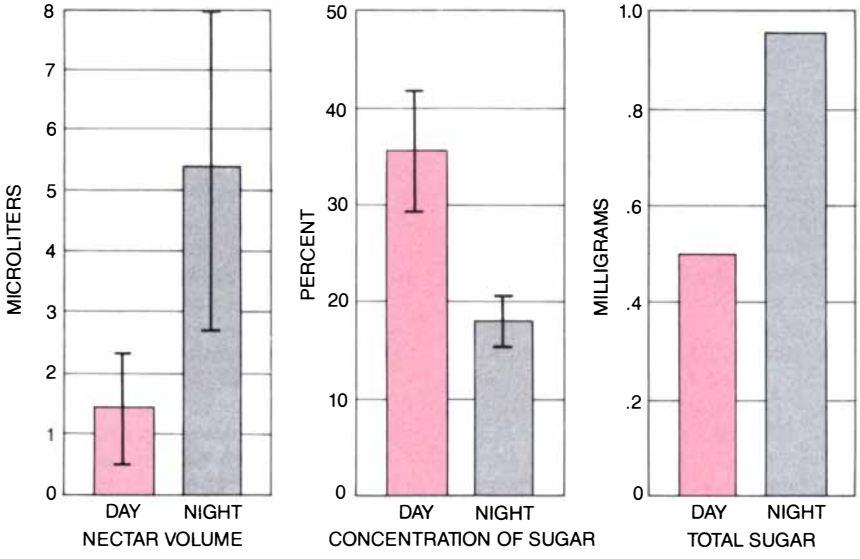
Other animals that frequent milkweed are attracted not by the plants themselves but by the nectar-seeking insects that are their prey. The commonest predators, such as the crab spider *Misumena vatia,* hunt from ambush among the flowers. Many crab spiders on milkweed grow rapidly by exploiting the heavy insect traffic. The most successful adult females increase their mass tenfold in as little as two weeks, attaining a weight of up to 400 milligrams (the size of a queen bumblebee). Almost two-thirds of this prodigious weight gain goes into producing a single clutch of eggs. The spider turns under the end of a milkweed leaf and deposits her eggs, protected by an egg sac, on the inner surface of the leaf. She then tightly secures the sides of the leaf with silk and in most cases guards the clutch until the eggs hatch.

The spider's success is by no means assured, however. Prey are abundant only while the plants are in bloom—generally less than three weeks—and so the spider must choose a good site and exploit it quickly or the opportunity will be lost. As might be expected, spiders hunting at bigger flower clusters grow larger and lay more eggs; about 20 percent of the crab spiders on milkweed apparently do not attain the minimum size needed to lay eggs at all.

Since the spider's predation is at the expense of the milkweed's pollinators, it could in principle diminish the plant's reproductive success. The spiders consume no more than 1 or 2 percent of the visiting insects, however, and so they are not likely to have a major effect on the welfare of either the prey species or the milkweed. I have been unable to demonstrate that bees respond to the presence of the spiders: bees show no tendency to avoid plants on which spiders are hunting.

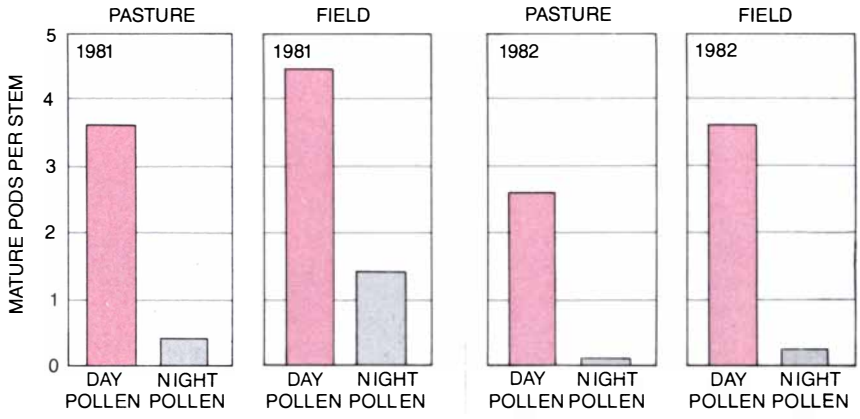
The crab spider itself is vulnerable to parasites and predators, which add yet another dimension to the fauna of the milkweed community. Flies and small wasps of the family Ichneumonidae parasitize between 10 and 15 percent of the crab spider egg sacs in my study area. Predation on adult spiders, however, seems to be rare, even though spider wasps (the Pompilidae) occasionally visit milkweed flowers to feed on nectar. The pompilids capture spiders as food for their young; the wasp paralyzes the spider with a sting, puts it in a nest and lays eggs there, so that the young can feed on the spider when the eggs hatch. In eight years of observation on the Maine coast I have seen a crab spider taken by a pompilid only once, although in other areas crab spiders are among the chief prey of both the pompilids and the mud-dauber wasps (the Sphecidae).

The crab spider digests its prey ex-



SECRETION OF NECTAR by common milkweed is greater at night, suggesting that nocturnal insects may have a role in pollination. The volume of nectar is almost five times greater at night, and although the concentration of sugar in the nectar is higher during the day, twice as much sugar is available to nocturnal foragers. The measurements were made by the author and Robert S. Fritz of Vassar College on a clone of milkweed growing in a pasture.

ternally: it injects its saliva, which contains enzymes, and then sucks out the liquefied tissues of the insect. The empty exoskeleton is discarded, but it seldom goes to waste. If it falls to the ground, ants or other scavengers, such as the harvestman, or daddy longlegs, (*Phalangium opilio*) quickly claim it. At night the harvestmen also gather carcasses lodged in the leaf axils of the milkweed stems, and I have even seen them trying to wrest prey from the spiders. In addition the harvestmen feed regularly on nectar in the flowers and capture small prey there, including young crab spiders. They thus figure at several levels in the food web.



DAYTIME AND NIGHTTIME POLLINATION were measured by covering plants with netting to exclude flying insects either during the day or at night. Although nocturnal insects visited the plants and consumed most of the nectar secreted at night, they were inefficient pollinators, responsible for only 5 to 25 percent of mature pods produced at the site.

With the departure of the nectar-seekers at the end of flowering, activity at milkweed plants declines. The chief insect occupants of the milkweed for the rest of the summer are monarch larvae, aphids in intermittent infestations, milkweed beetles (*Tetraopes tetraophthalmus*) and milkweed bugs (*Oncopeltus fasciatus*). The pods grow quickly, reaching their maximum length of from eight to 12 centimeters within a month. They are subject to depredation, however. The monarch larvae occasionally eat entire pods, but milkweed bugs, which suck on the young seeds, cause the greatest destruction in many areas.

Pods that survive open in October or early November, each releasing between 100 and 200 seeds. In spite of their silk parachutes most of the seeds do not get far. Release experiments done with my colleague Johanna M. Schmitt and my field-biology classes at Brown University have established that only 1 or 2 percent of the seeds travel far enough on the wind to escape the two-hectare, or about five-acre, field in which the plants grow.

Seeds from some clones, and even seeds from various pods of the same clone, travel farther than others. The differences in dispersal distance reflect variations in both the size of the coma (the parachute) and the weight of the seed. A large coma with a small seed, naturally, travels farthest; on the other hand, small seeds are less likely to germinate, and when they do, the seedlings grow more slowly and have a lower survival rate. Hence there is a compromise between seed size and dispersal distance. The observed variation may well have a genetic basis, in which case natural selection should tend to perpetuate those combinations of seed and coma size that are best suited to the distribution of colonization sites in a given area.

Milkweeds depend for their success on the rapid colonization of open ground. Sites opened to invasion as a result of natural events, such as tree falls or floods, generally last for only a year or two, and many manmade disturbances, such as tractor ruts in fields and pastures, are also transitory. Once a clone becomes established it initiates vigorous growth, although the plants may not flower for three years or more. When the flowering does come, the display is usually large enough to attract many insects, improving the chances of pollination. In subsequent years the number of flowering stems may increase rapidly.

In nature there is generally a limit to the proliferation of a milkweed clone. Open land in wooded areas seldom remains unshaded for long, and so milkweed must give way to the normal succession of more enduring plants. Under these conditions a clone is not likely to become very large. In particular, few clones are large enough to claim all the attention of the nearby large insects to the exclusion of other clones. Human activity, however, can delay succession indefinitely. Today extremely large clones grow in pastures, along power-line rights of way and in similar altered environments. The conditions the plants find—and to some extent create—in such areas must be rather different from those under which the milkweeds evolved.

What generalizations can be made about the relations of milkweed with its numerous visitors? Clearly the plant directly affects the lives of the pollinators, and indirectly it has an influence on the predators and parasites by gathering together prey and host species. Of course the pollinators in turn are essential to the reproductive success of the plant, and a few herbivore species can also have a major impact on the milkweed. In contrast, the predators of the pollinators and the parasites of the predators seem to have little control over the numbers of their prey and hosts, and so they would not seem to have much bearing on the welfare of the milkweed.

Such a conclusion may be premature. The importance of indirect relations among the members of a community can be obscured by more conspicuous interactions between plants and pollinators, plants and herbivores and plants and the physical environment. For example, Svata Louda of the University of Nebraska has shown that goldenweed flower clusters occupied by green lynx spiders produce more viable seeds than unoccupied clusters. Flowers occupied by the spiders have a relatively low rate of pollination and fertilization, in part no doubt because the spiders prey on pollinating insects. The spiders are even more effective, however, in controlling certain seed-eating flies, so that their overall effect on seed production is positive.

Interactions among organisms at three or more levels in a food web are now getting increased attention. Studies of them may reveal subtle relations that are important to the success of the organisms and also to an understanding of the overall structure of the community. Simple communities such as milkweed and its associated insects offer an excellent opportunity to explore multilevel relations of this kind. The milkweed community has three traits that Peter W. Price of Northern Arizona University has identified as essential for such studies: the membership of the community is readily defined, the resources of the community can be measured and the system is amenable to experimental manipulation. In addition, with the milkweed community observations are of the most direct kind: the biologist can see the organisms in action.

Almost priceless.

Hewlett-Packard presents the first 32-bit UNIX™ system u der $5,000.

It's the one computer for both the technical and administrative sides of your job.

Introducing the Integral Personal Computer.

Don't let the small size fool you. This is definitely a full-fledged member of the HP 9000 family of engineering workstations.

With a 16/32-bit MC68000. A graphics co-processor. 800KB of standard memory expandable to 5.8MB.

And a multitasking operating system based on UNIX System III. It's a UNIX system that's easy to use. Its powerful but friendly user interface means even novices can now tap the power of this ideal operating system for technical tasks. There are no cryptic commands to learn. And there's an optional mouse.

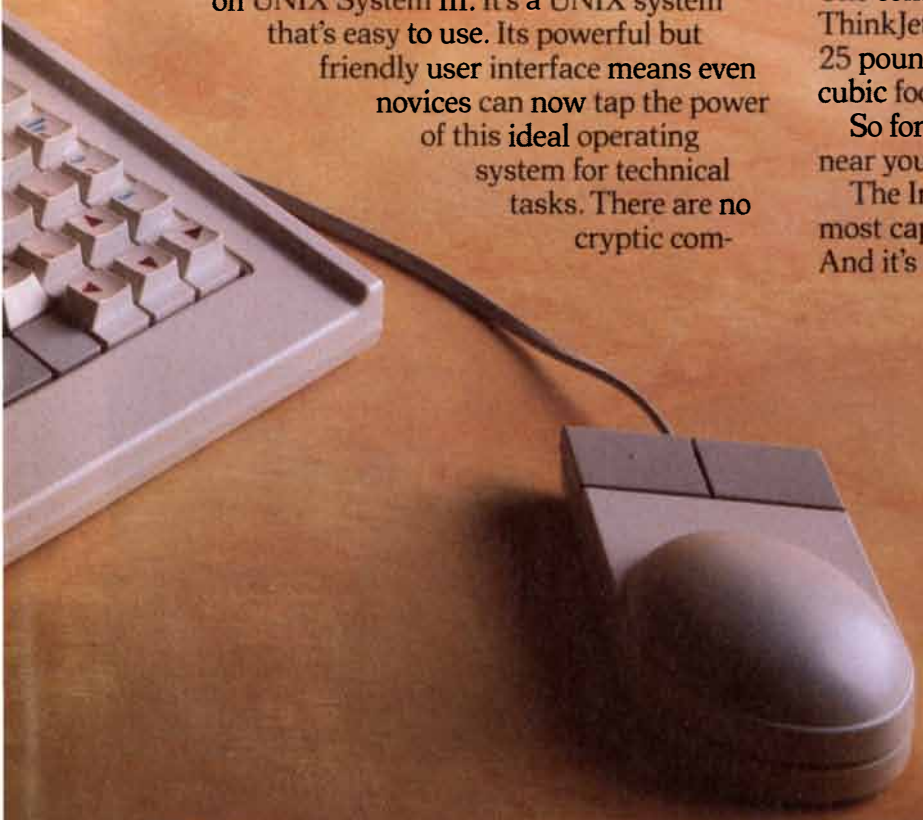
It's a personal computer without the limitations. Like other PCs, it runs a variety of popular spreadsheet, graphics, database and word processing software. But unlike other PCs, it has the multitasking operating system, powerful microprocessor and programming flexibility for your specialized applications. Just what you'd expect from Hewlett-Packard.

Developing custom solutions on the Integral PC is easy, with optional programming tools such as HP Technical BASIC, HP-UX "C" and Real-Time Extensions.

And the Integral PC goes where you need it. The complete system—including our built-in ThinkJet graphics printer—packs into a single 25 pound package that takes up less than a cubic foot.

So for the name of the dealer or HP sales office near you, call 1-800-FOR-HPPC.

The Integral Personal Computer. Perhaps the most capable personal computer you can buy. And it's only $4,995*



HEWLETT PACKARD

PG02505

UNIX is a trademark of AT&T Bell Laboratories.

optional. *$4,995 is suggested U.S. list price; prices elsewhere may vary.

THE AMATEUR SCIENTIST

Fly casting illuminates the physics of fishing

by Jearl Walker

When a fisherman casts a fly, how should he move the rod and throw the line? The fly is strongly affected by air drag; how then can it continue to move forward after the rod stops? Does the bending of the rod propel the fly as a bow propels an arrow? Why does one change rods and lines according to the kind of fish one is seeking? How does the angler resist the pull of a hooked fish and reel it in?

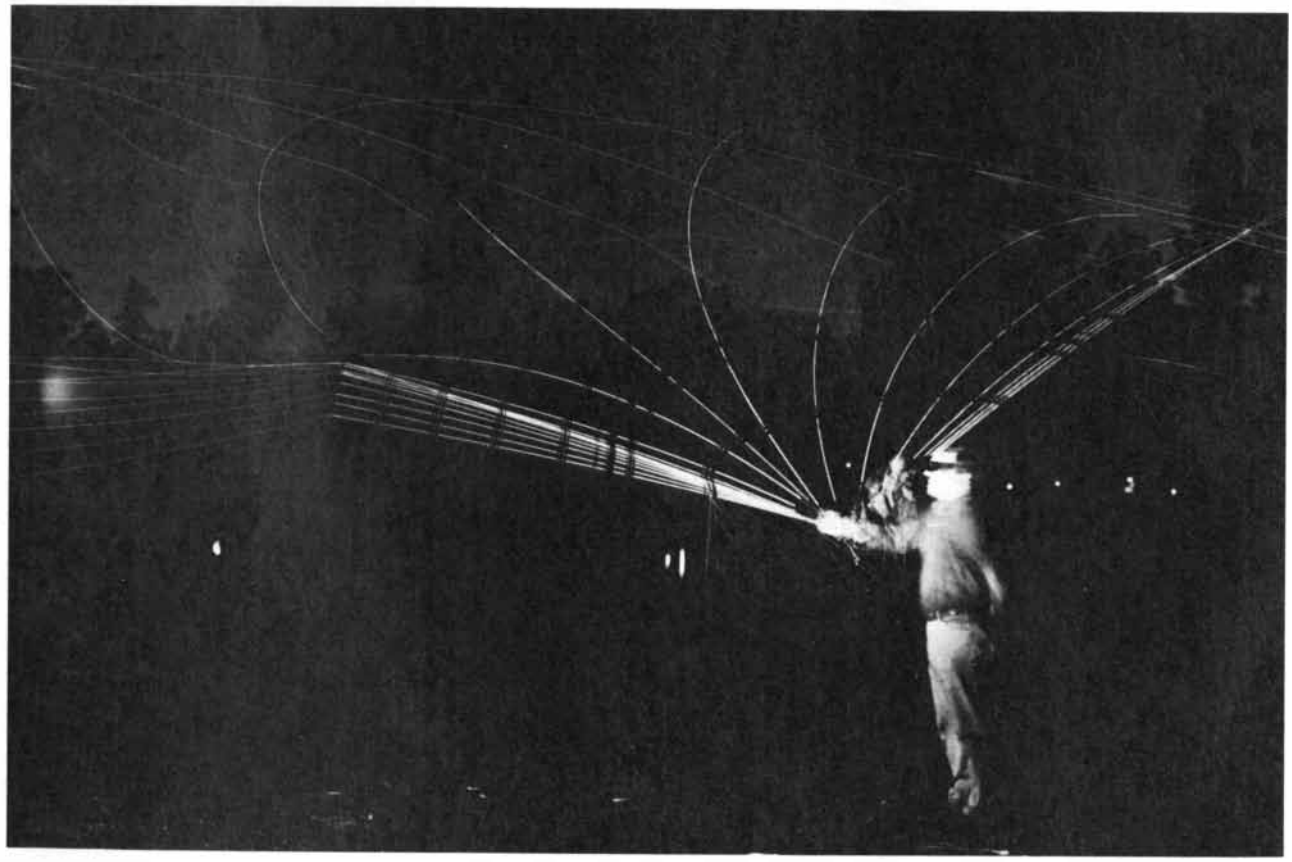
I have looked into these questions and others on the physics of fishing, drawing on manuscripts from several researchers. Edward Mosser of Burbank and William W. Buchman of Los Angeles published a paper on the dynamics of fly casting. Mosser has also written several unpublished papers on the subject. Graig A. Spolek of Portland State University and Steve C. Fry of Hawthorne, Calif., have independently sent me manuscripts in which a fly cast is studied mathematically.

Spolek analyzed a forward, overhead cast of a fly. Initially the length of line extending from the tip of the rod is short. To lengthen the line and to prepare for the final cast the angler makes a series of false casts forward and back in which the fly does not touch the water. In each false cast the angler moves the tip of the rod overhead and somewhat to his rear and then rapidly forward. The line follows. Although the rod may not pass appreciably beyond the vertical in the back cast, the line is stretched out horizontally (because of its rearward motion) just as the forward cast begins.

In each false forward cast the angler prevents the fly from touching the water by pulling backward on the rod. During the false back cast a right-handed fisherman uses his left hand to remove additional line from the reel and to hold it in a loop. During the false forward cast he releases the extra line as the forward momentum of the moving line pulls it through the line guides on the rod. In this way the total length of moving line increases.

Once enough line has been released the angler again brings the rod tip and the line rearward and then makes a final forward cast, this time stopping the rod in a forward direction and letting the fly land on the water. Spolek analyzed the flight of the line and fly during this final forward cast. The photograph below shows a view in stroboscopic light of a forward cast made by Mosser.

The ability of the angler to propel a



Edward Mosser makes a cast in stroboscopic light

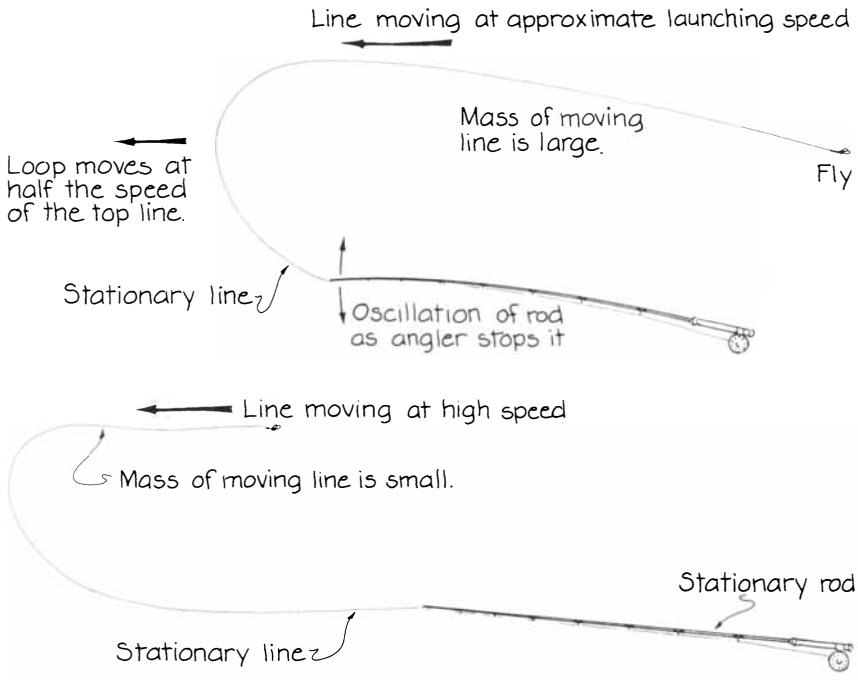
fly for any significant distance is puzzling because of the air drag on the fly. Imagine trying to hurl a fly without a line. To equal the distance commonly achieved with a cast you would have to make a heroic throw. Spolek calculates that to throw a typical fly 65 feet from an initial height of five feet the launching speed must be more than 300 miles per hour. Since the fly is attached to a line that also encounters air drag, the theoretical value for the initial speed is seemingly higher. Even if one argues that the long rod amplifies the throw, the air drag on the line and fly should still eliminate the possibility of a long cast.

This puzzle has been solved by many people who have studied the physics of casting. A forward cast involves a subtle mechanism that accelerates the line and fly after the rod has launched them. All skilled anglers know about the mechanism, but Spolek has shown by mathematical analysis why it is so successful.

To explain the mechanism and Spolek's analysis I shall follow the steps of the forward cast. The analysis begins at the end of the back cast, a point at which the line extends straight toward the rear. As the angler rotates the handle and moves it forward the flexible rod bends toward the rear, delaying the forward motion of the rod tip, line and fly. Once they begin to move the line and fly have the same speed as the rod tip until near the end of the rod's forward motion. As the angler slows the rod to stop it in its most forward position the line travels at the speed it had when the speed of the rod tip was highest. This is the launching speed.

As the rod slows, the line gains on it and sails over the tip, forming a loop that is convex in the forward direction. Between the loop and the tip the line is stationary because the tip is stationary. Between the loop and the fly the line continues to travel forward. Initially most of it is above the loop and moving, but because the loop moves at half the speed of the top part of the line, that part gradually passes through the loop and becomes stationary. When the end of the line reaches the loop, the fly flips over and the line straightens. From that point the fly and line fall to the water.

The loop is the propelling mechanism in a long cast. It concentrates kinetic energy in the continuously decreasing portion of line above the loop. The transfer of energy can be analyzed by starting with the line just before the loop forms. It has a kinetic energy equal to the product of half of the line's mass and the square of the line's speed (the launching speed). This energy has been supplied by the angler.



How a loop in the line generates acceleration

To simplify the analysis ignore for a moment the air drag on the line and fly. Once the line is launched its kinetic energy cannot change because no work is done on it. The kinetic energy remains constant. Since the amount of moving mass decreases, the speed must increase in order to conserve the kinetic energy. Therefore the part of the line above the loop and the fly attached to it accelerate forward. The acceleration ceases only when the end of the line passes through the loop. Fly casting is actually an exercise in the conservation of kinetic energy that yields an acceleration as the amount of moving mass decreases.

Spolek considered how air drag alters this simple model. Drag diminishes the acceleration because the moving line and the fly lose energy as they do work on the passing air, but the loss is not sufficient to prevent a long cast.

Much of the air drag on the line derives from the broad cross section the loop presents to the air. Skilled anglers know that a long, accurate cast requires forming a loop with a small diameter (about three feet). Larger loops mean greater losses of energy and shorter, less accurate casts. Can the angler reduce the diameter of the loop to less than three feet? Probably not, because the loop is enlarged by the overshoot of the rod when it oscillates around its final resting position.

To investigate the mechanics of a cast Spolek calculated the launching speed required when the line is 65 feet long and the final speed of the fly is 98 feet per second before it flips over. He postulated a uniformly dense line that had a mass typical of fishing line. He also assumed the line develops a three-foot loop. He found that if air drag is neglected, the line and fly must have a launching speed of approximately 20 feet per second.

Spolek also made a calculation for a rig called a double-taper line. It is narrow and uniform in diameter for the first two feet, increases steadily in diameter for the next 10 feet and then becomes uniform again. It has the same mass as the uniform line for the first 30 feet. Spolek found the launching speed for the double-taper line and fly to be less than 20 feet per second. Because the diameter of the double-taper line decreases toward the end, the amount of moving mass declines faster than it would decline if the line were uniform. If the final speed is 98 feet per second for both lines, the greater acceleration afforded by the double-taper line requires less launching speed.

When Spolek included air drag in these calculations, he discovered some interesting changes in the acceleration. If the fly is attached to a uniform line, it must be launched at more than 98 feet per second to attain a speed of 98 feet per second. More interesting, he found that the speed gradually decreases at first and then, near the end of the path, rapidly increases to its final value. Spolek explains these changes in terms of the competition between the concentration of kinetic energy in the decreasing amount of moving mass and the removal of energy by air drag. Just after the loop forms, the rate of

energy being lost to air drag is greater than the rate at which energy is being concentrated. The line and fly slow down. Near the end of the cast the rate of concentration begins to dominate, accelerating the line and fly.

A double-taper line encountering air drag requires a higher launching speed than a uniform line. As in the case of a line of uniform diameter, the speed of the line and fly initially decreases, but near the end of the cast it suddenly increases and then decreases. Some anglers describe the acceleration as a "kick" because the fly pulls on the line.

For the sake of calculation Spolek invented two experimental lines tapered along their full length. Each has the same mass as the uniform line for the first 30 feet. One of them, called a long-taper line, has the same rate of change in diameter as the tapered part of the double-taper line. The other one, called an experimental-taper line, has a more gradual taper. According to Spolek's calculations, the speed of a fly on the end of a long-taper line increases at first and then decreases near the end of the trip. If an experimental line is used, the speed of the fly increases throughout the trip.

The fly requires the lowest launching speed with the long-taper line and only slightly more with the experimental line. Each is lower than the final speed of 98 feet per second. With the uniform and double-taper lines the launching speeds exceed the final speed. The differences may not matter because all the launching speeds are within the angler's capability.

A more important input might be the initial energy the fisherman must give the line. If he makes many casts during an outing, he wants to minimize the energy required for each one. Spolek calculates that the uniform line requires the least initial kinetic energy followed (in order) by the experimental-taper, the double-taper and the long-taper lines. The long-taper line requires almost twice as much launching energy as the uniform line.

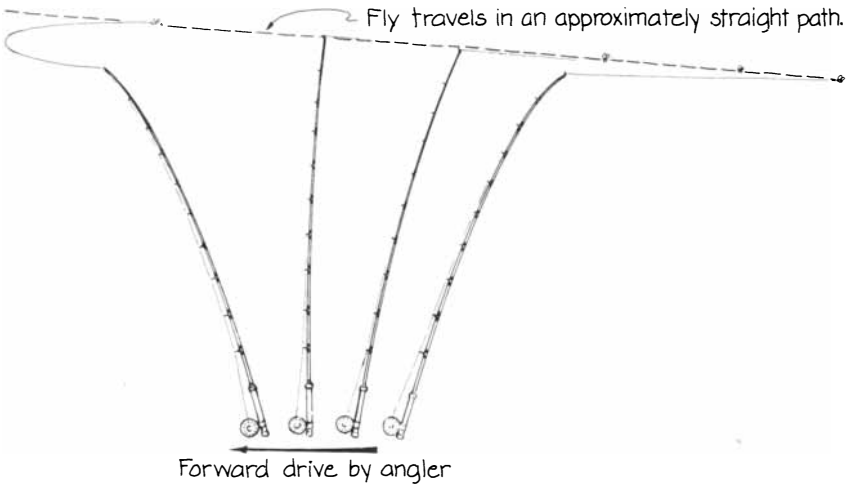
Fry's analysis dealt with the dynamics of the line and rod during the forward drive. In his model an external force (from the angler) compresses a spring (the rod) attached to a mass (the mass of the rod). The spring drives forward a second mass (the line).

The dynamics of the model can be ascertained only approximately. Much of the complication derives from air drag, which affects the rod as it moves forward and also as it oscillates at the beginning and end of that movement.

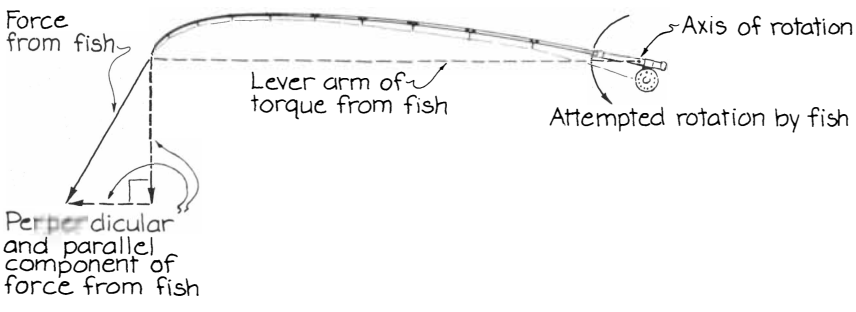
The force on the line rises rapidly and then falls during the forward drive of the rod. Initially the drag on the system is small because all components are moving slowly. The drag increases as their speed rises. Fry found that midway through the forward drive much of the energy supplied by the angler goes into accelerating the line whereas somewhat less goes into accelerating the rod. Much less energy is lost to the air drag on the rod and an even smaller amount of energy is lost to the air drag on the line.

Mosser and Buchman studied how an angler should execute a forward cast to give the line and fly a large amount of energy. If the back cast has not directed the line straight back, part of the rod's forward motion is wasted in straightening it, leaving less time for the angler to do work to give it kinetic energy. Another common error is to rotate the rod without also driving it forward. Then the line and fly move in an arc instead of a straight line. Again less work is done on them and they derive less kinetic energy.

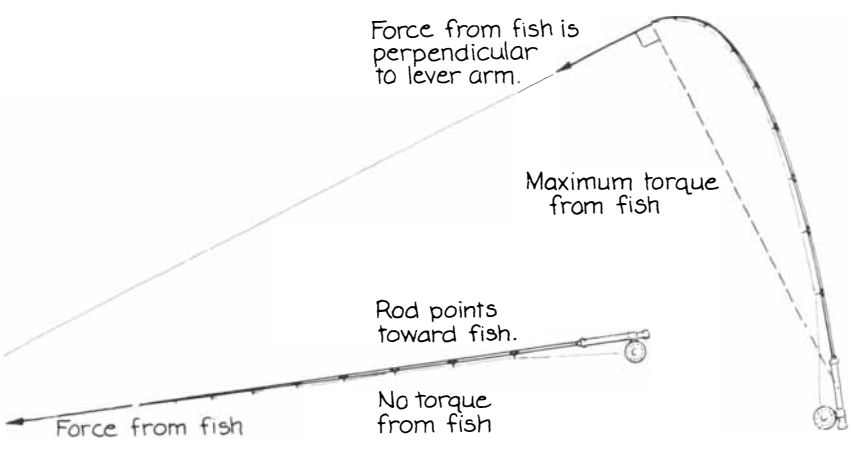
Some anglers believe the rearward



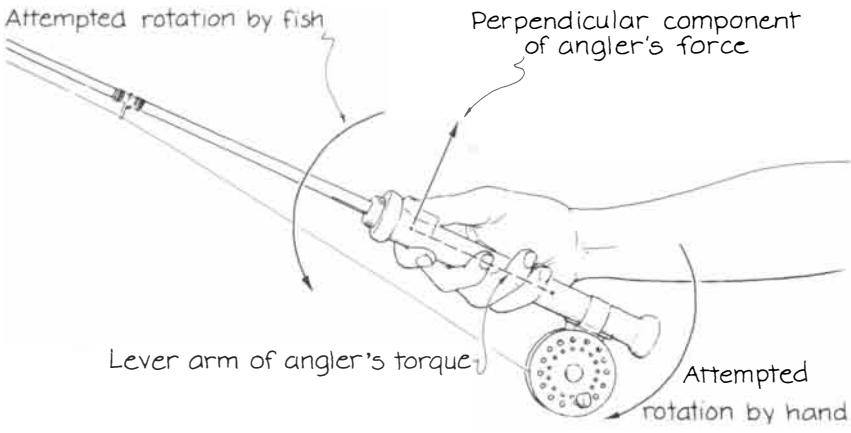
The movement of the fly in a forward cast



Torque generated by a hooked fish



How the position of the rod affects the torque



Torque generated by the angler

bending of the rod during the final back cast stores energy in it, enabling it to function like a bow. According to studies done by Mosser, Buchman and Fry, not much of this stored energy ends up in the kinetic energy of the line and fly.

The bending of the rod does serve a subtle purpose in a forward cast. It allows the line and fly to follow a path that is almost horizontal and therefore increases the efficiency of the angler's effort. If the fly moves in an arc, part of the angler's effort goes into pulling the fly upward and then downward along the arc, a trajectory that deprives it of forward speed.

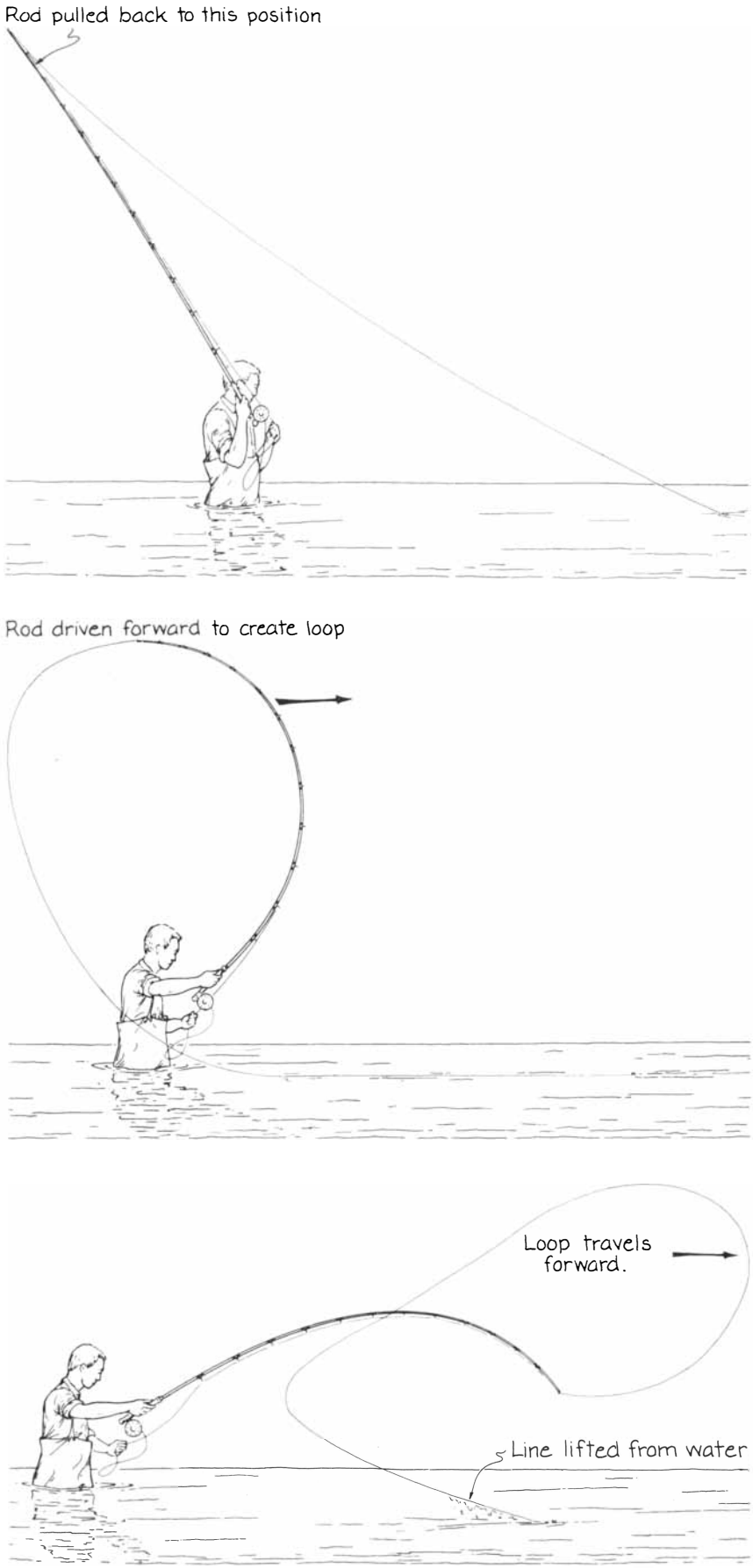
The proper size of rod is partially determined by the size of the fish being sought and also by what is comfortable. The advantage of a long rod is that a given forward motion of the angler's hand moves the rod tip through a greater distance than the same motion would move a shorter rod. The angler may then have more time in which to work on the line and fly, giving them more kinetic energy. The disadvantage of a long rod is that it is harder to move and rotate. If the rod is too long, the angler is forced to move and rotate it too slowly to make a cast. An angler usually winds up choosing a rod that achieves a suitable compromise between the extremes.

When a fish is on the line, what determines the force on the hook? Does it depend on the length or bending of the rod? Mosser argues that the bending does not determine the force on the hook and that a long rod may be a disadvantage in a fight with a strong fish. If the pull by the fish exceeds the total resistance of the line, additional line unwraps from the reel. At such a time the fish is said to be "running." When a fish is running, the force on the hook exceeds the resistance of the line, but otherwise the force is never greater than the resistance no matter how much the rod is bent.

With a fish on the line the angler feels a direct pull and a torque that attempts to rotate the rod about an axis near his hand. The lever arm for the torque is a line that runs from the axis of rotation on the grip to the rod tip, where the force from the fish is applied. The force from the fish can be split into two components, one component parallel to the lever arm and the other perpendicular to it. The perpendicular component produces the torque, which is the product of that component and the lever arm.

Two factors determine the size of the components: the force from the fish and the orientation of its pull in relation to the lever arm. If the angler points the rod toward the fish, the torque vanishes because no perpendicular component of force is exerted on the rod. In this case the entire force exerted by the fish is delivered parallel to the lever arm, and the angler pulls in the opposite direction to keep the rod stationary. If the angler rotates the rod vertically in such a way as to make the force from the fish entirely perpendicular to the lever arm, the torque is at its maximum.

To hold the rod stationary the angler must resist both the pull and the torque created by the fish. To counter the torque he must generate a torque that works to rotate the rod in the opposite direction. The torque is produced by a force from the angler's hand. The lever arm associated with the force is short because the hand is near the point about which the torques attempt to rotate the rod. Because of the short lever arm, the angler must provide a large force if his torque is to match that of the fish. Here the bending of the rod serves a purpose because it reduces the length of the lever arm affected by the fish's torque and therefore also re-



Making a roll cast

duces the torque the angler is required to supply.

If the angler must fight a large fish pulling strongly, a long rod means the torque from the fish is large. To reduce the torque the angler should use a shorter rod so that the lever arm of the torque exerted by the fish is reduced.

To reel in a fish an angler usually rotates the rod vertically so that the fish is drawn toward him. Then he winds line onto the reel as he lowers the rod tip. The procedure is repeated, with the fish being pulled closer when the tip is raised and remaining in place when the tip is lowered. In this way the angler provides the work needed to move the fish through the water.

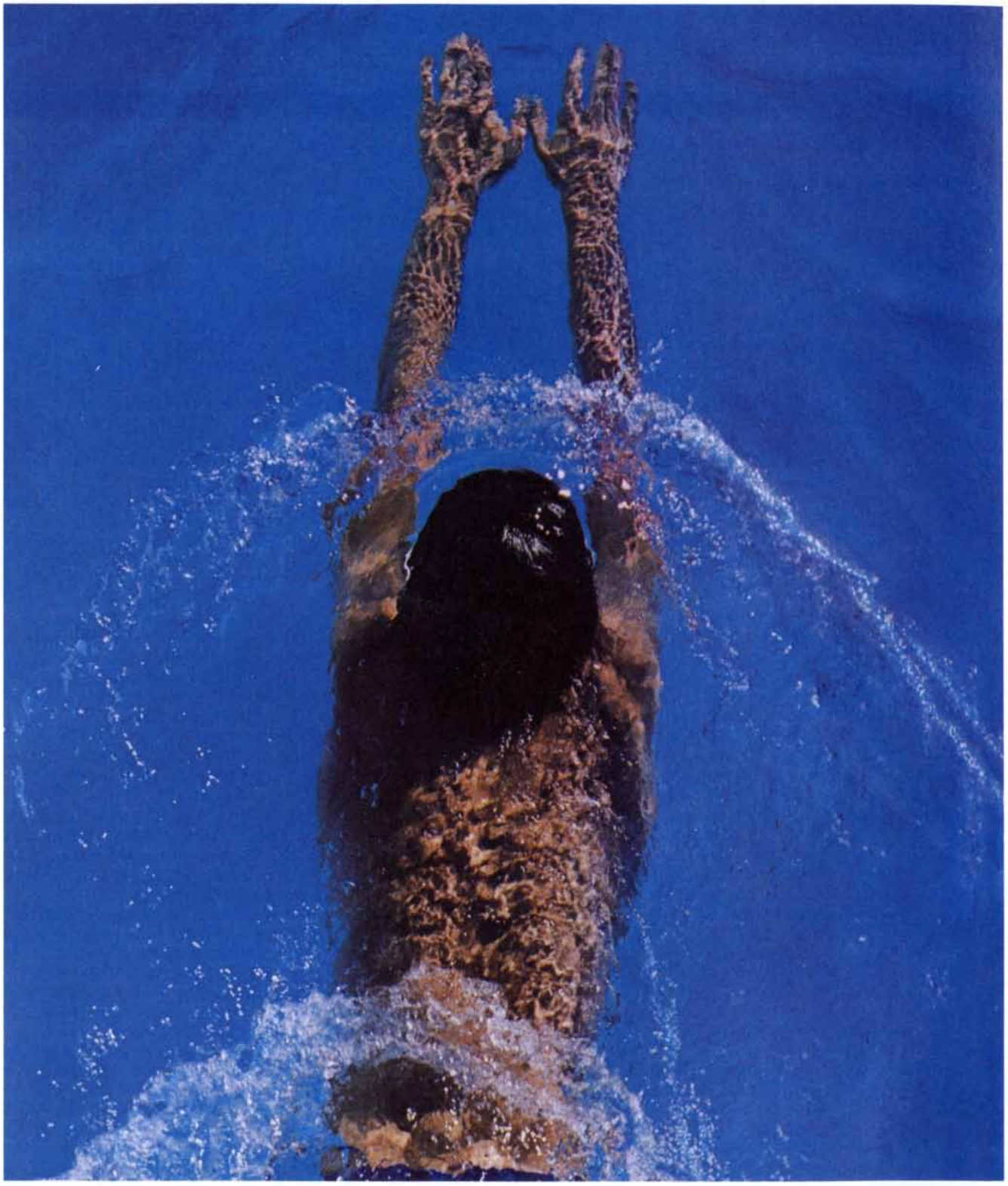
Many questions remain to be answered about the physics of fishing. You might want to investigate other types of fishing lines that are available commercially or to invent new ones, as Spolek did. You might also follow the analysis by Mosser and Buchman on how the flexibility of the rod, which is called the rod action, changes the cast. For example, if the rod is highly flexible and vibrates at the end of its forward motion, is the transfer of energy to the line altered?

You could also study several other types of cast. Mosser and Buchman have analyzed a lure cast in which the forward momentum of a somewhat heavy lure pulls line from the reel. Does the bending of the rod during the back cast supply kinetic energy for the launching of the lure? What is the proper method of moving the hand to make this cast?

I am also intrigued by the roll cast, which no one seems to have analyzed mathematically. This cast is employed when obstacles to the rear prevent the angler from making the back cast I have described. After he plays out about 15 feet of line on the water he brings the rod tip from a forward, eye-level position to a high, slightly rearward position. As the line is partially pulled from the water, it begins to pass by the side of the angler, bowed rearward. The angler then quickly moves the rod tip forward. This motion creates a complete loop in the line that travels toward the fly. After the loop reaches the fly the line is straight on the water in front of the angler.

How is the loop maintained as it travels along the line? What becomes of the energy the angler gives the line? How does the speed of the line at the top of the loop change as the loop moves? When the loop reaches the fly, is more line pulled from the rod? What role does the surface tension of the water play in the cast? Mosser, Buchman, Fry, Spolek and I should like to hear what you find out.

TAKING THE FIRST STROKES TOWARD 1988.



Preparation has, in fact, already begun. In pools across the United States, promising young swimmers and divers are now working toward the special goal of representing their country in 1988. That's why Phillips Petroleum has been a national sponsor of United States Swimming since 1973, and of United States Diving since 1979. We will continue our support for a long time to come. Not only to help our best athletes, but to introduce many thousands of young people to the skills, joys and disciplines of competitive swimming and diving. Because we think that, no matter what they ultimately achieve in the pool, the things they learn can make them all lifelong winners.

PHILLIPS 66

BIBLIOGRAPHY

Readers interested in further explanation of the subjects covered by the articles in this issue may find the following lists of publications helpful.

INFANT MORTALITY IN THE U.S.

INFANT MORTALITY AND THE HEALTH OF SOCIETIES. L. Newland. Worldwatch Paper No. 47, Worldwatch Institute, 1981.

PREVENTING LOW BIRTH WEIGHT. Institute of Medicine, National Academy of Sciences. National Academy Press, 1985.

THE WORLD ECONOMIC CRISIS AND THE CHILDREN: UNITED STATES CASE STUDY. C. Arden Miller, Elizabeth J. Coulter, Lisbeth B. Schorr, Amy Fine and Sharon Adams-Taylor in *International Journal of Health Services,* Vol. 15, No. 1, pages 95–134; 1985.

THE MOONS OF URANUS, NEPTUNE AND PLUTO

OUT OF THE DARKNESS: THE PLANET PLUTO. Clyde W. Tombaugh and Patrick Moore. Stackpole Books, Harrisburg, Pa., 1980.

THE OUTER SOLAR SYSTEM. David Morrison and Dale P. Cruikshank in *The New Solar System,* edited by J. Kelly Beatty, Brian O'Leary and Andrew Chaiken. Cambridge University Press and Sky Publishing Corporation, 1983.

PHYSICAL PROPERTIES OF THE URANIAN SATELLITES. R. H. Brown in *Uranus and Neptune,* NASA Conference Publication No. 2330, 1984.

OUTER PLANET SATELLITES. D. P. Cruikshank and R. H. Brown in *Natural Satellites,* edited by J. A. Burns and D. Morrison. University of Arizona Press, in press.

THE FUNDAMENTAL PHYSICAL LIMITS OF COMPUTATION

IRREVERSIBILITY AND HEAT GENERATION IN THE COMPUTING PROCESS. R. Landauer in *IBM Journal of Research and Development,* Vol. 5, No. 3, pages 183–191; July, 1961.

CONSERVATIVE LOGIC. Edward Fredkin and Tommaso Toffoli in *International Journal of Theoretical Physics,* Vol. 21, Nos. 3/4, pages 219–253; April, 1982.

THE THERMODYNAMICS OF COMPUTATION—A REVIEW. Charles H. Bennett in *International Journal of Theoretical Physics,* Vol. 21, No. 12, pages 905–940; December, 1982.

FUNDAMENTAL PHYSICAL LIMITATIONS OF THE COMPUTATIONAL PROCESS. Rolf Landauer in *Annals of the New York Academy of Sciences,* Vol. 426, pages 161–170; 1985.

THE EVOLUTION OF DARWINISM

IS A NEW EVOLUTIONARY SYNTHESIS NECESSARY? G. Ledyard Stebbins and Francisco J. Ayala in *Science,* Vol. 213, No. 4511, pages 967–971; August 28, 1981.

PERSPECTIVES IN EVOLUTION. Edited by Roger Milkman. Sinauer Associates, Inc., 1982.

EVOLUTION OF GENES AND PROTEINS. Edited by Masatoshi Nei and Richard K. Koehn. Sinauer Associates, Inc., 1983.

EVOLUTION: AN INTEGRATING PRINCIPLE FOR THE MODERN SYNTHESIS. G. L. Stebbins in *Experientia,* Vol. 39, No. 8, pages 823–834; August 15, 1983.

THE CHEMOSENSORY RECOGNITION OF GENETIC INDIVIDUALITY

CONTROL OF MATING PREFERENCES IN MICE BY GENES IN THE MAJOR HISTOCOMPATIBILITY COMPLEX. K. Yamazaki, E. A. Boyse, V. Mike, H. T. Thaler, B. J. Mathieson, J. Abbott, J. Boyse, Z. A. Zayas and L. Thomas in *The Journal of Experimental Medicine,* Vol. 144, No. 5, pages 1324–1335; November 1, 1976.

DISTINCTIVE URINARY ODORS GOVERNED BY THE MAJOR HISTOCOMPATIBILITY LOCUS OF THE MOUSE. M. Yamaguchi, K. Yamazaki, G. K. Beauchamp, J. Bard, L. Thomas and E. A. Boyse in *Proceedings of the National Academy of Sciences of the United States of America,* Vol. 78, No. 9, pages 5817–5820; September, 1981.

THE SENSORY PERCEPTION OF GENOTYPIC POLYMORPHISM OF THE MAJOR HISTOCOMPATIBILITY COMPLEX AND OTHER GENES: SOME PHYSIOLOGICAL AND PHYLOGENETIC IMPLICATIONS. Edward A. Boyse, Gary K. Beauchamp and Kunio Yamazaki in *Human Immunology,* Vol. 6, pages 177–183; 1983.

SENSORY DISTINCTION BETWEEN $H\text{-}2^{b}$ AND $H\text{-}2^{bm1}$ MUTANT MICE. K. Yamazaki, G. K. Beauchamp, I. K. Egorov, J. Bard, L. Thomas and E. A. Boyse in *Proceedings of the National Academy of Sciences of the United States of America,* Vol. 80, No. 18, pages 5685–5688; September, 1983.

MINOAN PALACES

THE PALACES OF CRETE. James W. Graham. Princeton University Press, 1969.

ZAKROS: THE DISCOVERY OF A LOST PALACE OF ANCIENT CRETE. Nicholas Platon. Charles Scribner's Sons, 1971.

THE AEGEAN CIVILIZATIONS. Peter Warren. Elsevier-Phaidon, 1975.

PALACES OF MINOAN CRETE. Gerald Cadogan. Methuen, Inc., 1980.

FOSSIL WATER UNDER THE SINAI-NEGEV PENINSULA

ON THE ANCIENT WATER OF THE UPPER NUBIAN SANDSTONE AQUIFER IN CENTRAL SINAI AND SOUTHERN ISRAEL. A. Issar, A. Bein and A. Michaeli in *Journal of Hydrology,* Vol. 17, No. 4, pages 353–374; December, 1972.

DESERT ISOTOPE HYDROLOGY: WATER RESOURCES OF THE SINAI DESERT. Joel R. Gat and Arie Issar in *Geochimica et Cosmochimica Acta,* Vol. 38, pages 1117–1131; July, 1974.

CHEMICAL COMPOSITION OF RAINWATER AND FLOODWATERS IN THE NEGEV DESERT, ISRAEL. R. Nativ, A. Issar and J. Rutledge in *Journal of Hydrology,* Vol. 62, Nos. 1/4, pages 201–223; April, 1983.

MILKWEEDS AND THEIR VISITORS

THE NORTH AMERICAN SPECIES OF *ASCLEPIAS.* Robert E. Woodson, Jr., in *Annals of the Missouri Botanical Garden,* Vol. 41, No. 1, pages 1–211; February, 1954.

CONTRIBUTIONS OF DIURNAL AND NOCTURNAL INSECTS TO THE POLLINATION OF COMMON MILKWEED (*ASCLEPIAS SYRIACA* L.) IN A POLLEN-LIMITED SYSTEM. Douglass H. Morse and Robert S. Fritz in *Oecologia,* Vol. 60, No. 2, pages 190–197; 1983.

THE PARTITIONING OF POLLINATORS AMONG THREE SPECIES OF *ASCLEPIAS.* Susan R. Kephart in *Ecology,* Vol. 64, No. 1, pages 120–133; February, 1983.

THE AMATEUR SCIENTIST

ON THE DYNAMICS OF A BULL WHIP. B. Bernstein, D. A. Hall and H. M. Trent in *The Journal of the Acoustical Society of America,* Vol. 30, No. 12, pages 1112–1115; December, 1958.

THE DYNAMICS OF A FLYCAST. Ed Mosser and William W. Buchman in *The Flyfisher,* Vol. 13, No. 4, pages 5–9; 1980.


Inside Deep Rover, undersea vehicle.

Aquanaut Sylvia Earle probes earth's other space.

It's Maui. A woman hangs suspended in the Pacific depths. Eye to eye with the great humpback whales of Hawaii.

She holds her space. Intent. Unflappable. Dr. Sylvia Earle never turns from a challenge.

Marine biologist. Oceanographer. Botanist. Dr. Earle led America's first team of women aquanauts. Leadership has been her lot ever since.

She has directed or participated in more than forty expeditions. Few have exceeded her record of five thousand hours under waters as diverse as the Pacific Ocean and the Red Sea.

But not all her fame has been achieved submerged. She has lectured in thirty countries. Written more than fifty articles. Authored the National Geographic book, *Exploring the Deep Frontier.*

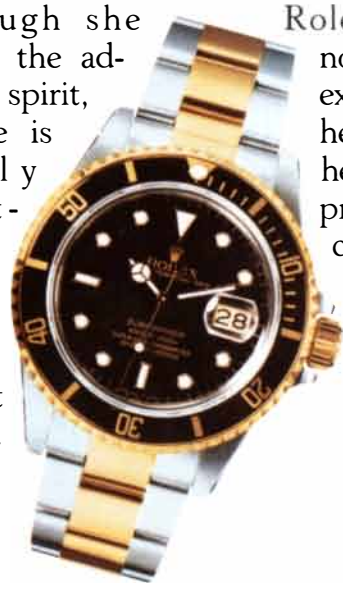

Descent off Hawaii in the JIM diving suit.

Though she embodies the adventurist spirit, Dr. Earle is equally committed to the preservation of the seas. And ardent in her defense of all life within them.

Today, she searches out methods of ocean conservation. Rallies support for the endangered whale. Catalogs unknown underwater life. Creates devices to explore virgin sea worlds.


Confrontation with a Humphead Wrasse in the Red Sea.

That she relies on a Rolex is not unexpected. Dr. Earle asks no less of her timepiece than she does of herself. Performance must be predictable in unpredictable circumstances. And high purpose must always be artfully executed. The accomplished make no compromise with excellence.



ROLEX

Submariner Oyster Perpetual Date in stainless steel and 18kt gold.
Write for brochure. Rolex Watch, U.S.A., Inc., Dept.798, Rolex Building, 665 Fifth Avenue, New York, New York 10022-5383.
World headquarters in Geneva. Other offices in Canada and major countries around the world.

A POWER PLANT THAT RECHARGES HUMAN BATTERIES.

Drivers, as a general rule, start cars.

What you see here, however, is an automobile capable of returning the favor. An automobile that transforms gasoline into that very exhilarating human fuel called adrenaline.

The BMW 535i.

Beneath the hood is BMW's new 3.5-liter in-line six. A racecourse-developed power plant that converts a seemingly modest 0.2 liter increase in engine capacity into an anything-but-shy boost in torque and acceleration. Such that the 535i now moves from 0-60 in a mere 7.4 seconds on its way to an ultimate speed in excess of 130 mph.

Of course, mated to this exceptional performance is an equally exceptional means of management. Namely, the same fully independent suspension that earned the 535i's predecessor much acclaim as "a car that loves bumpy roads, patchy pavements and unexpected alterations of course" (Car & Driver).

And, as you might expect in a car that's "on the cutting edge as far as European performance sedans go" (Motor Trend), the BMW 535i also comes to a halt with unquestionable authority, using an electronic sensor-based antilock braking system.

Still, despite its breathtaking performance, riding in the 535i could hardly be described as a calisthenic exercise. Cradled as you are in infinitely adjustable seats. Surrounded by virtually every power-assist and luxury option conceivable. Amenities such as a two-position sun-roof. Electronic climate control. And an eight-function onboard computer.

In sum, the BMW 535i is an automobile perfectly designed for those drivers who can think of few things more enjoyable than seeing their pulse rate clearly indicated on the speedometer.

An experience best explored at your nearest BMW dealership. Where you're invited to plug into a 535i for a thorough, and thoroughly invigorating test drive.

THE ULTIMATE DRIVING MACHINE.