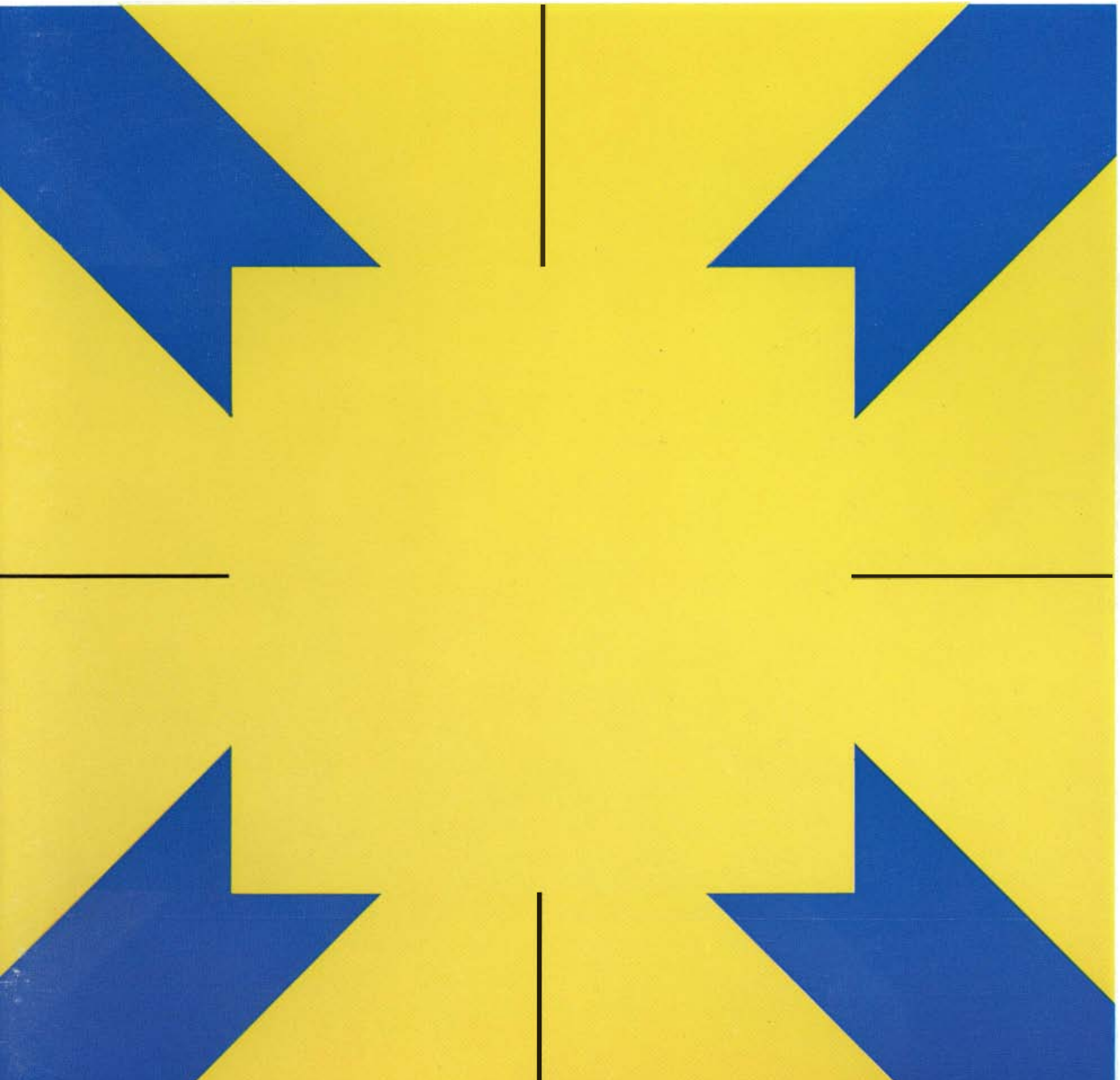
SCIENTIFIC AMERICAN



SUBJECTIVE CONTOURS

$1.25

April 1976

The Saab Philosophy, No 3 in a Series:

"WE BELIEVE THAT SAFETY MEANS AVOIDING ACCIDENTS. NOT JUST SURVIVING THEM."

You never know what's around the next corner until you get there. That's why we designed the Saab to give you every chance to avoid whatever surprise you might find.

We do it with what we call *"active safety devices"*: front-wheel drive for stability, rack-and-pinion steering for precise control, four-wheel power-assisted disc brakes for quick, fade-free stops, and a powerful, fuel-injected 2-liter engine for fast acceleration. Together, they give every Saab exceptional performance.

Of course, if you can't avoid an accident, we've also built every Saab with the *"passive safety devices"* that have made Saab a leader in automotive safety for years. There's roll-cage construction, six solid steel posts to support the roof, door impact panels, and a body designed to absorb impact to protect the driver and passengers.

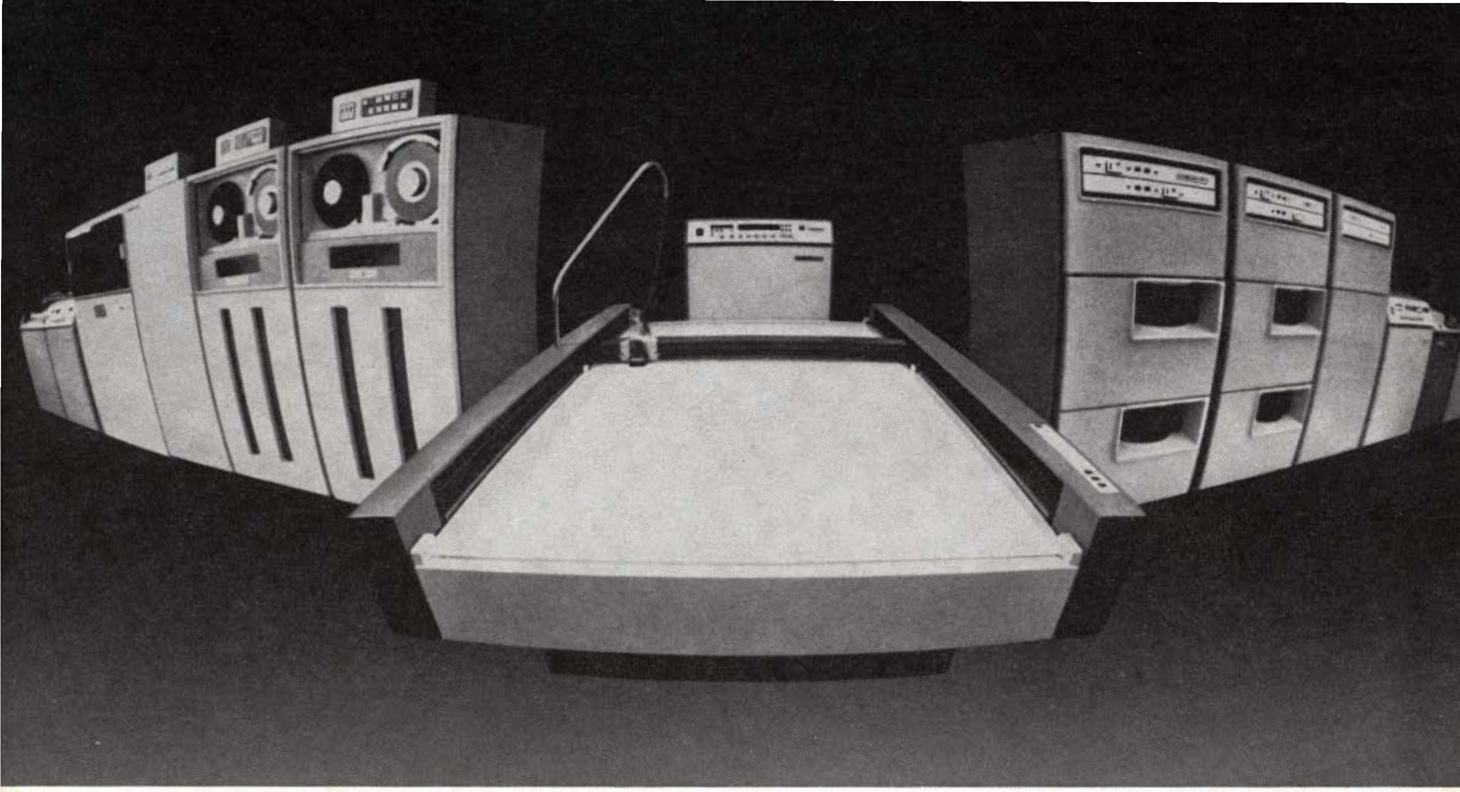
Every part of the interior of a Saab has also been designed with the safety of the driver and passengers in mind. For example, Saab was first with seats with integral, see-through head restraints. First with a heated driver seat. And first with the ignition key protectively hidden between the front seats.

But you can't really appreciate the confidence and security you get in a Saab, until you drive one.

So take a test drive. It's one thing you'll be glad you didn't miss.

SAAB

IT'S WHAT A CAR SHOULD BE.



Peripheral vision.

CalComp looks at the future as if it were wrapped around the computer.

The computer remembers an incredible amount of data, but no matter how large its brain becomes it will always need help. It needs more "brain cells." Our memory systems provide that.

In the mind of the computer there are pictures. Man can see these pictures because we taught the computer to draw. Today, CalComp is a world leader in the manufacturing of computer graphic equipment.

The information that spews out of the computer fills rooms with paper that is hard to retrieve. Our Computer Output Microfilm systems can reduce a roomful to a drawerful, and puts the information at your fingertips.

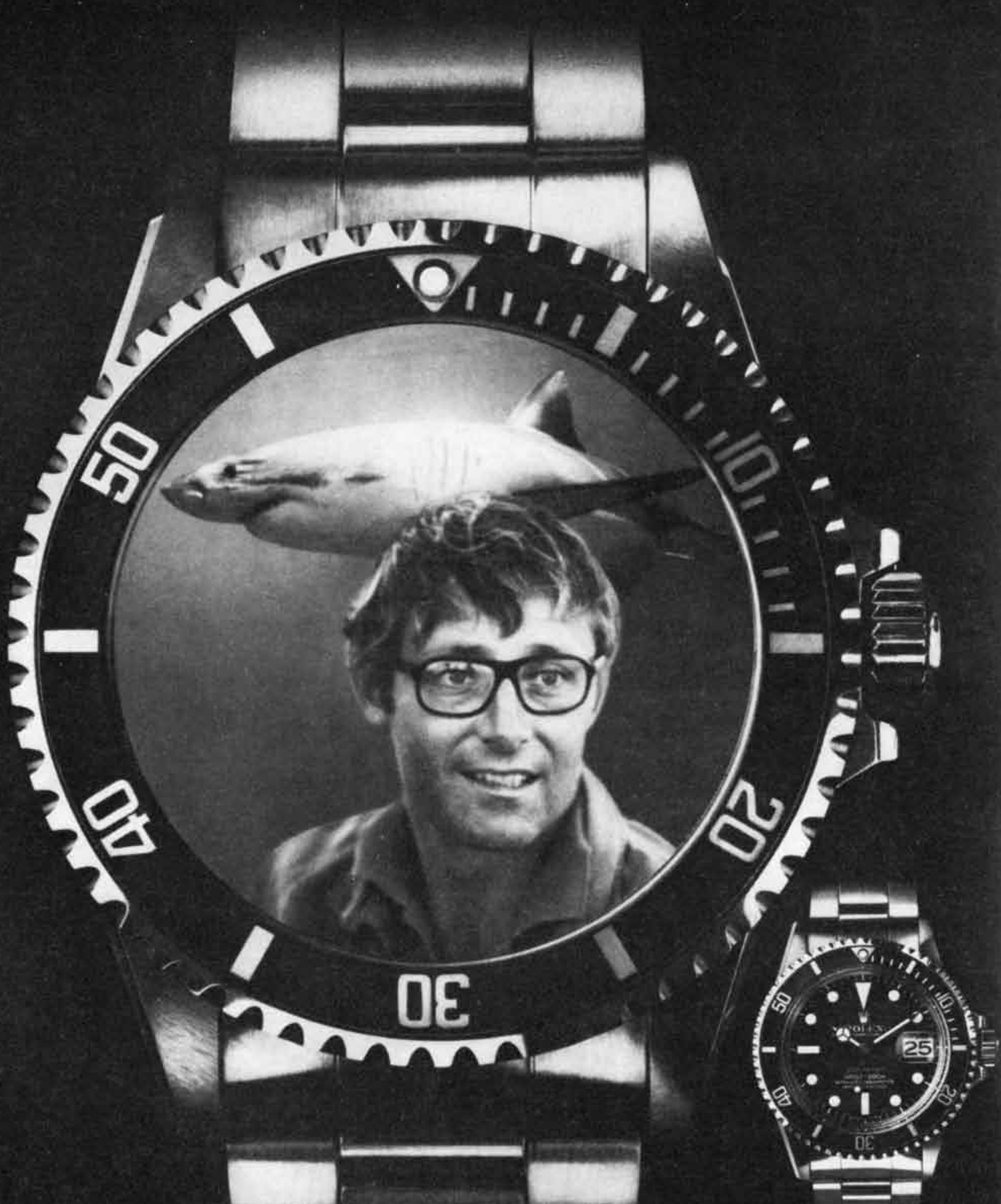
The information that is stored on computer tape is only as useful as the computer's access to that information. Our Automated Tape Library is the first system to give the computer automatic access to thousands of reels of tape—in seconds.

Our sales for fiscal 1975 were in excess of $120,000,000. And our backlog of orders at the beginning of fiscal 1976 was the largest in company history.

We've spent 17 years perfecting our peripheral vision.

Contact us for our view of the future. Write or call California Computer Products, Inc.,SA-M4-76,2411 West La Palma Avenue, Anaheim, California 92801. (714) 821-2011.

CALCOMP


50
10
20
30
40
25

Established 1845 **SCIENTIFIC AMERICAN** April 1976 Volume 234 Number 4

ARTICLES

DEPARTMENTS

SCIENCE/SCOPE

The first recipient of Sweden's new L M Ericsson International Prize for "an especially important scientific or technological contribution to telecommunications engineering" will be Harold Rosen of Hughes. King Carl XVI Gustaf will award the prize of 100,000 kronor to Dr. Rosen on May 5 "for his foresightedness -- against an originally contrary current opinion -- in proposing the introduction of geostationary communications satellites, and for his eminent scientific and technological contributions to their development, design, and operation."

The first scanning optical microscope that will inspect large-scale integrated circuits while they are being operated has been developed by Hughes. It has demonstrated its ability to detect, localize, and identify flaws in complex devices. Unlike other scanning microscopes, which scan only one logic state at a time, the Hughes microscope will effectively superimpose many logic states at one time to "characterize" or inspect the microcircuit. The completely non-destructive instrument scans with a modulated laser and was specifically designed under sponsorship of NASA's Marshall Space Flight Center to meet the high throughput requirements of manufacturers of high reliability microcircuits.

A Roland missile intercepted a jet drone at White Sands Missile Range, N.M., recently, as the U.S. Army began testing two West German-built Roland all-weather short-range air defense systems. Roland, first major foreign-designed weapon system selected for deployment with U.S. forces, protects battlefield troops and equipment and high-value rear-area emplacements against high-speed, low-level air attack. Hughes is prime contractor to the Army Missile Command for the U.S. Roland program.

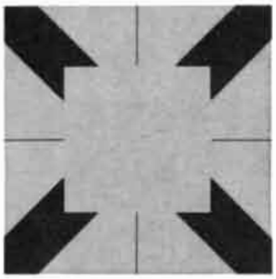
Hughes needs systems-level engineers. Sonar Systems: design and develop surface and subsurface ASW tactical and surveillance sonar systems for Underseas Systems Laboratory....Communications Systems: define architecture and signal processing concepts for next-generation TDMA communications systems. Requirements: BS or higher degree, U.S. citizenship. Please send resume to: Engineering Employment, Hughes Aircraft Company, P.O. Box 3310, Fullerton, CA 92634. An equal opportunity employer.

New products from Hughes include a CMOS/LSI digital frequency synthesizer that provides up to 1,021 output frequencies from a single crystal; it is designed to increase the performance and reduce the cost of industrial, military, and citizens' band communications equipment; circuit contains an adder for transmit/receive frequency offset adjustment....an X-band traveling-wave tube for military satellite communications earth terminals; using modulation anode control techniques, it operates efficiently at 600 and 1200 watts -- highest power to date for a PPM-focused, X-band CW, air-cooled TWT -- and is available as a component or in a power supply package.

Creating a new world with electronics

HUGHES

HUGHES AIRCRAFT COMPANY



THE COVER

The illustration on the cover is an example of a visual phenomenon in which contours and shapes are perceived where none actually exist. The central square appears as an opaque surface covering the two crosses. The borders of the square are clearly visible from a distance, but when they are examined closely, they disappear. In addition the area inside the square appears to be brighter than the background, but in reality the color is exactly the same in both regions. The conditions that give rise to this visual phenomenon are described by Gaetano Kanizsa in his article "Subjective Contours" (*page 48*).

THE ILLUSTRATIONS

Cover illustration by Jerome Kuhl

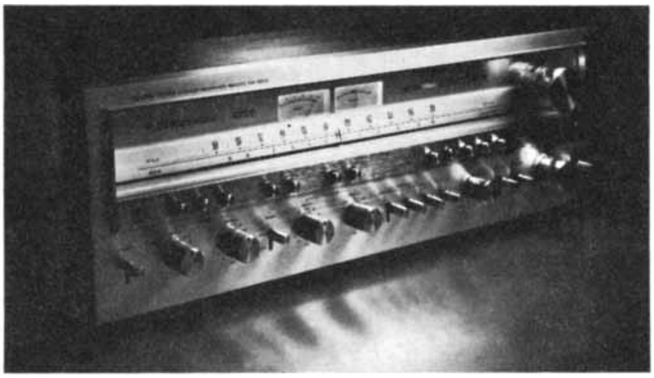
Page	Source
34–35	National Science Foundation
36–37	Total Communications Industries, Ltd.
38	California Board of Education
41	Julius Adler, University of Wisconsin
42	Ilil Arbel (*top and middle*); Julius Adler, University of Wisconsin (*bottom*)
43	Julius Adler, University of Wisconsin (*top*); Ilil Arbel (*bottom*)
44–47	Ilil Arbel
48–52	Jerome Kuhl
61	Air Weather, U.S. Air Force
66	Alan D. Iselin and George V. Kelvin
67–68	Alan D. Iselin
69–77	Alan D. Iselin and George V. Kelvin
78	Alan D. Iselin (*top*) Alan D. Iselin and George V. Kelvin (*bottom*)
79–82	Alan D. Iselin and George V. Kelvin
83	Alan D. Iselin
85	P. J. Darragh, A. J. Gaskin and J. V. Sanders
88	John Cubitto, Gemological Institute of America (*top*); P. J. Darragh, A. J. Gaskin and J. V. Sanders (*bottom*)
89	P. J. Darragh, A. J. Gaskin and J. V. Sanders
90–91	P. J. Darragh, A. J. Gaskin and J. V. Sanders (*top*), George V. Kelvin (*bottom*)
92	P. J. Darragh, A. J. Gaskin and J. V. Sanders
93	George V. Kelvin
94	P. J. Darragh, A. J. Gaskin and J. V. Sanders
95	George V. Kelvin
96–98	Allen Beechel
99	Daniel E. Appleman, Smithsonian Institution
100–103	Allen Beechel
105–106	Ben Rose
107	Ilil Arbel
108–110	Ben Rose
112–113	Ilil Arbel
115–116	Anthony C. Pooley
118	Tom Prentiss
119	Total Communications Industries, Ltd.
122–124	Tom Prentiss
126–130	Ilil Arbel
133	© 1975 Crane, Russak & Company, Inc.
134	*Structural Materials in Animals,* by C. H. Brown. © 1975 John Wiley & Sons, Inc., used by permission
137	*The Flying Circus of Physics,* by Jearl Walker. © 1975 John Wiley & Sons, Inc., used by permission

PIONEER HAS DEVELOPED A RECEIVER EVEN THE COMPETITION WILL ADMIT IS THE BEST.

One look at the new Pioneer SX-1250, and even the most partisan engineers at Marantz, Kenwood, Sansui or any other receiver company will have to face the facts.

There isn't another stereo receiver in the world today that comes close to it. And there isn't likely to be one for some time to come.

In effect, these makers of high-performance receivers have already conceded the superiority of the SX-1250. Just by publishing the specifications of their own top models.



As the chart shows, when our best is compared with their best there's no comparison.

To begin with, the SX-1250 is at least 28% more powerful than any other receiver ever made. Its power output is rated at 160 watts per channel minimum RMS at 8 ohms from 20 to 20,000 Hz, with no more than 0.1% total harmonic distortion.

And, for critical listening, no amount of power is too much. You need all you can buy.

To maintain this huge power output, the SX-1250 has a power supply section unlike any other receiver's, with a large toroidal-core transformer and four giant 22,000-microfarad electrolytic capacitors.

But power isn't the only area in which the SX-1250 excels. The preamplifier circuit has an unheard-of phono overload level of half a volt (500 mV). This means that no magnetic cartridge in the world can drive the preamp to the point where it sounds strained or hard. And the equalization for the RIAA recording curve is accurate within ±0.2 dB. A figure unsurpassed by the costliest separate preamplifiers.

Turn the tuning knob of the SX-1250, and you'll know at once that the AM/FM tuner section is also special. The tuning mechanism feels astonishingly smooth, precise and solid.

FM reception is loud and clear even on weak FM stations because the tuner combines extremely high sensitivity with highly effective rejection of spurious signals.

Of course, the Pioneer SX-1250 carries a price tag commensurate with its position at the top. But if you seek perfection you won't mind paying the price.

If, on the other hand, you'd mind, look into the new Pioneer SX-1050 or SX-950. They're rated at 120 and 85 watts, respectively, per channel (under the same conditions as the SX-1250) and their design is very similar. In the case of the SX-1050, virtually identical.

That means you don't just come to Pioneer for the world's best.

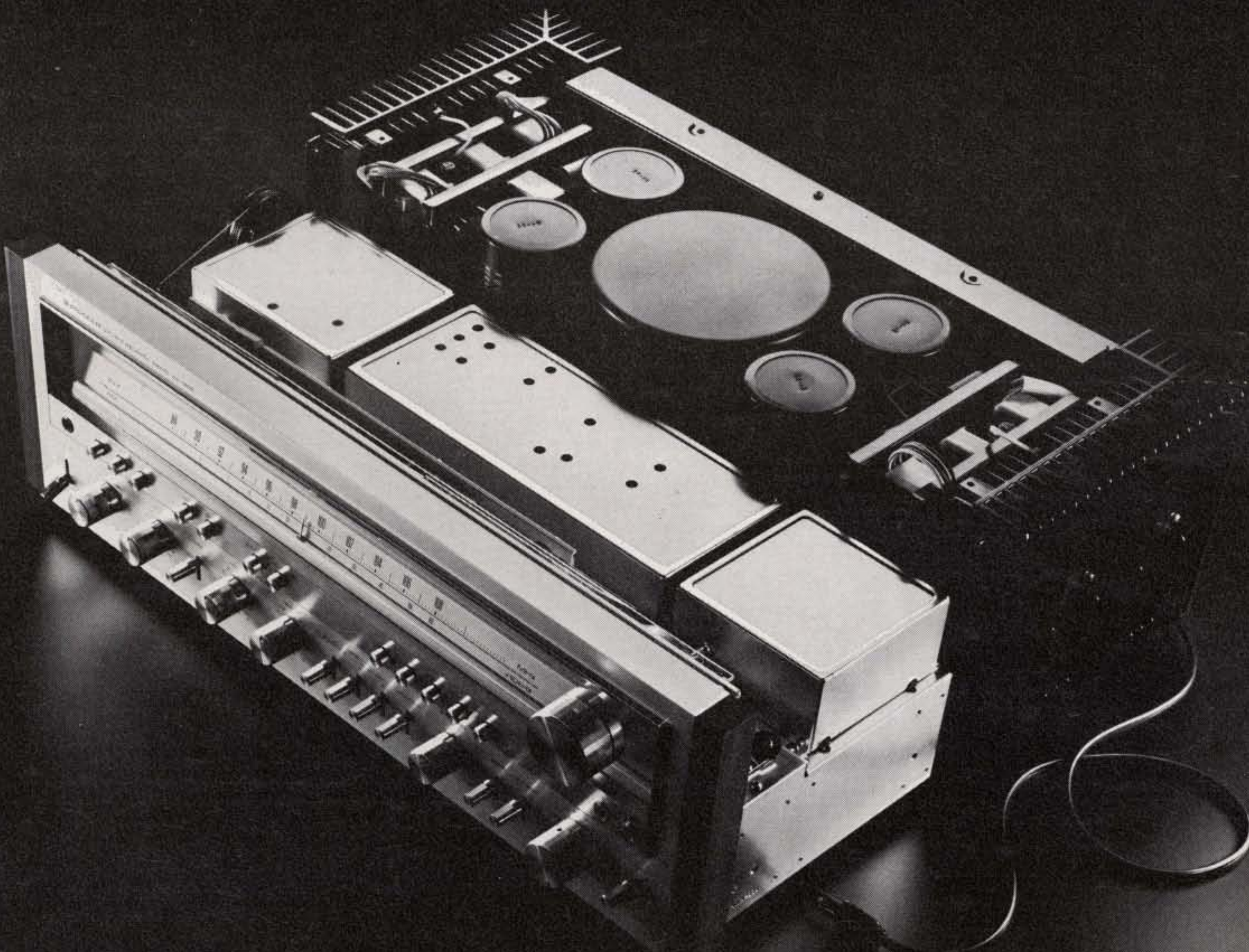
You also come to us for the next best.

For informational purposes only, the SX-1250 is priced under $900. The actual resale price will be set by the individual Pioneer dealer at his option.

PIONEER®
Anyone can hear the difference.

U.S. Pioneer Electronics Corp., 75 Oxford Drive, Moonachie, New Jersey 07074.

	PIONEER SX-1250	MARANTZ 2325	KENWOOD KR-9400	SANSUI 9090
POWER, MIN. RMS, 20 TO 20,000 HZ	160W+160W	125W+125W	120W+120W	110W+110W
TOTAL HARMONIC DISTORTION	0.1%	0.15%	0.1%	0.2%
PHONO OVERLOAD LEVEL	500 mV	100 mV	210 mV	200 mV
INPUT: PHONO/AUX/MIC	2/1/2	1/1/no	2/1/mixing	1/1/mixing
TAPE MON/DUPL.	2/yes	2/yes	2/yes	2/yes
TONE	Twin Tone: Bass-Bass-Treble-Treble	Bass-Mid-Treble	Bass-Mid-Treble	Bass-Mid-Treble
TONE DEFEAT	Yes	Yes	Yes	Yes
SPEAKERS	A,B,C	A,B	A,B,C	A,B,C
FM SENSITIVITY (IHF '58)	1.5μV	1.8μV	1.7μV	1.7μV
SELECTIVITY	90 dB	80 dB	80 dB	85 dB
CAPTURE RATIO	1.0 dB	1.25 dB	1.3 dB	1.5 dB



LETTERS

Sirs:

H. A. Bethe's article "The Necessity of Fission Power" [SCIENTIFIC AMERICAN, January] is an admirably clear and concise statement of the technical aspects of nuclear energy as it is related to the U.S. energy program. Unfortunately the article lacks perspective and may be misleading to the readers of *Scientific American.*

Perhaps the major omission in Professor Bethe's article is the failure to consider the *net* energy contribution of the proposed ambitious U.S. fission-energy program to the energy "account" of the economy. One approach to nuclear energy is to calculate the coal or oil equivalent of a pound of uranium 235. The ratio on a weight basis is more than a million to one and is the source of the popular view that fission energy is "free."

An alternative and more complete approach is to estimate the total energy consumed in the construction of the power plant and in the extraction, shipping and processing of the fuel. The net energy delivered to the society is obviously negative during the construction period and then becomes positive as the power plant functions as expected....

It is particularly important to apply such an "energy analysis" to the nuclear option, since nuclear energy is so often considered free. Consider a nuclear plant with an electric-generating capacity of 1,000 megawatts. Estimates made by Peter Chapman (*New Scientist,* December 19, 1974) and W. Kenneth Davis (address given to the Atomic Industrial Forum, March 3, 1975) agree that the initial fueling requires 1×10^9 kilowatt-hours of electricity per year and that refueling requires $.4 \times 10^9$ kilowatt-hours per year. To the energy associated with plant construction Chapman assigns 1.8×10^9 kilowatt-hours of electricity and Davis assigns $.7 \times 10^9$ kilowatt-hours. The disagreement in this last item is well within the uncertainty of such figures. If we assume a capacity factor for the plant of .62 and properly incorporate the energy costs of refueling and other energy losses, we discover, utilizing Chapman's figures, that the effective output of the 1,000-megawatt plant is 523 megawatts.

Suppose we express one year's output of the plant not in electrical energy but in the thermal-energy equivalents. (Using the standard rate of 10,000 B.t.u.'s per kilowatt-hour of electricity, this leads to a multiplication of all electrical-energy units by a factor of 2.8.) This view of the energy balance examines the fossil-fuel account of the economy and the nuclear impact on it....

In these terms we derive from Chapman's data that in 25 years a nuclear plant produces about 27 times as much "thermal equivalent" energy (kilowatt-hours of electricity $\times$ 2.8) as went into its construction and initial fueling. At that level the nuclear-reactor programs appear to be as advertised: a short- and medium-range solution to U.S. energy shortages and a solution to the hazardous dependence on foreign imports.

This picture breaks down drastically when one includes the 10-year delay between conception and birth for nuclear plants, and the required growth in plant starts in order to achieve the hoped-for additional capacity in the year 2000 of 430×10^3 megawatts. The present nuclear capacity is 31×10^3 megawatts, with 59×10^3 megawatts under construction. Estimates by the Energy Research and Development Administration (ERDA) and by others suggest 520×10^3 megawatts as the nuclear capacity in 2000.

A simple analysis, based on six new starts in the first year, 18 percent growth in starts up to 1990 (no starts after 1990) and the thermal-equivalent return ratio of 27, then shows that the net energy returned to the society would still be negative (-3.3×10^{15} B.t.u.'s) by 1990, and that by 1995 only 11.3×10^{15} B.t.u.'s (net thermal equivalent) would have been derived from such a massive nuclear program. Indeed, by the year 2000, 520 gigawatts of electric capacity would be at hand, but over the preceding 25 years the net thermal-equivalent benefit to the society would be only about 70×10^{15} B.t.u.'s....

There are other growth scenarios. Half of the additional capacity could be started now and half 10 years from now. With this growth pattern about 10×10^{15} B.t.u.'s (thermal equivalent) would be lost to the society in the first 10 years and a net of 90×10^{15} B.t.u.'s would be gained in the next 10 years (1995). The overall result of such a program would be about 7×10^{15} B.t.u.'s per year over a 25-year period....

These growth scenarios assume an increase in electrical-energy demand by about a factor of two in the year 2000, with the additional generating capacity being produced by nuclear fission. If we ignore the energy deficits and the long payback period, it is interesting to touch a little on the economics. If we ignore fission-fuel costs and assume a capital cost for nuclear plants of about \$1,000 per kilowatt of electricity, we can estimate the cost of nuclear energy on the basis of the above two scenarios over the next 25 years. An addition of 430×10^3 megawatts costs about \$430 billion. Under the first scenario (18 percent growth in starts, 71×10^{15} B.t.u.'s net in 2000) the energy cost is about \$6 per 10^6 B.t.u.'s (thermal equivalent). Under the second scenario (half of the capacity started now, half 10 years from now and 180×10^{10} B.t.u.'s net in 2000) the cost is \$2.40 per 10^6 B.t.u.'s. The comparison should be oil at about \$2 per 10^6 B.t.u.'s (\$12 per barrel) and coal at about \$1 per 10^6 B.t.u.'s (\$25 per ton).

There may be arguments for or against the nuclear option, and reasonable men may differ. However, it does seem to me that it cannot be argued that fission power can make a net major contribution to U.S. sources of energy before the turn of the century. There is only one short-term solution to the U.S. energy problem and that is conservation. Coal, oil and gas will continue to make the most positive net contributions.

Solar power, geothermal power, wind power, bioconversion power, fusion power and even fission power fit into relatively long-term solutions. For fission the technology has been largely developed by heavy Government investment. The point of this letter is to emphasize that it is not the energy panacea, and its future priorities have to be carefully evaluated. The other technologies have only recently been getting similar attention. I personally believe that these alternative techniques are not as inherently impractical as Professor Bethe suggests....

GEORGE D. CODY

Princeton, N.J.

Sirs:

There may indeed be a necessity for fission power, as Professor Bethe says, but that does not mean the fission power is necessarily going to be there. In fact, there is good reason to believe that light-water reactor technology is so inefficient and the fuel supply is so limited that any reactor completed after 1976 will not have a lifetime supply of fuel.

The most recent figures on the availability of uranium show that the reserves total approximately 600,000 tons. Probable resources (that is, undiscovered material thought by some to be available if enough exploration effort is expended) are estimat-

Scientific American, April, 1976; Vol. 234, No. 4. Published monthly by Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017; Gerard Piel, president; Dennis Flanagan, vice-president; Donald H. Miller, Jr., vice-president and secretary; George S. Conn, treasurer; Arlene Wright, assistant treasurer.

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

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

Offprint correspondence and orders should be addressed to W. H. Freeman and Company, 660 Market Street, San Francisco, Calif. 94104. For each offprint ordered please enclose 30 cents.

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

Name ____________________

New Address ____________________

____

____

Old Address ____________________

____

____

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

The most comprehensive encyclopedia of philosophy ever published...

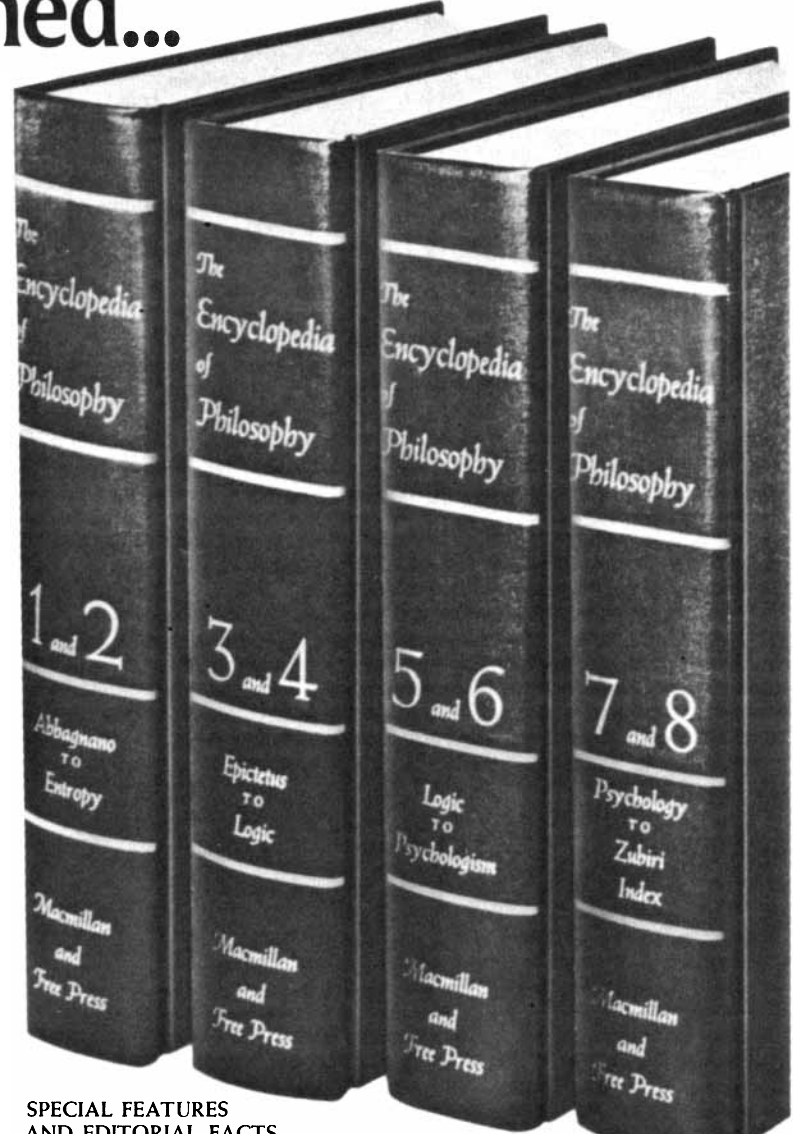
FOR ONLY

$17.50

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

PUBLISHER'S
LIST PRICE: $99.50

The complete contents of the original 8-volume $219.50 edition now in a handsome 4-volume reader's edition



ONE OF THE GREATEST repositories of philosophical information available anywhere, *The Encyclopedia of Philosophy* is the only major philosophical reference work published in English since 1901.

In this unique 5,000,000-word work can be found clear, authoritative answers to almost any question concerning philosophy — ancient, medieval and modern, Eastern as well as Western. And because the work represents the collaboration of more than 500 philosophers and philosophical scholars from twenty-four countries, every shade of opinion is represented. No thinker or movement is dismissed as too radical, no issue is scanted as too controversial. There are articles, for example, that discuss such contemporary matters as the ethics of suicide and euthanasia and the benefits and dangers of computing machines.

Clearly, *The Encyclopedia of Philosophy* belongs in every personal library, particularly in homes where there are high school or college students. Up to now its high price has kept it out of the reach of most families. But, as a new member of the Book-of-the-Month Club, you may own it for less than a tenth of the cost of the original eight-volume edition. Because only a limited number of sets have been set aside for this offer, you are advised to act quickly.

BOOK-OF-THE-MONTH CLUB, INC.
Camp Hill, Pennsylvania 17012

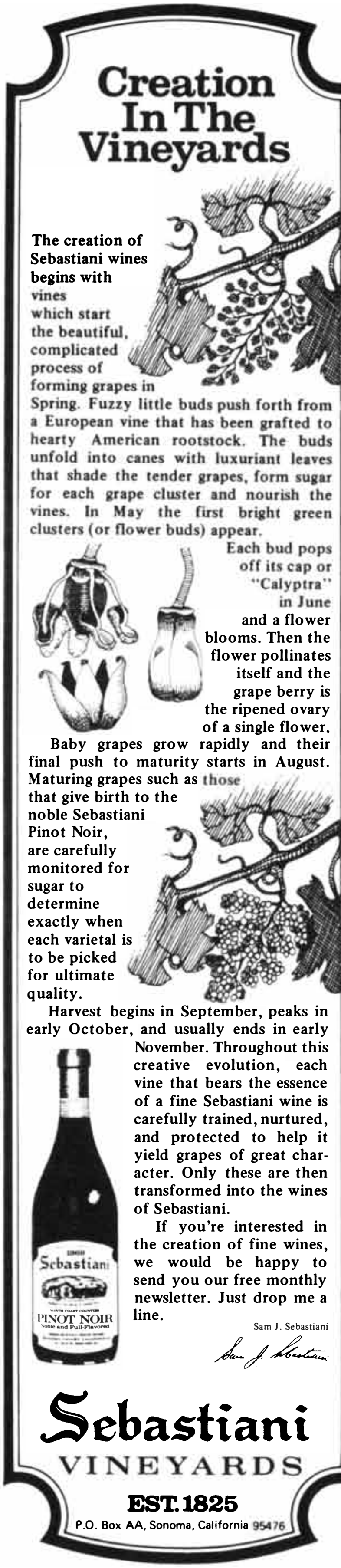
SPECIAL FEATURES
AND EDITORIAL FACTS

- 5,000,000 words
- 158-page index
- Paul Edwards, editor in chief
- Nearly 1500 articles, including 900 on individual philosophers
- 4200 pages
- Editorial Board of 153 international scholars
- Contributors: 500 specialists from twenty-four countries

A SMALL SAMPLING OF THE IDEAS
AND PHILOSOPHERS IN THESE VOLUMES

Atheism	Semantics	Good and Evil	Nonsense	Pain	Nothing
Free Will	The Absolute	Existentialism	Humanism	Truth	Linguistics

CONFUCIUS	SANTAYANA	DESCARTES	KANT	RUSSELL
ERASMUS	SPINOZA	HEGEL	NIETZSCHE	EINSTEIN
SARTRE	THOMAS AQUINAS	CAMUS	ARISTOTLE	NEWTON
LOCKE	WITTGENSTEIN	KIERKEGAARD	PLATO	PIAGET



ed, in an arcane way, to amount to possibly another 1.2 or 1.3 million tons. A very substantial exploration effort has been under way in the U.S. for the past four years to discover this postulated material. Additions to reserves have averaged about 12,000 tons per year, whereas the fuel used has equaled, or has slightly exceeded, that rate. The record shows that it is easy to total up postulated resources based on geological analogies but that it is difficult to convert these undiscovered resources to fuel you can count on. If you are optimistic and assume that in the next 30 years we shall add an average of 20,000 tons per year to the reserves, you will then have a total of 1.2 million tons as the total stock available.

Then the question is: How much uranium do you use during the 30-year (not 40-year) lifetime of a reactor? The published estimate is approximately 230 tons of U_3O_8 (yellowcake) per year, excluding initial fueling, for a 1,000-megawatt plant with a capacity factor of 75 percent. This means that you can generate 32×10^6 kilowatt-hours per short ton of U_3O_8. This figure implies that in its 30-year lifetime a plant will use about 6,900 tons of yellowcake (without uranium or plutonium recycling) and that if you go on the basis of existing reserves, you need only build 100 plants to preempt supplies, and that if exploration is successful, fewer than 200 nuclear plants will preempt the total supply.

However, it is not that simple. If you look into the operation of existing plants, you will find very little data that show how much uranium was burned and how much power was produced. Excluding such inefficient plants as Dresden 2 and Yankee Rowe, the average power production by light-water-reactor technology is about 18×10^6 kilowatt-hours per short ton of U_3O_8. This increases the fuel requirement to about 410 tons per year and makes the needed 30-year lifetime supply almost 12,500 tons. Hence 57 1,000-megawatt plants will preempt the supply, and the planned capacity of units now in operation or under construction and due to be in service by the end of 1976 may be slightly in excess of this figure. The additional fuel needed for the initial core in each reactor has not been included and increases the total required for each unit. But that is close enough.

After this year (1976) it is doubtful that a lifetime supply will be demonstrably available for any reactor just coming on line. And a finding rate of 20,000 tons per year implies that not more than two new 1,000-megawatt plants can come on line during any 18-month period and have a lifetime supply in view—and this means about twice the success in converting resources to reserves as has been the average experience over the past four years. If you agree with Professor Bethe that a reactor lifetime of 40 years is reasonable, you can reduce the preceding figures proportionately.

As for the breeder, the first commercial breeder, under the most optimistic estimates, such as one made by John Patterson of ERDA in May, 1975, will not be on line until 1993 at the earliest. And if it has the same doubling time as the much praised French Phénix (30 to 60 years), it will have little significance in the energy picture. And Professor Bethe says that without the breeder there is no justification for the mining of low-grade ores (less than 1,000 parts per million). Thus the figure of 600,000 tons of reserves becomes very significant.

The only conclusion that can be reached, based on the available information, is that any reactor that will not be on line by the end of 1976 (1977 at the latest) should be redesigned for the use of coal. Hence the thesis of Professor Bethe's article, that there is a nuclear source of energy that could be brought in were it not for the critics (safety, economy, danger of sabotage or nuclear blackmail, etc.), is unsound because the resource base is absent. . . .

Nuclear fission, therefore, is not a solution to the energy shortage. It is part of the problem: it diverts men, materials and money from the only long-term supply of fuel we have—coal.

RAPHAEL G. KAZMANN

Professor of Civil Engineering

JOEL SELBIN

Professor of Chemistry
Louisiana State University
Baton Rouge

Sirs:

It is surprising that so knowledgeable an authority as Professor Bethe should ignore the very significant contribution to our energy resources that could come from the direct conversion of sunlight into electricity. The feasibility and reliability of photovoltaic conversion has been conclusively established by its successful use as the power source in many space vehicles. The solar cells used for space vehicles are much too expensive for terrestrial use, but current research indicates that, with mass-production techniques being developed, the cost of manufacture will drop to the extent that this inexhaustible and pollution-free source of electrical energy will become competitive with nuclear fission.

Current estimates suggest that the cost of a photovoltaic generating plant per kilowatt peak output will be about $500. As was noted by Professor Bethe, an average output (24 hours per day, 365 days per year) of one kilowatt would require an installation providing a peak output of four kilowatts, which would cost about $2,000, if the projections mentioned above are realized.

The cost of a fission generating station per installed kilowatt-hour is quoted by Professor Bethe as being about $750, and this would seem to indicate that the photovoltaic generation of electricity will never be economically attractive. Three considerations suggest otherwise. In the first place, the cost of the installation for nuclear or

One of a series of reports on the first hundred years of the telephone.

Now that you've invented the telephone, Professor Bell, how can you make it so people can afford it?

Alexander Graham Bell's fingers were all thumbs in mechanical matters. He had a sensitive ear, an original mind, and a deep knowledge of speech and hearing. These assets enabled him to invent the telephone "mentally." But translating the concept into a working model required mechanical gifts he did not possess. And manufacturing telephones for use by the public posed still other problems.

Alexander Graham Bell's first telephone

To be fair, very few people had any experience fabricating electrical devices in 1873, when Bell began the investigations that led to the telephone. Samuel F. B. Morse's telegraph, invented in 1835, was the only important commercial use yet made of electricity. The Massachusetts Institute of Technology was only 14 years old. Thomas Edison's electric light was still some years in the future.

In the United States, some of the most concentrated work on new applications of electricity was being done in the workshop of Charles Williams, Jr. at 109 Court Street, Boston. Inventors including Edison took their ideas to Williams, who translated them into working models—or more often, into models that did not work.

Thomas A. Watson

Bell sought Williams' aid in 1874. The helper assigned to him in January 1875 was Thomas A. Watson, aged 20. Bell was 27. The two complemented each other marvelously. Watson had his own sort of genius—for the practical business of putting together metal and wood and glass to form devices that would do what he wanted. The collaboration between the two men produced, on March 10, 1876, that famous first telephone call: "Mr. Watson, come here. I want to see you."

Bell foresaw a time when telephone service would link the cities of the world. But that required vast improvements in the telephone and in telephone connections. The newborn Bell Telephone Company set out to make those improvements. Watson hired two assistants, and began what would be called today a research and development program. He made the experimental phone sturdier, and devised a hand-cranked magneto to ring a bell. Williams began manufacturing Watson's designs. The invention's financial backers organized telephone companies in various cities, raised additional capital, and hired a General Manager, Theodore N. Vail.

The Western Electric Manufacturing Co., Chicago 1881

By 1879 demand for telephones exceeded the capacity of Williams' shop. Manufacturers in Indianapolis, Cincinnati and Baltimore were licensed to make telephones under the Bell patents. More were needed. One company in Chicago had gained considerable expertise in making telegraph equipment. Years later, Watson reminisced, "When a piece of equipment built by Western Electric came into our shop...we always used to admire the superlative excellence of the workmanship." In 1881, Bell Telephone acquired a controlling interest in Western Electric, and in 1882 made that company the manufacturer of Bell equipment.

This arrangement was desirable for a number of reasons. Western Electric workmanship improved the quality of voice reproduction by telephone, a step essential to winning wide public acceptance of the new invention. It ensured reliability. When repairs were needed, standardized parts from Western Electric made for speedy restoration of service. A fourth advantage concerned the evolving nationwide network: a single manufacturer could see to it that telephone equipment throughout the country would work together compatibly, thus assuring the "interconnectedness" of the network—its capability of connecting any two phones.

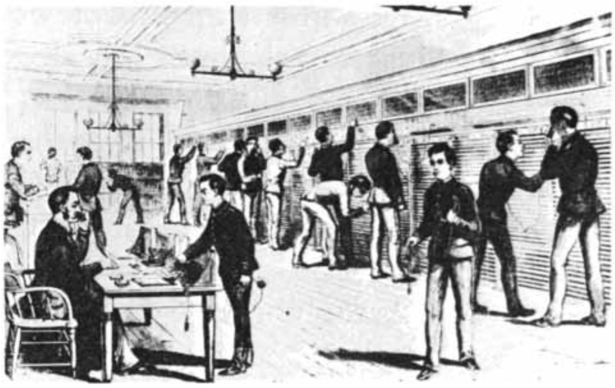
The Engineering Department of Western Electric joined in the search for telephone improvements. Managers of local companies made suggestions based on day-to-day experience with customers. Outside inventors worked out refinements. Bell Telephone management weighed all the ideas, in terms of value to customers and practicality in manufacture. The best ideas were incorporated into the phones being made.

Organizational "feedback"—the subtle flow of engineering information and understanding within a technologically oriented enterprise—was a novel concept in those days. But the young telephone industry had already achieved a union of the successive stages of effort essential to the development of a coherent telephone system. Today Bell Laboratories is responsible for research and development. Western Electric looks for—and finds—better ways to make things. The 23 regional Bell companies provide telephone service and report back their needs and the needs of their customers. And American Telephone and Telegraph Company management provides overall coordination and guidance. Some of the names are different, but the functions had been established by 1882.

Vertical integration is one name economists give to this form of corporate organization. In a typical case, raw materials change to finished product with successive stages of manufacture integrated under one company. For example, a printing company might own paper mills and even its own forests. But there is a difference, because the Bell company was not, and is not, primarily a manufacturer of products. From the beginning, the partners in the enterprise—whether engaged in invention or manufacture or operations—sold telephone service, not telephone equipment.

For instance, there was the fundamental item of connecting the lines of two subscribers who wanted to talk. This was handled in central offices by a corps of operators, using cords, plugs and jacks. Setting up a call could take as long as seven minutes. In 1884, Ezra T. Gilliland, working for the Bell company, devised a mechanical system that would allow a subscriber to reach up to 15 lines without the help of an operator. In 1891, Almon B. Strowger, a Kansas City undertaker, patented a dial machine constructed in a round collar box. It connected up to 99 lines. But the big city offices already handled thousands of lines, and the numbers were growing rapidly. The connection problem was growing much more rapidly, because of some basic geometry: it takes one line to interconnect two telephones, three lines for three telephones, six lines for four, 28 lines for eight, and 4,851 lines for 99 telephones. In connections, added telephones were just the opposite of "cheaper by the dozen."

The Bell company set out to develop a machine that would connect any of 10,000 telephones—49,995,000 possible connections. The search was costly, but necessary for continued good service, and the various parts of the company joined to pursue it to a successful conclusion. (Today in the



Young men manually connecting phone calls in 1879

United States a telephone can be connected to any of 140 million others. There are 10 quadrillion—10 million billion—possible connections.)

There was also the problem of financing the nationwide conversion of central offices to dial. Service improvements on the scale required are enormously expensive—so expensive as to be impossible without the most careful attention to economy. Here again the integrated corporate structure shows its values. Western Electric, because its prime objective is to benefit telephone service, has become a world champion in cost control, and a pacesetter in the improvement of productivity. Data issued by the Federal Bureau of Labor Statistics show that overall the productivity of the telephone industry has increased by 50% since 1965. That is two and a half times the productivity increase in the United States economy as a whole.

Savings of that sort continue, as a recent example shows. Bell scientists, building on the semiconductor research that helped them invent the transistor, also aided in the development of the light-emitting diode or LED. These solid-state lamps, now familiar as displays in pocket calculators and watches, can replace incandescent lamps in many pieces of telephone equipment. They will last the lifetime of the phone, operate with much less electrical power, and help hold down the cost of installation and maintenance. Over the next five years, LEDs should save the Bell System about $120 million. Bell Labs semiconductor research also resulted in the invention of another solid-state light source, the tiny semiconductor laser.

Future uses of these solid-state light sources may be even more important than today's. They will almost certainly be used

For the nation's future communications needs, Bell engineers are today developing systems to transmit telephone calls on lightwaves.

in systems transmitting telephone calls and other communications over lightwaves. Lightwave communications could mean an enormous increase in the capacity of the phone network, making it possible to meet the need for increased call volumes and new communications services economically in the years ahead. And to do it while conserving energy and scarce raw materials.

Innovation, productivity, advance planning—all mean improved service and reasonable costs. And in the telephone industry, cost savings benefit not only shareholders; they are passed along to the public as well.

In the decade 1965-1975, the cost of living rose 74%. Telephone rates for local service went up only 40%. And interstate long distance rates went up about 4%. Now 95% of all American homes have telephones. The quality of the service is the envy of the world.

Numerous studies have been made of the role of the Bell organizational structure in achieving those results. One of the most recent, concluded in 1974, was made by the independent auditing and management consulting firm of Touche, Ross & Co., acting as consultants for the staff of the Federal Communications Commission. According to their report:

> "Western Electric's efficient performance has resulted in lower costs than otherwise would have been incurred. Because of Western's pricing policies and practices, these lower costs have not increased profits, but have been passed on to operating companies in the form of lower prices...The effect of the interrelationship between Bell and Western Electric is to operate Western, not as a manufacturing concern, but as an integral part of a vertically integrated communications firm. These interrelationships result in a favorable impact upon Western's costs, prices and service to operating companies."

The best telephone service in the world didn't just happen. It was planned that way.

One Bell System. It works.

Bell System

fossil-fuel generating stations is only part of the cost; fuel, maintenance and obsolescence must also be paid for. A photovoltaic system uses no fuel, and the absence of moving parts minimizes maintenance costs and maximizes life expectancy. Secondly, it is misleading to compare costs on the basis of the "round the clock, round the calendar" operation of a nuclear installation. On the average, electric-generating plants in the U.S. generate only 50 percent of their rated output, averaged over the entire year. Thirdly, it should be remembered that in those parts of the country that have the most sunshine the heaviest demand for electricity comes when the sun is shining, because the air-conditioning load is superimposed on the industrial and commercial demand. During those periods the output of a solar generating plant would be close to four kilowatts for every $2,000 worth of installation. Since that peak load must be provided by one means or another, the photovoltaic option may be the most economical. Finally, it is not unlikely that combined photovoltaic and heating units will become available for home installation. A unit giving an average output of one kilowatt of electrical energy would produce about 8,700 kilowatt-hours per year, which is about the amount consumed by the "average family." The cost of such an installation, including battery storage, would be less than $3,000. If a lifetime of 15 years is assumed, the annual cost, including interest at 8 percent and repayment, would be about $360 per year. The cost per kilowatt-hour would then be just over four cents. This calculation is for the most highly insolated part of the country (the Southwest), but even in the Northeast, where there is only six-tenths as much sunshine, the cost would be seven cents, which is comparable to the cost of the electricity delivered by the utility company. It should be added that the cost of heating would be very substantially reduced, because the solar heat would provide at least half of the energy required for space heating and hot water.

BRUCE CHALMERS

Division of Engineering
and Applied Physics
Harvard University
Cambridge, Mass.

Sirs:

The problem of the energy needed to build a nuclear reactor has been much discussed, and I certainly agree with Dr. Cody that it is essential for a judgment of the usefulness of nuclear reactors. My appraisal of this problem differs from Dr. Cody's on three points.

1. Reactor building is a steady process. Many reactors are under construction now and will come on line in the next five or 10 years. It is therefore unrealistic to assume that construction starts now and then stops at some later time. Instead I shall assume

that the number of reactors being committed, and also the number being completed, increases exponentially with time.

At present we have about 31 gigawatts of nuclear reactors. I shall take the same numbers as Dr. Cody, that is, 520×10^3 megawatts, as the capacity in 2000. This corresponds to a yearly growth rate of 11.2 percent. Alternatively I shall take President Ford's aim of 225 reactors in 1985; this corresponds to a growth rate of 19.8 percent. To bracket these two numbers I shall use alternative growth rates of 10 percent and 20 percent.

2. The energy input is not uniform during construction. The fuel needs to be provided only shortly before the reactor goes into operation, let us say one year. The first two or three years of reactor building are occupied by paperwork, obtaining a license and so forth. I shall assume that the energy needed in construction is spent in equal amounts from two to seven years before the reactor is completed.

3. Of the energy assigned by Chapman to plant construction, 1.8×10^9 kilowatt-hours of electricity, about a third is required to obtain the heavy water that British reactors use. U.S. reactors do not use heavy water, so that Chapman's figure is reduced to 1.2×10^9 kilowatt-hours. I shall take the average between this and Davis' figure, namely 1.0×10^9 kilowatt-hours of electricity.

I then calculate the ratio of the energy spent in new reactor construction and fueling to the net energy produced by the reactors in operation, making the same assumptions on capacity factor and so forth that Chapman made. The result is that the new construction absorbs 5.8 percent of the produced energy if the growth rate is 10 percent per year and 16.5 percent of the produced energy if the growth rate is 20 percent per year. Neither of these figures is alarming. In an economy with steady construction of nuclear reactors the energy consumed in construction is less than 20 percent of the energy produced by the reactors.

Concerning the letter from Professors Kazmann and Selbin, the problem of uranium availability is clearly very important. Realizing this, ERDA established a Fuel Cycle Task Force early in 1975 that made a detailed assessment of uranium-ore prospects and published it in ERDA-33.

The important point is that one should consider not only proved reserves but also total probable resources. (If we had taken only proved reserves, oil would have run out many decades ago.) The proved reserves of uranium oxide are indeed about 600,000 short tons, as Kazmann and Selbin state. However, the probable resources are 2.4 million short tons, including ores of concentration down to 200 parts per million U_3O_8 (yellowcake). (The resources include copper leach residues and phosphates.)

ERDA-33 states the following: "Based on statements from experts in the field, it seems reasonable to assume that, with-

Laser Beam Digital Watch

Never press another button, day or night, with America's first digital watch that glows in the dark.

Announcing Sensor's new Laser 220—the first really new innovation in digital watch technology.

It's ingenious, it's simple and it makes every other digital watch obsolete. Scientists have perfected a digital watch with a self-contained automatic light source—a major scientific breakthrough.

SELF-CONTAINED LIGHT SOURCE

The Laser 220 uses laser beams and advanced display technology in its manufacture. A glass ampoule charged with tritium and phosphor is hermetically sealed by a laser beam. The ampoule is then placed behind the new Sensor CDR (crystal diffusion reflection) display.

The high-contrast CDR display shows the time constantly—in sunlight or normal room light. But, when the room lights dim, the self-contained tritium light source automatically compensates for the absence of light, glows brightly, and illuminates the display.

No matter when you wear your watch—day or night—just a glance will give you the correct time. There's no button to press, no special viewing angle required, and most important, you don't need two hands to read the time.

Replace the battery yourself by just opening the battery compartment with a penny. Free batteries are provided whenever you need them during the five-year warranty.

A WORRY-FREE WATCH

Solid-state watches pose their own problems. They're fragile, they must be pampered, and they require frequent service. Not the Laser 220. Here are just five common solid-state watch problems you can forget about with this advanced space-age timepiece:

1. Forget about batteries The Laser 220 is powered by a single EverReady battery that will actually last years without replacement—even if you keep the 220 in complete darkness. In fact, JS&A will supply you with the few batteries you need, free of charge, during the next five years. To change the battery, you simply unscrew the battery compartment at the back with a penny and replace the battery yourself.

2. Forget about water Take a shower or go swimming. The Laser 220 is so water-resistant that it withstands depths of up to 100 feet.

3. Forget about shocks A three-foot drop onto a solid hardwood floor or a sudden jar. Sensor's solid case construction, dual-strata crystal, and cushioned quartz timing circuit make it one of the most rugged solid-state quartz watches ever produced.

4. Forget about service The Laser 220 has an unprecedented five-year parts and labor warranty. Each watch goes through weeks of aging, testing and quality control before assembly and final inspection. Service should never be required. Even the laser-sealed light source should last more than 25 years with normal use. But if it should require service anytime during the five year warranty period, we will pick up your Sensor, at your door, and send you a loaner watch while yours is repaired—all at our expense.

5. Forget about changing technology The Sensor Laser 220 is so far ahead of every other watch in durability and technology that the watch you buy today, will still be years ahead of all others.

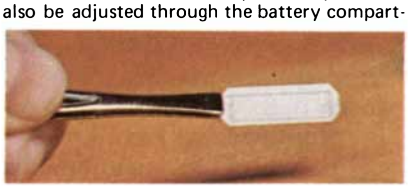
THE ULTIMATE ACHIEVEMENT

Other manufacturers have devised unique ways to produce a watch you can read at a glance. The new $300 LED Pulsar requires a snap of the wrist to turn on the display, but the Pulsar cannot be read in sunlight. The new $400 Longine's Gemini combines both an LED and liquid crystal display. (Press a button at night for the LED display, and view it easily in sunlight with the liquid crystal display.) But you must still press a button to read the time. All these applications of existing technology still fail to produce the ultimate digital watch: one you can read under all light conditions without using two hands. Until the introduction of the Sensor.

PLENTY OF ADVANCED FUNCTIONS

Sensor's five time functions give you everything you really need in a solid-state watch. Your watch displays the hours and minutes constantly, with no button to press. But depress the function button and the month and the date appear. Depress the button again and the seconds appear. To quickly set the time, insert a ball-point pen into the recessed time-control switch on the side. It's just that easy.

Sensor's accuracy is unparalleled. All solid-state digitals use a quartz crystal. So does the Sensor. But crystals change frequency from aging and shock. And to reset them, the watch case must be opened and an airtight seal broken which may affect the performance. In the Sensor, the crystal is first aged before it is installed, and secondly, it is actually cushioned in the case to absorb tremendous shock. The quartz crystal can also be adjusted through the battery compartment without opening the case. In short, your watch should be accurate to within 5 seconds per month and maintain that accuracy for years without adjustment and without ever opening the watch case.



The new exclusive laser-sealed tritium and phosphor light source is a thin solid-state tube that automatically illuminates the display when the lights dim.

Would you do this with your solid-state watch? Of course not. Most solid-state watches require care and pampering but not the Sensor. You can dunk it, drop it and abuse it without fear during its unprecedented five-year parts and labor warranty.

STANDING BEHIND A PRODUCT

JS&A is America's largest single source of digital watches and other space-age products. We have selected the Sensor Laser 220 as the most advanced American-made, solid-state timepiece ever produced. And we put our company and its full resources behind that selection. JS&A will warranty the Sensor (even the batteries) for five full years. We'll even send you a loaner watch to use while your watch is being repaired should it ever require repair. And Sensor's advanced technology guarantees that your digital watch will be years ahead of any other watch at any price.

Wear the Laser 220 for one full month. If you are not convinced that it is the most rugged, precise, dependable and the finest quality solid-state digital watch **in the world**, return it for a prompt and courteous refund. We're just that proud of it.

To order your Sensor, credit card buyers may simply call our toll-free number below or mail us a check in the amount indicated below plus $2.50 for postage, insurance and handling. (Illinois residents add 5% sales tax.) We urge you, however, to act promptly and reserve your Laser 220 today.

Stainless steel w/leather strap $129.95
(Add $10 for matching metal band)

Gold plated w/leather strap. $149.95
(Add $10 for matching metal band)

JS&A NATIONAL SALES GROUP

Dept. SA JS&A Plaza
Northbrook, Illinois 60062
CALL TOLL-FREE . . 800 323-6400
In Illinois call . . . . (312) 498-6900

©JS&A Group, Inc., 1976

Digital thinks a computer should save you more than it costs.

These three men think we're right on the money.

"Digital saved us 500 man-hours a year. And that was just the beginning."

At its Fullerton, California research lab, Hunt-Wesson Foods, Inc. conducts exhaustive product analyses in support of its extensive research program. Tests are performed to check on nutritional value, chemical composition, flavor, color and a number of other factors.

Since installing a Digital computer, Hunt-Wesson has increased the efficiency of its analytical effort dramatically.

"We're saving $20,000 a year, year after year."

Though still a small business, Field Concrete Pipe Company has grown to be one of New England's major suppliers of concrete culvert and sanitary sewer pipe.

"Our warehouse moves 10% faster thanks to two PDP-11 computers."

David DeWaters is the Staff Manager for Giant Food's Jessup Grocery Warehouse.

The warehouse is one of the largest and busiest in the U.S. They've got 10 acres—over 400,000 square feet—under one roof. With 32 truck shipping docks, 46 truck receiving docks and 700 feet of rail docks, they handle over 750,000 cases of groceries per week—in and out.

To keep everything running smoothly, Jessup uses a huge automatic order picking system driven by two Digital 11/20 computers. Says DeWaters: "Since

Saving managers millions.

Low cost computers that pay for themselves over and over again is what has made Digital one of the two computer companies in the world to install over 50,000 computers.

And what this means to you is that right at the time you need it most you have access to an enormous reservoir of experience and ideas for cutting costs while getting more done.

Digital is the pioneer in minicomputers, the leader in the development of small- to large-scale timesharing systems, and the company that has made many of the major breakthroughs in computerizing techniques.

Because of Digital, interactive computing, computer hierarchies and networks, data communications systems and new uses of computer systems are growing at an unprecedented rate.

And every year Digital places a remarkably comprehensive selection of computer tools within the financial reach of thousands and thousands of new users.

Today anyone who has a good applications idea for a computer can afford one. And just as important, they can get the help they

"Right off, we started saving 500 man-hours a year. In addition, we were able to reduce the number of analyses and improve the accuracy of the results," says Ken Stoub, R&D Section Head. "Moreover, data reduction takes 90% less time."

As a result, Hunt-Wesson's lab has been able to take on additional jobs without additional expense. The low cost of Digital equipment helped.

"Our system cost $20,000. And it's already paid for itself."

How does Mr. Stoub rate Digital service? "One of the things I've admired most about Digital is the company behind the computer. Their people have shown a consistently high interest in our operations and problems. In fact, they've been on the scene whenever we needed them right from our initial contact." ■

Robert Field, president and founder, gives a lot of credit for their recent business improvement to a small computer from Digital. "Our Digital Datasystem performs like one of our staff. It helps our dispatcher keep completely up to date with inventory, production, shipping, customer credit and payment records. It's amazing that, with the touch of a button, all the different functions of our business come together in a way that we know helps us make better decisions. The overall flexibility and 'usability' of this machine is really impressive. Our sales people set average price figures, trace business patterns and enter sales calls.

"We even use it in management meetings to get up to date financial and production information instantly.

"We figure our $50,000 investment in a DEC Datasystem will save us at least $20,000 a year, year after year." ■

---

switching over to an automated system, we increased control over our operation, improved service to all 104 of our stores, and increased our warehouse productivity by about 10%."

There are fringe benefits, too. Mr. DeWaters explains: "By taking the most physical part of our operation and automating it, our Digital computers helped increase employee morale and create a cleaner, healthier working environment."

Right now 60% of the total inventory is under automatic control. Eventually the plan is to put 70% of the inventory under computer control.

"We see the improvement in receiving, inventory control, and shipping as a way of reducing our operating costs."

Mr. DeWaters sums up Digital's contribution this way: "We installed DEC computers to increase efficiency. And they increased morale, too. I guess you could say we got double our money's worth." ■

---

need from a company with over 4000 field representatives in 25 countries around the world.

To find out how Digital can help you save money, send for your free copy of "50,000 Computers Saving Managers Millions." Fill out the coupon below, or call or write Digital Equipment Corporation, Maynard, Mass. (617) 897-5111, Ext.4089. European headquarters: 81 route de l'Aire, 1211 Geneva 26. Tel: 42 79 50. Digital Equipment of Canada, Ltd.

Please send me your free book "50,000 Computers Saving Managers Millions" to show me how I can get more work done at lower cost.
My computer application is ______________.
Name ______________ Title ______________
Company ______________ Address ______________
City ______________ State ______________ Zip ______________

digital

50,000 computers saving managers millions.

out going into such low-grade deposits as shales, lignites or seawater, and with reasonable assurance of a market for this product, the uranium mining industry could, with proper incentives, carry out an exploration, mining and milling program which, while taxing the money market and the ability of the equipment suppliers, could produce on the order of 2,400,000 tons of U_3O_8 equivalent." As is clear in the letter from Professors Kazmann and Selbin, the rate of exploration for uranium must increase. Again quoting ERDA-33: "A comprehensive Government program (National Uranium Resource Evaluation—NURE) is currently under way to evaluate domestic uranium resources and to identify areas favorable for uranium exploration. A preliminary evaluation is to be completed by the end of 1976 and a complete evaluation around 1980."

At present we are unfortunately in a vicious circle: public acceptance of nuclear power is uncertain, therefore there is not enough drilling for yellowcake, hence the proved reserve of uranium is insufficient, and this again diminishes public acceptance.

Professors Kazmann and Selbin make a rather low estimate of the energy that can be extracted from uranium and accordingly state that in its 30-year lifetime a 1,000-megawatt plant will use about 6,900 tons of yellowcake. According to the AEC evidence in the report WASH-1139, you need only 5,900 tons without recycling, or 5,200 tons with plutonium recycling. The difference is small but not negligible.

Professors Kazmann and Selbin are also concerned because actual plants have extracted only about 55 percent of the available energy from the fuel. However, as ERDA-33 states, "the causes of these difficulties have been identified, and both manufacturers now maintain they are confident their fuel will meet the guarantees of about 25,000 MWd/t for Boiling Water Reactor fuel and approximately 33,000 MWd/t for Pressurized Water Reactor fuel."

Moreover, even if energy extraction is inadequate, the remainder of uranium 235 in the fuel can be recovered in chemical processing plants. Although such reprocessing is expensive, it will become economically advantageous when the price of uranium ore increases. With reprocessing we may also regain the uranium 235 that remains in fully used fuel, which will reduce the ore requirement by another 10 or 15 percent.

Using the probable resource of 2.4 million short tons, we can supply 350 reactors with their lifetime fuel if we adopt the Kazmann-Selbin figure, 405 if we adopt WASH-1139 and 460 if we also include the recycling of plutonium. This is about the number of reactors now expected to be built by 1995. Moreover, there is an independent assessment by Milton F. Searl of the Electric Power Research Institute, who estimates that the total available yellowcake in the U.S. is about 10 million tons.

Nevertheless, I wholeheartedly agree

with Professors Kazmann and Selbin that a nuclear plant making a better use of the available uranium is urgently needed. In my article I mentioned the high-temperature gas-cooled reactor and a modification of the CANDU reactor using uranium 233. If this version of the CANDU reactor were adopted generally, our uranium resources could be stretched by a factor of about five.

Professors Kazmann and Selbin mention the breeder, which I also discussed. If breeders come on line in the 1990's, they will be just in time to relieve the uranium-ore situation. Accordingly no time should be lost in developing the breeder. A breeder with a doubling time of 30 to 60 years, like the French Phénix, would already be very useful, because after the first fuel investment no further fuel would be needed for the lifetime of the breeder. Indeed, such a breeder would provide for a gradual expansion of the nuclear-power establishment. On the other hand, most people in the business are confident a breeder can be built that will have a doubling time of about 20 years. This should be sufficient to allow expansion of the nuclear-power industry at a reasonable rate after the year 2000.

The letter from Professor Chalmers is good news, namely that photovoltaic generation of electricity from the sun's energy may be possible at a very reasonable price: $500 per kilowatt peak output. It is to be hoped that industrial realization of the scheme will keep the price at this low level, and that this development is receiving full support from ERDA.

Professor Chalmers is also correct in his economic analysis. Solar electricity will indeed be available just at the time of the heaviest demand for electricity and is therefore doubly welcome. However, the use of this peaking-power installation evidently presupposes that there are also generating plants for the base load of electricity. Nuclear plants are particularly suitable for this, so that a combination of nuclear and solar installations would be most advantageous economically.

H. A. BETHE

Cornell University
Ithaca, N.Y.

ERRATUM

In "The Ethics of Experimentation with Human Subjects," by Bernard Barber (SCIENTIFIC AMERICAN, February), it was stated that experiments involving the injection of cancer cells into geriatric patients had been conducted at the Jewish Hospital and Medical Center of Brooklyn. The experiments were conducted at the Jewish Chronic Disease Hospital.

50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

APRIL, 1926: "The wealth and general prosperity of the country are largely due to the intelligence and energy of its people, but it can hardly be disputed that it is equally due to the natural wealth of the country. It seems as if Nature must have said, I shall prepare a land that in respect to the necessities of life and the conditions for the upbuilding of a highly civilized people, shall be absolutely self-contained. I shall cover its surface with the richest of soils, watered by majestic rivers and clothed in great part by forests, which in the variety and magnificence of their timber will be unapproached elsewhere on the planet. Below the surface I shall lay up practically inexhaustible stores of fuel, both coal and oil. And in readiness for the great industrial age, which will come in due course, I shall enrich this favored land with vast accumulations of metalliferous rocks so readily accessible that in the case of iron the ores can be shoveled up by the millions of tons and drawn away in trainloads to the smelting furnaces. It is a priceless patrimony, but alas, with all the extravagance of the thoughtless youth who comes suddenly into a vast fortune, we have expended our riches as though they were absolutely without limit—which they are not. It is only of late years that we have been moving to a full realization of our youthful extravagance. We have learned these facts a little late, but not too late. It remains for us to put into practice those principles of thrift and conservation that have long been practiced by the older civilizations of Europe."

"The oxygen requirement of running at various speeds has been measured by Professor A. V. Hill of University College London. The subject runs about 100 yards at any required speed, holding his breath the while, and at the end lies down at once upon the ground and allows all the air that he breathes out to be collected in a bag. The long-distance runner may be breathing twice his own volume of air per minute. From this he may be taking into his blood some four litres of oxygen, about one gallon. This oxygen is utilized in the combustions that provide him with energy. Energy indeed is required: each member of the crew of an eight-oar boat may be developing .6 horsepower of actual mechanical work throughout a three-mile race. In a 100-yard sprint the muscular effort may be so intense that it would require energy equivalent to about 15 horsepower to maintain it."

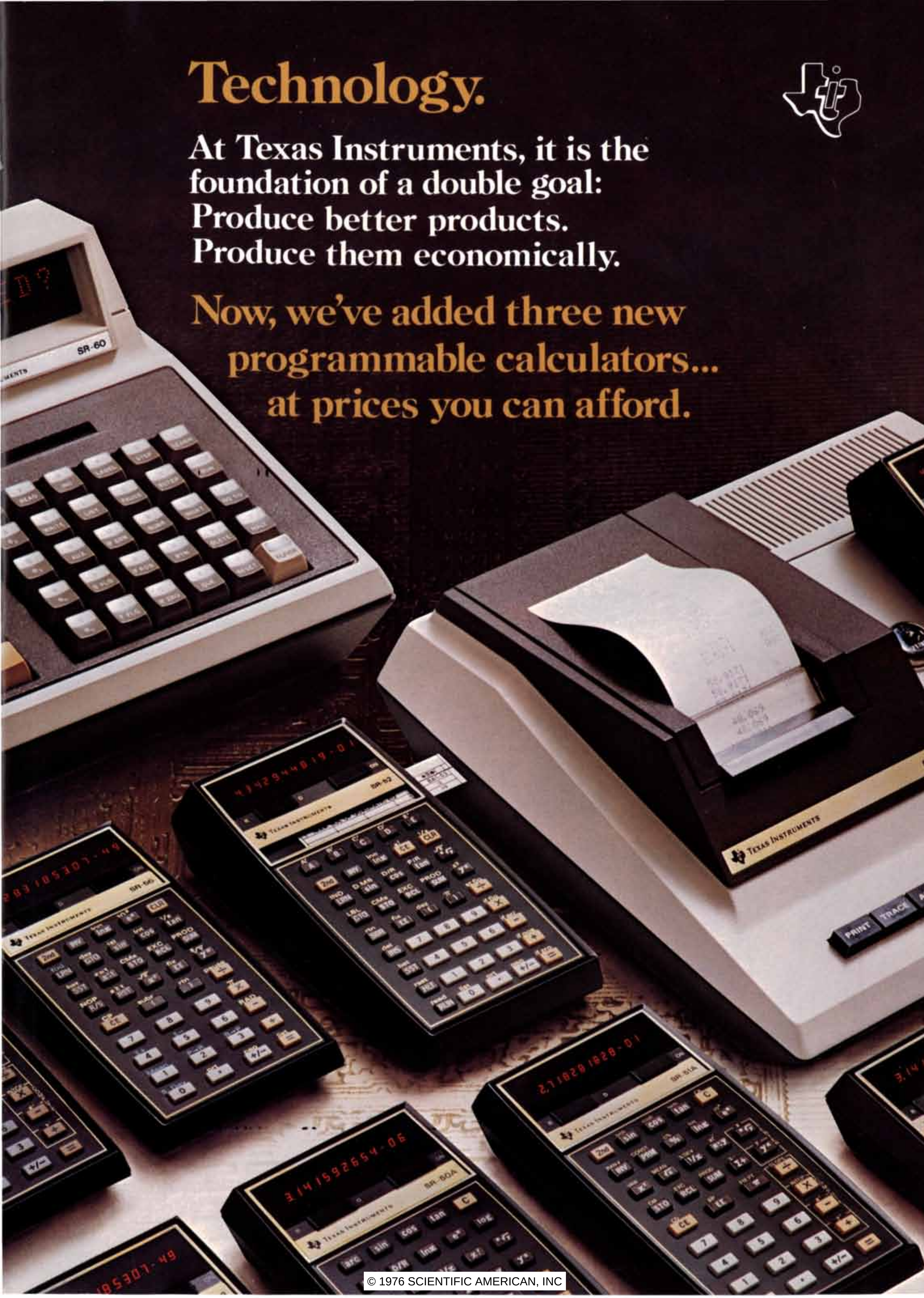
"For perhaps 100 years boats have been propelled in shallow waters by pumping water to the top of a tank and then allowing it to escape at a fair rate of speed from the bottom of the tank in a narrow jet. The reaction of the escaping jet was sufficient to drive the boat forward, even though the efficiency of the mechanism was poor. Is it possible that the principle of jet propulsion can be applied to the airplane? The French Service Technique de l'Aeronautique is testing out an airplane power plant invented by H. F. Melot that is based on the principle of jet propulsion, the energy of the gas being converted directly into propulsive effort without the mediation of connecting rods, crankshaft and screw propeller. American investigations on the subject have put in doubt the possibility of obtaining any great efficiencies, but for exceedingly high-speed planes the simplicity of the apparatus and the elimination of the propeller lead us to think that further developments are worth watching."

SCIENTIFIC AMERICAN

APRIL, 1876: "The President has recently sent to the Senate for ratification a treaty the object of which is to establish an international uniformity and precision in the standard of weights and measures. The treaty is between the United States and the governments of the Argentine Republic, Austria, Belgium, Brazil, Denmark, France, Italy, Peru, Portugal, Russia, Spain, Sweden and Norway, Switzerland, Turkey and Venezuela. It contains an agreement between all the parties to maintain in Paris, at the common expense, a permanent bureau of weights and measures, to be under the control of an international committee."

"The *Germanic,* mean displacement 8,525 tuns, recently made the trip from New York to Queenstown, Ireland, 2,894 nautical miles, in seven days 15 hours 17 minutes, being an average of 15.8 knots. This is the fastest time on record."

"Caroline Lucretia Herschel, the sister of Sir William Herschel, offers a bright example of what a woman's work may be when an intense personal sympathy and affection enlist her powers. It was to the advantage of Science that those powers were directed to its furtherance, for had Sir William Herschel, unstrengthened in his purpose by her, remained the humble music teacher, he would have passed to posterity as the composer of a few mediocre symphonies and not as the discoverer of Uranus. She herself, at the close of a vocalist's career, would have sunk into oblivion, and Sir John Herschel, the son of the one and nephew of the other, doubtless would have followed his father's lowly path. It is only necessary to remember the inestimable value of the labors of these three persons in the cause of the grandest of the natural sciences to realize how great would have been the loss to mankind had


Technology.
At Texas Instruments, it is the foundation of a double goal: Produce better products. Produce them economically.
Now, we've added three new programmable calculators... at prices you can afford.
SR-60
PRINT
TRACE

Choose the right programmable ...and problems that once took hours can be solved in seconds.

SR-56 Key programmable. Full programming capability at an economical price.

SR-52 Pocket-card programmable. Records your programs for instant use. Anytime, anywhere.

SR-60 Prompting, printing, card programmable. Bridges the gap between calculators and computers.

SR-60

A prompting, programmable, printing calculator with an alphanumeric display that communicates with the user. $1695.00*

The SR-60 card-programmable, prompting, printing calculator is designed to bridge the gap between simple desktop calculators and computers. A powerful asset to business and technical operations alike. Delivering capability found only in programmable desktop calculators costing far more.†

Its business capability ranges from solving intricate financial analyses and long-range forecasting, to simpler operations like payroll and amortization.

For technology there are 46 scientific functions on the keyboard and 480 program steps for complex programming. This capacity can be expanded to 1,920 steps and 100 data memories with its optional module.

Prompting: The SR-60's unique 20-character display lets the user run alphanumeric programs which "ask" for information at successive stages of the problem. The SR-60 then waits for your response before continuing. This dialogue allows even a novice to work with complicated problems immediately.

Programming: Is easy and straightforward yet flexible for the user with: 78 labels, 10 flags, 10 branches, 4 levels of subroutines, and 2 modes of indirect operation. Plus, complete program editing capability. And, by using the printer you can list and trace the actual program execution.

Programs are easy to write and record on magnetic cards. With alphanumeric prompting, the cards can be used by assistants or secretaries. A person merely needs a minimum amount of instruction and a general concept of what's to be solved to have answers to you in seconds.

Ten prerecorded cards are included in the SR-60's Basic Library: Power transformer and filter design. Add-on rate installment loans and compound interest. Polynomial evaluation, cubic and quadratic equations. Basic statistics. Random number generator and diagnostics. Well over 100 optional additional programs are available, including many on business.

Printing: The SR-60's quiet printer provides a scaled replica of what appears on the alphanumeric display on 2½-inch thermal paper. You can get a hard copy of any keyboard calculation that appears on the display, a complete program list of the contents of the data registers, whether entered from the keyboard or run from a program card.

SR-52

An easy-to-use card programmable. Bringing you exceptional power wherever your work happens to be. At an exceptional price. $395.00*

The SR-52 is a card programmable calculator offering twice the capability of the only other programmable in its class – at half the price.† TI's advanced technology and start-to-finish quality control is the key to this exceptional value.

With an SR-52, complex repetitive problems or lengthy calculations that once took hours can now be solved in seconds. Chances for error are dramatically reduced.

The SR-52 allows you to record up to 224 keystrokes. Programming is simple and straightforward, even if you've never programmed before. Programming cards are available which can be integrated into your problem solving routines. Repeat your program as often as needed. Change values of known quantities. Explore "what-if" possibilities. Solve for different unknowns. Optimize designs.

Enter calculations exactly as you would state them – left-to-right. The SR-52's nine levels of parentheses, plus its 11-register stack, allow you to enter problems containing up to 10 pending operations.

Operating versatility: You literally teach the SR-52 your own calculating methods. Key in your program directly from the keyboard. If you wish, record your program on a magnetic card to use again and again.

Used manually, the SR-52 is one of the most powerful handheld, slide-rule calculators available today.

The Basic Library that comes with the SR-52 includes these prerecorded programs: Conversions. Solution of Quadratic Equations. Hyperbolic Functions. Prime Factor of an Integer. Complex Arithmetic. Checkbook Balancing. Compound Interest. Ordinary Annuity. Trend Line Analysis. Permutations and Combinations. Statistical Means and Moments. Random Number Generator. High and Low Pass Active Filter. Dead Reckoning. Lunar Landing Game. Diagnostics.

Optional libraries are also available: Statistics (with 25 different programs). Math (31 programs). Electrical Engineering (22 programs). And Finance (19 programs). And more are on the way.

*Suggested retail price.
†Based on suggested retail prices of models available at the time of this printing.

SR-56

A key programmable that provides tremendous mathematical power and value. And at an economical price. $179.95*

With TI's new SR-56, you get an easy-to-use, yet powerful state-of-the-art calculator that reflects Texas Instruments state-of-the-art technologies. It's able to handle extremely difficult computational problems with 100 programming steps and nine levels of parentheses that handle up to seven pending operations. Yet it is simplicity itself to key program.

With the SR-56's internal 8-memory stack, you can store and recall data. Add, subtract, multiply or divide within a register without affecting the calculation in progress. Now you can optimize mathematical matrices. Explore multiple "what-if" options. Solve lengthy iterative and repetitive problems with speed and efficiency. And much more.

Six logical decision functions and four levels of subroutine permit branching to appropriate program segments automatically – without interrupting the program. You may also write-over errors, erase unneeded keystrokes. Reviewing a program is easy with single and backstep capability.

Two unique features: A special test register permits comparison with the displayed value at any point in a calculation – without interfering with the calculation in progress. This means you can make quick checks of intermediate results for possible pass along to subroutine operations.

A pause key causes the display to be visible for a half-second during program execution. You may also use it to single step through your entire program.

Easy to use: Supply your input data then automatically execute the solution of your stored sequence with a single key. Get answers without the tedium of remembering and executing repetitive keystrokes.

Iterative and repetitive problems, statistical reduction, mathematical modeling, optimization, etc., are entered directly into the SR-56's program memory from the keyboard. Two looping control instructions give you single-key control. There are also four levels of subroutines. They execute instructions from the main program, or from another subroutine. On completion, control returns to the calling routine, automatically.

Operated manually, the SR-56 easily handles your day to day problems using the 27 arithmetic and transcendental functions including: Trig, logs, conversions and statistics.

PC-100: New optional printer turns an SR-52 or SR-56 into a quiet, high-speed printing calculator. $295.00*

The PC-100 operates with TI's handheld programmables – the SR-52 and SR-56. It delivers hard copy right on the spot. Perfect for printing out a businessman's long amortization schedule, or each step of a scientist's iterative problem.

Printing can be controlled by keys on the PC-100 or by keys on the calculator. Simply remove the calculator's battery pack. Then press the calculator firmly on the PC-100's connectors. Lock it in place and you're ready to print whatever appears in the SR-52 or SR-56 display register.

The PC-100 prints a "list" of your entire program step-by-step, including the program code. You may halt it whenever you wish, or begin printing from any point in the program. This makes the PC-100 invaluable for checking whether you have keyed-in the instructions correctly – match tape against your coding sequence. Edit and debug your program. Or, verify that your results are based on a correctly formulated program.

Using the PC-100's "trace" mode delivers a complete audit of every number and function you've used.

The quiet, reliable electronic printer uses thermal tape 2½-inches wide and prints out characters in a five by seven dot matrix.

The technological achievement beneath the keyboard is the reason TI's programmable calculators offer so much value for the price.

A programmable calculator is a state-of-the-art product reflecting state-of-the-art technologies. It's logical, then, to look first to the manufacturer known worldwide for both – Texas Instruments.

TI has long been a leader in solid-state technology and has pioneered a series of landmark developments relating directly to calculators: The original integrated circuit. Key patents in basic MOS/LSI technology. The "calculator-on-a-chip" integrated circuit which became the heart of miniature calculators. And the basic patent on the miniature calculator itself.

TI is steeped in calculator technologies from start to finish, making all critical parts and controlling quality every step of the way. And that's the key to the exceptional quality and value of TI programmable calculators.

The programmable calculator ...it can help you make the best choices, the right decisions...day after day.

Personal programming is here.

Economical programmable calculators may well be more significant to business and industry than were slide rule calculators introduced just a short time ago.

Why? Because the programmable calculator introduces a new dimension in problem solving. It decentralizes and personalizes the decision-enhancing power of the computer – bringing to the individual what before was only available to the organization.

Now you can cope with more data, explore with more insight, far more successfully than ever before. Right at the source. On the spot, and right at the moment it is most important. Immediately.

So you make better decisions. In the conference room. In the laboratory. In the field. Wherever decisions have to be made. Better decisions chosen from more options – better decisions founded on a broader data base. Better decisions from more fully optimized trade-offs. Better decisions in a profession where better decisions are the name of the game.

Indeed a programmable calculator is a powerful personal mathematical resource. And you don't need to know programming to put it to work. There's no special language to learn. The entry system is easy to use, and so flexible that you can apply it to your own personal problem-solving techniques and style.

Chances are, you already own a calculator – perhaps a sophisticated one. Chances are, too, that you found it exceeded your expectations right from the start... that you grew into it, and it magnified your professional capability far in excess of its cost.

Now personal programming is here. A step-function increase in capability over sophisticated slide-rule calculators. Capability you can put to work now to further strengthen your contribution. Capability you won't fully discover until you've owned one and explored its potential for yourself. Capability to enhance decisions of far greater importance than the cost of the model you choose. You will find your programmable is a high-leverage investment.

Most of the important functions found on computers are available to you on TI programmable calculators: Iterative and repetitive problem solving techniques. Looping. Conditional and unconditional branching. Flags. Subroutines.

Consider for a moment the advantages, in terms of increased productivity, achieved with this capability of: Developing broad what-if matrices. Optimizing mathematical models. Making trend and risk analyses. Projecting and forecasting more accurately. Performing statistical reductions. Automating time-consuming "number crunching". The list could go on and on.

The programmable calculator's capability is in the very mainstream of today's fast-paced, competitive world. A pivotal means of responding to the pressures of making accurate, objective, cost-effective decisions. Faster.

Texas Instruments offers three choices of programmable calculators. This allows you to more precisely match your programming requirements to capability and price.

Each is compact. Easy to use. And of great value for the price. The direct result of leading edge technology developed and practiced at Texas Instruments.

AOS...the new choice that makes it easy.

Most handheld professional calculators use either algebraic entry or Reverse Polish Notation (RPN). Texas Instruments chose algebraic because it is the most *natural* to use with easy left-to-right entry.

The user can put the calculator to work immediately... there's no new language to learn.

Now there's a new dimension: TI's *full* Algebraic Operating System – AOS. Whether you currently own a calculator with algebraic entry, or Reverse Polish Notation – or no calculator at all – you can move into programming smoothly and naturally with TI's full AOS. No system is easier to master.

The case for algebraic is straightforward: It lets you key the problem just as you would state it. TI programmables with full AOS combine full algebraic hierachy with nine levels of parentheses.

Full algebraic hierachy means that sequence of entry is left-to-right, while the sequence of operations is in the accepted convention of the way mathematical operations are ordered: Functions are performed first, then powers and roots, then multiplication or division, then addition or subtraction. For example, the SR-52's nine levels of parentheses, plus its 11-register stack, allow you to enter problems containing up to 10 pending operations. (That's more than three times the capability of its nearest competitor.)

With TI's Algebraic Operating System, you don't have to presolve a problem or search for the most appropriate, efficient order of execution.

The case for TI's Algebraic Operating System is strong – that's why Texas Instruments chose it. If you evaluate the alternatives, we think you'll agree you'll prefer AOS. But, even if you are conditioned to Reverse Polish Notation, the added value and power of TI's programmable calculators with full AOS is well worth the easy transition.

the three lives passed away unmarked and unknown."

"Eosine is the latest addition to the list of artificial dye-stuffs. The two substances that form the starting-point of its preparation, *phthalic acid* and *resorcine,* can be obtained directly from the coal-tar hydrocarbons. There can be no doubt that this is the finest red color that has yet been produced. At first it seemed doubtful whether it could be produced cheaply enough for industrial application, and the first specimens offered about a year ago cost more than $100 per pound, but by successive improvements in its manufacture it has been gradually cheapened, and it is already offered in the market at such a price that it can be used on silk and the finest classes of woollen fabrics."

"Mr. Carl Myers writes: 'So many inquiries have been made by persons evidently seeking truth regarding the possibilities or probabilities of spirit photography that I am tempted to contribute some of the results of my own experience in such researches for general information. In this field almost any desirable effect can be obtained through the modern system of retouching, masking and double printing. By such methods as these can be furnished by wholesale the floods of so-called spirit photographs forwarded in our mails in response to letters containing the customary fee. The facts of spiritualism may prove spirit photography, but such apparent productions do not prove spiritualism. My own position is that of an unprejudiced investigator who has discovered enough to make him cautious as to assertions while seeking to know more for science's sake. I seek conviction, not conversion, and in this position I offer no technical testimony in response to the clamors of spiritualists that science should investigate its phenomena and declare upon its merits.'"

"A rather neat swindle was recently perpetrated on a bank in Dallas, Texas, by three scamps who evidently possessed considerable knowledge of telegraphy. Scamp No. 1, in the character of a wealthy New York cotton buyer, presented himself at the bank with a check for $10,000 to be cashed. He brought strong letters of endorsement, and the check, which had been drawn by the bank's New York correspondent, appeared all regular. In order the more thoroughly to assure the bank the stranger required the officials to telegraph to New York for advice. An answer speedily came back, saying that both check and man were good, and the cash was paid. Scamps Nos. 2 and 3 had gone a few miles out of town, rigged a battery and the necessary instruments and tapped the wires of the telegraph line. When the bank's message was sent, they received it and sent back the false answer, thus assuring the bank's officers and of course victimizing the unfortunate institution."

Hubert Entrop, who achieved this superlative photograph of the deep-sky object M8 with his Questar 3½, helped us develop our smooth-tracking Starguide that now makes such photography possible for all Questar owners.

M8 was taken at Table Mountain in central Washington. It is a one hour twenty-five minute exposure at f/12.2. The high-speed printing process cannot reproduce here the delicate detail that is so beautiful on the original photograph.

A LETTER TO QUESTAR ON PERFORMANCE

Recently we received the following letter from Questar owner, Dick McCarrick, who lives in Arizona:

"It has been a little over a year since I received my Questar, so I thought I'd send you this note on its performance.

"My home is one of the poorest observing sites in the State. The city of Phoenix lies four miles to the west; Tempe, with 50,000 inhabitants, is two miles due south. Immediately to the north is Scottsdale, population 80,000, while eight miles to the east is Mesa, 70,000. You can then understand why artificial skyglow is such a problem, and when smog sets in the situation is much worse.

"Considering these handicaps the Questar has performed remarkably well. My favorite objects are deep-sky clusters and nebulae: despite the light sky I have managed to view forty Messier objects, with the dimmest being tenth-magnitude M100. The hours before dawn, when ground lights and smog are at a minimum, are the best time for this sort of observing.

"Whenever the opportunity arises, I take the Questar with me to observe in really dark skies. On one trout fishing trip I viewed M42 and was amazed to see the faint trailing nebulosity run off the field of view in the 24-mm. eyepiece. On another trip, to Mexico, I saw Omega Centauri in brilliant splendor.

"The moon and planets are usually good sights, even here. My favorite is Saturn. On one particularlly steady night I boosted the magnification to 400x with no loss of detail.

"To sum up, the Questar's obviously excellent optics, combined with its astounding portability (essential to me with such poor sky conditions here) make it the 'scope you claim it is. Keep up the fine craftsmanship."

Ever since we first brought the Questar telescope to the market, back in 1954, we have stressed the point that in anything less than perfect seeing conditions, Nature favors the small aperture, particularly when a set of optics is as fine as the hand of man can make it. The letters that have come to us over the years, even from those living in the glare, haze and smog of large cities, confirm this over and over. We think you would enjoy a photovisual Questar whatever your location, and for whatever purpose you wish to use it. In fact a Questar need never be idle; you can take it along with you for terrestrial viewing or solar observation by day, then turn it on the skies at night. With many people it is an inseparable companion.

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

QUESTAR

Box ML20, New Hope, PA 18938, Phone (215) 862-5277

THE AUTHORS

DOROTHY NELKIN ("The Science-Textbook Controversies") is associate professor at Cornell University, where she holds a joint appointment in the program on science, technology and society and in the department of city and regional planning. A 1954 graduate of Cornell, she has in recent years devoted her attention primarily to the study of "controversies that reflect the social impact of technology and in particular the problems of democratic values in an expertise-based society." Her investigations so far have resulted in numerous articles and books, on such topics as migrant labor, nuclear power and its critics, the politics of housing, the relation between the university and military research, the effects of methadone maintenance and the siting of jetports. This article, she notes, arose out of her current research on "anti-science movements as expressed in textbook controversies"—the subject of her next book.

JULIUS ADLER ("The Sensing of Chemicals by Bacteria") is professor of biochemistry and genetics at the University of Wisconsin at Madison. Born in Edelfingen in Germany, he came to the U.S. (Grand Forks, N.D.) in 1938 at the age of seven. He acquired his A.B. from Harvard College in 1952 and his Ph.D. in biochemistry from Wisconsin in 1957. He spent the next three years as a postdoctoral fellow at Washington University and Stanford University before returning to Wisconsin to join the faculty in 1960. He writes: "When I was a youth, my greatest interest was collecting butterflies. Now I prefer to watch their behavior instead. Nature remains my deepest fascination, and observing living things is my greatest source of inspiration for doing science."

GAETANO KANIZSA ("Subjective Contours") is professor of psychology at the University of Trieste. He received his doctorate in psychology from the University of Padua, where, he notes, he was "imbued with the European Gestalt tradition." Since then he has dedicated himself "to the phenomenology of visual perception and thought processes." He carried on his work in Florence and Milan before returning to his native city of Trieste in 1953 to take up his present post. In addition to his psychological research he is also "interested in the relation between science and the visual arts" and is "actively engaged in painting."

E. C. ZEEMAN ("Catastrophe Theory") is professor of mathematics at the University of Warwick, where he founded and directs the Mathematics Research Centre. Zeeman began his study of higher mathematics in 1947, after four years of service as a flying officer in the Royal Air Force. He went on to obtain his degrees from the University of Cambridge: a B.A. in 1948, an M.A. in 1950 and a Ph.D. in 1954. He taught at Cambridge for 10 years before moving to Warwick in 1964. Prior to 1970, he reports, his main interest was topology, a field to which he contributed some 50 papers. (His favorite dealt with "the unknotting of spheres in five dimensions. After struggling with the problem for seven years I suddenly managed to prove it one day with a 10-line proof.") Meanwhile, "as a hobby," he adds, "about 1960 I wrote three or four papers on the brain, which were not very good but had the virtue of turning René Thom's attention to biology and being partly responsible for his creation of catastrophe theory." For the past five or six years Zeeman's main energies have been devoted to developing the mathematics of catastrophe theory and exploring its applications "to all branches of science."

P. J. DARRAGH, A. J. GASKIN and J. V. SANDERS ("Opals") work in Australia for the Commonwealth Scientific and Industrial Research Organization (C.S.I.R.O.), Darragh and Gaskin in the Division of Mineralogy, headquartered in Perth, and Sanders in the Tribophysics Division at the University of Melbourne. According to Sanders, their collaboration on "the opal project" began because "we had a common interest in solving the problem, but each had different (although complementary) skills and knowledge." Darragh's current professional interest is industrial minerals, particularly paper-coating clays; nonetheless, he reports, he has maintained "an active interest in all gem minerals... and this has included work on diamonds, emeralds and sapphires." Gaskin, who is chief of the C.S.I.R.O. Division of Mineralogy and whose "main working interests include the origin of ore deposits and the nature and utilization of nonmetallic minerals," notes that ever since his student days at the University of Melbourne in the early 1940's he has "made spasmodic attempts to unravel the problem of the nature of precious opal, in the later stages with my colleague Darragh, but it was not until the collaboration with Sanders began that we understood what the stuff really is and how it does what it does." Sanders, who is chief research scientist in the Tribophysics Division, heads a group that is engaged in studying the reactivity of metallic surfaces, relying mainly on electron microscopy. His work on opals, which he characterizes as "sort of a hobby," arose out of "the possibility of applying ideas of diffraction as used for X-ray and electron crystallography to a system where visible light is diffracted."

EDWARD A. STERN ("The Analysis of Materials by X-Ray Absorption") is professor of physics at the University of Washington. He was educated at the California Institute of Technology, obtaining his B.S. in 1951 and his Ph.D. in 1955. Before moving to Seattle in 1965, he was professor of physics at the University of Maryland. In addition to his work in solid-state physics, he writes, "my main interest has been the responsibility of scientists to society as a whole and to maintaining their scientific rights.... In the past few years, with détente and the growth of scientific exchange between the U.S. and the U.S.S.R., I have been most active in encouraging contact with scientists in the U.S.S.R. and promoting their right to emigrate. I have kept in contact with those scientists who have lost their positions in the U.S.S.R. because of their desire to emigrate even though this right has been denied. I have helped to have several of their papers published in Western journals and was the major organizer outside the U.S.S.R. of their International Seminar on Collective Phenomena and the Application of Physics to Other Fields of Science, a meeting that was scheduled to be held in Moscow in 1974 but was prevented by the Soviet authorities from ever taking place."

STILLMAN DRAKE ("Galileo and the First Mechanical Computing Device") is already familiar to readers of SCIENTIFIC AMERICAN as the author or coauthor of three previous articles on aspects of Galileo's prolific curiosity: "Galileo's Discovery of the Law of Free Fall" (May, 1973), "Galileo's Discovery of the Parabolic Trajectory" (with James MacLachlan, March, 1975) and "The Role of Music in Galileo's Experiments" (June, 1975). Drake is professor of history at the Institute for the History and Philosophy of Science and Technology at the University of Toronto.

ANTHONY C. POOLEY and CARL GANS ("The Nile Crocodile") approach their subject from quite different directions. Pooley, who was responsible for most of the field observations reported in their article, works in Africa for the Natal Parks Board. He writes: "One of my form masters at junior school once told me that one day I would go far—but the farther from his school, the better. Taking his advice, I joined the Natal Parks Board in 1957 and was sent to Mkuzi Game Reserve and began watching crocodiles instead. In the mid-1960's I was able to establish a formal breeding station at the Ndumu Game Reserve in Zululand, and I am now installing a series of research and exhibit tanks at the St. Lucia Estuary Reserve in Natal. Besides the Nile crocodile I am now attempting to keep and breed members of the other species of African crocodilians. The behavioral work carried out there is intended to facilitate restocking of these animals in areas in which they have become hunted to extinction." Gans, who is an expert on the functional morphology of vertebrates, particularly reptiles, is professor of biological sciences at the University of Michigan. His interest in biomechanics, he says, provides him "with a nice mixture of field and laboratory research, the former currently centering on India and Sri Lanka, where I am studying the endemic burrowing snakes."

Vivitar Series 1™ 70-210mm f3.5 automatic zoom lens with macro focusing.

This is the original automatic zoom lens for 35mm SLR cameras that also provides macro focusing capability. A unique mechanical design permits both zooming and focusing with the same control. Although this lens has a large maximum aperture, it is surprisingly short and light. These seemingly contradictory features were reconciled by the extremely sophisticated computer programs with which the designers worked. Most significant of all in practical picture taking situations is the high performance, accomplished by a novel approach to zoom optical design. This was another computer suggested departure from conventional zoom lens configuration. The quality of performance thus produced was previously unavailable in zoom lenses and compares favorably to the finest original equipment telephoto lenses.

The versatility of macro focusing combined with a three-to-one telephoto zoom make this a remarkable all-purpose lens; particularly for the photographer who does not care to compromise on performance. A specially anodized finish on the lens mount and deeply checkered vinyl non-slip grips provide for absolutely positive focusing. The Series 1 70-210mm zoom, like the other Series 1 lenses, cannot be manufactured quickly in large numbers. For this reason, current demand for this lens somewhat exceeds supply. Put in your order, be patient, and your Vivitar dealer will get you one as quickly as possible.

Optical Specifications

Construction:	15 elements 10 groups
Angle of view:	11-32 degrees
Minimum focus distance:	Normal: From focal plane—6′6½″ Macro: From front of lens—3″
Maximum magnification in Macro:	1:2.2

Mechanical Specifications

Filter size:	67 millimeters
Weight:	31 ounces
Length:	6⅛ inches
Max. diameter:	3⅛ inches
F/number range:	f3.5-22 (EE coupled lenses to f16 only)

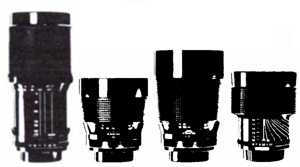
Vivitar Series 1 Program

In recent years, remarkable progress has been made in solving some of the classic problems of optical design. Intensive work in the field was spurred on by the demands of space exploration and military applications and vastly aided by the growing sophistication of computer technology.

Vivitar optical designers, working with programs devised for highly specialized optical tasks, have used computer generated designs to develop for Vivitar a new series of lenses capable of performance unreachable until now.

Each of the four Vivitar Series 1 lenses so far introduced has represented a breakthrough in the optical design of close focusing lenses. The 135mm f2.3 and the 200mm f3 are among the fastest automatic telephoto lenses in their respective focal lengths. They are unusually light and compact. Each has a uniquely positioned rear compensating element that automatically corrects aberrations at all points from the closest focusing point to infinity.

The 35-85mm f2.8 auto variable focusing lens is likewise a definite advance in lens design.



To come? Look for a 600mm f8 solid catadioptric lens of fantastically reduced size, a 90mm f2.5 macro, and a 28mm f1.9. Eventually there will be a complete optical system of more than twenty Vivitar Series 1 lenses, each representative of advanced technology and demonstrably superior to other lenses currently available.

Vivitar®

Marketed in the U.S.A. by
Ponder & Best, Inc.
Corporate Offices: 1630 Stewart St., Santa Monica, CA 90406.
In Canada: Precision Cameras of Canada, Ltd., Montreal.

Trademarks don't grow on trees.

An orange doesn't get to be a SUNKIST® orange just by being an orange.

It has to be the pick of the crop or we won't stamp it "SUNKIST."

SUNKIST is a trademark of Sunkist Growers, an association of citrus producers in California and Arizona. It's a trademark we're proud of. And it only gets stamped on the citrus we're proud of. The freshest, juiciest, and ripest.

So when you see the SUNKIST trademark, you know you're getting our best in fresh citrus and citrus products.

You have our word on it.

Sunkist®

Sunkist and [logo] are trademarks of Sunkist Growers, Inc. © 1976.

Established 1845 **SCIENTIFIC AMERICAN** April 1976 Volume 234 Number 4

The Science-Textbook Controversies

Over the past few years groups in various parts of the country have urged that special creation be given "equal time" with evolution. What accounts for these efforts and what are their future prospects?

by Dorothy Nelkin

In 1969 the California Board of Education issued new guidelines for the biology curriculum of the state's public schools. The guidelines included a statement that the Book of Genesis presents a reasonable explanation of the origin of life and that the concept of special creation should be taught as an alternative to the concept of organic evolution. It was only fair, it was asserted, that "equal time" should be given to the two concepts and that students should be allowed to choose between them.

Why is it that objection to the concept of evolution, dormant as a public issue since the Scopes trial of 1925, has gathered new impetus? How have comparatively small groups of people who believe in special creation been able to insist that their belief have a place in public education and even in the deliberations of science policy at the Federal level? What are the issues that have forced public recognition of concerns long ignored as being merely the private views of religious fundamentalists?

The California guidelines dramatically manifested the wide public interest in the values conveyed to students through the teaching of biology and the social sciences. The evolutionary concepts of modern biology are attacked by publications distributed in the millions by Jehovah's Witnesses and other missionary sects. The issue even reached the courts in 1972, when the religion editor of the *Washington Star,* who maintained that he was acting in the interest of 40 million evangelical Christians in the U.S., sued the National Science Foundation for using public funds to support education that violated religious beliefs. School boards and curriculum committees in many communities have been effectively prevented from recommending textbooks that discuss evolution.

The controversy over the science curriculum, which threatens to end a 20-year effort to modernize the precollege science curriculum in the public schools, has been bitter and filled with paradoxes. Although the evolutionary concepts of biology are among the most firmly established generalizations of modern science, public demands for the teaching of alternative explanations have exerted remarkable political influence. Although public education in science has fostered the widespread belief that natural phenomena can be rationally explained, there are vigorous efforts to reintroduce traditional religious explanations. Although the complexities of modern technology call for specialized knowledge, there is an increasing demand for participation by lay groups in the teaching of science on the grounds that education should reflect the values of the community. Perhaps the greatest paradox of all is that the critics of science textbooks do not represent the poorly educated or deprived sectors of the population only. Most textbook controversies issue not from rural folk in Appalachia but from middle-class citizens, many of whom are technically trained.

After the Scopes trial antievolutionary sentiment had a quiet but pervasive influence on the teaching of biology in the public schools. Textbooks for the most part avoided the topic: the word "evolution" or the name of Darwin can scarcely be found in books published in the 1930's. A national survey made in 1942 indicated that fewer than half of all high school biology teachers even mentioned evolution in their courses. As late as 1959, the centennial of the publication of Darwin's *Origin of Species,* the geneticist Hermann J. Muller could state that the teaching of biology in public schools was dominated by "antiquated religious traditions" and that it barely referred to modern research in population genetics and related areas that increasingly refined and supported the concept of evolution.

To help remedy the situation the National Science Foundation set up a program called the Biological Sciences Curriculum Study (BSCS) and gave it $7 million to create modern biology courses for the public schools. Three textbooks were developed, each emphasizing a different aspect of current biological research: molecular biology, cell biology and ecology. All three reflected the fact that modern biological research is based on evolutionary assumptions, which were described as "the warp and woof of modern biology."

The BSCS series was actually part of a larger National Science Foundation program that had been inaugurated in the late 1950's after the U.S.S.R. had launched its first Sputnik. The program was intended to build up the scientific and engineering manpower of the U.S. by bringing modern scientific concepts, methods and knowledge to the nation's public schools. The program as a whole had a considerable impact on precollege science education. A new curriculum in physics was devised by the Physical Science Study Committee (PSSC). It was soon followed by new curricula in mathematics, chemistry, biology (BSCS) and finally the social sciences. In all, the National Science Foundation financed 53 such projects.

Initially there was little problem with public acceptance of the new curricula in physics, chemistry and mathematics. A shadow of future problems appeared, however, when the BSCS biology series was introduced in 1964. That immediately touched off several intense, although short-lived, disputes. For example, in Texas critics of the curriculum declared that the teaching of such "atheistic" material was associated with the kind of "godless behavior" that had led to the assassination of President Kennedy the year before. The fact that the assassin fired from the Texas School Book Depository Building was not neglected by those who fought the adoption of the BSCS textbooks. All three books were finally approved in Texas, however, and the series was soon in service throughout the country.

The development of the social-science curriculum, titled *Man: A Course of Study* (MACOS), was launched in 1963 when a group of scholars from the Education Development Center, Inc. (then Educational Services, Inc.), of Cambridge, Mass., received a grant from the National Science Foundation to develop an integrated program of precollege social-science courses. Until MACOS was introduced the teaching of social science in the public schools had consisted mostly of descriptive presentations of American history. MACOS, designed for children in the fifth and sixth grades, asks three questions: What is human about human beings? How did they get that way? How can they become more so?

The MACOS curriculum relies on studies of animal behavior and of the culture of the Netsilik Eskimos to explore questions about the nature of human beings, patterns of social interaction and child rearing, and the development of a culture's total view of the world. To the social scientists who worked on the MACOS curriculum the study of animal behavior provided a provocative metaphor to illuminate features of human behavior. The study of a traditional tribal culture showed how human beings as well as animals adapt to a particular environment; in order for the Netsilik to survive in an environment with limited food resources they practice infanticide and senilicide as means of controlling the population. MACOS suggested that in some societies such practices, disturbing as they would be in our own culture, were functional, and that neither behavior nor beliefs have an absolute value apart from their social and physical context.

It was difficult to find a publisher for a curriculum that was innovative and costly and required special training for teachers. Accordingly the National Science Foundation provided credit to back the publication of MACOS, and it set up workshops for training the teachers who would be adopting the program. In all, the National Science Foundation granted $4.8 million to develop MACOS and $2.16 million to implement it. By 1974 the MACOS curriculum, widely praised as an imaginative contribution to education, had found its place in some 1,700 schools in 47 states.

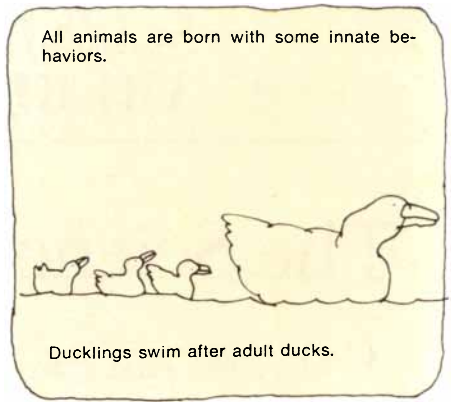
Elementary education in the public schools has always been one of the more volatile areas of public policy. Any effort at innovation can count on provoking conflict for two reasons: first, parents are persistently anxious about what influences their offspring, and second, parents and teachers alike are ambivalent about the proper role of public education in disseminating values as well as knowledge. An army of "textbook-watchers" has long been on guard for changes in textbooks that threaten various concepts of basic education, public morality or patriotism. More recently textbook-watchers have also been concerned with eliminating ethnic and sex bias. With the notable exception of the Scopes trial, however, there had been relatively few disputes over the teaching of science. Science was generally perceived as being morally neutral and associated with material progress. Thus it was dissociated from the questions of values that concerned the textbook-watchers.

By the late 1960's attitudes appeared to shift. The change became evident in the growing criticism of scientific rationality and in the proliferation of cults and sects based on Eastern mysticism. Less visible, but perhaps more important in the light of subsequent events, was a remarkable growth in the membership of fundamentalist churches, particularly in urban Texas and southern California—the very centers of industry based on high technology.

In those areas some citizens were disillusioned with what they called the "decadence of scientism" and with political authority that seemed to remove their sense of local power. They expressed their resentment by attacking science courses in the local schools. In fact, they formed the financial, social and political base for a nationwide movement to challenge science textbooks. Their answer to the uncertainties of a technological society was not to reject technology but to return to fundamentalist religion and traditional beliefs. It was in this atmosphere that the National Science Foundation precollege curricula in biology and the social sciences became the focus of extended and bitter controversy.

The activists at the core of the antievolution movement are the "scientific creationists," people with degrees in science who work out of "creation research centers." They maintain that they are scientists who are engaged not in a controversy between religion and science but in a debate about the validity of two scientific theories. Their organizations and activities are patterned on those of organized science. Titles and credentials are offered as proof of their legitimacy. Research projects are designed to examine their hypotheses. Journals and textbooks disseminate their findings. They believe "all basic types of living things, including man, were made by direct creative acts of God during the creation week" and they seek to reinterpret the evolution of organisms according to biblical authority.

It was creationists of this kind who in 1969 convinced the California Board of Education that creation and evolution should be taught as alternative theories. Bills with similar provisions were introduced in state legislatures throughout the country. In Tennessee legislation requiring equal time for the concept of special creation was passed in 1973. The California creationists eventually failed, however, to implement the teaching of the creation concept in the public schools, and even in Tennessee the law requiring equal time for the creation concept was declared unconstitutional and was repealed in 1975. The creationist movement has nonetheless retained a strong base of support among people who think that their traditional values are in some way threatened by the rational explanation of natural phenomena.



STUDIES OF ANIMAL BEHAVIOR are used to illuminate aspects of human behavior

The strength of that fear became apparent in the reaction against MACOS. The controversy involved many of the same people who had earlier fought the BSCS series of biology textbooks. Initially the MACOS materials sold well and were enthusiastically received in many communities. The only early resistance was a protest in Florida in 1970 and another a year later in Phoenix, Ariz., where the superintendent of schools banned the materials throughout the state.

In 1973, however, a rash of disputes broke out across the nation. MACOS suddenly became a symbol for local frustrations. The specific nature of the protests varied from one region to another. In university communities strains in the "town-gown" relationship became apparent when groups of local citizens leveled protests at "those experts on the campus who try to determine our values." In the South religious issues were predominant. In small towns there was criticism of the message of cultural relativism in MACOS suggesting that beliefs and values unacceptable in our culture might be acceptable in others. In urban centers a major concern was that such "irreligious teaching" would cause "immoral behavior."

These seemingly isolated disputes were linked by a network of communications. The same activists showed up in many different communities disputing MACOS, and they repeatedly issued identical documentation of the problems the course allegedly presented. Several organizations based in Washington provided legal assistance and circulated reports and out-of-context inflammatory material from the course to arouse parental concern. By 1975 the sales of MACOS had declined by 70 percent.

The moral content of MACOS caused some to question the propriety of Federal financing of a controversial curriculum. Soon the controversy became a national issue. Representative John B. Conlan of Arizona carried the issue to the House Committee on Science and Technology. He found little support for his suggestion that

in the National Science Foundation curriculum *Man: A Course of Study* (MACOS). Four panels are reproduced from the booklet *Innate and Learned Behavior,* which shows how behavior of all animals, including human beings, is a mixture of both instinct and learning.

Congress censor the curriculum, but he did find that his colleagues sought some control over "unaccountable executive bureaucracies" such as the National Science Foundation. Congressional attitudes had been primed in part by Senator William Proxmire's criticism of "those damn fool projects in the behavioral sciences," by the perennial concern over the spending of public funds and by the post-Watergate mistrust of executive agencies.

Last April, Conlan proposed to the entire House of Representatives that the bill authorizing the National Science Foundation to finance the development of science curricula for public schools be amended to require Congressional approval for specific science-curriculum projects. The amendment was defeated by a vote of only 215 to 196. A different amendment was passed, requiring that all material pioneered by the foundation (including the guidelines for training the teachers) be open to parental inspection. Funds for MACOS, however, were terminated, and further support of science-curriculum projects was suspended pending a review of the entire National Science Foundation educational program.

Several review committees were formed, including a group within the foundation and the Science Curriculum Implementation Review Group, under the chairmanship of J. M. Moudy, chancellor of Texas Christian University. The Moudy group was appointed by Representative Olin E. Teague of Texas, chairman of the House Committee on Science and Technology. The committee of the National Science Foundation recommended that the system for reviewing specific science-curriculum projects before they are implemented in the public schools be tightened and that the review system include laymen such as parents and community leaders. It further recommended that the National Science Foundation itself remain "at arm's length" from the implementation process.

The Moudy study group represented a spectrum of opinion on the issue and agreed on only a few recommendations. Nevertheless, it recommended that the National Science Foundation continue to fund the implementation of MACOS, provided that initiative for implementation projects came from local institutions. Such separation would help to shield local communities against being unduly influenced by the Federal Government. The Moudy study group also agreed that laymen should be included in the review system. Specifically it suggested that representative parents who are "innocent of professional and scholarly bias" be involved in decisions regarding curricula that impinge on widely observed customs and long-held religious beliefs.

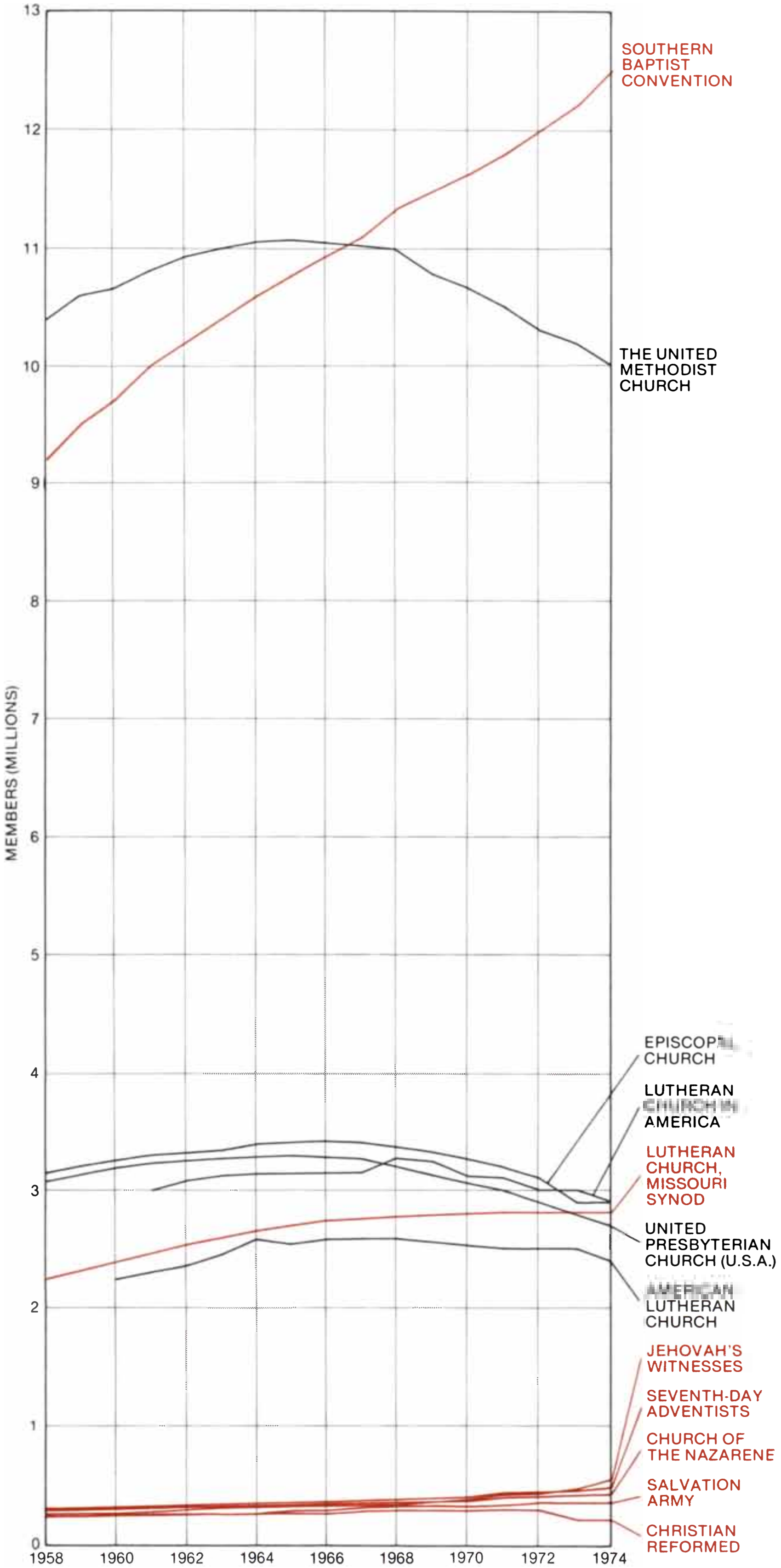
As the Moudy study group was considering whether or not the National Science Foundation's system of reviewing educational materials was adequate, Representative Conlan extended his criticism of the review system to include the foundation's other policies. His criticism helped to stimulate a Congressional examination of the foundation's peer-review system of evaluating proposals for grants for scientific research, and also a call for more public accountability for funds expended on such research.

The National Science Foundation is faced with the task of how to evaluate and monitor its educational projects, a task that presents many dilemmas. If the foundation finds it cannot approve a course of study for dissemination to the public, it is open to accusations of censorship. If it does approve a course for dissemination and parents and community leaders object to it, then it is even more open to accusations of trying to impose Federal standards on education.

It is easy to label those who question the validity and limits of modern science as ignorant, irrational or crackpot. Those labels throw no light on the social and political tensions that sustain objections to the teaching of science in the public schools. Three themes pervade the science-textbook controversies. First, the protests reflect the fact that a non-negligible fraction of the population is disillusioned with science and is concerned that it threatens traditional religious and moral values. Second, the protests reflect the fact that many people clearly resent the authority represented by scientific dogmatism, particularly when that authority is expressed in an increased professionalism of the school science curriculum. Third, the protests reflect the fact that many people are afraid that the structured, meritocratic processes operating within science threaten more egalitarian, pluralistic values.

Let us examine those three themes in more detail. First, for many fundamentalists the social disruptions of the 1960's are evidence that the quality of American values is declining. These people associate what they perceive as an expanding immorality with the dominance of "liberal" secular and scientific values, and they blame the decline of religion on the rise of a technological society. They do not react by opposing technology, however. They focus instead on the uncertainties and disruptions that they feel characterize modern life as a result of secularism and scientific rationality. They seek to alleviate these uncertainties and disruptions by turning toward a traditional and fundamentalist religious outlook. The creationist who wrote the California guidelines proposing equal time for the theory of the creation in biology textbooks associates the concept of evolution with "a campaign of secularization in a scientific-materialistic society—a campaign to totally neutralize religious convictions, to destroy any concept of absolute moral values, to deny any racial difference, to mix all ethnic groups in cookbook proportions, and finally [to destroy] the differences between male and female."

Fundamentalist textbook critics are particularly distressed by the teaching of modern biology and social science because of the emphasis on the similarities between man and other animals, because of the implication that moral values are relative and because of the denial that an omnipotent and omniscient force determines human development and behavior. They argue that emphasizing the genetic similarities between human beings and other animals may encourage "animal-like," socially dangerous behavior. One creationist stated: "If man is an evolved animal, then the morals of the



GROWTH OF MEMBERSHIP for 12 Protestant groups is plotted for the years from 1958 through 1974. Membership of groups with a fundamentalist or related viewpoint (*color*) has increased steadily, while membership of nonfundamentalist groups (*black*) peaked in the mid-1960's and has declined since then. Membership data are from National Council of Churches.

barnyard or jungle are more natural . . . than the artificially imposed restrictions of premarital chastity and marital fidelity. Instead of monogamy, why not promiscuity and polygamy? . . . Self-preservation is the first law of nature; only the fittest will survive. Be the cock-of-the-walk and the king-of-the-mountain. Eat, drink and be merry, for life is short and that's the end. So says evolution." One woman even blamed the "streaking" fad of 1974 on the concept of evolution. "If young people are taught they are animals long enough, they'll soon begin to act like them."

Many of the disturbing values that are implicit in the concept of evolution became explicit in MACOS. Some critics asked: Why should a discussion of man concentrate in such detail on animals? Others argued that the curriculum was a pernicious attempt to spread the religion of secular humanism, an anthropocentric view of the world that emphasizes man's capacity to achieve self-realization through reason and denies the importance of God and a spiritual order. Critics were particularly disturbed by the MACOS material on the adaptation of the Netsilik Eskimos to their harsh environment. MACOS suggested that values and behavior were based not on God's law but on specific environmental pressures. To some critics this point of view implied that scientists were expounding a philosophy of relativism that denies absolute standards. Such relativistic assumptions, it was thought, were morally destructive and harmful to family life. One critic remarked: "Already we condone feticide and abortion and devalue human life, all in the name of biological principles." To strengthen their case some textbook-watchers cite examples in which science has been misused, thus associating themselves with widely held concerns about the social impact of modern science and technology.

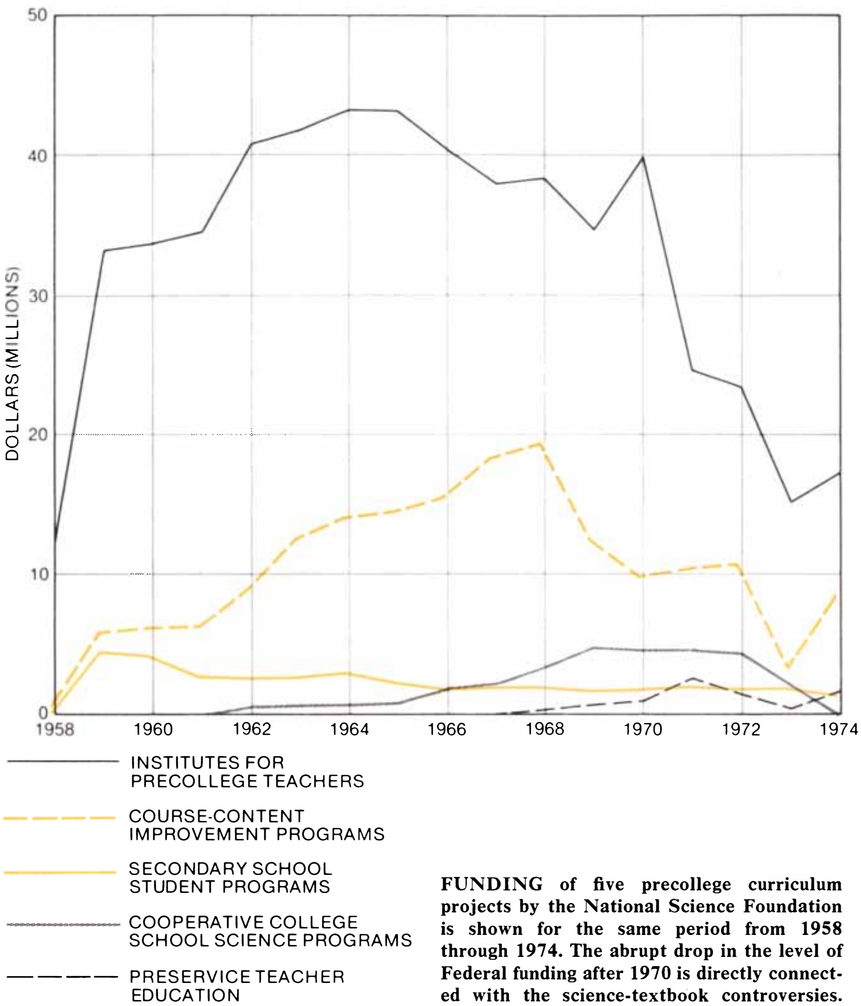
The second theme in the textbook controversies is the resentment of the role played by scientists and other professionals in the development of science curricula for the public schools. Representative Conlan declared: "An elite corps of unelected professional academics and their government friends run things in the schools." Creationists spoke of arrogance and the absence of humility among scientists. One sympathetic journalist wrote of his joy in "seeing science humbled" and of witnessing a break in the "monopoly of truth." A creationist observed: "After all, scientists put on their trousers in the morning one leg at a time, just like the rest of the world." And a Jehovah's Witness wrote about the "arrogant authoritarianism required by evolutionists to sustain what they cannot prove." Such concerns are not, of course, unique to textbook-watchers; resentment against professional authority is expressed at nearly every public confrontation between experts and laymen.

In California the attack on scientific authority resulted in a number of changes in textbooks that were intended to remove dogmatic statements and to indicate the limits of scientific explanation. The creationists also attacked the National Science Foundation for its support of the BSCS materials. The theme of "a Federal takeover of education" emerged in full force during the MACOS disputes and during the 1975 Congressional hearings on the annual appropriation for the National Science Foundation. At that time Representative Conlan found that one way to arouse interest in MACOS among his Congressional colleagues was to describe the curriculum as an "insidious attempt to impose particular school courses ... on local school districts, using the power and financial resources of the Federal Government to set up a network of educator lobbyists to control education throughout America."

Textbook disputes are organized around demands for the increased participation of laymen in decisions about the school curriculum. The demands for local control have been stimulated by the decade-old ruling of the U.S. Supreme Court to the effect that decisions on obscenity and pornography were to be made locally. Textbook-watchers, who had been concerned with "smut in the schools," had watched the obscenity case with interest. They have used the Supreme Court ruling to back their argument that any issue impinging on local values, including how science is to be taught, should be judged within the community. One organizer in Texas who is active in anti-MACOS and pro-creationist protests throughout the country summarized the prevailing philosophy of the movement: "Unless people take an active voice in assisting the authorized units of government in a process of selecting textbooks, the selection will continue to deteriorate."

The public education system has been one of the last grass-roots institutions in America. School systems were traditionally decentralized, run by local school boards composed of elected nonprofessional citizens. Local control has gradually been eroded, however, through court decisions, through the reliance of schools on nonlocal funds, through the merging of school districts and through the trend toward professionalism. The curriculum in a public school today is guided more by national testing standards and college entrance requirements than by local values. The influence of individuals who expect to retain local control over educational policy is threatened by the growing power of statewide curriculum committees and by the increasingly professional departments of education. Federally funded programs devised by professionals are an additional threat.

The decline of local control has been a major source of tension in the controversies over ethnic balance and the busing of students as well as in the controversies over curricula and textbooks. Hence when textbook-watchers reject professional control of the curricula that influence their children, they can count on support from diverse



FUNDING of five precollege curriculum projects by the National Science Foundation is shown for the same period from 1958 through 1974. The abrupt drop in the level of Federal funding after 1970 is directly connected with the science-textbook controversies.

groups ranging from the "radical right" to the "new left," all concerned with Government infringement on local powers and all sympathetic to any movement that seeks to increase individual political activity.

Demands for increased nonprofessional participation in curriculum decisions raise difficult practical questions. Does the state have a responsibility to require that certain textbooks be used in all classes? What kinds of constraints or standards can be imposed to balance academic interests with local and individual religious concerns? What happens if local values deny students access to widely accepted knowledge and wisdom?

From the professional perspective designing a school curriculum is a technical enterprise that is best organized by experts, so that the curricula will provide the student with the best available information. From a local perspective, however, public education also transmits values and beliefs. Since such values and beliefs are very much family matters, parents must be involved. Clearly science education is no longer exempt from that perspective.

The third theme in the textbook controversies is a defensiveness toward the egalitarian, pluralistic processes that operate outside science, processes that lean heavily on the notion that there are two sides to every question and that each side should be treated fairly and given equal time.

In January, 1973, Henry M. Morris, Jr., the director of the Institute for Creation Research, wrote to the director of the BSCS project and challenged him to a public debate. The debate would take up the statement: "Resolved, that the special-creation model of the history of the earth and its inhabitants is more effective in the correlation and prediction of scientific data than is the evolution model." Morris proposed that the winner of the debate would be determined by the applause from the audience. According to Morris, the issue was one of free public choice, of equality and of fairness. Creationists argue that since the biblical account of the origin of man and other living things is scientifically valid, it deserves equal time with the concept of evolution. Whenever there are two equally valid hypotheses, it is only fair that students be exposed to both of them and be allowed to choose for themselves.

The modern concept of equal time originated with the Fairness Doctrine of the Federal Communications Commission, which states that broadcasters are responsible for affording a "reasonable opportunity for the presentation of contrasting view-

Original Version	Changed Version
CHANGES IN DEFINITIONS	
Science is the total knowledge of facts and principles that govern our lives, the world, and everything in it, and the universe of which the world is just a part.	Science is one way of discovering and interpreting the facts and principles that govern our lives, the world and everything in it, and the universe of which the world is just a part. The scientific way limits itself to natural causes and to descriptions that can be contradicted, at least in principle, by experimental investigation.
Evolution is a central explanatory hypothesis in the biological sciences. Students who have taken a biology course without learning about evolution probably have not been adequately or honestly educated.	Evolution is a central explanatory hypothesis in the biological sciences. Therefore, students need some knowledge of its assumptions and basis concepts.
QUALIFICATIONS TO "REDUCE DOGMATISM"	
Scientists can reconstruct the (prehistoric) animal.	Scientists do their best to reconstruct the (prehistoric) animal.
Scientists believe that these species were ancestors...	According to the evolutionary view, these species were ancestors.
A short description ...	A short approximate description ...
Modern animals that are descendants...	Modern animals that seem to be direct descendants.
Some fish began to change.	Some fish began to change, although we don't know why.
The earth had spore-bearing plants long before the first seed plants appeared.	There is evidence that the earth had spore-bearing green plants long before the first seed plants appeared.
Paleontologists...have reconstructed past history...	Paleontologists...have done their best to reconstruct the past history...
How do we know...	On what basis has it been concluded that...
The evidence that shows how...	Evidence that is often interpreted to mean...
They would have to change.	They would have to be different.
Fishes adapted...	Fishes were adapted...
Paleontologists have been able to date the geological history of North America.	Paleontologists have assembled a tentative outline of the geological history of North America.
CHANGES TO AVOID EVOLUTIONARY ASSUMPTIONS	
Slowly, over millions of years, the dinosaurs died out.	Slowly, the dinosaurs died out.
As reptiles evolved from fishlike ancestors, they developed a thicker scaly surface.	If reptiles evolved from fishlike ancestors, as proposed in the theory of evolution, they must have developed a thick scaly surface.
Many scientists believe that the universe had a beginning similar to that of a snow fort. They believe that the stars and the galaxies of the entire universe in the beginning were in the form of very small scattered particles.	Science, by definition, cannot say anything about where the first matter and energy of which our universe is made came from. That is because there cannot be any science without matter and energy to deal with. When scientists speak of the beginning of the universe, therefore they mean the first interactions of matter and energy. Many scientists believe that these first interactions were like those in making a snow fort...
Shortly after the flying reptiles took to the air, the early birds developed.	Birds appear in the fossil record shortly after flying reptiles.
The constant rate at which radioactive elements give off particles enables scientists to determine how long it will take for one ball of any sample of a radioactive element to form another element.	Scientists know that radioactive elements give off particles today at a constant rate. If they assume that this rate has remained constant back in time to the date of interest, they can determine how long a...
Scientists believe life may have begun from amino acids or viruses, neither of which is usually considered living. Scientists believe life may have been transported from another planet.	Scientists do not know how life began on earth. Some suggest that life began from non-living material. Others suggest that life may have been transported...
Plants took to the land and conquered it.	Plants appeared on the land.

SOME OF THE CHANGES IN PASSAGES in high school biology textbooks that were accepted by the California Board of Education are compared with the original passages. The changes were intended to reduce "scientific dogmatism" and to eliminate "evolutionary assumptions." The author obtained the examples from a manuscript notebook belonging to the Board of Education that contained single pages from a number of unidentified textbooks with the changes typed on slips of paper and then superposed on the original passages.

points on controversial issues of public importance." The Fairness Doctrine was based on the rationale that television, as a costly and monopolistic medium, did not sufficiently allow dissenting viewpoints to be expressed. In practice the doctrine has been extraordinarily difficult to define and to implement on television. In principle, however, it has had an enormous appeal in American society because it is equated with fairness and justice and is rooted in the democratic attitude.

The concept of equal time has been particularly influential in American education. Over the past decade textbooks have been challenged by minority groups and by women who see the presentation of history as ethnically and sexually biased and who seek to have their interests more fairly represented in the educational process. The concept of equal time, of pluralism, implies that all groups have a right to maintain their cultural and religious traditions in the face of pressures for conformity. The textbook-watchers have extended the concept to the teaching of science; they believe the struggle against cultural conformity is paralleled by their struggle against scientific conformity.

The theme of equal time appeared in the disputes over MACOS in the form of demands that diverse values be represented in social-science courses. MACOS, it was argued, presented a one-sided view. The attorney for groups protesting against MACOS in Kanawha County, W.Va., in 1975 commented: "Under the banner of science, value systems are being marched into the schoolroom with a shameless disregard of the will of the polity.... The truth dawning on parents—be they creekers and red-necks from West Virginia or Gold Coasters from Connecticut—is that education is a sectarian occupation."

One report on MACOS from a conservative organization in Washington that provides legal counsel for textbook controversies specifically asserted that science education should either avoid all questions of values or provide equal time for the presentation of alternative values. The Moudy study group said that MACOS implicitly promotes "an evolutionary and relative humanism" because it does not allow any "hints of patriotism, theism, creationism or other explicit values." An aide to Representative Conlan added: "Diverse facts should be taught fairly and the student allowed to make up his own mind."

The suggestion that questions of scientific fact and scientific education should be settled by public debate has left most scientists amazed. Would the lay community really want to give quack doctors equal time with licensed doctors? To include the doctrines of Christian Science in books on health? To include astrological lore in books on astronomy? For that matter, would the community entertain putting a paragraph in the Book of Genesis to indicate that the scientific method rejects supernatural explanations of the universe?

Concepts of pluralism, of equity and of participatory democracy, as they are defined in the political context, are incongruous in science. Scientific concepts are taught when they are generally accepted by the scientific community. In fact, it is precisely that acceptance by the scientific community that acts to validate one concept and reject another; acceptance by those outside the scientific community is irrelevant. Moreover, although science is an open system in terms of social class, the reputation of a scientist depends on his achievements and their rigorous evaluation by his peers. Indeed, the internal standards of science may run counter to egalitarian principles. Hence scientists are particularly distressed by the proposal that laymen participate in defining the nature of an appropriate education in science.

The suggestion that egalitarian principles be extended to science indicates that the character and the content of science, as they are understood by scientists, are perceived quite differently by many laymen. Conversely, when scientists dismiss the critics of science as simply being "irrational," they fail to grasp the differences between the structured, meritocratic processes that operate within science and the more egalitarian, pluralistic processes that operate outside it.

The differences between the attitudes of scientists and those of nonscientists naturally hinder efforts to communicate scientific information to the public effectively. Indeed, the way in which such information is communicated may partly explain why public resistance to the authority of science is so persistent. Where scientists themselves understand that their work is approximate, conditional and open to critical examination, many nonscientists believe science is authoritative, exact and definitive. The organized skepticism toward scientific findings that is tacitly understood by those who practice science contrasts sharply with the external view of science.

Perhaps the most difficult concept for scientists to convey to those who are not scientists is the delicate balance between certainty and doubt that is so essential to the scientific spirit. Textbooks in particular tend to convey a message of certainty to the nonspecialist. In the process of simplifying concepts, findings may become explanations, explanations may become axioms and tentative judgments may become definitive conclusions. Few textbooks are careful to stress the distinction between fact and interpretation or to suggest that intuition and speculation actually guide the development of scientific concepts.

Authoritarian public representations of science are reinforced by scientists who deeply desire to avoid challenge and criticism from people outside their own profession. They tend to respond to criticism with a kind of scientific fundamentalism: by citing the value-free character of their work or the weight of the factual evidence that supports their conclusions. To those whose religious faith is challenged, however, the scientific merits of a concept that defines man's place in the universe may be less to the point than the concept's social and moral implications. The concept of evolution, after all, has had a remarkable social impact. It was described as "a secure basis for ethics" by C. H. Waddington. It was called a naturalistic religion by Julian Huxley. It was a justification of laissez faire economics for the 19th-century entrepreneurs who called the growth of large companies "the survival of the fittest." The moral implications that can be drawn from the concept of evolution and the threat it presents to absolute ethical values are clearly far more important to many laymen than the details of the concept's scientific verification.

Those details can, in fact, easily be ignored or rationalized by such laymen. The persistence of creationism and the controversy over MACOS are reminders that many people are reluctant to surrender their deepest convictions to the authority of science, and that intuitive expectations and beliefs can easily take precedence over scientific explanations. Furthermore, simply increasing the amount of available technical information is unlikely to change well-rooted beliefs. Selection operates within the mind of the individual to determine the way in which such information is interpreted, particularly when the information is not fully understood.

Nevertheless, it is not accurate to dismiss the critics of science textbooks as being merely an antiscience fringe group. Creationism is an unlikely combination of religion and science in which theological beliefs are conveyed in a context of research monographs and scientific societies. And most of the people who have been working against MACOS do not deny that science is a useful activity. They object primarily to an impersonal educational bureaucracy that fails to represent their interests and that insults their personal beliefs. They are not reacting against science so much as resisting its image as an infallible source of truth that denies their sense of place in the universe.

In these respects the textbook critics are part of the romantic resistance to science that is reflected in the popularity of astrology, mystical cults and the imaginary cosmologies of Immanuel Velikovsky and Erich Von Daniken. They are also part of a political resistance to science that is reflected in increased social action against innovation and in the demands for lay participation in scientific and technical decisions. As questions that are normally resolved by professional consensus are brought into the political arena, and as democratic values such as freedom of choice, equality and fairness enter into science policy, the consequences of such resistance to science may be painful.

The Sensing of Chemicals by Bacteria

Motile bacteria are capable of chemotaxis, or movement toward some chemicals and away from others. Chemosensors in the cell envelope can detect the chemicals and signal the flagella of the cell to respond

by Julius Adler

Of all the attributes of living things the one that most fascinates me is the ability of organisms to detect a stimulus and respond to it by moving—what is called irritability or, in a precise sense of the word, behavior. Such behavior is a universal property of biological systems: plants grow toward light, the Venus's-flytrap closes at a touch, butterflies congregate around a mud puddle, children are attracted to the kitchen by the smell of good food. And down at the elemental level of the single cell the presence of life is evinced by the capacity for the same kind of behavior: motile bacteria are attracted by certain chemicals and are repelled by other chemicals. We are beginning to learn something about how bacteria detect such stimuli and how they move in response to them. The bacterial version of a "nervous system" that links detection and response, however, remains a mystery.

The capacity of bacteria for positive and negative chemotaxis, as such behavior is called, has been known to biologists for about a century. Indeed, it was intensively investigated from about 1880 until about 1920, after which, for some reason, it was studied very little and was apparently all but forgotten. Some 15 years ago I came across descriptions of bacterial chemotaxis as I was searching for a simple system in which to study the elements of behavior in the hope some underlying principles might be discovered that are applicable at many different levels of biological complexity. Clearly bacteria would be advantageous subjects. They are the simplest organisms of all, and in them detection and response take place within a single cell. Much is known about the genetics of bacteria, and genetic experiments are effective for analyzing behavior, all the more so when large numbers of identical organisms can be grown quickly, as they can in the case of bacteria. The biochemical conversions that bacteria carry out are well known. Finally, information gained from bacterial studies of molecular genetics and of both the breakdown of nutrients and the synthesis of cellular components has been applicable in higher organisms. The same may be true of behavior.

In my laboratory at the University of Wisconsin I chose to study the intestinal bacterium *Escherichia coli,* whose genetics and biochemistry are particularly well known. *E. coli* cells are about a micrometer (a thousandth of a millimeter) wide and twice as long. They propel themselves at as much as 20 micrometers a second by means of four to eight flagella: thin helical appendages some seven micrometers long that arise at various places on the cell surface. My first task was to convince myself and others, by objective methods, that *E. coli* are in fact attracted and repelled by various compounds.

In the 1880's the German botanist Wilhelm Pfeffer had shown that if a thin capillary tube filled with an attractant substance is dipped into a suspension of bacteria, within a few minutes the bacteria congregate near the mouth of the capillary as a result of the diffusion of the chemical out of the tube. A little later, moving up the gradient of the attractant toward the region of greater concentration inside the tube, they accumulate within the capillary itself. In order to convert Pfeffer's observation into an objective, quantitative assay it was necessary only to plate the contents of the capillary on a culture medium and count the visible colonies of proliferating cells that arose overnight. It was easy to show that many thousands of bacteria accumulate inside a capillary that contains an attractant, whereas only a few hundred cells happen to enter a capillary in which there is no attractant.

Another method I developed to demonstrate positive chemotaxis is to put bacteria at the center of a plate of agar medium (at a concentration low enough to allow them to swim) in which there is an attractant substance. Soon the bacteria have used up the local supply of attractant, and then they follow the attractant gradient they themselves have produced. At the same time they multiply. The result is an ever expanding ring of bacteria, easily visible to the unaided eye, that marks the boundary between the region that has been depleted of attractant and the region still rich in attractant.

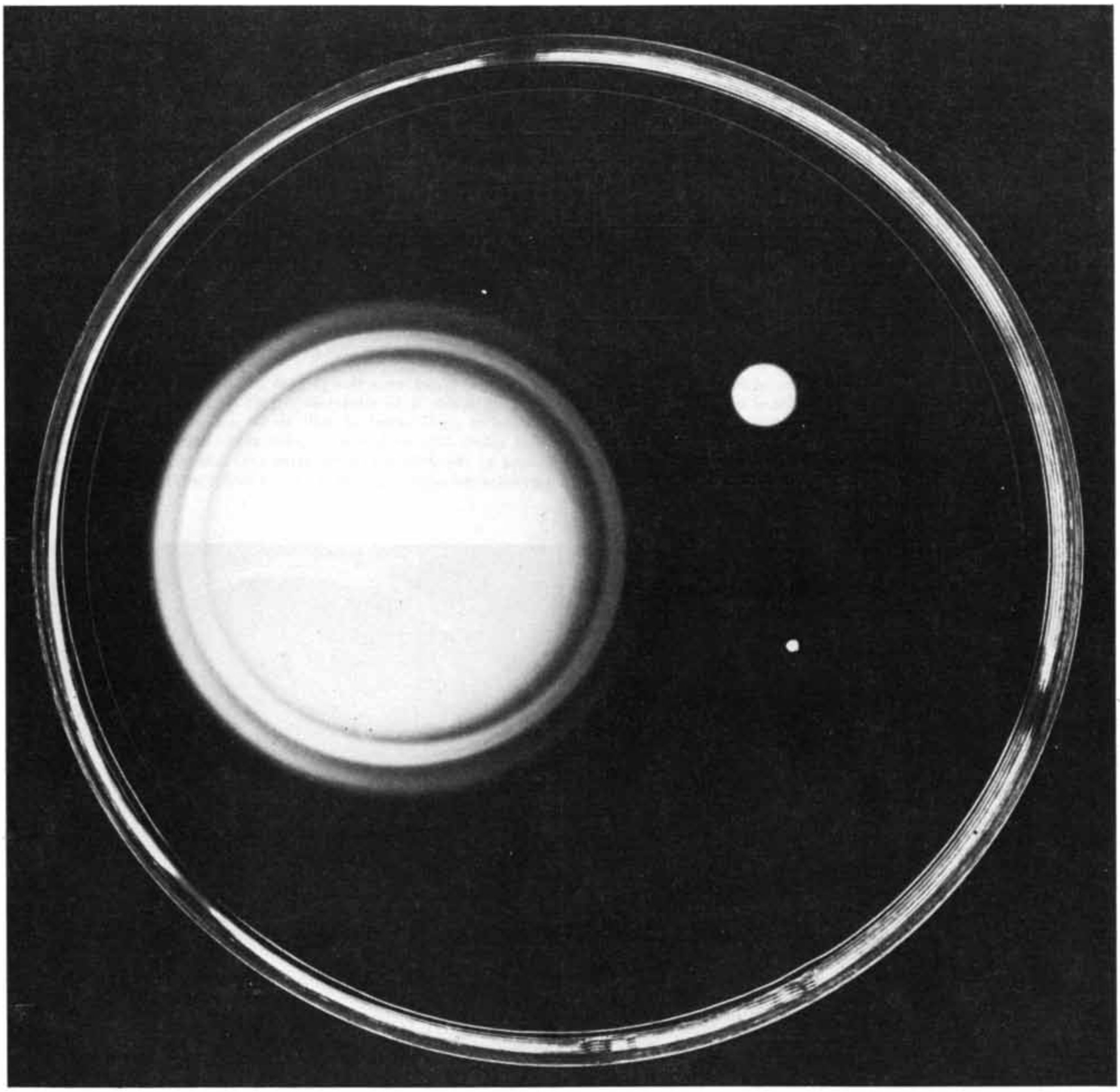
Wung-Wai Tso and I developed a method for measuring negative chemotaxis. We place a capillary in which there is no repellent in a bacterial suspension containing repellent. The bacteria swim into the capillary for refuge and can be counted by letting them grow into visible colonies on a plate of agar. Negative chemotaxis can also be demonstrated, we found, by placing a source of repellent in a plate containing agar (dilute enough for the bacteria to swim in) and a suspension of bacteria concentrated enough to be visible to the unaided eye. Soon the bacteria vacate the region around the source of repellent.

These procedures made it clear that *E. coli* can respond chemotactically to a large number of attractants, most of which are nutritious sugars and amino acids. The bacteria also move away from a large number of repellents, which are mostly harmful substances or bacterial excretory products such as acids and alcohols. The question now became: How do they do it?

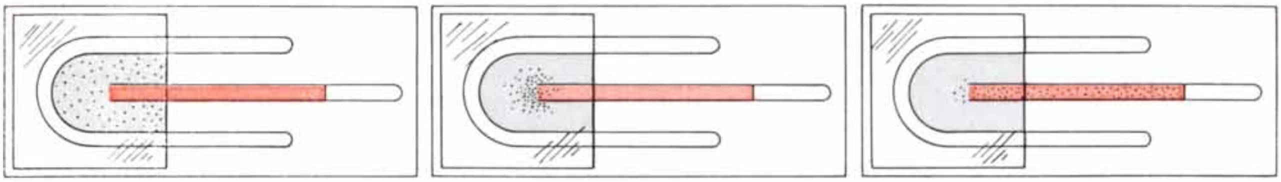
One of modern biology's most powerful methods for determining the mechanism of a process is genetic analysis. The investigator isolates mutant organisms that are unable to conduct some activity or metabolize some substance. If there are several steps in the process, there will be several different classes of mutants, each unable to accomplish one of the steps because the gene governing that step is defective. In this way any process can be dissected into its component steps. In the case of bacterial chemotaxis, mutants unable to carry out the behavior can be found by looking for cells that remain at the center of a positive-chemotaxis plate or in the vacated region around the source of repellent in a negative-chemotaxis plate.

In 1966 John B. Armstrong and I reported the first isolation of a bacterial behavioral mutant. Behavioral mutants have also been described by others for phototaxis, or attraction by light, in the fruit fly *Drosophila* [see "Genetic Dissection of Behavior," by Seymour Benzer; SCIENTIFIC AMERICAN, December, 1973]; for phototropism, or growth toward the light, in the mold *Phycomyces;* for swimming in the protozoon *Paramecium,* and for swimming and chemotaxis in nematode worms.

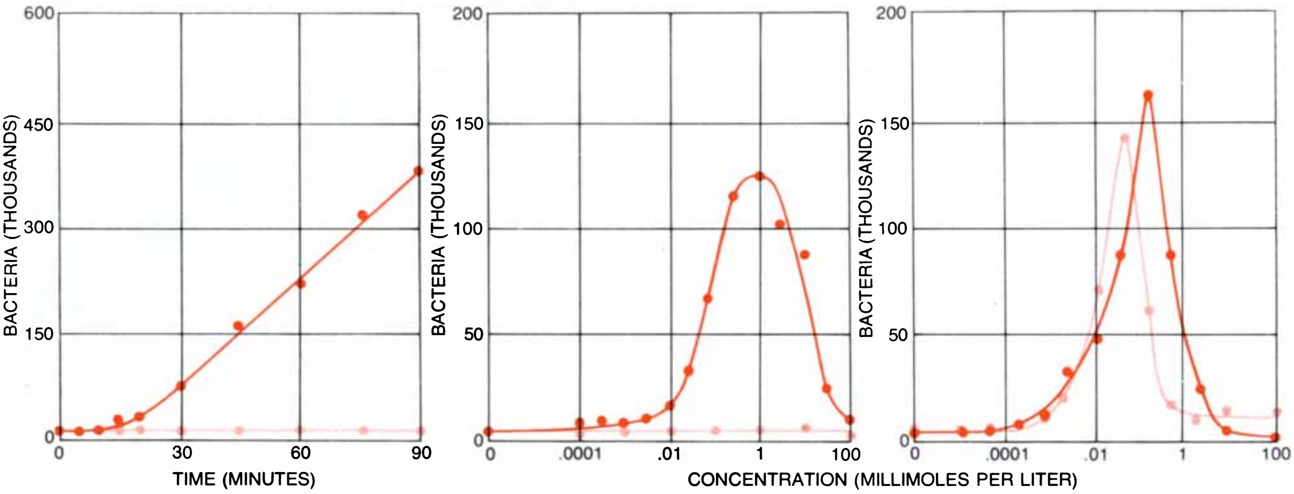
The behavioral mutants of bacteria were of several types. Apart from mutants that simply fail to swim because they lack flagella or the flagella do not function (and these had been known for years) we found mutants that were "specifically" nonchemotactic and mutants that were "generally" nonchemotactic. The specifically nonchemotactic mutants either have lost the ability to be attracted by one chemical (or a group of closely related chemicals) while retaining the ability to be attracted to all other attractants and to be repelled by all repellents, or they have lost the ability to be repelled by just one chemical (or one group of closely related chemicals). The generally nonchemotactic mutants, on the other hand, fail to be attracted or repelled by all the attractants and all the repellents, even though they are quite able to swim. This pattern indicated that there can be "lesions" in the detection of each of the attractants or repellents. That would explain the specifically nonchemotactic mutants, in each of which a different detection system is presumably defective [*see bottom illustration on page 43*]. Then, apparently, the sensed information flows from the several detection systems through a final common pathway to the flagella; lesions in this common pathway would give rise to the generally nonchemotactic mutants. Armstrong and I and John S. Parkinson, Jr. (who is now at the University of Utah), showed there are at least four genes that can mutate to cause a general loss of chemotaxis, and so the common



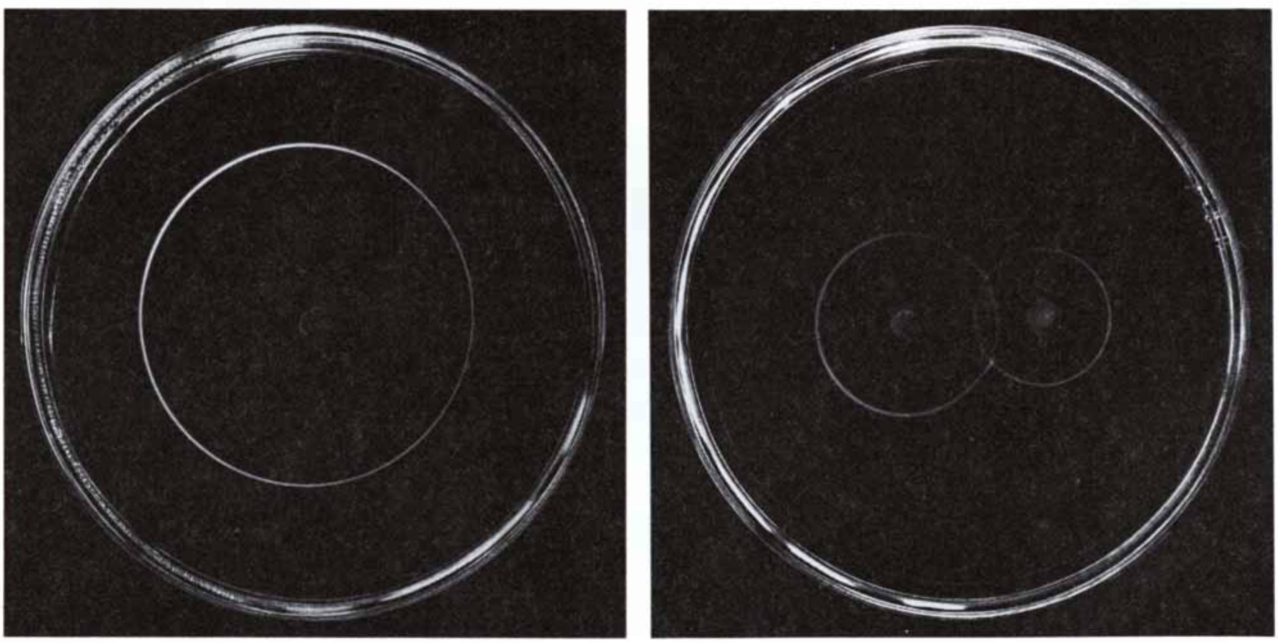
CHEMOTAXIS AND ITS ABSENCE are demonstrated on a tryptone culture plate, which contains a number of nutrients and attractant chemicals. At left is a series of rings formed by *Escherichia coli* bacteria that are chemotactic. The bacteria were originally plated at the center of the rings. Some immediately consumed serine, a preferred attractant in the dish, and followed the resulting serine gradient, forming an expanding ring (as is explained in more detail on the next page). Others made do with aspartate, a less preferred attractant, forming a second ring. Still others were left behind to proliferate at the center of the pattern. At upper right on the plate is a colony of "generally nonchemotactic" mutants: they are not chemotactic but can swim, and so they proliferate and spread out by random movement but do not form a ring. At lower right is a colony of nonmotile cells whose members remain exactly where they were plated.



CHEMOTACTIC BEHAVIOR is demonstrated and measured by placing the tip of a thin capillary tube that contains an attractant solution in a suspension of motile *E. coli* bacteria. The suspension is on a slide, in a chamber created by a U-tube and a coverslip. At first the bacteria are distributed at random through the suspension (*left*). After 20 minutes they have congregated at the mouth of the capillary (*center*). In about an hour many cells have moved up into the capillary (*right*). They are counted by plating them on a culture medium.



NUMBER OF CELLS ATTRACTED can be studied as a function of time at a single concentration or as a function of concentration in the course of one hour. In 90 minutes (*left*) 400,000 cells enter a tube containing aspartate (*dark colored curve*) whereas a negligible number enter a tube lacking attractant (*light color*). Concentration curve (*center*) shows glucose (*dark color*) is an attractant whereas glycerol (*light color*) is not, even though both are metabolized by *E. coli*. Similarly, galactose is an attractant (*right*) for mutants that cannot metabolize it (*light color*) as well as for normal bacteria, which can (*dark color*). (The decreased response at the highest concentrations is caused by the diffusion of the attractant into the suspension, which saturates the cells so that they cannot detect attractant still in tube.)



RINGS OF BACTERIA in pursuit of attractant demonstrate positive chemotaxis. Bacteria are put at the center of an agar plate containing attractant. On a plate containing galactose (*left*) cells consume galactose where they are deposited, proliferate and follow the gradient they create, forming an expanding ring at the edge of the attractant-rich zone. Two mutant strains are plated on agar containing ribose and galactose (*right*). One strain cannot metabolize ribose and the other cannot metabolize galactose. Each consumes the attractant it can metabolize, forming a gradient and an expanding ring. The two rings expand independently across the agar culture medium.

pathway must have at least four components.

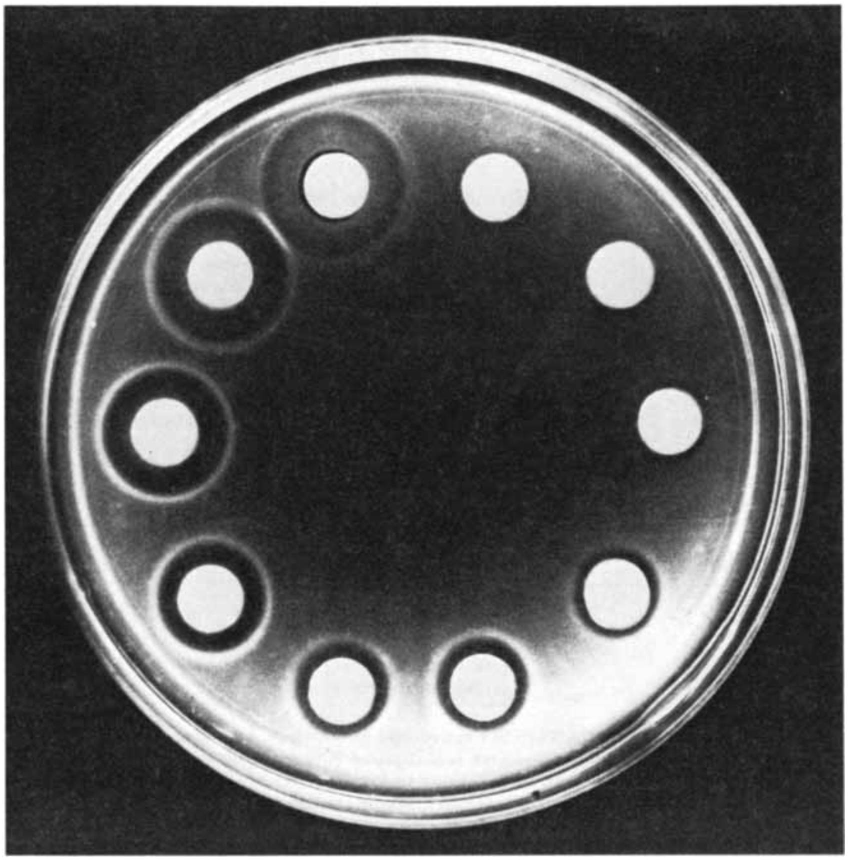
Do the bacteria detect the attractant and repellent chemicals themselves or do they detect something that is manufactured from those chemicals? Nutritious attractants, for example, provide energy. Conceivably the bacteria could be recognizing a higher concentration of attractant indirectly, by virtue of the fact that more energy is available where there is more attractant. Indeed, that was widely considered to be the case when in 1969 I first raised the question of just what the bacteria detect.

The answer is that they detect the attractant itself. Many chemicals that are nutritious for *E. coli* and provide energy fail to attract them; the fact that the organism can metabolize a chemical as a food substance is not sufficient to make that chemical an attractant. Moreover, some chemicals that *E. coli* are unable to utilize for food or even to alter in any way do attract the bacteria, demonstrating that a chemical need not be convertible into some other substance in order to serve as an attractant.

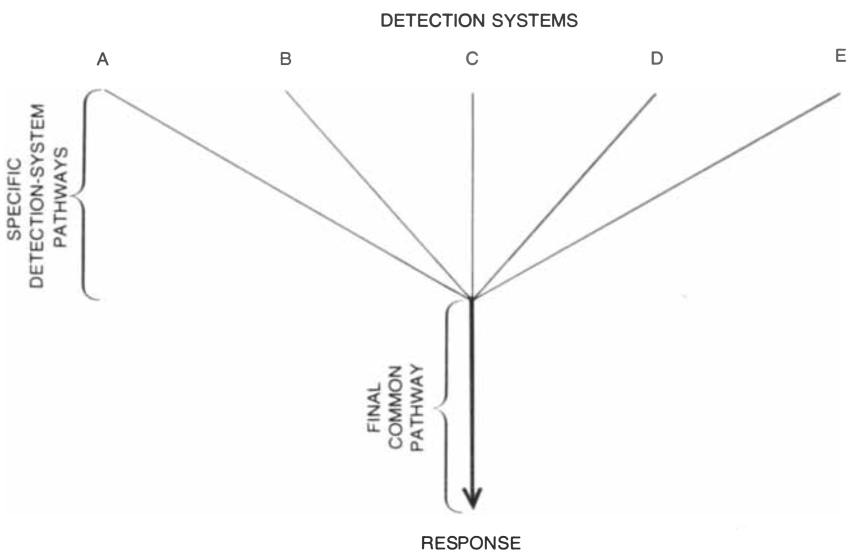
For example, normal bacteria are attracted to the nutritious sugar galactose, but mutant bacteria that fail to metabolize galactose are attracted just as readily. Even normal bacteria fail to utilize the related sugar 6-deoxygalactose, and yet they are attracted to it, presumably because it closely resembles galactose; indeed, bacteria can be "tricked" into leaving a nutritious environment containing a food substance that is not an attractant in favor of an energy-poor environment containing the attractant 6-deoxygalactose. (It should be pointed out, however, that attractants are normally nutritious substances, and so positive chemotaxis as a general phenomenon has survival value for the bacteria.)

Thus it was that in 1969 I discovered chemosensors in bacteria: if bacteria detect the attractants per se, then they must have detection devices. They simply lack such chemosensors for nutritious chemicals that fail to attract them; a specifically nonchemotactic mutant is a mutant that lacks one particular chemosensor. Just as bacteria detect attractants by means of chemosensors rather than indirectly through some secondary benefit, so the bacteria are equipped with other chemosensors that detect repellents directly; they do not simply sense the repellents by experiencing some harm they do. We know this because some repellents that are harmful can be detected at concentrations too low to be harmful, not all harmful substances are repellents and not all repellents are harmful.

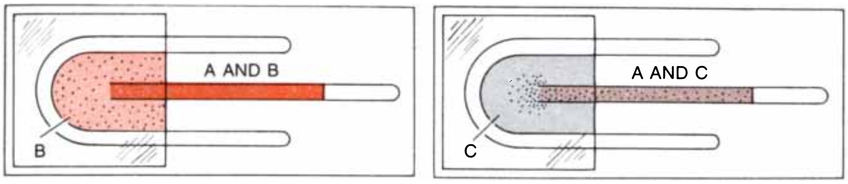
How many different chemosensors does a single bacterial species such as *E. coli* have? One approach is to isolate a specifically nonchemotactic mutant for each sensor, and that has been done in a number of cases. A second approach is a competition experiment [*see upper illustration on next page*]. Attractants are tested against one another in pairs. Each time chemical *A* is in the



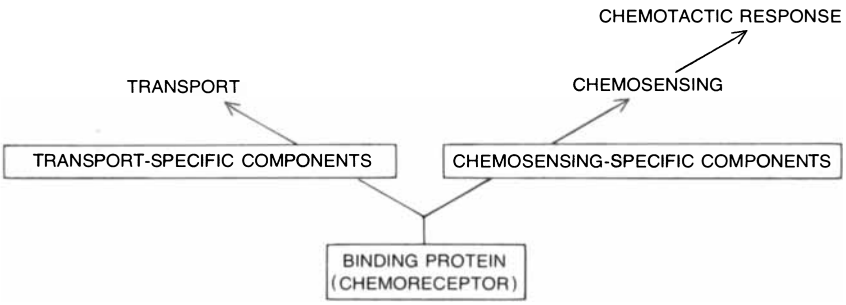
NEGATIVE CHEMOTAXIS is demonstrated by visualizing the ring of bacteria fleeing from a repellent. Plugs of concentrated agar (*bright disks*) containing the repellent acetate were placed in a dish of dilute agar containing a suspension of *E. coli*. The repellent is present in concentrations ranging from zero (*top right*) to three moles per liter (*top left*). After 30 minutes the bacteria have vacated the region (*darkest circles*) around each plug of repellent. The light rings are composed of the bacteria that have moved out of vacated region in dish of dilute agar.



TWO KINDS OF MUTANT, specifically nonchemotactic and generally nonchemotactic, can be accounted for by the two-stage process suggested here. Information flows from a number of specific detection systems (five are arbitrarily shown) into a four-component final common pathway that elicits a response from flagella that propel bacteria. In a specifically nonchemotactic mutant one of specific pathways is defective; in a generally nonchemotactic mutant a component of final pathway is defective. (Branching of tree is more complex than it is here.)



COMPETITION EXPERIMENT determines whether two chemicals are detected by the same chemosensor or by different ones. One attractant (*A*) is in the capillary tube only; another attractant (*B*) is in both the capillary and the suspension (*left*), and at a concentration high enough to saturate its sensor. The cells do not enter the capillary, showing that *A* and *B* must be detected by the same sensor, which, once saturated by *B*, is blocked for attraction to *A*. When *A* is tested against a different chemical (*C*), bacteria do enter tube (*right*). That means *A* and *C* are detected by different sensors, so that *A* can attract in spite of saturation of the *C* sensor.



CHEMOSENSING AND TRANSPORT are related but distinct. A binding protein in the cell envelope acts as a receptor for a chemical; it is required both for detection of the chemical for chemotaxis and for transport into cell. Sensing further requires a set of specific components (*right*) and transport requires another set (*left*). Mutations can affect branches independently.

capillary only; chemical *B* is both in the capillary and in the bacterial suspension, and it is present at such a high concentration that it will saturate its chemosensor: it will fill all the receptor sites, leaving no more room for an attractant to be bound. Because of this saturation attractant *A* can no longer attract bacteria if *A* and *B* are detected by the same chemosensor. *A*, however, will still attract bacteria normally if *A* is detected by a sensor different from *B*. All the repellents are similarly tested pair by pair. The results of mutant studies and competition experiments conducted by Margaret M. Dahl, Gerald L. Hazelbauer, Robert E. Mesibov, Tso and me indicate that *E. coli* has at least 20 different chemosensors, about 12 for attractants and about eight for repellents.

What are these chemosensors? In 1971 Herman M. Kalckar of the Harvard Medical School suggested that we test a certain *E. coli* mutant that lacks what is called the galactose-binding protein to see if it had lost its ability to be attracted to galactose. The protein had been discovered in 1968 by Yasuhiro Anraku, who was then working in Leon A. Heppel's laboratory at the National Institutes of Health; Winfried Boos of the Harvard Medical School had subsequently shown that it is required for the transport of galactose from the outside of the bacterial cell to the inside. Sure enough, Hazelbauer and I found that this mutant failed to be attracted to galactose. We had previously isolated our own mutants that were nonchemotactic specifically for galactose; on testing them we found that most of them lacked the galactose-binding protein. We found that in revertants of those nonchemotactic bacteria, selected for a reverse mutation that had restored galactose chemotaxis, the galactose-binding protein was restored too. There was further proof of the role of the galactose-binding protein in the sensing of galactose: the strength with which a number of structural analogues of galactose bind to the protein was found to exactly parallel their strength as attractants for the bacteria. And so the galactose-binding protein became the first material to be identified as the component that actually recognizes a chemical in chemotaxis: what I call the chemoreceptor.

At the same time we recognized specific binding activity for the attractant maltose, another sugar. The maltose-binding protein has since been purified by Odile Kellermann and Sevec Szmelcman in Maxime Schwartz's laboratory at the Pasteur Institute in Paris, and at the University of Uppsala, Hazelbauer has studied its role as a receptor in maltose taxis. We also found binding activity for the sugar ribose, and ribose-binding protein has since been purified from *E. coli* by Clement E. Furlong of the University of California at Riverside and from the closely related bacterium *Salmonella typhimurium* by Robert Aksamit and Daniel E. Koshland, Jr., of the University of California at Berkeley; Aksamit and Koshland have demonstrated the protein's role as the chemoreceptor for ribose taxis. All three of these binding proteins are situated between the cell wall and the cytoplasmic membrane, in what is called the periplasmic space. For a number of other sugars Wolfgang Epstein of the University of Chicago and I have identified chemoreceptors that are instead tightly associated with the cytoplasmic membrane. These receptors are components of a sugar-transport system, discovered by Saul Roseman and Werner Kundig at Johns Hopkins University, known as the phosphotransferase system.

All the binding proteins, or chemoreceptors, identified so far as being involved in chemotaxis play a dual role: they are required for sensing chemicals for chemotaxis and they are also involved in transporting those chemicals into the interior of the cell. Chemotaxis depends on additional components that sense what fraction of a specific chemoreceptor is occupied by an attractant or a repellent, or, more exactly, sense any changes in the occupied fraction. These sensing components must then signal the information to the flagella. How that is done is still not known.

The sensing and the transport branches of this dual system are independent. Transport is not required for sensing and sensing is not required for transport. Hazelbauer, George W. Ordal and I found mutants that have a normal binding protein for a chemical and exhibit normal taxis in response to the chemical, but that nevertheless fail to transport the chemical into the cell. Such mutants presumably lack components required for the transport branch. Other mutants lack components of the sensing branch. They have a normal binding protein and they transport the bound chemical into the cell, but they do not carry out chemotaxis in response to the chemical.

To understand further how chemosensors work it is helpful to know about the swimming behavior of individual bacteria that are responding chemotactically. Bacteria such as *E. coli* move so fast under the microscope that it is very difficult for the human eye to follow their movement precisely. A great advance was made when Howard C. Berg of the University of Colorado constructed a microscope that automatically tracks individual bacteria [see "How Bacteria Swim," by Howard C. Berg; SCIENTIFIC AMERICAN, August, 1975]. He and Douglas A. Brown found that in the absence of any gradient an *E. coli* bacterium swims in gently curved "runs," or lines, each run lasting for a second or so. The runs are interrupted frequently by brief "twiddles": periods of thrashing about or tumbling, each lasting for about a tenth of a second. After each tumble the bacterium starts off in a new, random direction on its next run.

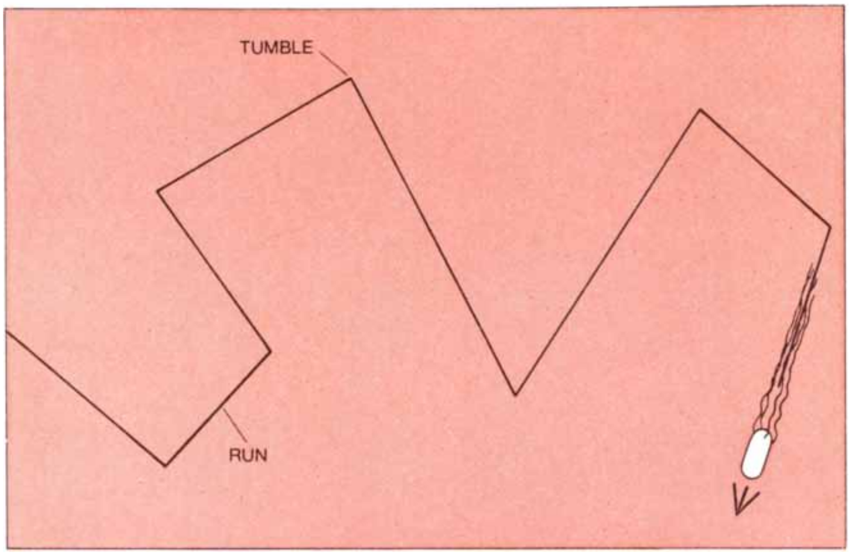
Berg and Koshland and their colleagues have shown that in a gradient of attractant or repellent it is the frequency of tumbling (and therefore the length of the runs) that is altered. If the bacterium happens to swim up the gradient of attractant (in the "right" direction), tumbling is inhibited; if it happens to swim down the gradient (in the "wrong" direction), the frequency of tumbling is increased. In the presence of repellents the effects are just the opposite. The end result is that the bacteria accumulate near the source of an attractant and away from the source of a repellent. (The response is asymmetrical: it takes much steeper changes in concentration to increase the frequency of tumbling than it does to decrease it, and the increase persists for a much shorter time.)

Robert M. Macnab, Nora Tsang and Koshland made the important discovery in 1972–1973 that these effects on tumbling are brought about not only by spatial gradients (for example a higher concentration of attractant to the right than to the left) but also by temporal gradients (for example a higher concentration now than a second ago). Koshland's group increased or diluted the concentration of an attractant or a repellent quickly and then looked through the microscope to determine the effect on the tumbling frequency of *Salmonella*. This work, and similar experiments in 1974 by Brown and Berg in which an enzyme changes the concentration of a chemical, made it clear that bacteria respond to temporal stimulation by attractants and repellents; indeed, the effect of a spatial gradient is probably the result of the fact that bacteria encounter temporal stimulation as they swim along a spatial gradient.

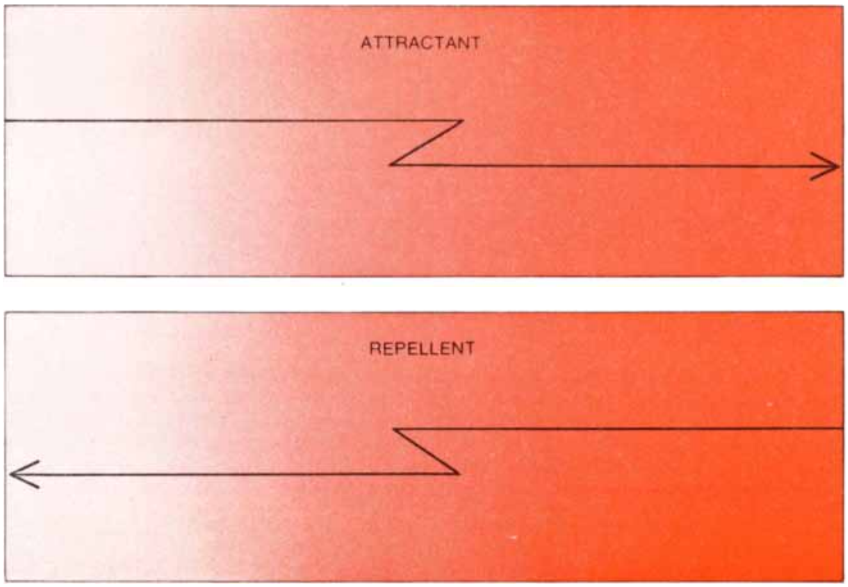
The work from Koshland's and Berg's laboratories emphasized the central role in chemotaxis of changes in the tumbling frequency of the bacteria. How, then, is tumbling generated and controlled? How do flagella bring about tumbling of bacteria?

Melvin L. DePamphilis and I had purified intact flagella from *E. coli* and had described their complex structure as it was revealed in electron micrographs. An *E. coli* flagellum consists of a basal body (a rod and four rings) that is inserted into the cell envelope, a short, curved hook just outside the cell and a long, helical filament attached to the hook. How bacterial flagella work remained unexplained, however. Two ideas had been proposed: the flagellum might flail like a whip, with a wave of contraction starting at the base and propagating along the filament, or it might rotate like a propeller. In 1973 Berg and Robert A. Anderson of the Sandia Laboratories argued, on the basis of previously uninterpreted evidence, that the flagellum rotates. Michael R. Silverman and Melvin I. Simon of the University of California at San Diego soon presented a striking and convincing experiment in support of that position. They tethered *E. coli* bacteria to a glass slide by using antibody prepared against flagellar filaments. (The antibody combines with the filaments and just happens to stick to glass.) Now the flagellum is no longer free to rotate, and instead the cell body rotates! The sense of rotation alternates between clockwise and counterclockwise.

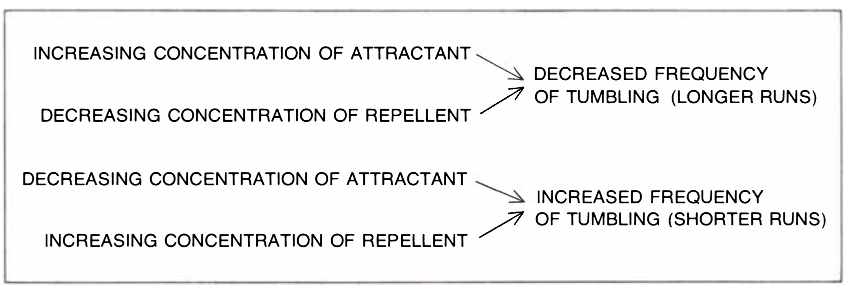
In 1974 Steven H. Larsen, Robert W. Reader, Edward N. Kort, Tso and I studied the effects of attractants and repellents on the direction of rotation of the flagella. The



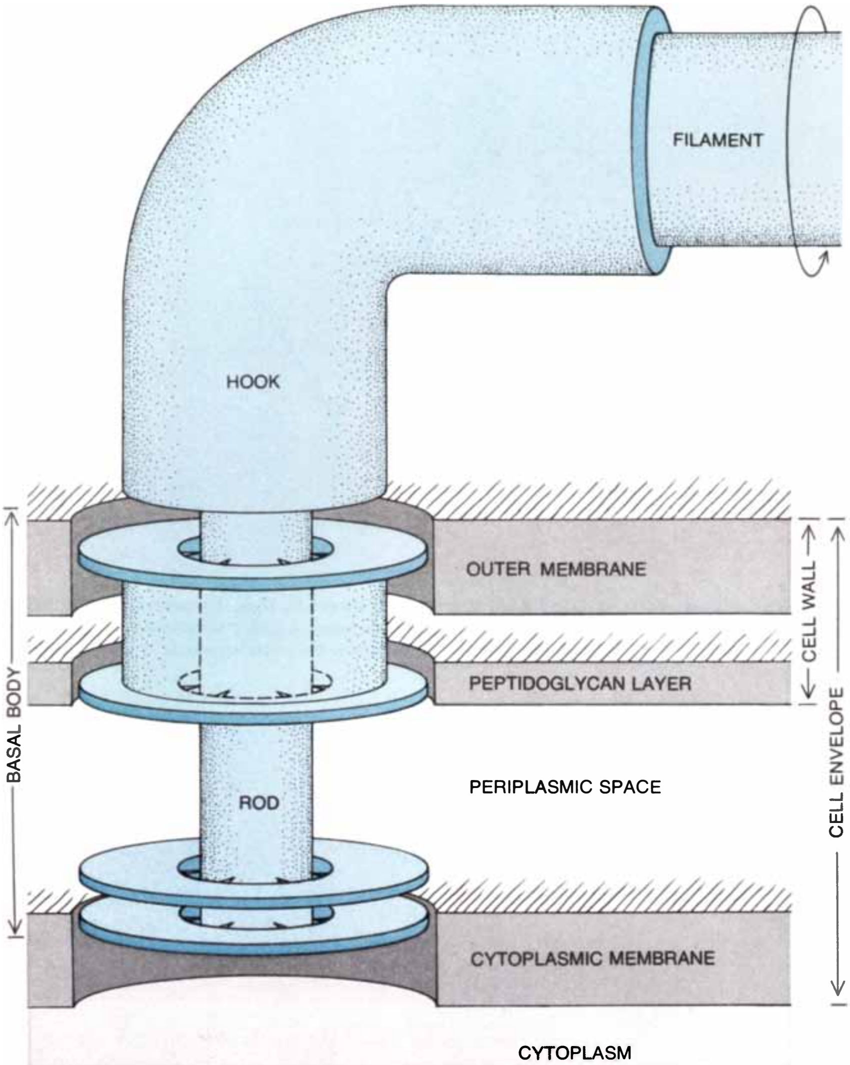
NORMAL PATH of *E. coli*, Douglas A. Brown and Howard C. Berg demonstrated, is a three-dimensional "random walk" (shown here in two dimensions). A series of almost straight runs is interrupted by brief periods during which cell tumbles before setting out in a new direction.



PATH OF CELL IN A GRADIENT is biased by a change in the frequency of tumbling, which is inhibited when the cell is moving in the "right" direction (up-gradient in an attractant or down-gradient in a repellent) and promoted when the cell is going in the "wrong" direction. (Here only runs that happen to be in the "right" or the "wrong" direction are shown.) The diagram is based on results from Berg's laboratory and that of Daniel E. Koshland, Jr.



EFFECTS ON CELL'S PATH of the presence of a gradient are summarized. The end result is that bacteria accumulate near source of an attractant and away from the source of a repellent.



BASAL BODY of the flagellum of *E. coli* and other gram-negative bacteria is inserted into the cell envelope as indicated in this drawing based on electron micrographs and other findings made by Melvin L. DePamphilis and the author. They found that the top ring is associated with the outer membrane and the bottom ring with the cytoplasmic membrane. Binding proteins have been identified in the periplasmic space and the cytoplasmic membrane. The flagellum rotates, generally (as seen looking along the filament toward the cell) counterclockwise (*arrow*).

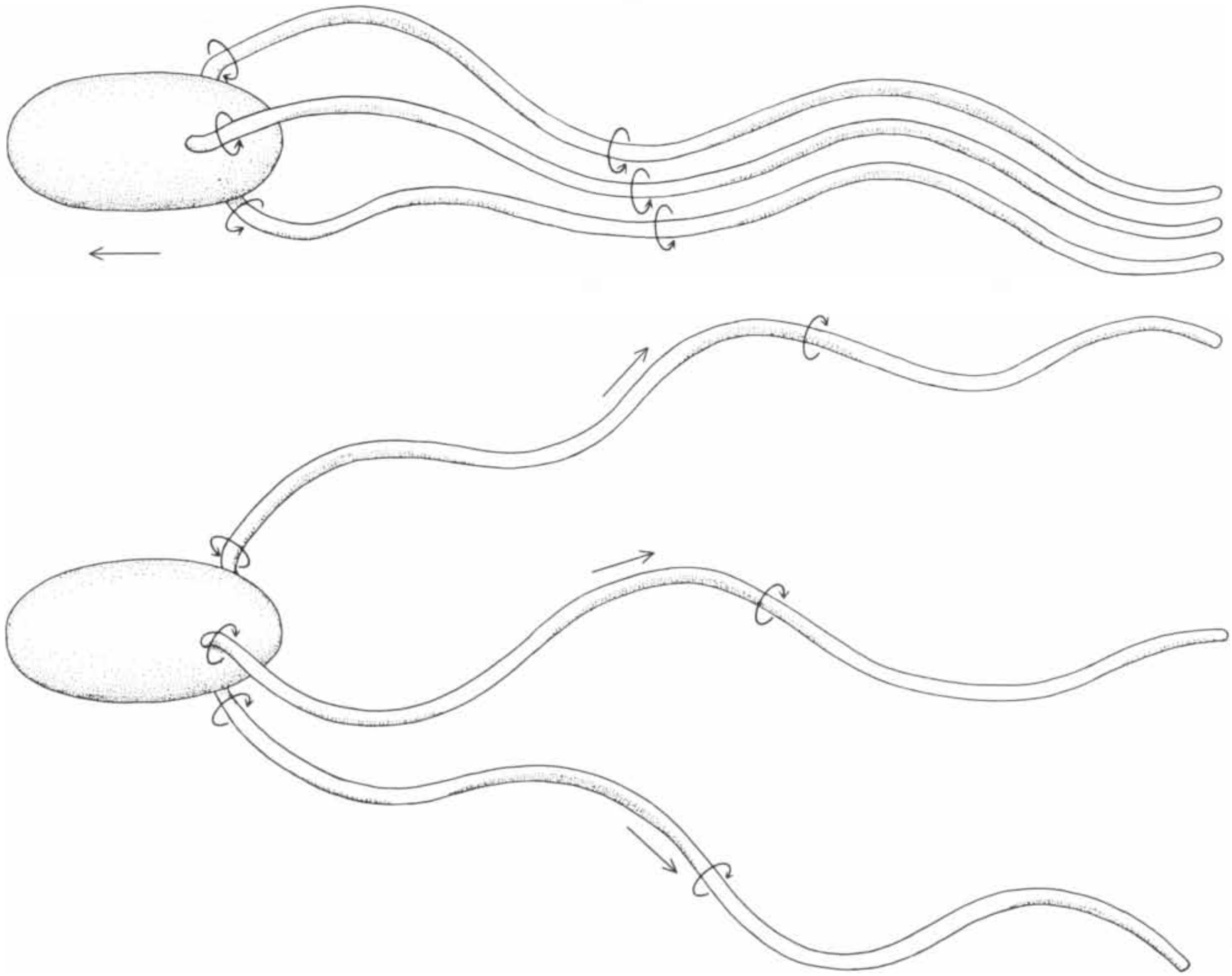
addition of attractant to tethered *E. coli* bacteria brought about counterclockwise rotation of the cells (which happens, by convention, to be the equivalent of counterclockwise rotation of a flagellum that is free to rotate); the addition of repellent caused clockwise rotation. Since the addition of attractant inhibits tumbling (that is, it lengthens the runs) and the addition of repellent increases tumbling, this indicated that sustained swimming in runs is brought about by counterclockwise rotation and tumbling is brought about by clockwise rotation. The effect of the direction of rotation on tumbling was confirmed by another finding: Generally nonchemotactic mutants that never tumble (that always swim in runs) always rotate counterclockwise when they are tethered; other generally nonchemotactic mutants that nearly always tumble nearly always rotate clockwise. This tells us, incidentally, that the common pathway into which the various chemoreceptors lead must deal with the generation and control of tumbling.

It has long been known from light microscopy that the flagella of *E. coli* work together in a bundle behind the bacterium when it is swimming forward, even though the flagella originate at various sites on the cell. Apparently when the flagella rotate counterclockwise, they can function together (although each is rotating separately) in the bundle that propels the cell, When the flagella rotate clockwise, however, the bundle flies apart. Macnab and Koshland were able to demonstrate, by making micrographs with a powerful light source, that the flagella fly apart when a cell tumbles. Now each flagellum is pulling instead of pushing and is fighting against every other one; the result, as one or another flagellum prevails, is a tumble.

The chemotactic response in bacteria is transient, that is, the cells adapt to a stimulus. The addition of attractant or repellent to a suspension of tethered cells brings about a change in direction of rotation that lasts for a period of time depending on the concentration. Then the bacteria adapt to the new concentration and return to their normal mode of rotating in both directions (but mostly counterclockwise). Similarly, in the case of untethered cells a change in the concentration of attractant or repellent produces a transient change in the frequency of tumbling. Such transience, or adaptation, is a characteristic feature of many sensory functions. The transience of the chemotactic response is in contrast to all other responses of bacteria to changes in the concentration of a chemical, responses that persist as long as the new concentration is maintained.

Some indication of how the frequency of tumbling is controlled comes from my discovery some years ago that bacterial chemotaxis requires the amino acid methionine, which most bacteria synthesize. Without methionine most bacteria swim in runs, never tumbling, and so methionine must somehow be required to turn on clockwise rotation of the flagella. Last year Kort, Michael F. Goy, Larsen and I found a protein in the cytoplasmic membrane of *E. coli* that becomes methylated by methionine, that is, it accepts CH_3 groups from the amino acid. Some generally nonchemotactic mutants fail to carry out this methylation. When such mutants revert to normal chemotaxis, the methylation system also returns to normal. The addition of attractants affects the methylation of the membrane protein in chemotactically normal cells. All of this provides evidence that the methylated membrane protein plays a central role in chemotaxis. More specifically, the chemosensors somehow influence the degree of methylation of the proteins; the degree of methylation somehow determines the direction of rotation of the flagella and thus the frequency of tumbling.

Finally, to get to the realm of bacterial "psychology" we presented bacteria with a capillary tube containing not only an attractant but also a repellent. Now they had to choose whether or not to enter the tube. Their "decision" turns out to depend on the relative concentrations of the attractant and the repellent. The mechanism for "decision making" in such a "conflict" situation is still not known, but one can say that bacteria are able somehow to integrate multiple sensory inputs. Koshland's group had reached a similar conclusion. Recently Larsen and I have found another example of integration. *E. coli* can carry out thermotaxis, or attraction and repulsion by solutions at various temperatures. The bacteria are ordinarily repelled by cold, but not if the cold solution contains a strong enough attractant. Likewise they are attracted by warmth, but not if the warm solution contains a strong enough repellent. In these cases the bacteria must be integrating, or

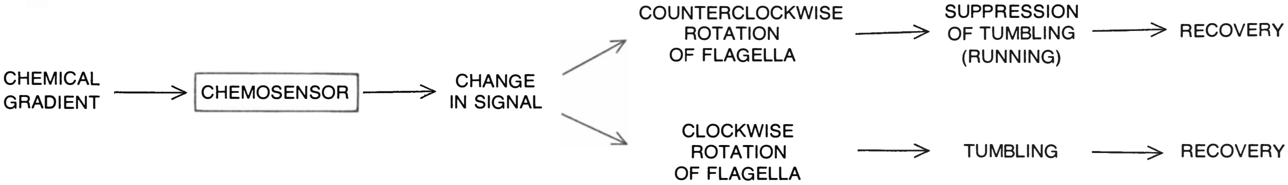


SWIMMING AND TUMBLING seem to be accomplished by *E. coli* as shown in these drawings. When the flagella (only three of the usual four to eight are shown) rotate counterclockwise (*top*), the helical filaments form a bundle and function as propellers that push the cell through the medium in runs. When the flagella rotate clockwise (*bottom*), on the other hand, the flagellar bundle flies apart. The filaments now tend to pull the cell, but each filament pulls in a different direction depending on its orientation, and so the cell tumbles about.

processing, information from two sensory modalities: temperature and chemicals.

The basic elements that make behavior possible in a higher organism are thus also present in a single bacterial cell; they are sensory receptors, a system that transmits and processes sensory information and effectors to produce movement. Whether the mechanisms of any of these elements in the bacterium are similar to those in more complex organisms remains to be established. Obviously there must be major differences; for example, since bacteria are independent cells, the cell-to-cell synaptic action that is so important in determining behavior in more complex organisms cannot possibly exist in bacteria, at least not at a cellular level. Still, it appears that the bacterial system may be a good model for the study of behavior.

By now the broad outlines of bacterial chemotaxis appear to have been sketched, but many questions remain to be answered, primarily at the biochemical level. What is the nature of the "motor" that rotates the flagellum? What is the nature of the "gearshift" that determines whether the motor rotates clockwise or counterclockwise, and what role does the methylation system play in shifting the gear? What is the nature of the signal sent by the chemosensor to indicate that there has been a change in the fraction of chemoreceptor occupied by attractant or repellent? How is the signal generated, and how does it act on the gearshift or on the methylation system? Do stimuli have their effect through a change in the membrane potential of the cell, as is the case in protozoa and higher animals? How does the cell process information from several stimuli, sometimes conflicting ones? And what is the mechanism of adaptation?



CHAIN OF EVENTS from stimulus to response and recovery is summarized. The chemosensor detects and signals a change in the fraction of the receptor occupied by a chemical, eliciting either counterclockwise or clockwise rotation of the flagella and thus either swimming in runs or tumbling. The response is transient: with adaptation, or recovery, there is a return to alternating senses of rotation.

Subjective Contours

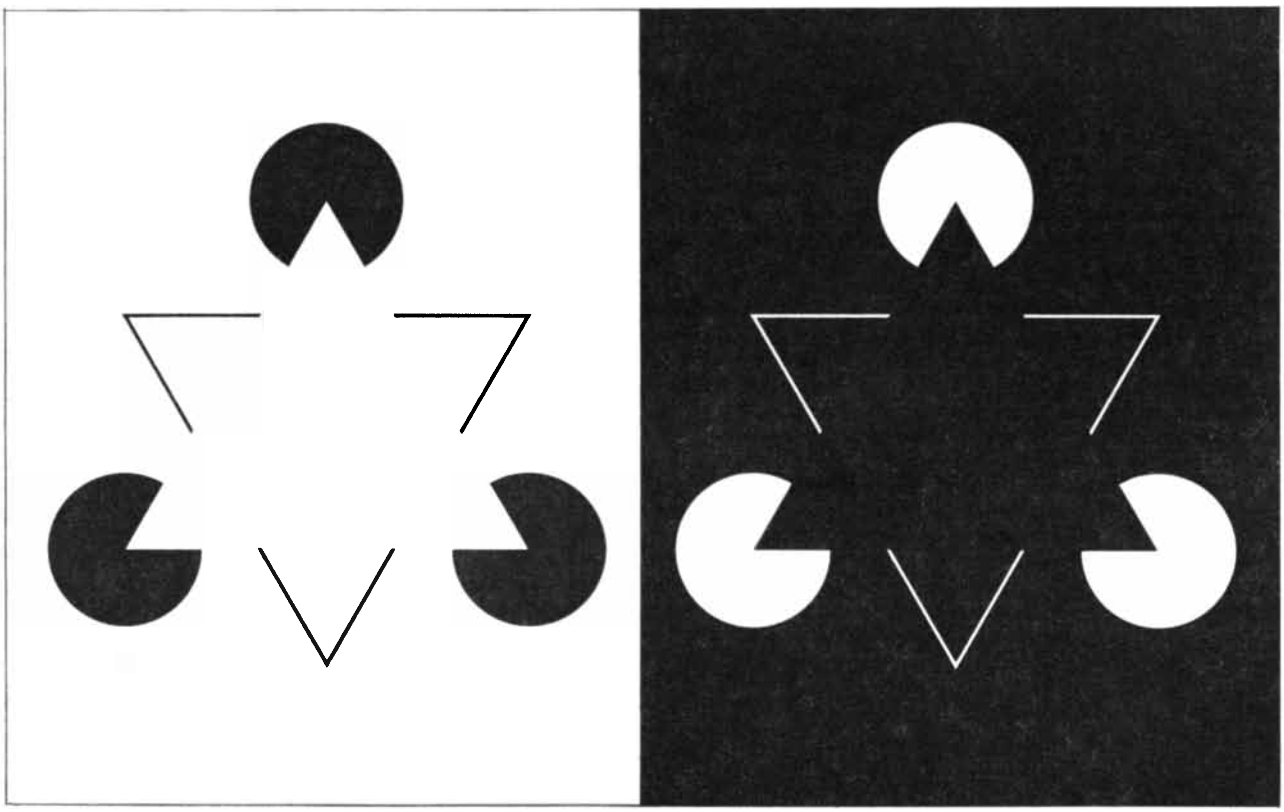
Certain combinations of incomplete figures give rise to clearly visible contours even when the contours do not actually exist. It appears that such contours are supplied by the visual system

by Gaetano Kanizsa

If we examine the conditions that give rise to visible contours, we usually find that a contour is perceived when there is a jump in the stimulation between adjacent areas. The jump may be due to a difference in brightness or a difference in color. There are conditions, however, that cause us to perceive contours in visual areas that are completely homogeneous. For example, in the illustration below the solid triangles in the center of each figure appear to have well-defined contours, but close examination of the contours where they cross an open area reveals that they have no physical basis. If you fix your gaze on one of these contours, it disappears, yet if you direct your gaze to the entire figure, the contours appear to be real.

The phenomenon of contours that appear in the absence of physical gradients has aroused considerable interest among psychologists on both the experimental and the theoretical level. A number of variants of the effect have been discovered, and several explanations have been proposed for it. Here I shall describe some of the more interesting properties of the effect and examine some of the attempted explanations. First, however, let us consider a related visual phenomenon: the phenomenon of virtual lines.

When we view three dots that are equidistant from one another and are not in a straight line, the visual system spontaneously organizes the dots into a triangle. In addition the three dots appear to be connected by three straight lines. These lines are called virtual, and although they are not actually seen, they are a real presence in our visual experience. They are far more compelling than other connecting lines that can be imagined. For example, the three dots



TWO SUBJECTIVE TRIANGLES, one whiter than white and the other blacker than black, appear to have distinct contours, but when the contours are examined closely, they disappear. The contours are subjective and have no physical reality. The region bounded by the subjective contours appears to be more intense than the background even though the color of the inner and the outer regions is identical.

could just as readily be points on a circle, but the curved connecting lines of the circle are more difficult to "see" than the straight lines of the triangle.

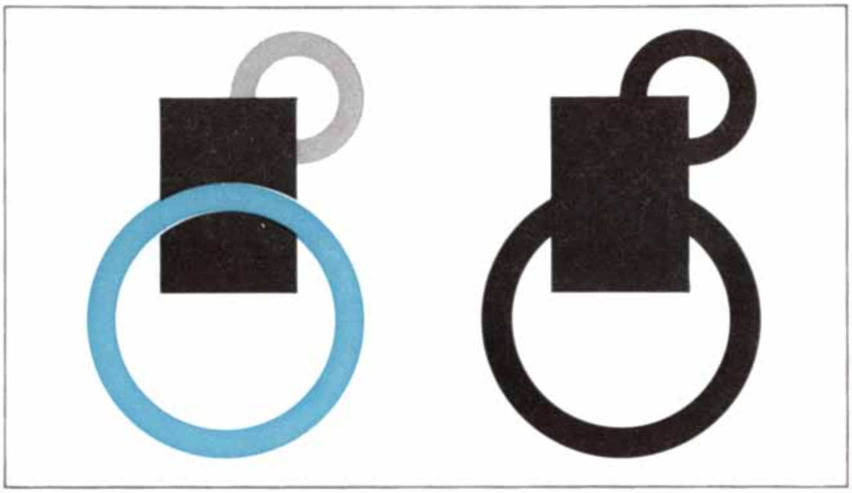
Because virtual lines are only phenomenally present and do not have a sensory modality, one may speak of them as being "amodal." Another kind of amodal contour is found in partially hidden figures [*see top illustration at right*]. Consider a black rectangle that has a gray ring behind it and a colored ring in front of it. Although the missing contours of the rectangle are not actually seen, they nonetheless have a strong phenomenal presence. If the two rings in the illustration are now made black, a new effect results. Both black rings complete themselves behind the black rectangle in an amodal manner, but the contours of the rectangle are visible in their entirety. Even in the homogeneous black regions where the rings overlap, the contours of the rectangle are visible. In other words, the contours have acquired a visual modality.

This "modal" presence is also found in the contours of the central triangles in the illustration on the opposite page. Since those contours appear in the absence of the gradients that normally produce modal, or visible, contours, the situation is clearly anomalous. For that reason I prefer to call such contours anomalous contours. In order to emphasize the fact that the contours have no physical basis over most of their length, other investigators have called them subjective contours. They are also known as illusory contours. Whatever term is used, the phenomenon is the same.

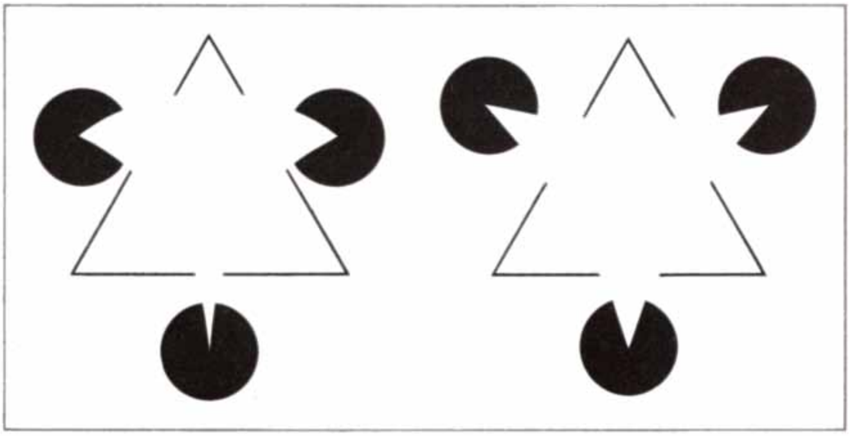
What factors are involved in the formation of subjective contours? Analysis of many examples of the phenomenon yields the following common characteristics. First, the region that is bounded by the subjective contours appears to be brighter than the background, even though the visual stimulation provided by both regions is exactly the same. Second, the region within the subjective contours appears as an opaque surface that is superposed on the other figures in the illustration.

The subjective contours we have considered up to this point have all been straight lines. Is it possible to create curved subjective contours? As the middle and bottom illustrations at the right demonstrate, there are a variety of ways for generating such subjective contours. Indeed, even amorphous subjective figures can be created.

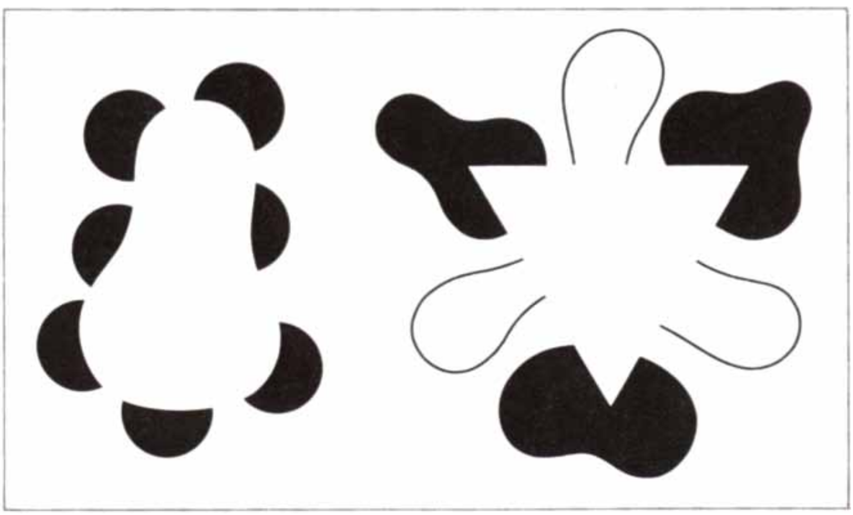
The strength of the phenomenon of subjective contours can be measured in part by determining the resistance such contours show to interference by real lines. When a real line intersects a subjective contour, the contour in that region disappears, indicating that it has a relatively low degree of resistance to interference. On the other hand, the opaque subjective surface displays surprising resistance: it appears to pass under lines that intersect it [*see top illustration on page 51*]. The subjective sur-



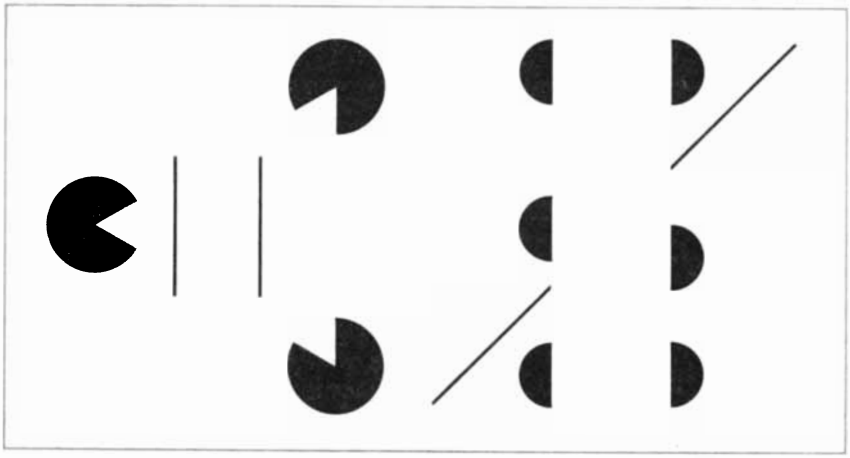
AMODAL AND MODAL CONTOURS are found in overlapping figures. Amodal contours are not actually seen, but they have a strong phenomenal presence. For example, the missing contours of the black rectangle complete themselves behind the colored ring in an amodal manner. Modal contours, on the other hand, appear to be visible. For example, the contours of the rectangle at the right are visible even in the regions where they overlap the black rings.



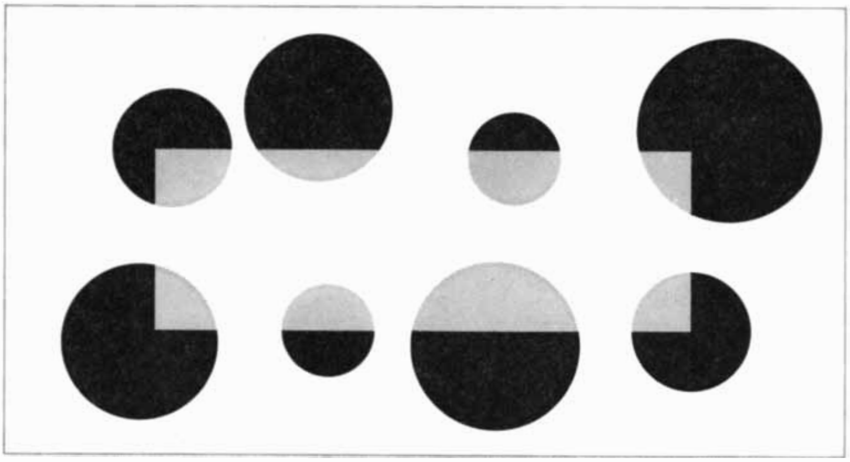
CURVED SUBJECTIVE CONTOURS are created by sectors with curved angles (*left*). Sectors with straight angles can create curved contours if angles are not aligned with one another.



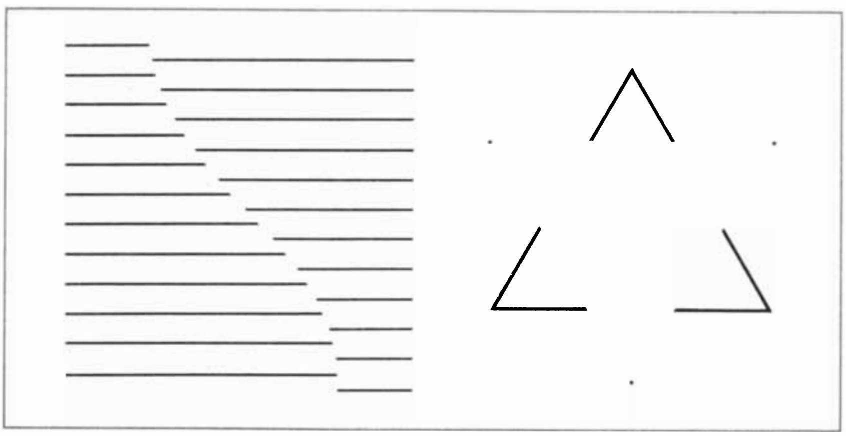
GEOMETRIC REGULARITY is not a necessary condition for the formation of subjective surfaces and contours. Amorphous shapes are possible and irregular figures can generate contours.



OPTICAL ILLUSIONS show that subjective contours have the same functional effects as real contours. In the Ponzo illusion (*left*), although both vertical lines are the same length, the effect of the subjective triangle is to make the line at the left appear to be longer. In the Poggendorf illusion the subjective surface gives rise to an apparent displacement of the slanted line.



TRANSPARENT SUBJECTIVE SURFACES, as well as opaque ones, can be produced. The transparent surface, with clearly visible contours, seems to lie in a plane in front of black disks.



CONTOUR-DETECTOR HYPOTHESIS states that subjective contours are generated by partial activation of contour detectors in the visual system by short line segments in the stimulus. Subjective contours, however, can have an orientation completely different from that of the line segments (*left*). Furthermore, line segments are not necessary for generation of subjective contours (*right*). Curved subjective contour formed by line segments also demonstrates that differences in brightness due to contrast are not needed for formation of subjective contours.

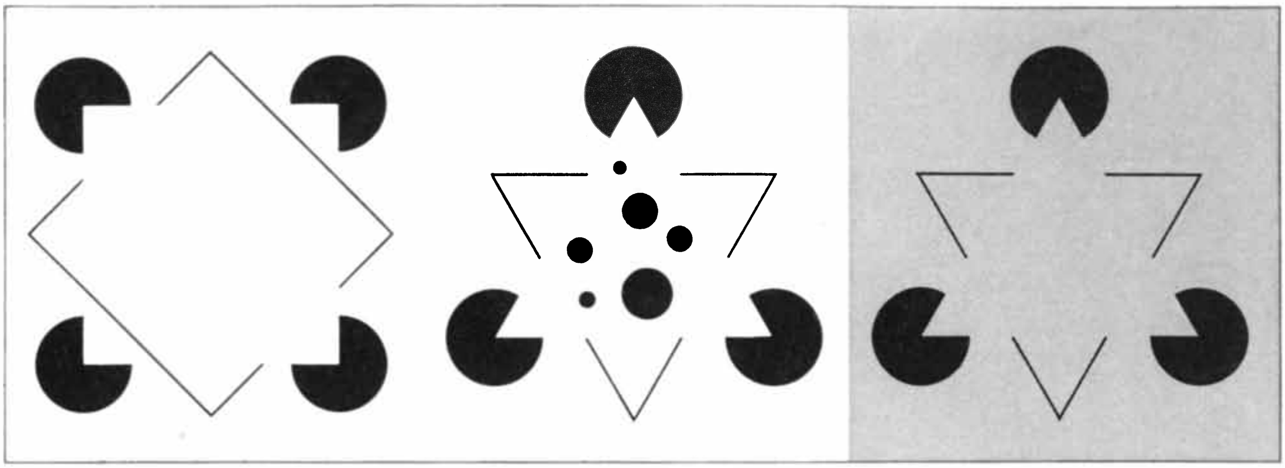
face also displays strong resistance to interference within its borders. If large dark spots are placed inside the borders, the spots do not become part of the background but rather appear to be on the subjective surface. What happens when the background, instead of being homogeneous, has a texture? It turns out that a texture does not impede the formation of subjective contours or surfaces.

A number of optical illusions are produced by the reciprocal action between lines and surfaces. These optical illusions offer an opportunity to ascertain whether subjective contours and shapes have the same functional effects as objective, or real, contours and shapes. In many instances subjective contours and shapes are able to duplicate the illusion created by objective ones. As the top illustration at the left demonstrates, subjective contours and surfaces will interact with physically real lines to give rise to familiar optical illusions.

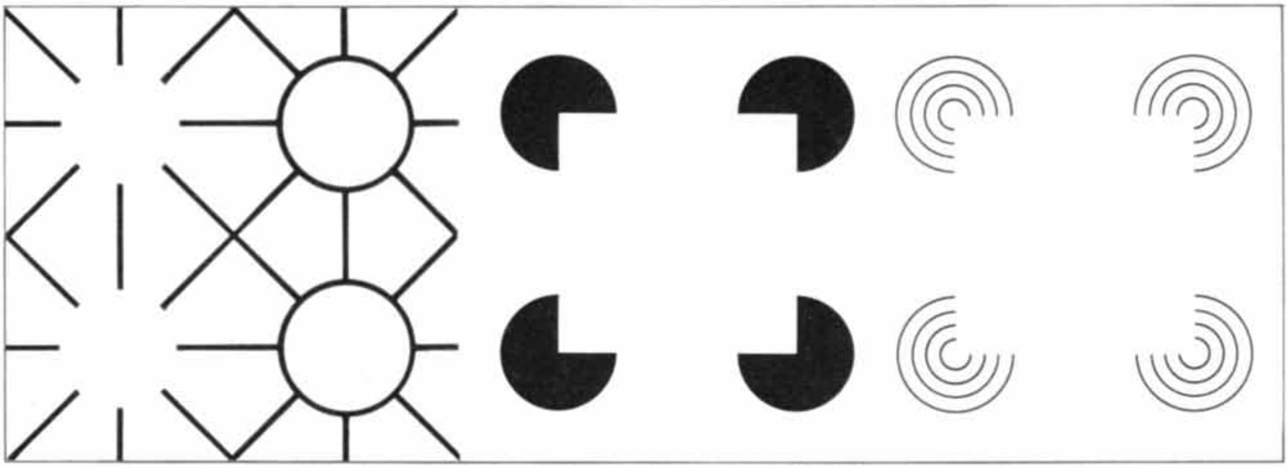
As we have seen, one of the characteristics of subjective surfaces is that they appear to be superposed on the other figures in the illustration. We have also seen that the subjective surface appears to be opaque. It is not difficult, however, to produce transparent subjective surfaces with distinct subjective contours [*see middle illustration at left*].

In most of the situations we have been examining the subjective surface appears to be brighter than the background, even though the two regions are identical in brightness and color. It is possible that the brightness of the subjective surface is due to contrast enhancement. Such enhancement is generally found when a light surface is adjacent to one dark surface or more. The intensity of the effect depends on the extent of the dark surface. Although the brightness-contrast effect may play a role in creating subjective surfaces, it is not a necessary condition for the formation of such surfaces or contours. This is readily demonstrated in the middle illustration on the opposite page, where a substantial reduction in the amount of black does not diminish the effect. A decisive item of evidence that contrast is not necessary for the formation of a subjective contour is presented in the figure at the left in the bottom illustration at the left. In this figure there are no differences in brightness that could be attributed to contrast, yet a curved subjective contour between the line segments is clearly visible.

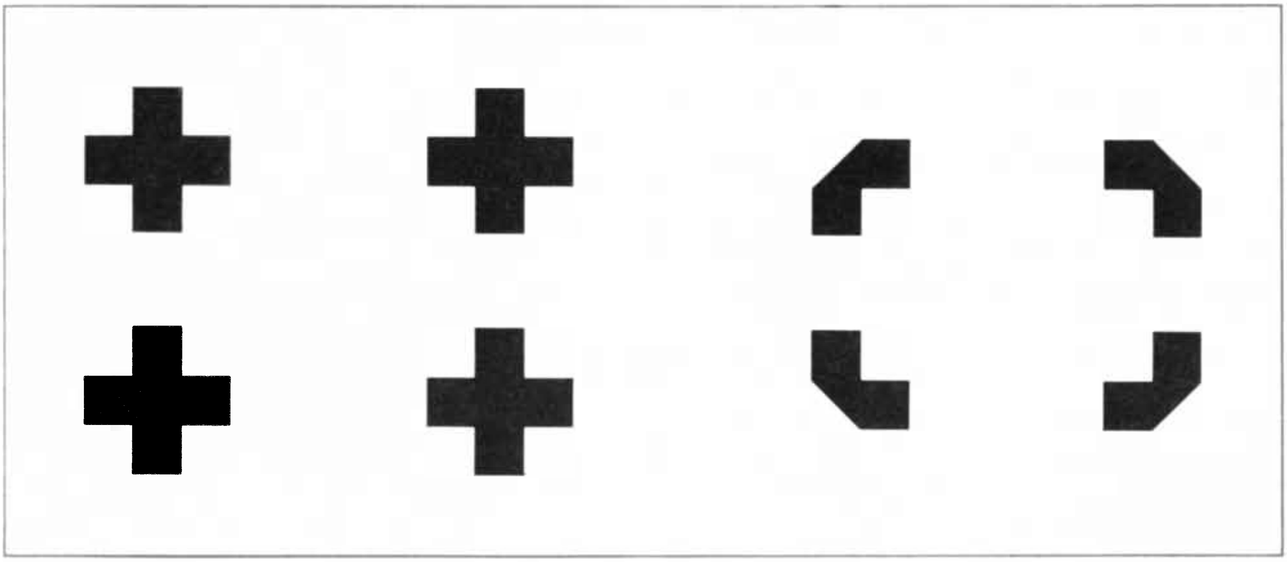
It has been suggested by some investigators that subjective contours can be explained in terms of the partial activation of contour-detector cells in the visual system. According to this hypothesis, the short line segments in the visual stimulus activate some of the contour detectors, and signals from the activated detectors are interpreted as being a stimulus from a continuous line. The hypothesis does not stand up to careful examination, however. In many cases a subjective contour does not continue in the same direction as the stimulus line seg-



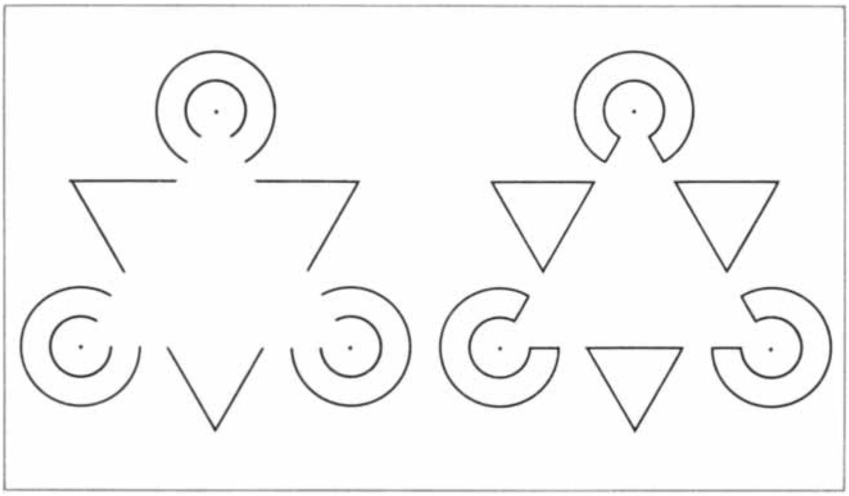
RESISTANCE TO INTERFERENCE is a measure of the perceptual strength of subjective surfaces. A subjective surface appears to pass under lines that intersect with it (*left*), but subjective contours are destroyed by the line. Spots inside the borders of the subjective surface become part of it (*middle*). The formation of subjective contours or surfaces is not impeded by the presence of a texture (*right*).



ENHANCED BRIGHTNESS of subjective surfaces is not due to contrast. If contrast were a primary condition of the effect, the black-ringed circles should appear to be brighter than the subjective circles (*left*). Reducing amount of black does not diminish the effect (*right*).



COMPLETE FIGURES do not generate subjective surfaces. Although crosses provide outlines of a rectangle, the rectangle is not perceived as a surface. When crosses are cut in half, a subjective rectangle is perceived and the half crosses are now seen as mutilated hexagons.



FIGURES WITH OPEN BORDERS appear to be incomplete. In order to complete the figures the visual system superposes an opaque surface that fills the gaps in the figures. Because the surface must have borders the necessary contours are also supplied by the visual system. If borders of figures are closed, there is no further need for completion and contours disappear.

ments. Moreover, line segments are not necessary for the generation of subjective contours. In some instances the line segments can be replaced by dots and subjective contours will still be perceived [*see figure at right in bottom illustration on page 50*].

There is one condition, I have found, that is always present in the formation of subjective contours. That condition is the presence in the visual field of certain elements that are incomplete, which on completion are transformed into simpler stable and regular figures. For example, it could be said that the figure at the left in the illustration on page 48 consists of three black sectors and three angles. Each of these figural elements is incomplete in some way. Most observers, however, report that they see a white triangle covering three black disks and another triangle with a black border. This perceptual organization has obvious advantages from the standpoint of simplicity and stability. The three angles become a triangle, a stabler and more balanced figure. The three circular sectors acquire completeness and regularity by becoming disks. In order for this perceptual organization to materialize, however, the white area in the center must be seen as an opaque triangle that is superposed on the other figures. And since the triangle must have a border, the necessary contours are supplied by the visual system. The contours are therefore the result of perceiving a surface and not vice versa. The subjective surface in turn is generated by the tendency of the visual system to complete certain figural elements.

If these assertions are correct, we should be able to demonstrate that subjective contours and shapes will not be perceived when the visual field does not contain incomplete figural elements. Since figures with open borders tend to appear incomplete, it is not difficult to create subjective contours with them. If we close the borders on these figures and make no other changes, the subjective contours disappear [*see top illustration on this page*].

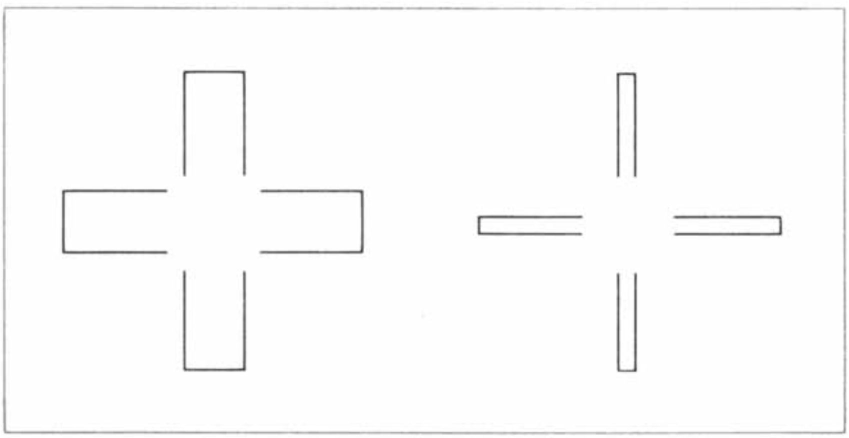
The following, I believe, offers further confirmation of the completion hypothesis. At the left in the bottom illustration on the preceding page there are four black crosses on a white field. In spite of the fact that the crosses provide the outlines of a rectangle in the central region, we do not perceive the rectangle as a subjective surface. The reason is that the crosses are balanced and self-sufficient figures and do not require completion. When the crosses are cut in half, however, a subjective surface appears in the central area. The half crosses are in this case more likely to be seen as mutilated hexagons.

We have seen that irregularly shaped subjective figures can be produced. In most of my examples the incomplete figures I have used to create subjective contours have been regular and symmetrical. Although geometric figures may enhance the effect, however, they are by no means necessary [*see bottom illustration on page 49*].

Finally, is it possible to generate subjective contours that meet and form a subjective angle? Paolo Sambin of the University of Padua found that an incomplete cross gives rise to such an effect [*see bottom illustration on this page*]. According to Sambin, the rectangular shape of the subjective surface that is perceived is produced by the resistance of the arms of the cross to invasion by the subjective surface. Without such resistance the subjective contour would assume the shape of a circle. The validity of his hypothesis can be demonstrated by narrowing the arms of the cross to the point where the invasion of the internal area is minimal. Under those conditions the subjective surface that is perceived has the form of a circle.

Another example of contour perception in the absence of brightness gradients is found in the random-dot stereograms created by Bela Julesz of Bell Laboratories. These stereograms do not reveal any contours when they are viewed monocularly, but when they are viewed with a stereoscope, they combine to form three-dimensional shapes and contours. Stanley Coren of the New School for Social Research has advanced the hypothesis that the perceptual mechanism giving rise to subjective contours and shapes is the same as the mechanism giving rise to three-dimensional depth perception.

Since the formation of subjective contours is usually connected with the generation of surfaces and their stratification, or apparent layering, the line of reasoning proposed by Coren may be valid. On the other hand, in all the cases that we have examined stratification depends on the completion of some figural elements. When there is no need for completion, stratification does not occur and there are no subjective contours. Once more the primary factor seems to be the tendency to completion. Stratification seems to arise as a function of this completion.



INCOMPLETE CROSS (*left*) gives rise to the unusual phenomenon of subjective angles. The angles are formed by the contours of the subjective square that covers central portion of the cross. Rectangular shape of the subjective surface is attributed to the resistance of arms of cross to invasion by the surface. If arms of cross are narrowed, circular subjective surface results.

Magnetic tape, they called it.
They said that some day it might rival phonograph records.
Maybe even carry pictures.
Hmmm.

Well, it happened. Photography acquired a sibling no less vigorous than itself. Nowadays, to be restricted to but one of the two media for professional (or even private) communication may seem like driving with one headlight—deplorable.

We have two things to say about that.

1. A slide/tape presentation on whatever it is that you and your colleagues wish to communicate can save costly, soul-trying travel. Perhaps you don't have to make the trip, and yet the valuable question period following the presentation may not need to be sacrificed. The telephone company may have some sensible suggestions on how to set it up. We on our part have some detailed suggestions on how to synchronize a slide/tape presentation to make it run a shade more smoothly, perhaps, than can be expected from a weary traveler, even when assisted by a well-meaning stranger or two. To arrange for a slide/tape presentation to your department on how to synchronize a slide/tape presentation, make interest known to Bob Scott, MP&AV, Dept. 640, Kodak, Rochester, N.Y. 14650.

2. A very respectable percentage of the active adult population have been imprinted to video since infancy. An organization that cannot risk being regarded as antiquated therefore invests quite a bundle in videotape equipment, perhaps for research, perhaps for training. Prudence suggests that it be used frequently, even though one may often wish that film could be shown on the videotape monitors. It can be. This is a KODAK SUPERMATIC Film Videoplayer:

Feed it super 8 film, and it feeds color and sound to those monitors. Nothing more required to get the advantages of both media. Less than $1700. Write Dick Schiavo at Kodak, MP & AV, Dept. 620, Rochester, N.Y. 14650.

Kodak

Price subject to change without notice.

SCIENCE AND THE CITIZEN

Out of Control

In a period of global economic recession and simultaneous price inflation one sector of the world economy—the arms race—appears to be exempt from the general squeeze on the public purse. According to a recent study conducted by a group of non-Government arms-control experts, world military expenditures rose to a record high of $270 billion in 1974 and continued upward in 1975, approaching the $300 billion mark. Measured in constant 1973 dollars the total outlay of public funds for military purposes in the past year was almost 45 percent higher than in 1960.

Of the total commitment of public funds to arms and armies the portion spent for military-related research and development represents about 10 percent, or some $25 to $30 billion annually. Military research and development is currently estimated to occupy about 25 percent of the world's scientific manpower and to consume an average of 40 percent of the public funds allocated for all kinds of research and development. The latter proportion is even higher for the two superpowers. In the fiscal year 1975 the U.S. spent 52 percent of its publicly raised research and development funds for military purposes; the corresponding proportion for the U.S.S.R. is believed to be between 50 and 60 percent.

These and other statistical indicators of the social costs of the world arms race are contained in the 1976 edition of "World Military and Social Expenditures," a report sponsored by the Arms Control Association, the Institute for World Order and the Members of Congress for Peace through Law Education Fund. The report was prepared by a group headed by Ruth Leger Sivard, an economist formerly associated with the U.S. Arms Control and Disarmament Agency, where she supervised a series of similar reports issued between 1962 and 1971, before the monitoring project was dropped by the Nixon Administration.

The most pronounced relative increase in military expenditures in recent years, the present report finds, has been in the developing countries of Asia, Africa and Latin America. In 15 years military spending in the developing countries more than doubled: from $15 billion in 1960 to $39 billion in 1974 (again measured in constant 1973 dollars). The relative increase in military expenditures varied considerably from region to region, ranging from a twofold increase in Latin America to an eightfold increase in the Middle East. In the developing world as a whole, however, military expenditures have increased twice as fast as the economic base supporting them.

Although the tempo of military growth is rising faster in the developing nations, a few developed nations still dominate military spending and determine the nature and course of the worldwide arms competition. Together the U.S. and the U.S.S.R. spend almost 60 percent of the world's military outlay, a proportion even greater than their share of the world economy. (Although the two superpowers are first in military strength, they rank lower than many other nations in indicators of social well-being, the report adds.) In addition the two superpowers account for 75 percent of the arms moving in international trade, a traffic that has grown in value between 1961 and 1974 from $1 billion per year to $10 billion.

The cumulative total of world military expenditures (at present prices) since 1960 is nearly $4 trillion—substantially more than the cumulative expenditure of public funds for education, health care or any other activity of government. In the 15 years from 1960 through 1974 the cumulative military total was more than 20 times the total of all foreign economic aid.

Summarizing their findings, the authors of the 1976 report point out that in addition to "the growing potential for cataclysmic destruction," the worldwide arms buildup represents "an immediate and heavy burden on the world economy. It is destructive whether or not the weapons are put to use in war. It contributes to inflation, retards economic and social development, and diverts resources urgently needed for human well-being. Until it can be put under control it undermines the national and international security which it is intended to protect."

Satellites' Progress

Not quite seven years ago a global satellite communications service was inaugurated when an Intelsat satellite was maneuvered into a synchronous (stationary) orbit 22,300 miles above the Indian Ocean. In conjunction with similar satellites previously placed in synchronous orbits over the Atlantic and the Pacific, the Indian Ocean relay station became operational just in time to provide worldwide television coverage of man's first landing on the moon on July 20, 1969. The Indian Ocean satellite was the eighth launched by the National Aeronautics and Space Administration for Intelsat, the International Telecommunications Satellite Organization. Of the first eight launchings, two were failures. Since then 15 more Intelsats (the name also given to the satellites) have been launched with only three failures. The latest model of Intelsat, the Intelsat IV-A built by the Hughes Aircraft Company, costs about $23 million. NASA charges another $23 million to place it in a synchronous orbit.

At present four Intelsats (either IV's or IV-A's) are in full-time service: two over the Atlantic, where the traffic is heaviest, one over the Pacific and one over the Indian Ocean. Each of the four satellites is accompanied by a full-time standby Intelsat that is ready to take over in an emergency. Unlike transoceanic submarine cables, which directly connect only two points, the synchronous satellites can simultaneously receive from and transmit to many ground stations, more than 40 on four continents in the case of the Atlantic Intelsats. Overall the global system provides service to 109 countries, territories or possessions on six continents through 407 different pathways. The system makes it possible for more than a billion people to see an international event "live" on television.

Although television transmissions represent the most dramatic use of the Intelsat system, they account for less than 2 percent of the system's use. During 1975, on the average, the international satellite system was engaged in relaying television programs only about nine hours a day. In comparison, at any given moment several thousand telephone conversations are being transmitted over more than 6,600 circuits leased on a full-time basis. (One color television transmission requires the equivalent of 400 voice circuits.) In the year of the telephone's 100th anniversary telephone service accounts for 80 percent of the global satellite traffic; 15 percent represents message and data transmission, and the remaining 3 percent consists of miscellaneous occasional uses. The satellite system is now carrying more than two-thirds of all transoceanic communications.

Intelsat's total expenditure on satellites to date, including the five that failed to go into orbit, has been about $600 million. The investment by all participants in ground stations and other supporting facilities adds another $600 million. The total investment in submarine cables, which also provide a global communications network, is not available. However, TAT-6, the sixth cable to be laid between the U.S. and Europe, will cost 17 participating nations $191 million by the time it is placed in service later this year. The existing transatlantic cables, including two terminating in Canada, now provide a total of 3,339 voice circuits. TAT-6 alone will carry 4,000. Each Intelsat IV-A, which costs a total of $46 million to build and place in orbit, carries some 6,600 voice circuits. It might seem, therefore, that the satellite investment, per circuit, is substantially less than the cable investment, even if a second Intelsat is needed as a standby. On the other hand, the minimum expected lifetime of TAT-6 is 25 years compared with a seven-year design lifetime for an Intelsat IV-A. The owners of TAT-6 calculate that the total annual carrying charge for one TAT-6 voice circuit will be only $5,700, whereas carriers now pay an annual leasing charge of $34,200 for one voice circuit.

The International Telecommunications Satellite Organization is jointly owned by 92 countries. (Although they are not members of Intelsat, the U.S.S.R. leases seven Atlantic voice circuits and the People's Republic of China leases 44 Pacific circuits, which they use in conjunction with their own ground stations.) With minor exceptions each country's investment is proportional to its fractional use of the network and is readjusted annually. The U.S. investment is made through the Communications Satellite Corporation (Comsat), a publicly owned company created by an act of Congress in 1962. Comsat's share in Intelsat, originally about 65 percent, has dropped to 33.61 percent as other countries have increased their use of the global system. The five biggest users, after the U.S., are the United Kingdom, France, Japan, West Germany and Australia.

When Early Bird (Intelsat I) made possible the first transmission of live television across the Atlantic in 1965, the charge for one hour of prime-time color transmission was $22,350. Today the charge is $5,100. Comsat's monthly charge to U.S. carriers for a voice-grade half-circuit to Europe was $4,200 in 1965; today it is $2,850. (A half-circuit is one country's share of a two-way communications link.)

This month the Comsat General Corporation, a wholly owned subsidiary of Comsat, is scheduled to place in orbit the first of three Comstar satellites built by Hughes Aircraft to the specifications of the Bell System. When all three are in orbit in 1978, one will serve as a standby for the other two. Each of the two operating satellites will provide 14,400 high-quality voice circuits, which will be integrated into the Bell System's domestic communications network. Two other American companies, the Western Union Corporation and the RCA Corporation, are already using satellites for domestic communications. Western Union's Westar I was placed in service in July, 1974, and was joined by its standby, Westar II, a few months later. RCA's Satcom I was launched last December; its standby, Satcom II, is awaiting launching.

A Quorum of Quarks?

The hypothesis that certain subatomic particles are made of simpler things called quarks was introduced more than a dozen years ago, when there was almost no evidence to support it. The concept rested mainly on an aesthetic argument: by assuming that all hadrons (a class of particles that includes the proton, the neutron and the pion) are constructed of three quarks or of a quark and an antiquark, the classification of particles was greatly simplified. Where there had been 100 or more kinds of particles, there were now merely 100 ways of combining three particles. The utility of the concept was immediately apparent; what was lacking was direct evidence that the quarks actually exist. In the past two years or so that evidence has begun to appear, and as a result quarks have begun to seem more substantial, even if they are still not quite tangible.

When two particles collide at high energy, a shower of other particles is created, including some that are hadrons. If quarks did not exist, the hadrons would be expected to fly away from the collision randomly in all directions. The quark theory predicts, however, that the hadrons will be emitted in narrow "jets" oriented perpendicularly to the line of motion of the colliding particles. Earlier experiments that attempted to test this prediction were inconclusive. It has recently been confirmed for three kinds of collision by extending the measurements to higher energy.

At the Stanford Linear Accelerator Center (SLAC) the jets have been observed following the annihilation of electrons and positrons and following the collision of electrons and protons. They have been seen in the aftermath of proton-proton collisions at the Fermi National Accelerator Laboratory (Fermilab) and at the European Organization for Nuclear Research (CERN) in Geneva. The hadrons are emitted in jets because they form from pairs of quarks and antiquarks, which in turn materialize from the energy of the colliding particles. In some of the SLAC experiments it has been possible to measure the spin angular momentum of the products of the collision, and thereby to learn the spin angular momentum of the quarks. It appears to be 1/2 unit, in agreement with the theoretical prediction.

Even among those who are convinced that quarks exist there has been little agreement about how many there are. In the original theory there were three kinds; only two of them appear in the protons and neutrons of ordinary matter, the third being reserved for those particles that bear the property called strangeness. In 1964 a fourth quark was proposed, bearing the property called charm. The charmed quark helped to explain why certain processes involving strange particles are not observed, but the charm conjecture too depended heavily on aesthetic principles. There are four known particles that are classified as leptons (the electron, the muon and two kinds of neutrino) and they resemble the quarks in some of their properties; it therefore seemed there ought to be four quarks also. However appealing that notion is, it will remain unconvincing until a particle containing a charmed quark has been identified.

The first candidates were the particles discovered in 1974 at the Brookhaven National Laboratory and at SLAC, which were named *J* or psi. It was immediately proposed that the psi consists of a charmed quark and a charmed antiquark, a combination called charmonium. The nature of the psi is by no means settled, and indeed some of its discoverers have recently pointed out that certain predictions of the charmonium hypothesis are contradicted by experiment. Nevertheless, there is now strong evidence in support of the charmonium interpretation. The theory predicts an entire spectrum of charmonium states, analogous to the spectrum of energy levels in an atom. At least seven of those states have been found, and their masses and other properties agree quite well with the predictions.

Even if the charmonium interpretation of the psi is correct, the psi would not exhibit its charm overtly because the properties of the charmed quark and antiquark cancel each other. As a result a great many experiments have been directed toward creating a particle made up of a charmed quark in combination with some other kind of quark. Such particles have now almost certainly been observed at several laboratories.

Over the past three years a substantial number of events have been recorded at Fermilab in which a neutrino interacts with a proton or a neutron and gives rise to an ensemble of particles that includes two muons. These "dimuon events" were recognized immediately as a signal that some new kind of particle is being created, but at first it was not certain what kind. Recent calculations show that the particle is probably a member of some new family of hadrons, and the new family could be the one made up of particles containing charmed quarks. The dimuon experiments were conducted by a group of investigators from Harvard University, the University of Pennsylvania, the University of Wisconsin and Fermilab; they have named the new particle *Y*. Similar events have since been observed in another experiment at Fermilab, conducted by physicists from Fermilab and the California Institute of Technology.

More direct evidence for the creation of a charmed particle has recently been found in photographs made with the 15-foot bubble chamber at Fermilab by investigators from the University of Wisconsin, CERN, the University of Hawaii and the Lawrence Berkeley Laboratory of the University of California. In this case too protons and neutrons were bombarded with neutrinos, and in eight cases a distinctive pattern of particles was detected. Among the many products of each interaction the significant ones were a muon, a positron and a single strange hadron. The appearance of the positron and the strange particle cannot be explained by any conventional process, but they could be the decay products of a charmed hadron. The charmed particle would have a mass about twice that of the proton and a lifetime of less than 10^{-13} second. Those values are within the range anticipated for the charmed particle of lowest mass.

A year ago events were reported at Brookhaven and at CERN that had characteristics expected of interactions involving charmed particles, but because in each case there was only one event their interpretation was ambiguous. Two more events of the same kind have now been found in bubble-chamber photographs made at CERN by a group of 56 investigators from seven European laboratories. As in the Fermilab experiment, the particles were observed in the aftermath of neutrino interactions, and the pattern thought to signify the creation of charm was again the presence of a muon, a positron and a strange particle. The three events were found by examining some 600,000 photographs.

The apparent agreement of these diverse experiments might seem to imply that the uncertainties of the quark theory have been resolved: quarks do exist, there are four of them, and all hadrons are composed of quarks selected from among those four. Not all the recently acquired data, however, bend so meekly to the support of the established theory. For example, a group of investigators from the Lawrence Berkeley Laboratory and SLAC have created a new particle in electron-positron collisions that they designate *U* (for "unknown"). Charmed particles ought to be created by this technique, and the *U* may in fact be one. The manner of its decay, however, suggests an alternative explanation: it may be a new member of the lepton family, one much more massive than any of the four known leptons. The existence of heavy leptons is not in conflict with theory, but of course it would destroy the attractive symmetry relating leptons and quarks.

The discovery of still another new particle, the heaviest ever observed, was announced in February by a group of investigators from Columbia University, Fermilab and the State University of New York at Stony Brook. The particle, which has been named upsilon, was made at Fermilab by bombarding protons and neutrons with extremely energetic protons. It has a mass six times that of the proton and is probably too heavy to be part of the conjectured series of particles made up of charmed quarks. Any explanation of its nature is therefore highly speculative. One possibility that cannot be excluded is that the spectrum of quarks does not stop at four. Perhaps, for example, there are six quarks, and the upsilon contains one or both of the new pair, which would be more massive than the other four. If there are also additional leptons, then the lepton-quark symmetry might be restored. On the other hand, if the list does not stop at three and does not stop at four, there is no obvious reason why it should stop at six.

Act of the Gods

The sudden collapse of the Minoan civilization in the middle of the second millennium B.C. after 1,500 years of prosperous growth is commonly attributed to an invasion by Mycenaean Greeks who put Knossos and other population centers on Crete to the torch and established their own rule. The fact that at about the same time the island of Thera, some 80 miles north of Crete, was shattered by a tremendous volcanic eruption has only in recent decades become firmly established, partly by excavations on Thera by the Greek archaeologist Spyridon Marinatos and partly by geological studies in the area. Marinatos was the first to suggest that the eruption rather than the Mycenaean invasion might account for the Minoan collapse. His hypothesis was widely dismissed on the ground that Thera was destroyed perhaps half a century before Crete fell.

Reviewing the evidence recently, J. V. Luce, professor of classics at Trinity College, Dublin, has now dismissed this chronological objection. It was based from the start on a fine point of the pottery classification first established by the pioneer excavator of Knossos, Sir Arthur Evans, in the 1920's. Evans assigned the time interval from 1550 to 1500 B.C. to pottery that he viewed as being representative of the culture known as Late Minoan *IA*, and the interval from 1500 to 1450 B.C. to pottery representative of the successor culture, Late Minoan *IB*. Because the first excavations at Thera yielded no pottery of Late Minoan *IB* style it was concluded that Minoan Crete had continued to flourish for several decades after the destruction of Thera. Writing in *American Journal of Archaeology,* Luce points out that at least one example of Late Minoan *IB* pottery has now been found under the volcanic ash on Thera. The discovery removes the key objection to the Marinatos hypothesis.

Equally important, Luce suggests, are a number of other considerations: the volcanic ash at many sites in northern and eastern Crete has been positively identified as having originated at Thera; there is a growing weight of evidence that whatever struck the Minoans did so simultaneously over a wide area of Crete; plots of the distribution of debris at certain Cretan sites suggest the action of tidal waves. Luce reconstructs the catastrophe, which probably came during the year 1470 B.C., as follows:

1. Earthquakes cause major damage on Thera and some damage on Crete.
2. The dormant volcano on Thera becomes active and ejects pumice; minor earthquakes produce small tidal waves, and some of the Thera pumice comes ashore on Crete. Collections of the pumice are offered to the Cretan earth goddess at shrines at Kato Zakro and elsewhere. The alarmed inhabitants of Thera evacuate the island.
3. A huge volcanic explosion buries all of deserted Thera deep in pumice and ash; the ash cloud spreads over hundreds of square miles. Strong earthquakes overturn lamps and braziers in many buildings on Crete and entire communities are set afire.
4. The volcanic cone on Thera collapses, initiating major tidal waves. Walls of water more than 100 feet high strike the northern and eastern coasts of Crete. Fleets are wrecked; thousands are drowned.
5. The ash fall in eastern Crete, several feet thick, chokes wells and streams and destroys all vegetation. Devastation is total on Thera, very severe in eastern Crete and substantial in the Knossos area. Survivors move to western Crete or overseas; the losses in population, shipping and stores are a mortal blow to Minoan power.
6. Five or 10 years later unopposed Mycenaean Greeks establish a dynasty at Knossos.

Ecological Staircase

Geological and biological processes have combined to produce an instructive "ecological staircase" on the Pacific coast north of San Francisco: a series of five marine terraces formed by uplift and erosion, each covered with increasingly depleted soils and each bearing a distinct flora that ranges in a few miles from grassland through towering redwood forest to a unique climax forest of undernourished pygmy pines and shrubs. The staircase and its vegetation were described by William W. Fox in a recent issue of *California Geology.*

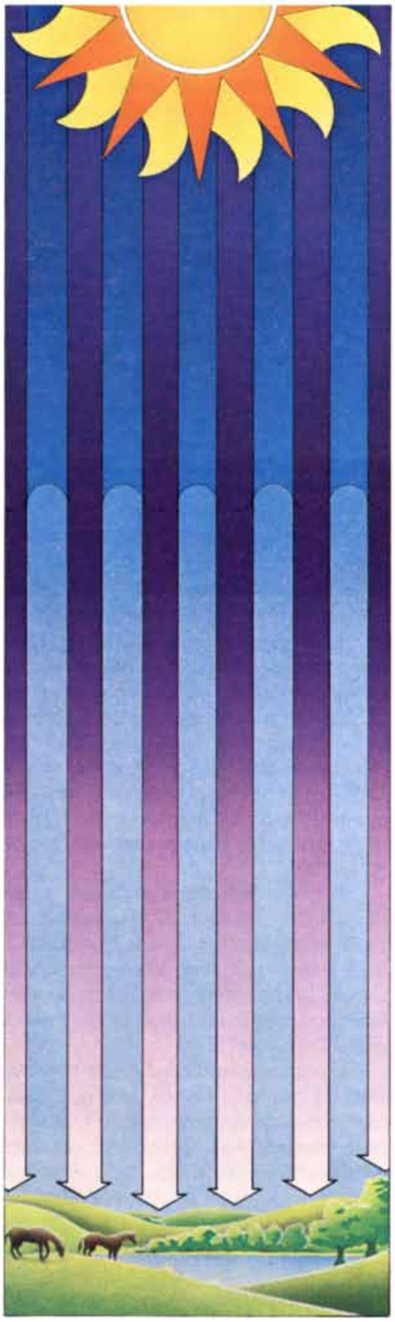
The bedrock in this segment of the Mendocino County coast is a marine sedimentary sandstone, largely quartz and feldspar. It has been rising quite steadily for some 500,000 years at the rate of two or three centimeters per century. Periodically, during times of high sea level, the coastal rim of the rising land was subjected to battering by ocean waves, which produced vertical cliffs and a fringe of rocky debris and sand. Then, during ice ages, the accumulated debris was elevated on its sandstone shelf, forming the tread of another step of the staircase; the wave-eroded risers range from about 100 to 150 feet in height. A fringe of wind-driven sand dunes formed along the seaward edge of each step.

The first tread, a broad terrace along the present coastal bluff, is carpeted by a rich black prairie soil that supports—and over the years has been created by—a luxuriant cover of perennial grasses and flowers. The wind-driven salt spray discourages the growth of trees. The second step, about a mile from the sea and 200 feet above it, drenched by heavy rainfall and out of the range of the salt spray, is very different. Much of its seaward rim of dunes has long been covered by forest, and the decaying compost of leaves and needles has altered the sandy soil, which now supports a towering rain forest of redwoods, firs and hemlocks. The flat inland region of the same step is more sparsely forested. Its old ocean-beach surface is waterlogged much of the year because the dunes obstruct runoff drainage and springs in the dunes add moisture; then, in the short summer season, the flatlands are parched.

These conditions are exaggerated on the higher steps, where the nutrients have been leached out of the soils in the process known as podzolization. The dunes are partly podzolized, and they sustain pines, dwarfed redwoods and shrubs. The flat ocean-beach deposits have become true podzols. Here the top layers of the soil are completely sterile. Even the feldspar has been washed out of the sand, leaving quartz, whose fertility is "no better than that of pulverized glass." In places the lack of drainage has produced permanent bogs that are the most southerly-known habitat of sphagnum moss. In other places some tree cover remains, but it is what Fox calls "nature's own bonsai forests": pygmy forests of Bolander pine, bishop pine, Mendocino cypress and shrubs such as manzanita, rhododendron and huckleberry. A 100-year-old pygmy pine may be from five to 10 feet high, with a two-and-a-half-inch trunk. The extreme pygmy-forest condition, according to Hans Jenny of the University of California at Berkeley, is "species-poor and space-unsaturated," with a quarter of the ground

DP Science Dialogue

Notes and observations from IBM that may prove of interest to the scientific and engineering communities



Computer Models Throw New Light on Ozone Depletion

Many scientists suspect that nuclear explosives, supersonic jets, and aerosol sprays may produce trace substances that could significantly reduce the supply of ozone (O_3) in the atmosphere. Should this occur, increased ultraviolet radiation would reach the earth with important effects on living conditions.

A thinner ozone layer could result in harm to plants and crops, affect life in the oceans and eventually change global weather patterns. Of most immediate concern is a possible rise in human skin cancer.

All this points to the growing need for greater knowledge of broad atmospheric phenomena . . . and scientists are hard at work to determine the validity of these projections. There is no practical way, however, to enclose the atmosphere in a laboratory for controlled experiments. So scientists are using the computer to manipulate model "atmospheres" mathematically for better understanding of the interplay of systemic forces.

A number of such studies have been made by IBM scientists, including Dr. J. V. Dave and Dr. Paul Halpern of the IBM Scientific Center at Palo Alto, California, and Dr. Norman Braslau of the IBM Thomas J. Watson Research Center at Yorktown Heights, New York. According to Dr. Dave, any approximation of realistic sky phenomena through mathematical models becomes extraordinarily complex.

"Even the most powerful computers can't provide a meaningful evaluation of actual atmospheric conditions over the earth at a given time," he says. "However, my colleagues and I are seeking to come closer to such a comprehensive model through the introduction of more and more detail."

He cites a recent study he conducted with Dr. Halpern on the effects of changes in ozone on the ultraviolet radiation received at sea level. Specifically, they sought to determine what effect given decreases in ozone would have both on direct solar radiation and on the radiation diffused through the sky as a result of "scattering"—the deflection of rays by dust and clouds.

Using the model atmosphere technique, they introduced a complex array of parameters. Among them were six different concentrations of ozone; the dust and ozone content of each of 50 layers of the atmosphere; eleven ultraviolet wavelengths (since different wavelengths produce different effects); nine positions of the sun, from zenith to low on the horizon; and two indices of the reflectivity of the earth's surface.

The "building blocks" of such a model atmosphere are mathematical equations computed with varying data. In this study, computations of the direct component of ultraviolet radiation followed a fairly straightforward equation, but those for the diffuse component were very tedious and time-consuming. In such situations, says Dr. Dave, the computer is particularly helpful. The computations were run on the Center's IBM System/370 Model 145 under the

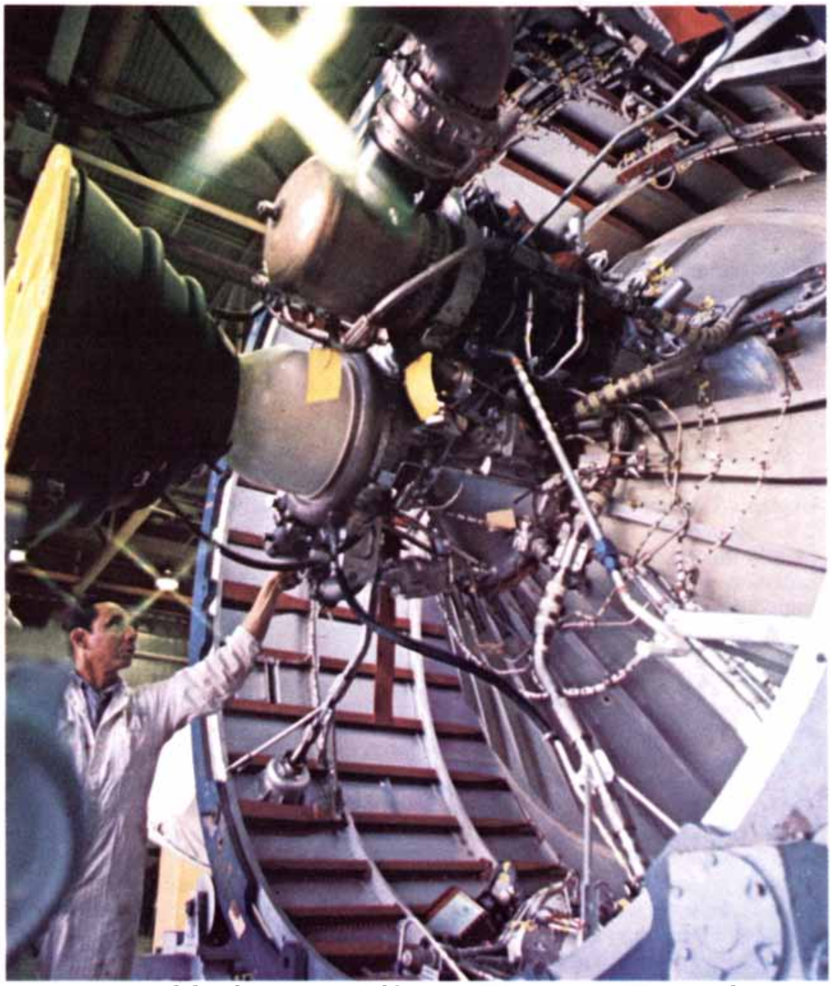
Ozone in the atmosphere acts as a screen against solar radiation. IBM scientists are simulating the process with a computer to determine the effects of varying amounts of ozone.

(Continued on next page)

A New Data Base Concept for Engineers

High personnel costs . . . stringent reliability requirements . . . expanding documentation needs. They all make it increasingly important for engineering firms to assure that up-to-date, accurate information is readily available to every user within the organization.

Traditionally, many engineering and manufacturing companies have treated every department as a separate entity. Each would have its own data base and computer programs. With that approach to information handling, however, data corrections or modifications made in one area might have taken a week or longer to reach other departments. The impact on the company's ability to respond to an urgent bid request, or to solve a production problem, could be critical.



An integrated data base approach like AEIMS helps Martin Marietta design and produce the Titan III Space Launch Vehicle.

To help improve the speed and accuracy of communication between departments, IBM has developed the Administrative Engineering Information Management System (AEIMS) concept. It is a terminal-oriented approach to storing and updating information in a central data base . . . from initial contract award to final product.

The AEIMS concept is designed to eliminate the need for redundant data files, enable many different users to share information when they need it, and help prevent the discrepancies that often crop up in separate data bases.

The Data Systems division of Martin Marietta's Aerospace Group in Denver has already established a system using a similar concept. Called the Technical Requirements Management System (TRMS), Martin Marietta's data base approach is made up of various subsystems corresponding to the functions of requirements identification, planning and scheduling, configuration verification, document status and parts data. The TRMS system uses IBM's Information Management System (IMS) which runs on Martin Marietta's System/370 Model 168 computer based in Orlando, Florida.

"Our greatest administrative problem used to be lack of crosstalk among departments," says David Lucero, chief of configurations management operations. "Now that data entry and corrections can be made directly at our IBM 3270 display terminals, we can operate with the assurance that all of us have exactly the same up-to-date information."

Another benefit is reduced paper generation. Instead of requesting extensive reports on a scheduled basis and then analyzing them manually, people in each department can now select the specific information they need, when they need it.

From an operational point of view, Don Edmunds, TRMS product manager for data systems, says: "Our data base system was developed in close consultation with users. As a result, it is flexible enough to be effective for everything — from our smallest to our most complex projects. Most importantly, the system costs are being recovered through productivity gains."

Ozone Depletion...

(Continued from first page)

interactive Conversational Monitor System (CMS) and required about 30 hours of computer time.

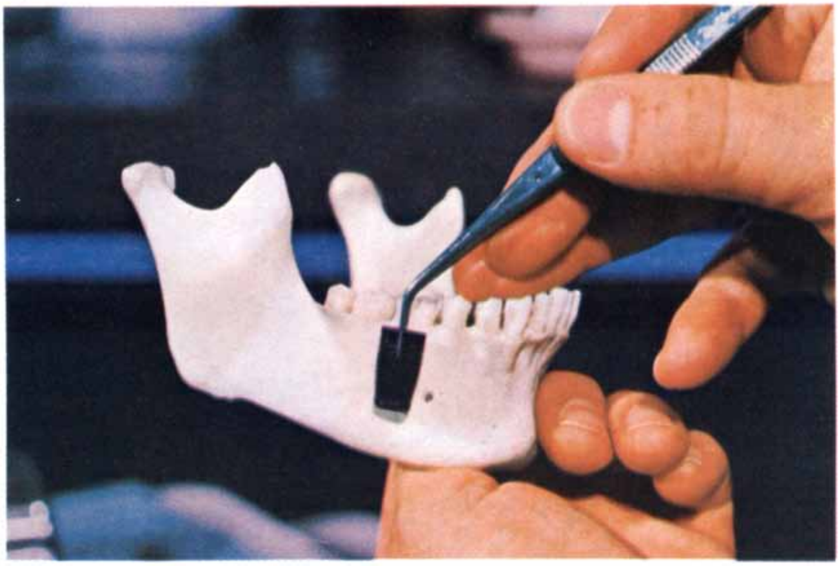
The study showed, in particular, that a given decrease in the atmospheric ozone content results in a strong increase in the direct ultraviolet radiation received at sea level. On the other hand, the increase in diffuse radiation is relatively moderate. Because a large part of the ultraviolet radiation received at the ground is diffuse, the relationship between the amount of atmospheric ozone and the solar energy received at ground level can be a complicated one.

Dr. Dave emphasizes the continuing need for accurate simulation of solar ultraviolet radiation through realistic models of the atmosphere. Results of such studies can then be used for the analysis of actual measurements. At the same time, the responses of plants and animals to changes in ultraviolet radiation can be determined. This can lead to a meaningful estimate of the possible biological damage that might be caused by the depletion of ozone as a result of human activities.

Scientists Perfect Dental Implantation Technique

Until recently, replacing a lost tooth almost always required bridgework—anchoring the false tooth to the healthy ones surrounding it. Now, using a new technique, dentists at the University of Southern California are implanting artificial teeth directly into the jawbone. Their successful procedure uses vitreous carbon, a material new to dentistry, whose characteristics were tested on the university's IBM computer.

"The jaw is subject to constantly changing stresses that are created as a person chews," says Dr. Ronald Voss, director of the clinical implantation program at USC. "We had to be sure that artificial teeth made of vitreous carbon would approximate the behavior of normal teeth very closely. If there is insufficient stress, the jawbone will start to deteriorate. On the other hand, if the stress is too great, either the bone or the implant itself will break down."



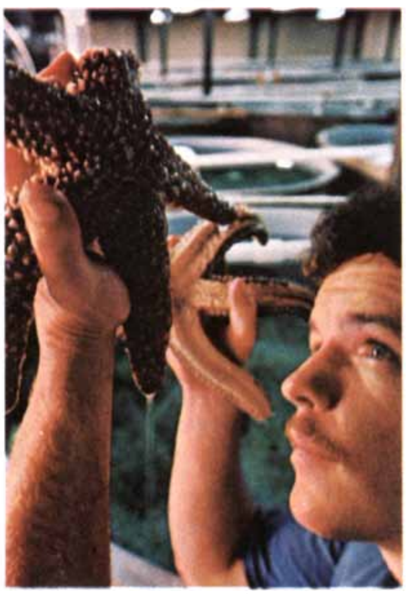
A vitreous carbon rod whose use in dentistry was developed with the aid of computer analysis, is implanted in the jawbone as the foundation of a dental crown.

Using an IBM System/370, Dr. Dale Grenoble, associate professor of dentistry, and biomechanics engineer Albert Knoell developed a mathematical model of the carbon implants to determine their maximum stress patterns. "This computer analysis focused on one part of the problem by helping us verify the structural integrity of the implant design without incurring the time and expense of building actual models," says Knoell.

The implant operation, which normally takes about half an hour, involves drilling a hole into the jawbone and inserting a grooved vitreous carbon rod. After a healing process of several months, a stainless steel post is anchored to the rod and a dental crown is attached.

So far, implants have been performed successfully on more than 300 patients. "Our oldest implants have been in place just over four years," says Dr. Voss. "We believe that in certain situations, implants can be more aesthetic and functionally efficient than any other dental replacements."

Environmental Research Aided by TSO



Alive and well in Redondo Beach, this starfish lives in a mixture of ordinary and recycled sea water.

For snails, starfish and sea urchins who live in the King Harbor area of the California coast, life has never been better. At least that's what studies of the sea creatures seem to indicate to scientists at Southern California Edison.

According to Kevin A. Muench, chief marine research scientist and director of the utility's Redondo Beach laboratory, preliminary results of a novel research program show that growth of marine life is expanding, due to the warmer temperatures of water that has been used to cool power plant condensers and has then been returned to the harbor.

Key to the study is the interactive IBM Time Sharing Option (TSO), together with PL/1, a powerful IBM language. Using them, scientists can speed data to the central computer for analysis and return. Throughout Southern California Edison, some 250 people —including scientists and engineers involved with construction and power supply, and financial and customer service personnel—can sit down at remote terminals and communicate directly with the computer whenever they need to.

Robert Umbaugh, Edison's data processing manager, reports, "By putting TSO on System/370 Model 168, we've eliminated outside timesharing charges of $400,000 a year. TSO, PL/1 and VS BASIC have made the computer as much a part of the engineer's tools as a desk or pencil. With them we're getting a lot more for less."

While long active in marine monitoring, Southern California Edison began its specific research program at Redondo Beach over a year ago to assess any potential disruption of biological patterns caused by water which returns to the harbor several degrees warmer.

The long-range goal of the program is predictive biological modeling, which will help the utility plan and operate future power plants in ways that will maintain a balanced marine life.

Advertisement

At Mobil's Paulsboro Research Laboratory, Dr. C. Dwight Prater (left) and Dr. Vern W. Weekman Jr., check the latest catalyst aging run as methanol is converted into gasoline.

Energy Option for the Future: Gasoline Without Oil

Turning coal into gasoline may sound like an impossible dream . . . but it can be done. Using a unique catalytic process, scientists at a Mobil Research and Development Corporation laboratory in Paulsboro, New Jersey, have successfully produced high-octane gasoline directly from methanol.

Coal, wood chips and even garbage can be converted into methanol by well-known techniques. Mobil's process completes the transformation into gasoline in one relatively simple operation.

Under a cost-sharing contract with the government's Energy Research and Development Administration, researchers at Mobil are operating several test units. With these "miniature refineries" they explore the effects of variations in factors like pressure, temperature and flow rates.

Products from thousands of test runs are analyzed in a chromatograph, which determines the chemical composition of the gasoline.

Data from the pilot plants are transmitted to Mobil's IBM System/370 Model 158 in Princeton, New Jersey.

"Gasoline is a mixture of hundreds of different chemical compounds, all basically hydrocarbons," explains Dr. C. Dwight Prater, manager of process research and development. "Our goal is to produce high quality gasoline from methanol while reducing production costs, which are quite high right now."

Using A Programming Language Shared Variables (APLSV), a powerful programming language developed by IBM, Mobil has created extensive data bases that contain information like the catalyst aging rate, the amount of gases in each test run, the octane and the overall quality of the gasoline produced in each run.

"APLSV is ideal for handling the long streams of numbers our kind of research produces," says Dr. Vern W. Weekman Jr., manager of systems research. "Its interactive features and ability to organize data give us a speed and flexibility to retrieve information unmatched by any other programming language.

"As a result, we estimate that our overall turnaround time is five times faster since we switched to APLSV. That means our engineers can pursue an idea while it is hot, instead of waiting days for the data to be assembled."

In addition to providing built-in control features, the language is easy to learn. Dr. Prater says, "APLSV's similarity to linear algebra makes it especially logical to mathematicians and scientists. It also appeals to them by making some of the more tedious aspects of their research—such as plotting graphs—automatic."

Dr. S. L. Meisel, vice president of research, looks at APLSV this way: "Any tool that gives our scientists more time for creative research puts us closer to our goal of translating ideas into useful products and processes. The specific goal in this case is the economic, large-scale production of gasoline from methanol."

DP Science Dialogue is concerned with topics that may prove of interest to the scientific and engineering communities. Your comments and suggestions are welcome. Just write: Editor, DP Science Dialogue, IBM Data Processing Division, 1133 Westchester Ave., White Plains, N.Y. 10604.

IBM®
Data Processing Division

area bare or covered only by lichens. The forest and its podzol, he says, come close to being a terminal steady-state ecosystem.

Women's Liberation

Technology has revolutionized almost every aspect of modern life, not excluding life in the home. In particular the introduction of laborsaving machinery into the home has altered the role of the housewife—not entirely in ways one might expect. In the journal *Technology and Culture* Ruth S. Cowan of the State University of New York at Stony Brook has analyzed the development of household technology and its impact on middle-class American women between the two world wars.

The traditional interpretation of the impact of technology on the modern family is as follows. Before industrialization the family was the basic social unit, which produced and processed virtually everything it needed for its own support. In preindustrial families women had much to do, and their time was almost totally occupied by household tasks. In an industrialized society the household was no longer the focus of production. As the function of the family diminished, the bonds that held it together loosened. In such postindustrial families the women have little to do, and many suffer from anxiety about their role.

Cowan has examined both preindustrial and postindustrial families as they were depicted in the articles and the advertisements in women's magazines in the 1920's and 1930's, and she concludes that the traditional interpretation is inadequate. Although the middle-class housewife in the preindustrial family had much to occupy her time, she commonly had paid or unpaid domestic help. As electrical appliances such as irons, stoves and washing machines were introduced into the home, however, the housewife came to do increasingly more of her own work. In the process she was reduced from the role of a personnel manager to that of a household worker.

Furthermore, as the number of household assistants declined, the number of household tasks increased. Although appliances reduced the amount of physical labor associated with many household tasks, they did not appreciably reduce the amount of time necessary to complete those tasks. The average middle-class housewife had to spend more time with her children, she had to cook most of the meals and typically she had to do most of the shopping for the family's food and clothing. Since these tasks are aided by expert knowledge, she was now expected to read about those subjects. In short, after World War I the middle-class housewife was expected to be competent at a wide range of tasks that had formerly not been in her domain, or that had not even existed before industrialization.

As the job of housewife changed, the ideologies connected with the job changed also. Cowan demonstrates that in the women's magazines there was a clearly perceptible difference in the attitudes that women brought to housework before and after World War I. Before the war doing housework without domestic help was portrayed as a chore that had to be tolerated until a qualified servant could be found. After the war it was portrayed not as a chore but as an expression of the housewife's love for her family and as a duty that could not possibly be entrusted to a servant.

When societies become industrialized, the structure of the work force becomes highly differentiated, individual workers become more specialized and the emotional context of the work disappears. According to the traditional interpretation of the impact of technology on the family, the same effects should be seen in the home. What is observed, Cowan concludes, is exactly the reverse. "The industrialization of the home was a process very different from the industrialization of other means of production," she writes, "and the impact of that process was neither what we have been led to believe it was nor what students of the other industrial revolutions would have been led to predict."

Super Slurper

One need no longer cry over spilled milk, provided that the source of sorrow is the prospect of cleaning up the mess rather than the loss of the cost of a quart of milk. With a supply of "super slurper" the miserable pool can be quickly transformed into a crumbly mass and swept up. Super slurper is the name the Agricultural Research Service of the U.S. Department of Agriculture has given to a hybrid material it has devised for absorbing aqueous liquids in prodigious amounts (for distilled water, 1,400 times the weight of the material).

The material is made from starch and acrylonitrile. The starch, obtained from corn and other farm crops, is a renewable resource; it therefore serves to extend the petroleum resources from which acrylonitrile is made. The combination, half starch and half acrylonitrile, is treated with lye to yield a product that is identified chemically as a hydrolyzed starchpolyacrylonitrile graft copolymer. Made as film, flakes, powder or a mat, the material after absorbing water is like a soft, rubbery ice, except that it is not cold.

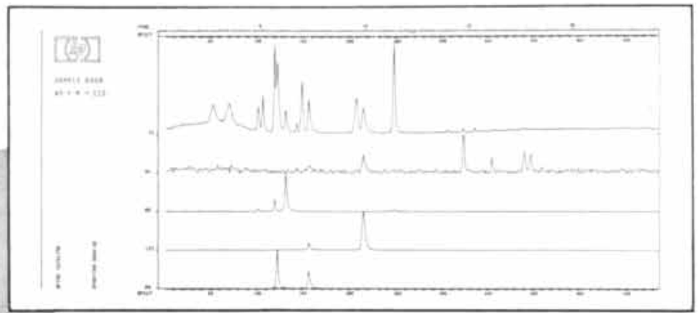
The product can be mixed with or added as a coating to such materials as sand, straw, sawdust, seeds, roots, fibers, flour, gelatin or starch. It can hold water in soils, diapers, surgical pads and many other things. In experiments at Iowa State University it was added to sand in which oats and grass were planted; the plants grew much better in the treated sand than in untreated sand. The Northern Regional Research Center of the Agricultural Research Service in Peoria, Ill., where the material was developed, has demonstrated that the product can gel water to make industrial and animal wastes, mud and sewage sludge easier to handle.

Top View

From the ground a meteor streaking across the sky is seen as a celestial event. From above, however, it is seen as the kind of event it really is: a phenomenon of the earth's atmosphere. This was made vividly apparent to Captain John S. S. Kim of the Air Weather Service of the U.S. Air Force as he examined pictures made by one of the satellites of the Defense Meteorological Satellite Program. He describes his discovery in the magazine *Weatherwise.*

In the satellite photograph below, made on November 19, 1974, a meteor is seen as the streak of light in the lower left corner. The bright spot at the upper right is the island of Oahu in the Pacific. At approximately the time when the satellite was overhead Air Force personnel stationed at Johnston Island saw the meteor from the ground as an unusually bright "falling star."

A meteor* (lower left) *is seen from a satellite



This chromatogram plot from GC/MS analysis of a water sample helps identify trace amounts of volatile organic compounds. Peaks in the single-ion chromatograms (the four lower lines) indicate where certain classes of compounds may be found in the corresponding total-ion chromatogram (upper line). On command, the data system recalls the appropriate spectrum from computer memory, compares it with reference spectra in its tape library, and identifies the compound with a stated probability of match.

A sensitive, versatile analytical tool for environmental chemistry.

Ultrasensitive measurement and identification of organic compounds can be accomplished by combining gas chromatography and mass spectroscopy. The Trace Organics Laboratory at the Foremost Foods Company Research Center, a major food industry laboratory, is centered around a Hewlett-Packard 5981A Gas Chromatograph/Mass Spectrometer (GC/MS) and an HP 5933A Data System for the analysis of organic compounds in drinking water, food, and biochemical systems.

Using a Hewlett-Packard GC/MS and data system, Foremost Foods' research lab can solve challenging problems for scientists within the company and for a variety of clients in its contract research services function. Analyses range in scope from the part-per-trillion identification of volatile organic compounds in drinking water to the identification of sterols in the lipids of antarctic krill.

New concern about environmental pollution has centered on a huge number of organic chemicals—for example the halogenated hydrocarbons such as chloroform that the Environmental Protection Agency has recently identified in drinking waters. Careful analyses have revealed several hundred other trace compounds in the part-per-billion and part-per-trillion ranges, and have focused attention on analytical techniques capable of identifying and quantifying trace amounts of these compounds. Effective analysis demands a versatile analytical tool that can provide both qualitative and quantitative answers, and data must be reproducible and easily managed to provide ready access and archival documentation.

Specialized analysis of volatile organics from aqueous media at the Foremost laboratory has been easily adapted to the HP 5981A GC/MS and the HP 5830 series gas chromatograph. The method involves "scrubbing" volatile organics from a water sample using a stream of helium in a special chamber. The gas stream is then passed through a small chromatographic column filled with a porous polymer material that traps the volatiles. Once "absorbed" in the trap, the volatiles are "desorbed" into the GC/MS for analysis.

Foremost's research group selected the HP system for its overall flexibility, citing in particular the GC/MS' dodecapole mass filter, which provides a scan range of 3-1000 atomic mass units, is very stable, and is easy to calibrate and tune. Sensitivity to picogram quantities of some materials is readily achieved in its single ion monitoring mode.

The HP 5933A Data System, with its 16K-core-memory minicomputer and a dual disc drive for input/output tasks and data storage, is fully programmed to control the functions of the mass spectrometer as well as to acquire data on its disc.

Foremost's Trace Organics Laboratory finds that HP's extremely powerful software provides impressive flexibility in data management and storage for a large number of in-house projects and contract clients. The libraries of reference mass spectra of chemical compounds that are available with the system, plus those accessible by using the data system as a time share system, allow the group to search approximately 100,000 spectra to find suitable matches.

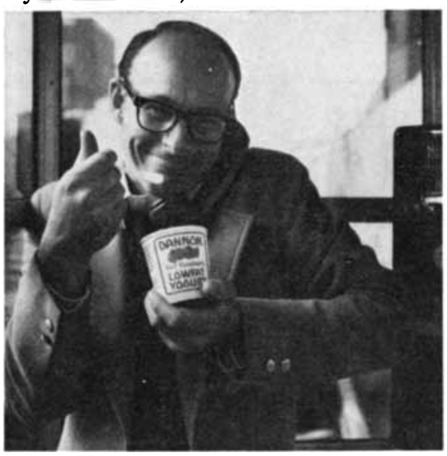
Dannon Yogurt. If you don't always eat right, it's the right thing to eat.

Every day, millions of people give up eating. For snacking.

Well, if you find yourself doing more eating on the run than at a table, make sure you're eating Dannon Yogurt.

Our label shows you that Dannon is high in protein, calcium and other things nutritionists say are good for you.

It also shows that, unlike so many snack foods, Dannon is low in fat, contains no starch, no gelatin or other thickeners. And none of those hard-to-pronounce additives. Because Dannon Yogurt is 100% natural. Not just "natural flavor," but natural everything. No artificial anything.



Dannon is reasonable in calories, too. Especially when you consider how satisfying and nutritious it is.

What's more, Dannon gives you the benefits of yogurt cultures. They make yogurt one of the easiest foods to digest, and have been credited with other healthful properties too.

In many other yogurts—mainly pre-mixed or Swiss style—the cultures are destroyed during processing.

We created a whole culture of yogurt lovers.

Dannon outsells all other brands. For a number of good reasons.

For example, we go out of our way to get the best natural ingredients: to Eastern Europe for strawberries, to the West Coast for boysenberries, and we go to Canada for blueberries. (Maybe the reason that other yogurts don't come close to the taste of Dannon is that other yogurt makers don't go quite as far as we do.)

And it's the yogurt delivered direct to your store "from Dannon to dairycase." So if it tastes fresher, that's because it *is* fresher.

Dieters aren't the only people who are big on Dannon®

Today, almost everybody's eating Dannon. It makes a quick, delicious breakfast, a light but filling lunch, and of course you can't beat it as a high nutrition dessert or snack. Spoon it out of the cup as is, or mix with cottage cheese, fresh fruit, peanut butter, honey, or what-have-you.

A suggestion for beginners: since plain yogurt may be a bit tart, start with Dannon fruit yogurts—strawberry, blueberry, red raspberry, and others.

For more facts, including some unexpectedly delicious ways to eat Dannon, write for our booklet, "Yogurt and You." Dannon, 22-11 38th Avenue, Long Island City, N.Y. 11101. It's free and it will give you more reasons why Dannon is the right thing to eat—even if you always eat right.

Catastrophe Theory

Things that change suddenly, by fits and starts, have long resisted mathematical analysis. A method derived from topology describes these phenomena as examples of seven "elementary catastrophes"

by E. C. Zeeman

Scientists often describe events by constructing a mathematical model. Indeed, when such a model is particularly successful, it is said not only to describe the events but also to "explain" them; if the model can be reduced to a simple equation, it may even be called a law of nature. For 300 years the preeminent method in building such models has been the differential calculus invented by Newton and Leibniz. Newton himself expressed his laws of motion and gravitation in terms of differential equations, and James Clerk Maxwell employed them in his theory of electromagnetism. Einstein's general theory of relativity also culminates in a set of differential equations, and to these examples could be added many less celebrated ones. Nevertheless, as a descriptive language differential equations have an inherent limitation: they can describe only those phenomena where change is smooth and continuous. In mathematical terms, the solutions to a differential equation must be functions that are differentiable. Relatively few phenomena are that orderly and well behaved; on the contrary, the world is full of sudden transformations and unpredictable divergences, which call for functions that are not differentiable.

A mathematical method for dealing with discontinuous and divergent phenomena has only recently been developed. The method has the potential for describing the evolution of forms in all aspects of nature, and hence it embodies a theory of great generality; it can be applied with particular effectiveness in those situations where gradually changing forces or motivations lead to abrupt changes in behavior. For this reason the method has been named catastrophe theory. Many events in physics can now be recognized as examples of mathematical catastrophes. Ultimately, however, the most important applications of the theory may be in biology and the social sciences, where discontinuous and divergent phenomena are ubiquitous and where other mathematical techniques have so far proved ineffective. Catastrophe theory could thus provide a mathematical language for the hitherto "inexact" sciences.

Catastrophe theory is the invention of René Thom of the Institut des Hautes Études Scientifique at Bures-sur-Yvette in France. He presented his ideas in a book published in 1972, *Stabilité Structurelle et Morphogénèse;* an English translation by David H. Fowler of the University of Warwick has recently been published. The theory is derived from topology, the branch of mathematics concerned with the properties of surfaces in many dimensions. Topology is involved because the underlying forces in nature can be described by smooth surfaces of equilibrium; it is when the equilibrium breaks down that catastrophes occur. The problem for catastrophe theory is therefore to describe the shapes of all possible equilibrium surfaces. Thom has solved this problem in terms of a few archetypal forms, which he calls the elementary catastrophes. For processes controlled by no more than four factors Thom has shown that there are just seven elementary catastrophes. The proof of Thom's theorem is a difficult one, but the results of the proof are relatively easy to comprehend. The elementary catastrophes themselves can be understood and applied to problems in the sciences without reference to the proof.

A Model of Aggression

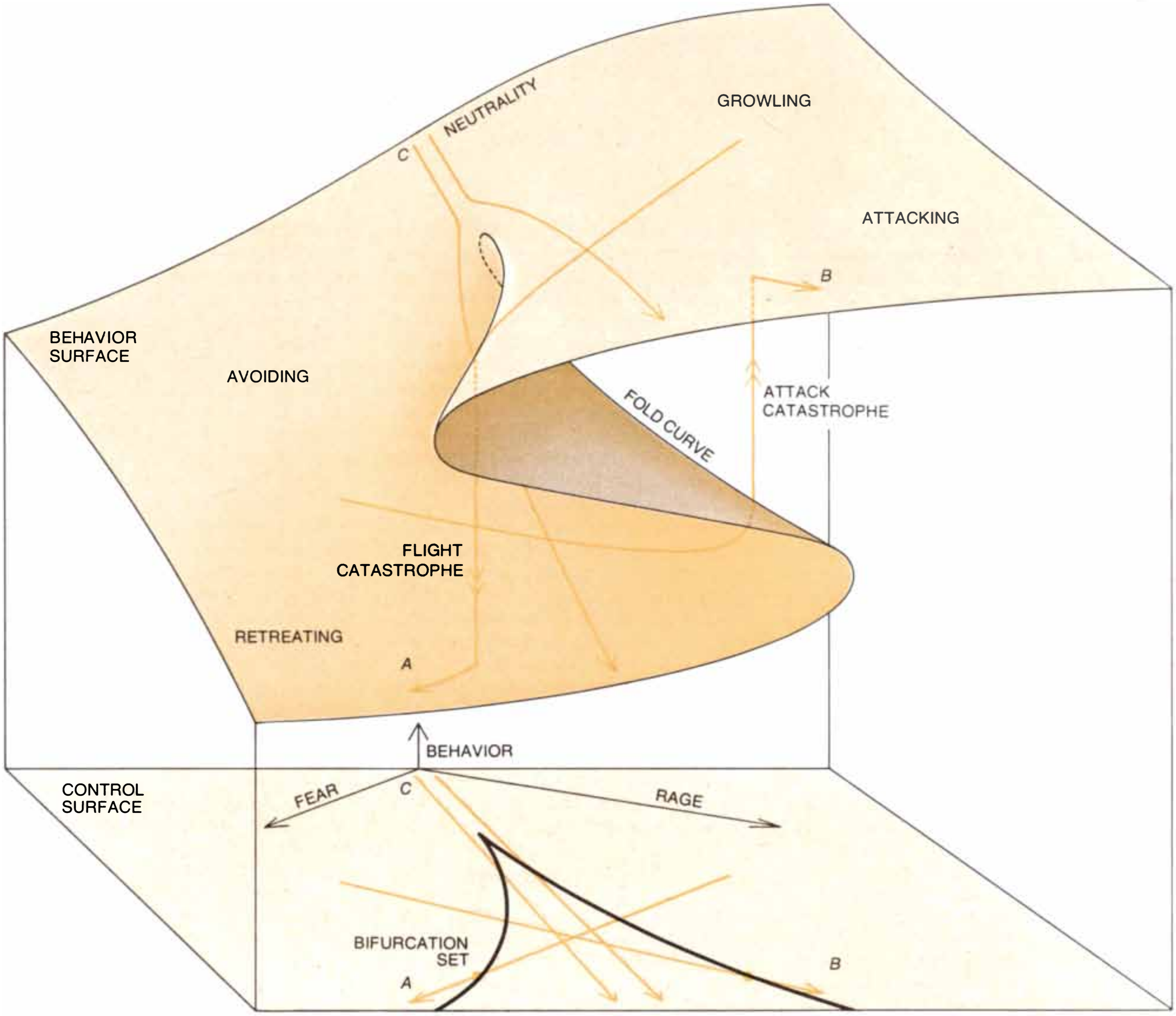
The nature of the models derived from catastrophe theory can best be illustrated by example, and I shall begin by considering a model of aggression in the dog. Konrad Z. Lorenz has pointed out that aggressive behavior is influenced by two conflicting drives, rage and fear, and he has proposed that in the dog these factors can be measured with some reliability. A dog's rage is correlated with the degree to which its mouth is open or its teeth are bared; its fear is reflected by how much its ears are flattened back. By employing facial expression as an indicator of the dog's emotional state we can attempt to learn how the dog's behavior varies as a function of its mood.

If only one of the conflicting emotional factors is present, the response of the dog is relatively easy to predict. If the dog is enraged but not afraid, then some aggressive action, such as attacking, can be expected. When the dog is frightened but is not provoked to anger, aggression becomes improbable and the dog will most likely flee. Prediction is also straightforward if neither stimulus is present; then the dog is likely to express some neutral kind of behavior unrelated to either aggression or submission.

What if the dog is made to feel both rage and fear simultaneously? The two controlling factors are then in direct conflict. Simple models that cannot accommodate discontinuity might predict that the two stimuli would cancel each other, leading again to neutral behavior. That prediction merely reveals the shortcomings of such simplistic models, since neutrality is in fact the least likely behavior. When a dog is both angry and frightened, the probabilities of both extreme modes of behavior are high; the dog may attack or it may flee, but it will not remain indifferent. It is the strength of the model derived from catastrophe theory that it can account for this bimodal distribution of probabilities. Moreover, the model provides a basis for predicting, under particular circumstances, which behavior the dog will choose.

To construct the model we first plot the two control parameters, rage and fear, as axes on a horizontal plane, called the control surface. The behavior of the dog is then measured on a third axis, the behavior axis, which is perpendicular to the first two. We might assume that there is a smooth continuum of possible modes of behavior, ranging, for example, from outright retreat through cowering, avoidance, neutrality, growling and snarling to attacking. The most aggressive modes of behavior are assigned the largest values on the behavior axis, the least aggressive the smallest values. For each point on the control surface (that is, for each combination of rage and fear) there is at least one most probable behavior, which we represent as a point directly above the point on the control surface and at a height appropriate to the behavior. For many points on the control surface, where either rage or fear is predominant, there will be just one behavior point. Near the center of the graph, however, where rage and fear are roughly equal, each point on the control surface has two behavior points, one at a large value on the behavior axis representing aggressive action, the other at a small value representing submissive action. In addition we can note a third point that will always fall between these two, representing the least likely neutral behavior.

If the behavior points for the entire control surface are plotted and then connected, they form a smooth surface: the behavior surface. The surface has an overall slope from high values where rage predominates to low values in the region where fear is the prevailing state of mind, but the slope is not its most distinctive feature. Catastrophe theory reveals that in the middle of the surface there must be a smooth double fold, creating a pleat without creases, which grows narrower from the front of the surface to the back and eventually disappears in a singular point where the three sheets of the pleat come together [*see illustration below*]. It is the pleat that gives the model its most interesting characteristics. All the



AGGRESSION IN DOGS can be described by a model based on one of the elementary catastrophes. The model assumes that aggressive behavior is controlled by two conflicting factors, rage and fear, which are plotted as axes on a horizontal plane: the control surface. The behavior of the dog, which ranges from attacking to retreating, is represented on a vertical axis. For any combination of rage and fear, and thus for any point on the control surface, there is at least one likely form of behavior, indicated as a point above the corresponding point on the control surface and at the appropriate height on the behavior axis. The set of all such points makes up the behavior surface. In most cases there is only one probable mode of behavior, but where rage and fear are roughly equal there are two modes: a dog both angry and fearful may either attack or retreat. Hence in the middle of the graph there are two sheets representing likely behavior, and these are connected by a third sheet to make a continuous, pleated surface. The third or middle sheet, shown in gray, has a different significance from the other two sheets: it represents least likely behavior, in this case neutrality. Toward the origin the pleat in the behavior surface becomes narrower, and eventually it vanishes. The line defining the edges of the pleat is called the fold curve, and its projection onto the control surface is a cusp-shaped curve. Because the cusp marks the boundary where the behavior becomes bimodal it is called the bifurcation set and the model is called a cusp catastrophe. If an angry dog is made more fearful, its mood follows the trajectory *A* on the control surface. The corresponding path on the behavior surface moves to the left on the top sheet until it reaches the fold curve; the top sheet then vanishes, and the path must jump abruptly to the bottom sheet. Thus the dog abandons its attack and suddenly flees. Similarly, a frightened dog that is angered follows the trajectory *B*. The dog remains on the bottom sheet until that sheet disappears, then as it jumps to the top sheet it stops cowering and suddenly attacks. A dog that is angered and frightened at the same time must follow one of the two trajectories at *C*. Whether it moves onto the top sheet and becomes aggressive or onto the bottom sheet and becomes submissive depends critically on the values of rage and fear. A small change in the stimuli can produce a large change in behavior: the phenomenon is divergent.

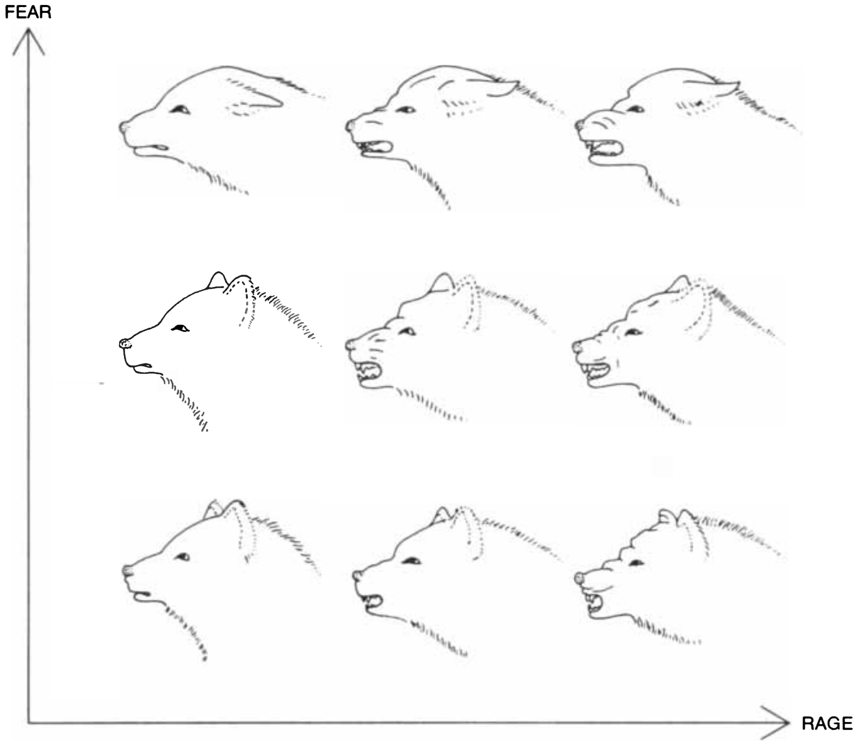
points on the behavior surface represent the most probable behavior of the dog, with the exception of those on the middle sheet, which represent least probable behavior. Through catastrophe theory we can deduce the shape of the entire surface from the fact that the behavior is bimodal for some control points.

In order to understand how the model predicts behavior we must consider the reaction of the dog to changing stimuli. Suppose that initially the dog's emotional state is neutral and can be represented by a point at the origin on the control surface. The dog's behavior, given by the corresponding point on the behavior surface, is also neutral. If some stimulus then increases the dog's rage without affecting its fear, the behavior changes smoothly, following the upward trend of the behavior surface, to more aggressive postures; if the rage is increased enough, the dog attacks. If the dog's fear now begins to increase while its rage remains at an elevated level, the point representing its emotional state on the control surface must move across the graph toward the center. The point representing behavior must of course follow, but because the slope of the behavior surface in this region is not steep the behavior changes only slightly; the dog remains aggressive.

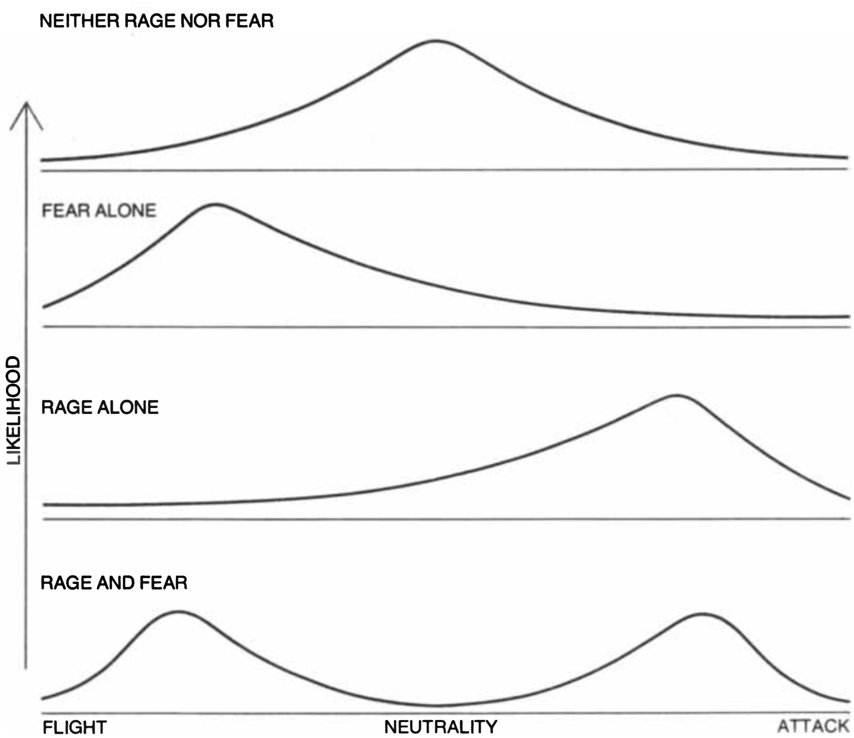
As fear continues to increase, however, the behavior point must eventually reach the edge of the pleat. The novel and illuminating properties of the model then become evident. At the edge of the pleat the sheet on which the behavior point has been traveling folds under and is thereby effectively annihilated; with any additional increase in fear the sheet vanishes. The behavior state must therefore fall directly to the bottom sheet of the graph, which represents quite different modes of behavior. The aggressive states of the top sheet are no longer possible; there is no alternative but a sudden, and indeed catastrophic, change to a meeker attitude. The model thus predicts that if an enraged dog is made progressively more fearful, it will eventually break off its attack and retreat. The sudden change in behavior might be called a flight catastrophe.

The graph also predicts the existence of an opposite pattern of behavior: an attack catastrophe. In an initial state dominated by fear the dog's behavior is stabilized on the bottom sheet, but with a sufficient increase in rage it passes the edge of the opposite side of the pleat and jumps up to the top sheet and a more aggressive frame of mind. In other words, a frightened dog, if it is placed in a situation in which rage steadily increases, may suddenly attack.

Finally, consider the behavior of a dog whose mood is initially neutral as its rage and fear are increased simultaneously. The behavior point is initially at the origin and under the influence of the conflicting stimuli it moves straight forward on the graph. At the singularity, however, where the behavior surface becomes pleated, the point must move onto the top sheet as the dog grows



CONTROL FACTORS in the model of aggression are rage and fear, which in dogs can be measured by facial expression. Rage is reflected by the extent to which the mouth is opened, and fear is revealed by the degree to which the ears are flattened back. From these indicators it is possible to judge the dog's emotional state, and through the model to predict its behavior.



LIKELIHOOD FUNCTION determines the behavior of the dog under the conflicting influences of rage and fear. When neither stimulus is present, the most likely behavior is neutrality; rage alone elicits aggression, fear alone submission. When the dog is made both angry and fearful, the likelihood graph becomes bimodal: attack and flight are both favored, and neutrality is the least likely response. The bimodality is reflected in the behavior surface of the cusp catastrophe, which has two sheets representing most likely behavior where both stimuli are present.

more aggressive or onto the bottom one as it becomes less aggressive. Which sheet is selected depends critically on the dog's state of mind just before it reaches the singularity. The graph is said to be divergent: a very small change in the initial conditions results in a large change in the final state.

The Cusp Catastrophe

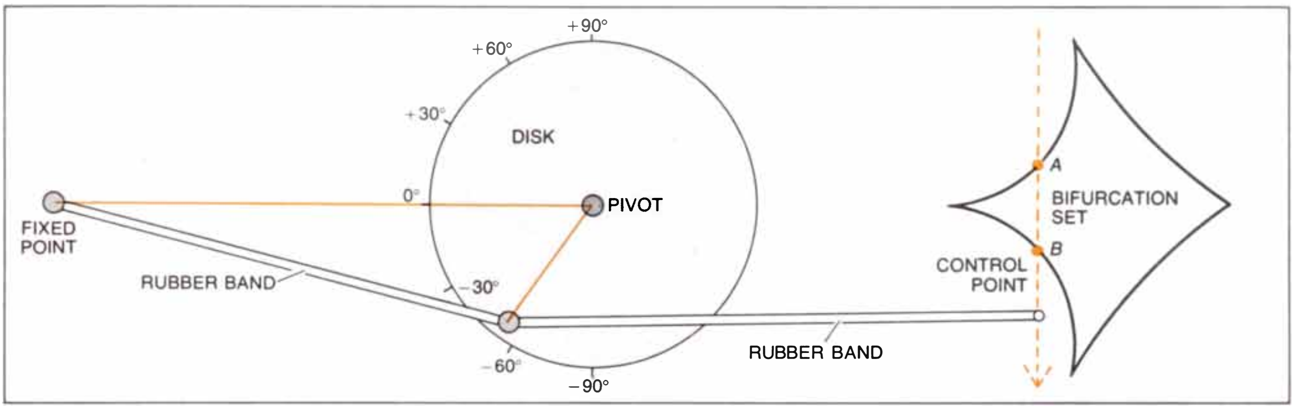
The line that marks the edges of the pleat in the behavior surface, where the top and bottom sheets fold over to form the middle sheet, is called the fold curve. When it is projected back onto the plane of the control surface, the result is a cusp-shaped curve. For this reason the model is called the cusp catastrophe. It is one of the simplest of the seven elementary catastrophes, and so far it has been the most productive.

The cusp on the control surface is called the bifurcation set of the cusp catastrophe, and it defines the thresholds where sudden changes can take place. As long as the state of the system remains outside the cusp, behavior varies smoothly and continuously as a function of the control parameters. Even on entering the cusp no abrupt change is observed. When the control point passes all the way through the cusp, however, a catastrophe is inevitable.

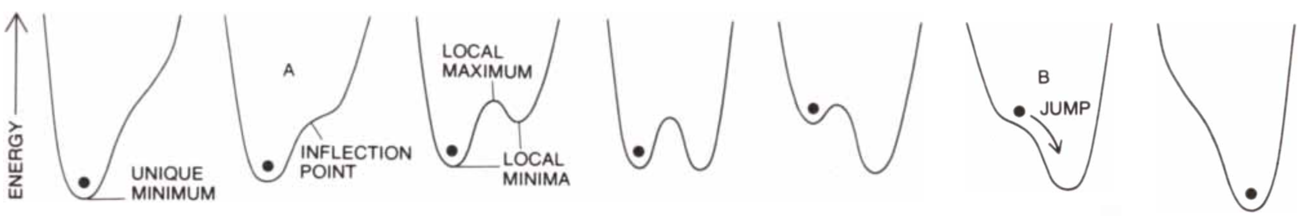
Everywhere inside the bifurcation set there are two possible modes of behavior; outside it there is only one mode. Moreover, in the cusp there are just two modes of behavior even though the behavior surface there has three sheets. That is because the middle sheet in the pleated region is made up of points representing least probable behavior. The middle sheet is included in the graph primarily so that the behavior surface will be smooth and continuous; the behavior point never occupies the middle sheet. Indeed, there is no trajectory on the control surface that could bring the behavior point onto the middle sheet. Whenever the fold curve is crossed, the point jumps between the top and bottom sheets; the middle sheet is said to be inaccessible.

The construction of this model began with an essentially determinist hypothesis: that the behavior of the dog could be predicted from its emotional state, as reflected in its facial expression. The bimodality of the resulting graph may seem at first to undermine that hypothesis, since the existence of two possible modes of behavior for a given emotional state makes unambiguous prediction impossible. Indeed, it is true that if we know only the present emotional state (and if that state falls within the bimodal region of the graph), we cannot predict what the dog will do. The determinism of the model is restored, however, and the model is made more sophisticated, when we consider an additional factor as we make predictions. The behavior of the dog can be predicted if we know both its present emotional state and the recent history of its emotions. It should come as no surprise that the effects of frightening an enraged dog are different from those of angering a frightened dog.

Aggressiveness is not, of course, a uniquely canine trait, and the model describes a mechanism that might operate in other species as well. Consider, for example, the territorial behavior of certain tropical fishes that establish permanent nesting sites on coral reefs. The parameters controlling aggression might in this case be the size of an invading fish and proximity to the nest; the behavior is once again described by a cusp catastrophe. A fish foraging far from

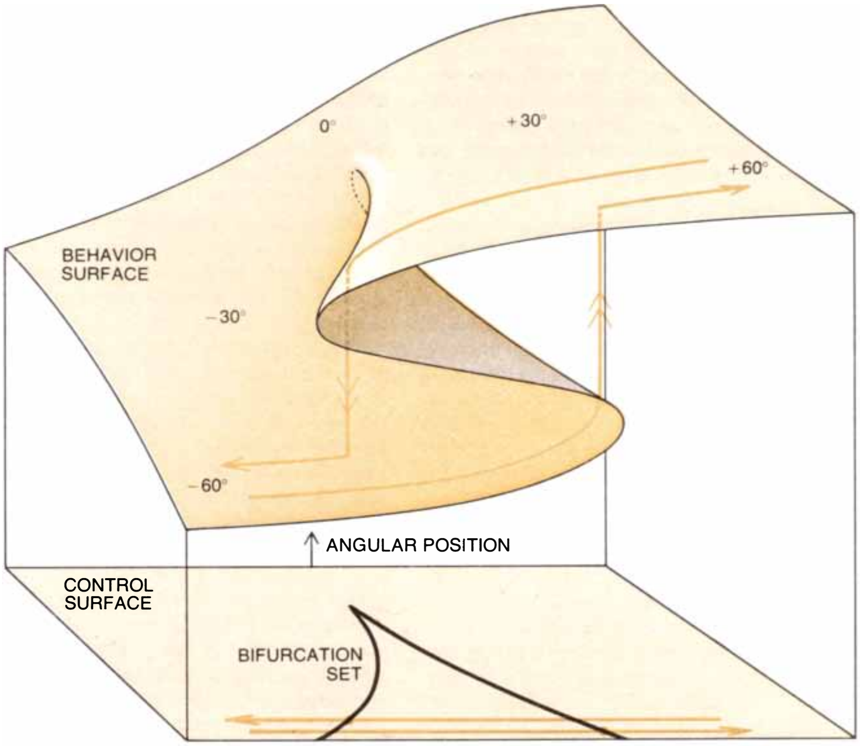


CATASTROPHE MACHINE invented by the author exhibits discontinuous behavior that can be described by a cusp catastrophe. The machine consists of a cardboard disk pivoted at its center, with two rubber bands attached at a point near the perimeter. The unstretched length of each rubber band is approximately equal to the diameter of the disk. The free end of one rubber band is fixed to the mounting board and the machine is operated by moving the other rubber band, the free end of which is designated the control point. The behavior measured is the angle formed by the fixed point, the pivot and the point at which the two rubber bands are attached to the disk. Many movements of the control point cause only smooth rotation of the disk, but in some cases the disk swings suddenly from one side to the other. If all the positions of the control point at which such sudden movements take place are marked, a diamond-shaped curve is generated. The curve is made up of four cusps, each forming the bifurcation set of a cusp catastrophe. Moving the control point along the trajectory shown in color, there is no movement at point *A*, but the disk turns suddenly at *B*. If the path of the control point is reversed, it passes *B* uneventfully, but the disk moves when the control point reaches *A*. Inside the cusp there are two stable positions of the disk.



ENERGY FUNCTION governs the behavior of the catastrophe machine. The machine tends always to assume a position of minimum energy, that is, the disk rotates to minimize the tension on the rubber bands. When the control point is outside the bifurcation set, there is only one position of minimum energy, corresponding to the one stable position of the disk. As the control point is moved across the bifurcation set a second local minimum develops at *A*, and eventually it becomes deeper than the original one. The machine cannot shift to the new local minimum, however, because it is separated from it by a local maximum. Only when the control point crosses the second line of the cusp at *B* is the local maximum eliminated; the equilibrium then breaks down and the machine moves suddenly to the new minimum.

its nest would flee on meeting a larger adversary; once it reached the "defense" perimeter of its own territory, however, its attitude would suddenly change and it would turn to defend its nest. Conversely, if the fish were threatened in its nest, it would chase the invader, but only until it reached the "attack" perimeter of its own territory, where it would abandon the chase and return to its nest. The distance from the nest at which the behavior would change would be determined by the cusp lines of the bifurcation set. Because of the shape of the cusp, the model makes the interesting prediction that the "defense" perimeter is smaller than the "attack" perimeter. Moreover, the size of both perimeters depends on the size of the adversary; a larger invader could approach the nest more closely before the fish would be provoked to fight. The model readily accounts for an observed feature of fish behavior: in mating pairs of territorial fish the partner that happens to be closer to the nest offers the more vigorous defense.



CUSP MODEL of the catastrophe machine erects a pleated behavior surface over one segment of the bifurcation set, such as the cusp nearest the disk. Each point on the top and bottom sheets of the behavior surface gives the position of the disk having minimum energy for that position of the control point. Within the bifurcation set, where there are two stable positions of the disk, there are likewise two local minimums, one on the top sheet and the other on the bottom sheet. The middle sheet represents the local maximum in the energy function. Catastrophic changes in angular position are observed whenever the control point moves all the way across the cusp.

Models of Human Behavior

The cusp catastrophe also provides an interpretation of certain kinds of human behavior. For example, an argument often involves a display of aggression, and its progress is strongly influenced by anger and fear. A cusp catastrophe can be constructed with these emotions as the control parameters and with the intensity of the conflict as the behavior axis.

At the origin on the behavior surface is the most unemotional behavior: rational discussion. As anger and fear increase, the behavior point moves forward on the graph into the pleated region of the surface, where the behavior is bimodal. The opponents are then denied access to sober discourse, and they must either make stronger assertions or make concessions. With an additional escalation of emotion the available modes of behavior diverge further, and the alternatives are invective or apology; finally the opponents must choose between fury and tears. Once the argument has become a heated one a small increase in either anger or fear can cause an abrupt shift in behavior. An aggressive advocate who begins to waver in his opinion may abandon his position and apologize; a timid opponent forced to make repeated concessions may suddenly lose his temper and become truculent. The model even suggests a strategy for effective persuasion. If an argument is likely to induce both anger and fear, then it is best to state one's case and leave, allowing emotions to subside and enabling one's opponent to regain access to rational thought.

Another human behavioral pattern that can be described by the cusp model is self-pity and the catharsis that sometimes relieves it. In this case the controlling parameters, analogous to anger and fear, are the less extreme emotions frustration and anxiety. Moreover, the behavior axis measures not overt behavior, which in animals is the only kind that can be observed, but underlying moods, which in man can be identified directly. A typical range of moods might extend from anger and annoyance through neutral moods to dejection and self-pity.

In the model a large increase in anxiety induces a persistent mood of self-pity; the point representing the mood is trapped on the bottom sheet of the pleated behavior surface [*see top illustration on page 75*]. Self-pity is a defensive attitude commonly adopted by children, and it often seems that sympathy is powerless to alleviate it. A sarcastic remark, on the other hand, may provoke a sudden loss of temper and by releasing tension may open a pathway back to a less emotional state. It is unfortunate that sarcasm should succeed where sympathy fails, but the cause of that irony is apparent in the model. The sarcasm brings an increase in frustration, and as a result the point representing mood travels across the behavior surface as far as the fold curve; having reached the extremity of the bottom sheet, it is forced to make a catastrophic jump to the top sheet, and self-pity is transformed into anger.

These examples of the cusp catastrophe offer an interesting and apparently successful model of certain modes of animal and human behavior, but it is a phenomenological model only; it cannot yet be said to explain the behavior. The question of why a dog or a fish behaves as it does has not been answered but merely recast at a level of greater abstraction. We must now ask why the model works. In particular, why does the behavior point follow the surface to the edge of a pleat, then catastrophically jump to another sheet? Why does it not tunnel smoothly from one surface to another so that there is a gradual transition? What mechanism holds the state of the system on the behavior surface? The answers to these questions can best be approached through another example of the cusp catastrophe, one dealing with the behavior of a much simpler system than a dog, a fish or a man.

A Catastrophe Machine

Elementary catastrophes can be generated at will with a simple device made from stiff cardboard, rubber bands and a few other materials. The heart of the machine is a disk of cardboard pivoted at its center and with two rubber bands attached at a point near the perimeter. The free end of one rubber band is attached at a point outside the disk; the other rubber band serves to control the motion of the disk, and the position of its free end is designated the control point [*see upper illustration on opposite page*].

The catastrophe machine is operated by

moving the control point in the plane of the disk. In many cases the result is smooth rotation of the disk; in one region, however, in the vicinity of a point diametrically opposite the anchorage of the fixed rubber band, smooth movement of the control point can cause abrupt motion of the disk. If the position of the control point is marked each time the disk jumps, a concave, diamond-shaped curve is generated. This curve is made up of four connected cusps, the bifurcation sets of four cusp catastrophes.

If we consider only one of the four cusps, such as the one closest to the disk, the corresponding behavior surface can be constructed by arranging the fold curve so that it lies directly over the cusp. For any position of the control point outside the cusp the behavior surface has only one sheet and the disk has only one stable position. If the control point is inside the cusp, the behavior surface has three sheets, but again the middle one is to be excluded because it corresponds to an unstable equilibrium. As a result there are two stable positions for the disk. That the behavior of the machine conforms to this model can be verified by moving the control point from left to right across the graph. The disk moves smoothly and only slightly until the control point reaches the right edge of the cusp; the disk then suddenly turns as the behavior point falls off the extremity of the bottom sheet and jumps to the top one. When the path of the control point is reversed, the point crosses the right edge of the cusp uneventfully and the disk continues to move smoothly until the left edge of the cusp is reached; this time the behavior point jumps from the top sheet to the bottom one.

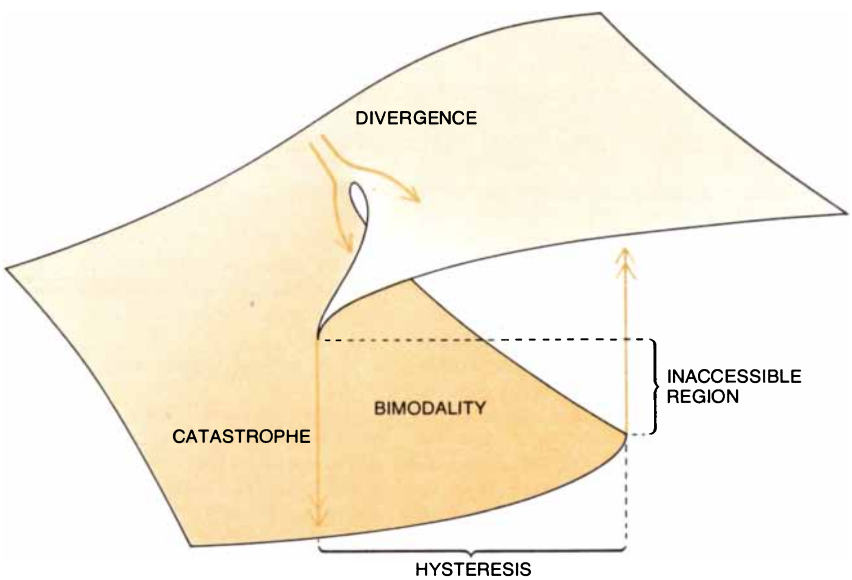
In the catastrophe machine the cause of this behavior is readily discovered: it is the tendency of all physical systems in which friction is important to assume a state of minimum energy. The energy to be minimized is the potential energy stored in the rubber bands, and the disk therefore rotates until the tension on the two rubber bands is at a minimum. At that position the machine is in stable equilibrium. Unless energy is added to the system the machine must remain at the equilibrium point; the process that keeps it there is called the dynamic.

The operation of the dynamic can be demonstrated by a series of graphs, each one showing for a single position of the control point the energy of the machine for all possible rotations of the disk [*see lower illustration on page 68*]. As long as the control point is outside the cusp the graph is a smooth curve with a single trough, or minimum, and the state of the machine will always move swiftly to the state of minimum energy at the bottom of the trough. As the control point enters the interior of the cusp a second trough, or local energy minimum, develops next to the original one. This second trough gradually grows deeper, but the machine cannot enter it because the two troughs are separated by a small peak, a local energy maximum. The state of the machine does not change until the second trough coalesces with the local maximum. This takes place as the control point crosses the second cusp line, and the state of the machine is then carried swiftly by the dynamic to a new unique position of minimum energy.

The significance of the dynamic becomes apparent when it is revealed that the behavior surface of the cusp catastrophe is the graph of all the minimums and maximums of the energy function. Outside the cusp there is a single energy minimum and there are no maximums; the behavior surface therefore has a single sheet. At the cusp line a new local minimum and maximum are created, and consequently two new sheets form in the behavior surface. The state of the machine can never lie stably on the middle sheet because that is a position of maximum energy.

The mathematical procedure for drawing the behavior surface comes from elementary calculus: the behavior surface is a graph of all the points where the first derivative of the energy function is equal to zero. It is not necessary to understand how this operation is performed; it is sufficient to know that the first derivative is zero wherever the graph of the energy function is horizontal (where its slope is zero). It is horizontal only at the minimums and maximums and at inflection points. The minimums form the stable top and bottom sheets, the maximums form the unstable middle sheet and the inflection points form the fold curve that marks the boundaries of the sheets.

The behavior associated with the cusp farthest from the disk could be analyzed in the same way that we have explained the nearer cusp. The two cusps at the sides, however, differ in a crucial respect: for them the energy function is inverted, so that inside the cusp there are two points—corresponding to two positions of the disk—with maximum energy and only one point with minimum energy. The dynamic compels the machine to remain in the single stable position of minimum energy. On the behavior surface also the positions of minimums and maximums are reversed: the middle sheet represents stable energy minimums, the top and bottom sheets represent unstable maximums. The behavior point therefore can lie only on the middle sheet. The graph is called a dual cusp catastrophe.



FIVE PROPERTIES characterize phenomena that can be described by the cusp catastrophe. The behavior is always bimodal in some part of its range, and sudden jumps are observed between one mode of behavior and the other. The jump from the top sheet of the behavior surface to the bottom sheet does not take place at the same position as the jump from the bottom sheet to the top one, an effect called hysteresis. Between the top and bottom sheets there is an inaccessible zone on the behavior axis; the middle sheet, representing least likely behavior, has been omitted for clarity. Finally, the cusp catastrophe implies the possibility of divergent behavior.

The Role of the Dynamic

The success of catastrophe theory in accounting for the behavior of the catastrophe machine can now be explained. The crucial concept is that of the dynamic, which has two functions. First, it holds the behavior point firmly on the top or bottom sheet of the behavior surface. If the disk is turned by hand against the force of the rubber bands and is then released, the dynamic brings it sharply back to equilibrium, that is, it brings the behavior point back onto the surface. Second, when the behavior point crosses the fold curve, it is the dynamic that causes the catastrophic jump from one sheet to the other.

The same principles can be applied to the psychological models considered above. The likelihood functions of these models are analogous to the energy function in the catastrophe machine, except that the roles of minimums and maximums are reversed. The top and bottom sheets of the behavior


The Silver Martini.
For people who want a silver lining
without the cloud.
SMIRNOFF SILVER
VODKA
90.4 PROOF
DISTILLED FROM GRAIN
Smirnoff Silver
Ninety point four proof. Smirnoff leaves you breathless®
SMIRNOFF SILVER® VODKA 90.4 PROOF DIST. FROM GRAIN STE. PIERRE SMIRNOFF FLS (DIV. OF HEUBLEIN) ©1971 HEUBLEIN INC. HARTFORD, CONN.

Do it all. Instantly.

"Polaroid", "Polacolor", and "SX-70"®

Polaroid has a camera or film to fit practically all your photographic needs.

And to do the job quickly and economically. On the spot and without a darkroom.

What can you do with Polaroid instant films and cameras? Almost everything—from superb black and white copies of original artwork to sharply detailed color records from sophisticated instruments.

For proof of our versatility, just look at these pages. You'll see examples of Polaroid self-developing films that deliver a positive with a fully-developed, reusable negative; a black and white projection transparency; a high contrast print; a high speed black and white recording of scientific data; Polacolor 2 and SX-70 photographs with brilliant colors and the most stable dyes in general purpose photography; highly secure full-color portrait identification cards.

And a great deal more.

For more information and assistance, send us the coupon below. Or if you have a photographic or recording need that can't wait, call us collect on our applications "hot line": (617) 547-5176.

You'll discover that there's an easy-to-use Polaroid Land camera or film (or an adapter for your own equipment) that can save you time and money.

And do it in just seconds.

21 Instant Films from Polaroid

Polaroid Corporation, Dept. A336
Cambridge, Mass. 02139

I'd like to know more about Polaroid instant films and cameras.

My application is________

Name________

Position________

Organization________

Address________

City________ State________ Zip________

Telephone________

"Unbelievable."

The Aspen wagon has everything going for it.

Even if you never wanted a wagon...you'll want the Aspen wagon. That's how perfect it is. Its size alone will make you want to drive it. And once you do, the ride will make you want it even more. So before everybody gets there first, you ought to go down and take a look at the small wagon that'll probably become the biggest hit in America.

The Aspen wagon has a special size all its own.

Smaller than the big and bigger than the small, the Aspen carries only 100 pounds less than a full-sized wagon and over 200 pounds more than a subcompact wagon. So you can move into a smaller wagon for economy without having to get tiny about it.

The Aspen wagon has a ride that rivals that of a full-sized car.

The unique Isolated Transverse Suspension is rubber isolated to reduce noise and vibration transmitted to the passenger compartment. This gives Aspen an extremely comfortable ride. And for a small wagon...that's unbelievable.

The Aspen wagon achieved EPA estimated highway mileage of up to 30 MPG.

According to EPA estimated mileage results, the Aspen wagon got 30 MPG on the highway and 18 city. The coupe and sedan got 27 MPG highway and 18 city. All were equipped with a 225 Six and manual transmission. (Your actual mileage depends on driving habits, condition of your car, and options. In California, see your Dealer for mileage results.)

The Aspen wagon offers the important convenience features of a luxury car.

A long list of options includes everything from power seats and windows to electric door locks and automatic speed control.

Here's "The Clincher."

"For the first 12 months of use, any Chrysler Corporation Dealer will fix, without charge for parts or labor, any part of our 1976 passenger cars we supply (except tires) which proves defective in normal use, regardless of mileage." The owner is responsible for maintenance service such as changing filters and wiper blades.

Dodge

ASPEN

The new Dodge Aspen.
For a small car at a small price,
it's unbelievable.

Winner of the 1976
Motor Trend Magazine
Car of the Year Award.

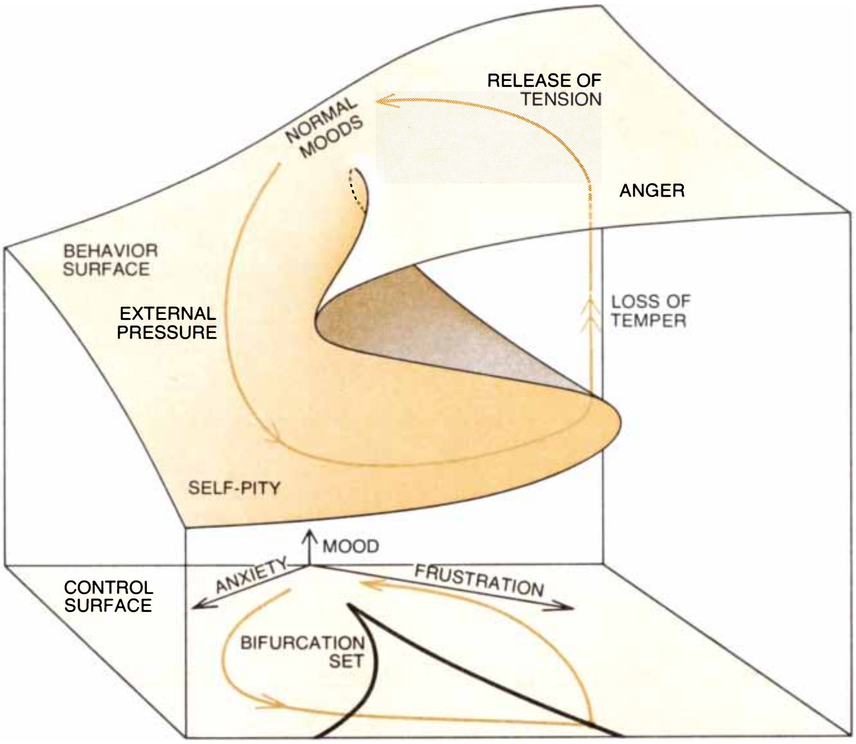
surface are made up of all the points representing maximum likelihood and the middle sheet is made up of those representing minimum likelihood. An important question remains: What is the dynamic? In the model of aggression what compels the dog to express the most likely behavior, and in the model of self-pity why is the "most likely mood" the one that is adopted?

An energy minimum in a physical system such as the catastrophe machine is a special instance of a concept called an attractor. In this case it is the simplest kind of attractor, a single stable state, and its effect is like that of a magnet: everything within its range of influence is drawn toward it. Under the influence of the attractor the system assumes a state of static equilibrium.

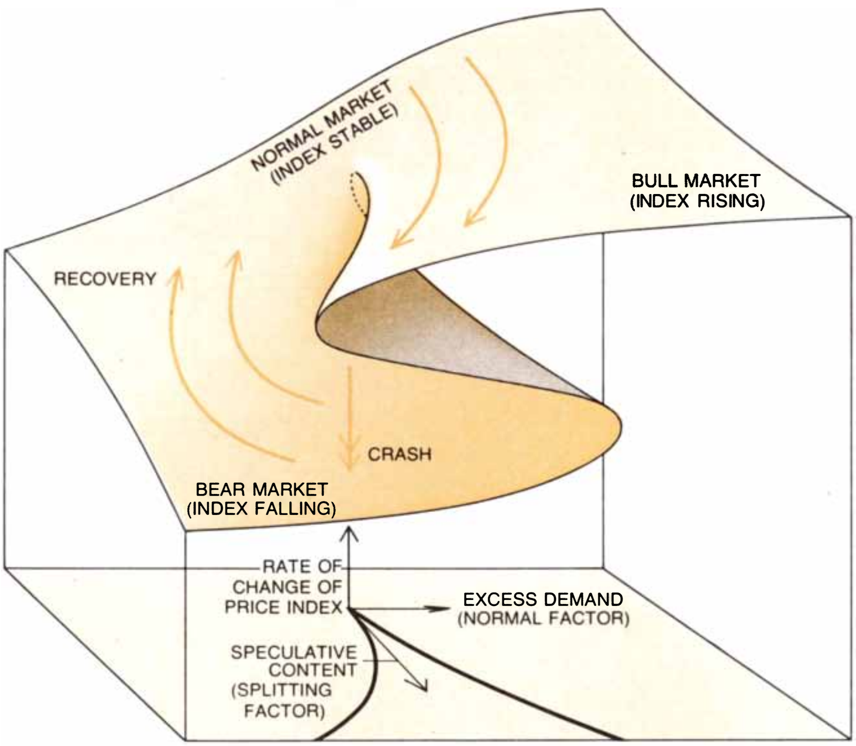
In the psychological models there must also be attractors, although they need not be as simple as this one. The attractor of a system that is in dynamical equilibrium consists of the entire stable cycle of states through which the system passes. For example, a bowed violin string repeats the same cycle of positions over and over at its resonant frequency, and that cycle of positions represents an attractor of the bowed string.

In the psychological models the obvious place to seek attractors is in the neural mechanisms of the brain. Of course the brain is far more complicated and less well understood than a violin string, but its billions of neurons are known to be organized in large, interconnected networks that form a dynamical system; the equilibrium states of any dynamical system can be represented by attractors. Some of these attractors may be single states, but others are more likely to be stable cycles of states, or higher-dimensional analogues of stable cycles. As various parts of the brain influence one another, the attractors appear and disappear, in some cases rapidly and in others slowly. As one attractor gives way to another the stability of the system may be preserved, but often it is not; then there is a catastrophic jump in the state of the brain.

Thom's theory states that all possible sudden jumps between the simplest attractors—points of static equilibrium—are determined by the elementary catastrophes. Thus if the brain dynamic had only point attractors, it could exhibit only the elementary catastrophes. That is not the case; one obvious item of evidence for more complicated attractors is the alpha rhythm of brain waves, a cyclic attractor. The rules governing jumps between cyclic attractors and higher-dimensional ones are not yet known; they must include not only elementary catastrophes but also generalized catastrophes, and their study is today an active area of research in mathematics. Hence as yet there is no complete theory for the description of all brain dynamics. Nevertheless, the elementary catastrophes provide meaningful models of some brain activities. The models are explicit and sometimes disarmingly simple, but the powerful mathematical theory on which they are based implicit-



CATHARTIC RELEASE FROM SELF-PITY is described by a cusp catastrophe in which anxiety and frustration are conflicting factors influencing mood. Self-pity is induced by an increase in anxiety; it can be relieved by some event, such as a sarcastic remark, that causes an increase in frustration. As the control point crosses the cusp the mood changes catastrophically from self-pity to anger; the resulting release of tension gives access to calmer emotional states.



BEHAVIOR OF THE STOCK MARKET is described by a model in which the controlling parameters are excess demand for stock and the proportion of the market held by speculators as opposed to that held by investors. The behavior itself is measured by the rate at which the index of stock prices is rising or falling. The control factors are oriented not as conflicting factors but as normal and splitting factors. A fall from the top sheet to the bottom represents a crash; the slow recovery is effected through feedback of the price index on the control parameters.

ly allows for the complexity of the underlying neural network.

The concept of an attractor of the brain dynamic provides what is needed in our models of human and animal behavior. The neural mechanism responsible for a mood such as self-pity is not known, but the existence of the mood as a stable state implies that that mechanism is an attractor. Indeed, in the model of self-pity every point on the behavior surface corresponds to an attractor for the system in the brain that determines mood. If that neural system is disturbed in some way, it is quickly returned, under the influence of an attractor, to the behavior surface, just as the catastrophe machine returns to equilibrium. Abrupt changes in mood are encountered when the stability of an attractor breaks down, allowing the mood-determining system to come under the influence of another attractor, toward which it immediately moves.

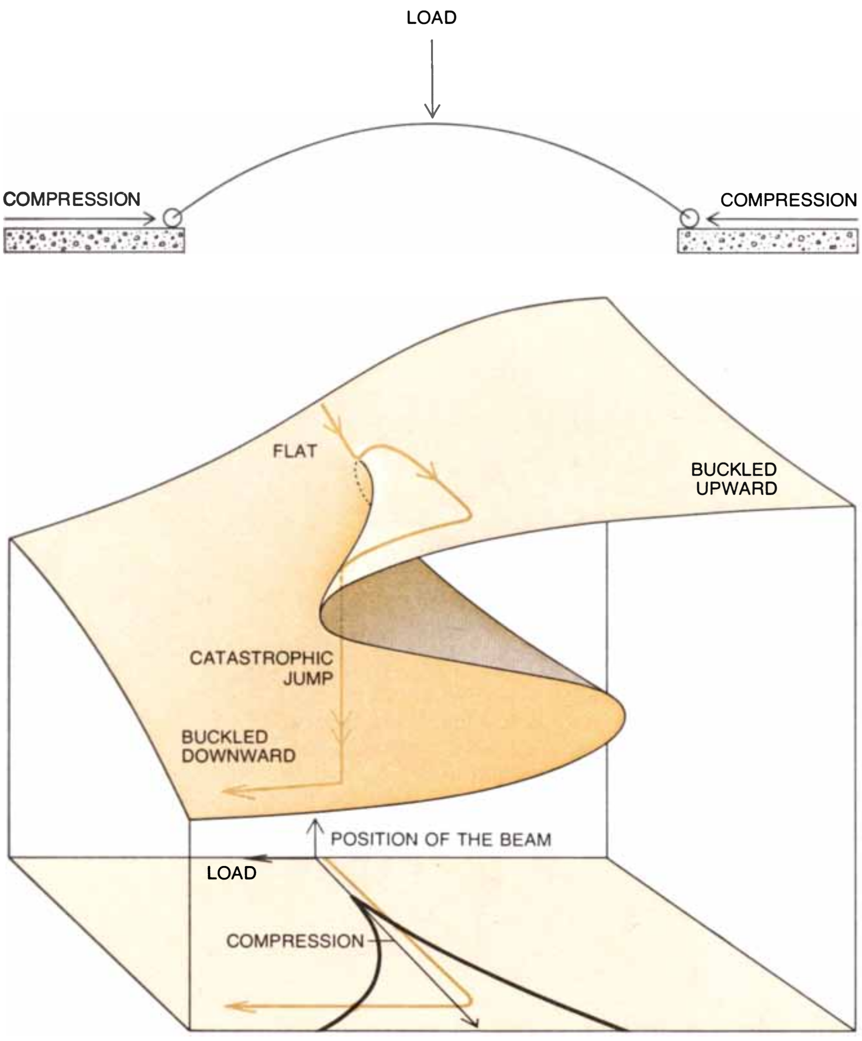
Through this hypothetical mechanism catastrophe theory provides a model not only of expressed behavior but also of the activity of the brain that directs the behavior. The model is probably most appropriate to primitive regions such as the midbrain, where the networks are highly interconnected and therefore may act as a unit. (In the phylogenetically younger cerebral cortex, patterns of activity are much more complex.) The psychological models we have considered are largely concerned with emotion or mood, and it is thought that the part of the midbrain called the limbic system is primarily responsible for generating mood.

Features of the Cusp Model

The preceding examples and analysis suggest several features common to all cusp catastrophes. One invariant characteristic is that the behavior is bimodal over part of its range, and sudden changes are observed from one mode of behavior to the other. In addition, the pattern of the sudden changes exhibits the effect called hysteresis, that is, the transition from the top sheet to the bottom one does not take place at the same point as the transition from the bottom sheet to the top one. The sudden change does not come at the middle of the cusp but is delayed until the bifurcation set is reached. Another characteristic is that inside the cusp, where the behavior becomes bimodal, the middle zone on the behavior axis becomes inaccessible. Finally, the model implies the possibility of divergence, so that a small perturbation in the initial state of the system can result in a large difference in its final state. These five qualities—bimodality, sudden transitions, hysteresis, inaccessibility and divergence—are related to one another by the model itself. If any one of them is apparent in a process, the other four should be looked for, and if more than one is found, then the process should be considered a candidate for description as a cusp catastrophe.

One process where the five qualities can be detected is the development of hostilities between nations, a situation with obvious analogies to the models of aggression and argument. The control parameters in those models were rage and fear; here we substitute threat and cost. The behavior axis describes the possible actions of the nation, ranging from full-scale attack to lesser military responses such as blockade through neutrality to appeasement and surrender. In a situation where threat and cost are both high public opinion often becomes bimodal as the nation is divided into "doves" advocating surrender and "hawks" advocating attack. The dynamic in the model is the sensitivity of the government to the will of its constituency; the government continuously adapts its policy in order to increase its support, and therefore it remains on the behavior surface. From the model we can deduce and recognize the possible catastrophes. A threatened nation may make repeated concessions, but there is a limit beyond which further threat elicits a sudden declaration of war. Conversely, as costs rise a nation may escalate a war, but there is a limit beyond which further costs may result in a sudden surrender. Hysteresis is recognizable in the delays observed before declaring war or surrendering. The inaccessible region of the behavior axis is the middle zone representing negotiation or compromise. Finally, divergence can be observed in a conflict between two equally strong nations, in which the distribution of public opinion is similar, but the response of the governments is quite different, one becoming increasingly aggressive as the other grows more submissive.

Another candidate for analysis as a cusp catastrophe is the behavior of the stock market, where the terms "bull market" and "bear market" suggest an obvious bimodality. Moreover, a crash, or collapse, of the market is readily explained as a catastroph-



BUCKLING OF AN ELASTIC BEAM is controlled by load and compression, which are respectively normal and splitting factors in a cusp catastrophe. If the beam is flat, an increase in compression forces it to buckle upward or downward. If it buckles upward, a subsequent increase in load drives the control point across the cusp, causing a catastrophic downward motion.

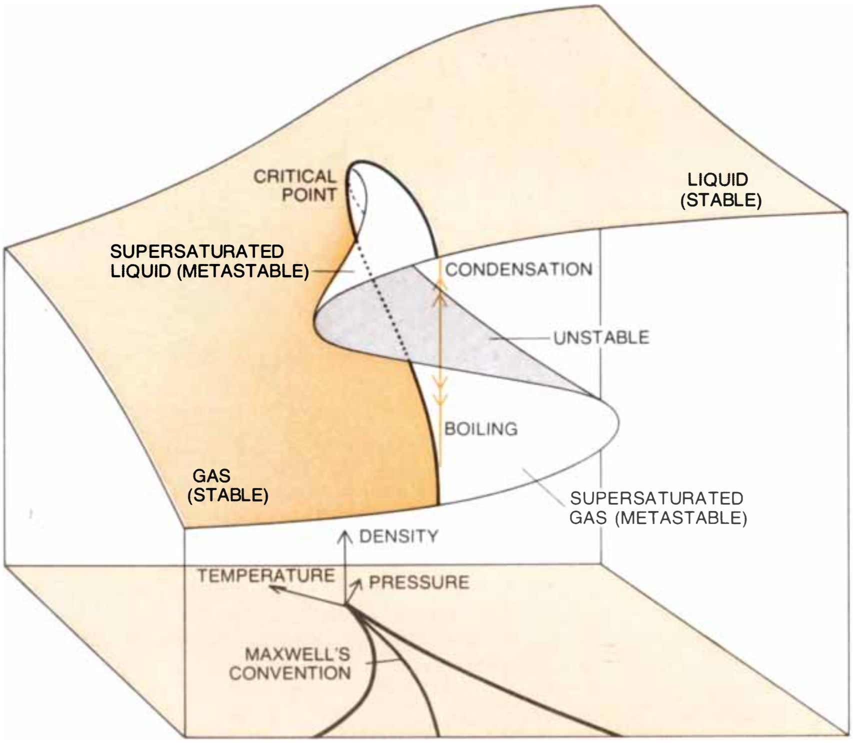
ic jump from one sheet of the behavior surface to the other.

In the construction of the model a small modification is necessary. In this case the control axes do not diverge on each side of the cusp, as they do in all the previous examples; instead one axis comes straight forward on the graph, bisecting the cusp, and the other is perpendicular to the cusp [*see bottom illustration on page 75*]. The parameter that bisects the cusp is called a splitting factor, since increasing it causes a progressively larger divergence between the top and bottom sheets. The other factor is the normal factor, since at the back of the behavior surface the behavior increases continuously with it.

In the stock-market model the normal factor is excess demand for stock; the splitting factor is perhaps more difficult to identify, but it might be related to the amount of stock held by speculators compared with that held by long-term investors. The behavior axis is best measured by the rate of change of the index of stock prices. A market with a rising index is a bull market, and its behavior point is on the top sheet; a falling index, or bear market, places the behavior point on the bottom sheet.

The mechanism of the crash can now be understood. A market with some excess demand and a high proportion of speculators is a bull market, on the top sheet of the behavior surface. A crash can be precipitated by any event that reduces demand enough to push the behavior point over the fold curve. The larger the share of the market held by speculators, the severer the crash. One is immediately prompted to ask why the subsequent recovery is usually slow, and why there is no "upward crash" from a bear market to a bull one. The probable answer is that the behavior axis (the rate of change of the index) has an influence on the control parameters through a feedback loop. A falling, bear market discourages speculation, but after a while the resulting undervaluation encourages long-term investment; as a consequence after a crash the splitting factor is reduced and the market moves back on the graph to the region where the behavior surface is no longer bimodal. As confidence grows and produces excess demand the index rises, but slowly and smoothly, without catastrophes. Speculation is now encouraged and investment discouraged, and the stage is set for another cycle of boom and bust.

As was pointed out above, most applications of catastrophe theory have been in biology and the human sciences, where other modeling techniques are often uninformative, but there are many situations in physics (a science with a highly developed mathematical language) where the theory can contribute to understanding. One such instance is the transition between the liquid phase and the gaseous phase of matter. We can rewrite the equations of J. D. van der Waals as a cusp catastrophe where temperature and pressure are conflicting control factors and density is the behavior axis. The top sheet is then the liquid phase and the bottom sheet the gaseous phase; the two catastrophes represent boiling and condensation. The vertex of the cusp is the critical point, where liquid and gas exist simultaneously. By going around the back of the cusp a liquid can be converted into a gas without boiling.



PHASE TRANSITIONS between the liquid and the gaseous state of matter conform to a modified cusp model in which temperature and pressure are the control factors. Ordinarily both boiling and condensation take place at the same values of temperature and pressure. Thus there are catastrophic changes, but there is no hysteresis. Under special circumstances, however, a vapor can be cooled below its dew point and a liquid can be heated above its boiling point, so that the behavior surface is followed all the way to the fold curve. The critical point, where liquid and gas exist simultaneously, is represented by the singularity where the pleat disappears.

Under exceptional circumstances the physical system can be made to follow the behavior surface all the way to the edge of each sheet. With due care, for example, liquid water can be heated well beyond its normal boiling point and water vapor can be cooled below its dew point before the phase transitions take place. Such superheating and supercooling are employed in the bubble chambers and cloud chambers used to detect subatomic particles. Ordinarily, however, a substance boils and condenses at the same temperature and pressure, so that a "cliff" forms in the behavior surface, cutting across the middle of the pleat [*see illustration above*]. The formation of the cliff is explained by a rule called Maxwell's convention, and it reflects the fact that the model is a statistical one, averaging the behavior of many particles.

Another cusp catastrophe in physics, which is derived from the work of Leonhard Euler in the 18th century, is the buckling of an elastic beam under horizontal compression and a vertical load. The compression is a splitting factor, the load a normal factor. An increase in compression causes the behavior point to move forward on the graph, into the region of the cusp, where the beam has two stable states, one buckled upward and the other downward. If the initial buckling is upward, increasing the load can move the behavior point across the cusp, causing the beam to snap suddenly downward. The effect can be observed with a piece of cardboard held between the fingers. When it happens to a girder supporting a bridge, the result is a catastrophe in both the mathematical and the mundane sense.

Another beautiful example in physics is provided by the bright geometric patterns called light caustics, which are created when light is reflected or refracted by a curved surface. A familiar caustic is the cusp-shaped curve that sometimes appears on the surface of a cup of coffee in bright sunlight; it is caused by the reflection of the sun's rays from the inside of the cup.

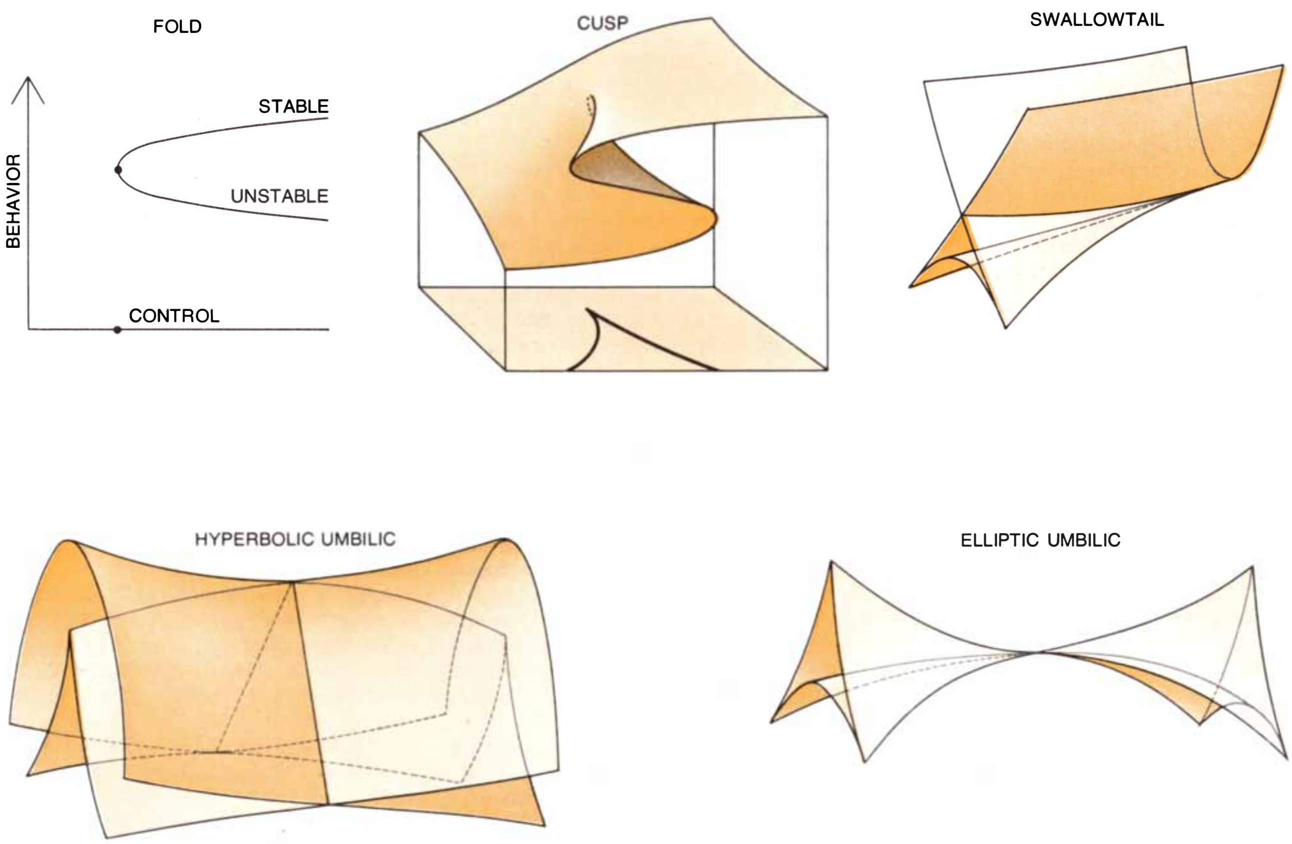
Another familiar caustic, which exhibits temporal as well as spatial discontinuities of brightness, is the changing pattern on the bottom of a swimming pool in sunlight. The rainbow is a family of colored caustic cones. More complex caustics can be produced by

CATASTROPHE		CONTROL DIMENSIONS	BEHAVIOR DIMENSIONS	FUNCTION	FIRST DERIVATIVE
CUSPOIDS	FOLD	1	1	$\frac{1}{3}x^3 - ax$	$x^2 - a$
	CUSP	2	1	$\frac{1}{4}x^4 - ax - \frac{1}{2}bx^2$	$x^3 - a - bx$
	SWALLOWTAIL	3	1	$\frac{1}{5}x^5 - ax - \frac{1}{2}bx^2 - \frac{1}{3}cx^3$	$x^4 - a - bx - cx^2$
	BUTTERFLY	4	1	$\frac{1}{6}x^6 - ax - \frac{1}{2}bx^2 - \frac{1}{3}cx^3 - \frac{1}{4}dx^4$	$x^5 - a - bx - cx^2 - dx^3$
UMBILICS	HYPERBOLIC	3	2	$x^3 + y^3 + ax + by + cxy$	$3x^2 + a + cy$ $3y^2 + b + cx$
	ELLIPTIC	3	2	$x^3 - xy^2 + ax + by + cx^2 + cy^2$	$3x^2 - y^2 + a + 2cx$ $-2xy + b + 2cy$
	PARABOLIC	4	2	$x^2y + y^4 + ax + by + cx^2 + dy^2$	$2xy + a + 2cx$ $x^2 + 4y^3 + b + 2dy$

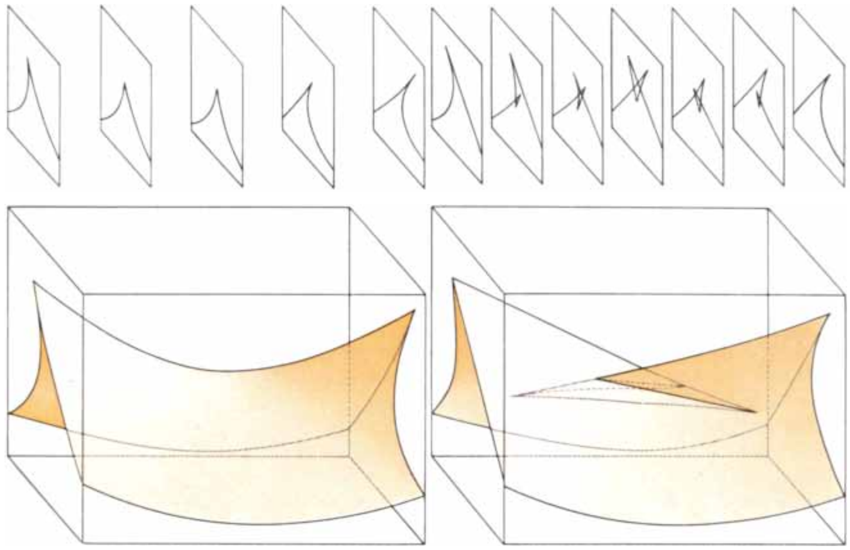
SEVEN ELEMENTARY CATASTROPHES describe all possible discontinuities in phenomena controlled by no more than four factors. Each of the catastrophes is associated with a potential function in which the control parameters are represented as coefficients (*a, b, c, d*) and the behavior of the system is determined by the variables (*x, y*). The behavior surface in each catastrophe model is the graph of all the points where the first derivative of this function is equal to zero or, when there are two first derivatives, where both are equal to zero.

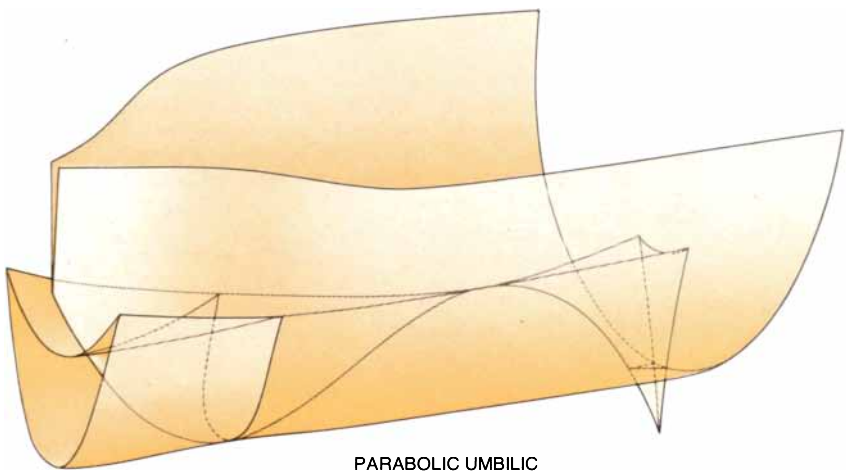
shining a beam of light in a concave mirror or through spherical or cylindrical lenses (such as a light bulb or a beaker filled with water). In this application catastrophe theory has led to a better understanding of the phenomenon; Thom has shown that stationary caustics can have only three kinds of singular point. A mathematical subtlety of the catastrophe-theory analysis of light caustics is that there is no dynamic; instead there is a variational principle that gives equal importance to both minimums and maximums.

The cusp catastrophe is a three-dimensional figure: two dimensions are required for the two control parameters and one more is required for the behavior axis. Actually the behavior axis need not represent a single behavior variable; in models of brain function, for example, it may represent the states of billions of neurons, all varying at the same time. Nevertheless, catastrophe theory shows that it is always possible to select a single behavior variable and to plot



GRAPHS of five of the elementary catastrophes suggest the nature of their geometry. The fold catastrophe is a transverse section of a fold curve of the cusp catastrophe, and its bifurcation set consists of a single point. The cusp is the highest-dimensional catastrophe that can be drawn in its entirety. The swallowtail is a four-dimensional catastrophe and the hyperbolic umbilic and the elliptic umbilic catastrophes are five-dimensional. For these graphs only the three-dimensional bifurcation sets can be drawn; the behavior surfaces are not shown.





SECTIONS are the only recourse for illustrating the remaining two catastrophes, since even their bifurcation sets have more than three dimensions. The four-dimensional bifurcation set of the butterfly catastrophe is shown in three-dimensional sections; the fourth dimension is the butterfly factor, and if it happens to be time, then one configuration evolves into the other. Moving from left to right in each drawing is equivalent to changing the bias factor; two-dimensional "slices" reveal the effect of this factor more clearly. The four-dimensional bifurcation set of the parabolic umbilic catastrophe is also shown in a three-dimensional section. It is based on a drawing prepared with a computer by A. N. Godwin of Lanchester Polytechnic in England.

the behavior surface with respect to that axis only, and so to obtain the familiar three-dimensional graph.

If the graph is reduced to two dimensions, an even simpler model results: the fold catastrophe. In the fold catastrophe there is only one control parameter; the control space is a straight line and the bifurcation set is a single point on that line. The behavior space is a parabola, half of which represents stable states, the other half unstable. The two regions are separated by a fold point directly above the bifurcation point.

The Classification Theorem

The fold catastrophe can be regarded as a transverse section of the fold curve of the cusp catastrophe. The cusp in turn can be regarded as a stack of many fold catastrophes, together with one new singular point at the origin. More complicated catastrophes of higher dimension can be constructed on the same plan: each consists of all the lower-order catastrophes together with one new singularity at the origin.

If the control space is made three-dimensional while the behavior space remains one-dimensional, a unique four-dimensional catastrophe can be constructed. The behavior surface becomes a three-dimensional hypersurface, and instead of being folded along curves as in the cusp catastrophe it is folded along entire surfaces, a configuration that cannot easily be visualized. The bifurcation set no longer consists of curves with cusp points in two dimensions but is made up of surfaces in three dimensions that meet in cusps at their edges. A new singularity appears at the origin, called the swallowtail catastrophe. It is impossible to draw the complete swallowtail catastrophe because we cannot draw four-dimensional pictures. We can, however, draw its bifurcation set, which is three-dimensional, and from that drawing it is possible to derive some geometrical intuition about the swallowtail, just as it would be possible to describe the cusp catastrophe by drawing its bifurcation set (the cusp) in two dimensions and bearing in mind that the behavior surface is bimodal over the inside of the cusp. The catastrophe is called a swallowtail because the bifurcation set looks somewhat like one; the name was suggested by Bernard Morin, a blind French mathematician.

If yet another control parameter is added, a five-dimensional catastrophe is created. The fold, the cusp and the swallowtail again appear as sections, and a new singularity is associated with a "pocket" formed by the interpenetration of several surfaces. The shape of this pocket, or of sections of it, has suggested the name butterfly catastrophe. In the butterfly even the bifurcation set is four-dimensional and therefore cannot be drawn. It can be illustrated only through two- or three-dimensional sections [*see illustration on this page*].

There are two more five-dimensional catastrophes, formed when the control space has three dimensions and the behavior space has two dimensions. They are called the hyperbolic umbilic and the elliptic umbilic catastrophes. As in the case of the swallowtail, their bifurcation sets consist of surfaces with cusped edges, and since they are three-dimensional they can be drawn. Finally, the six-dimensional catastrophe generated by a four-dimensional control space and a two-dimensional behavior space is called the parabolic umbilic. Its geometry is complex, and again only sections of its bifurcation set can be drawn.

By increasing the dimensions of the control space and the behavior space an infinite list of catastrophes can be constructed. The Russian mathematician V. I. Arnold has classified them up to at least 25 dimensions. For models of phenomena in the real world, however, the seven described above are probably the most important because they are the only ones with a control space having no more than four dimensions. One particularly common class of processes, those determined by position in space and by time, cannot require a control space with more than four dimensions, since our world has only three spatial dimensions and one time dimension.

Even the catastrophes that cannot be drawn can be employed in modeling phenomena. Their geometry is completely determined, and the movement of a point over the behavior surface can be studied analytically if it cannot be seen graphically. Each catastrophe is defined by a potential function, and in each case the behavior surface is the graph of all the points where the first derivatives of that function are zero.

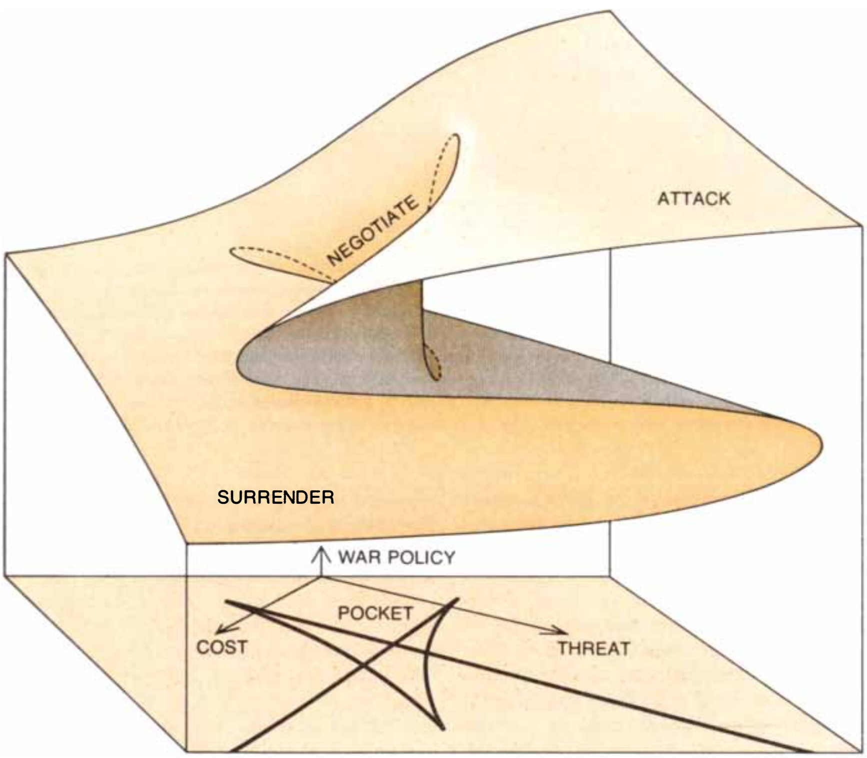
The power of Thom's theory lies in its generality and its completeness. It states that if a process is determined by minimizing or maximizing some function, and if it is controlled by no more than four factors, then any singularity of the resulting behavior surface must be similar to one of the seven catastrophes I have described. If the process is governed by only two control factors, then the behavior surface can have only folds and cusps. The theorem states in essence that in any process involving two causes the cusp catastrophe is the most complicated thing that can happen to the graph. The proof of the theorem is too technical and too long to be presented here, but its consequences are straightforward: whenever a continuously changing force has an abruptly changing effect, the process must be described by a catastrophe.

After the cusp, the catastrophe with the richest spectrum of applications is the butterfly. Just as bimodal behavior determines the cusp model, so trimodal behavior determines the butterfly. In the cusp model of war policy, for example, where public opinion is divided between "doves" and "hawks," the butterfly model provides for the emergence of a compromise opinion favoring negotiation. The new mode of behavior arises as a new sheet of the behavior surface, growing smoothly out of the back of the pleat [*see illustration below*].

The geometry of the butterfly is controlled by four parameters. Two of them are familiar from the cusp models: the normal factor and the splitting factor. The remaining two are new: the bias factor and the butterfly factor. The effect of the bias factor is to alter the position and shape of the cusp; it swings the main part of the cusp left or right, but the vertex of the cusp bends the opposite way. At the same time the bias factor moves the behavior surface up and down.

The effect of the butterfly factor is to create the third stable mode of behavior. As the butterfly factor increases, the cusp on the control surface evolves into three cusps, which form a triangular "pocket." Above the pocket is the new, triangular sheet on the behavior surface, between the top and bottom sheets.

In order to draw the butterfly catastrophe two of the four parameters must be suppressed, and ordinarily the bias and butterfly factors are chosen. Their influence on the graph cannot be ignored, however. One effect of the bias factor is to reduce one side of the pocket until it disappears in a swallowtail catastrophe; the bias factor therefore tends to destroy a compromise. Since the butterfly factor controls the growth of the intermediate behavior sheet it enhances the stability of a compromise.



BUTTERFLY CATASTROPHE provides for the emergence of compromise opinion in a model of the development of war policy. In the butterfly four controlling parameters are required, but here only two are shown (threat and cost) and the other two are assumed to remain constant. The bifurcation set is one of the sections on the preceding page; it is a complex curve with three cusps and a "pocket" in the middle. On the behavior surface above the pocket is a new sheet that provides for a new intermediate mode of behavior. If both the threat and the cost of war are high, the cusp model would allow for only the extreme positions advocating attack or surrender. The new sheet in the butterfly model represents a compromise opinion, advocating negotiation.

Anorexia Nervosa

A second application of the butterfly catastrophe, and an exceptionally fertile one, is to anorexia nervosa, a nervous disorder suffered mainly by adolescent girls and young women in whom dieting has degenerated into obsessive fasting. The model was developed by me in collaboration with J. Hevesi, a British psychotherapist who has introduced trance therapy in the treatment of anorexia. In a recent survey of 1,000 anorexic patients his were the only ones to state they had been completely cured.

In the initial phase of anorexia the obsessive fasting can lead to starvation, and sometimes to death. With the passage of time the patient's attitudes toward food and her behavior become progressively more abnormal. After about two years a second phase, called bulimia, usually develops, in which the victim alternately fasts and gorges herself. The bimodal behavior of this second phase immediately suggests a cusp catastrophe. The anorexic is caught in a hysteresis cycle, jumping catastrophically between two extremes, and she is denied access to the normal behavior in between. Catastrophe theory also suggests a theoretical cure: if a further bifurcation could be induced according to the butterfly catastrophe, a new pathway back to normality might be created.

The behavior surface in this model represents the overt behavior of the patient, ranging from uncontrolled gorging through normal eating to satiety and obsessive fasting. It also provides some indication of the underlying states of the brain; as in the models of aggression, we are concerned with emotional states that probably originate in the limbic system. Psychological evidence suggests that the behavior variable may actually be a measure of the relative weight given by the limbic system to inputs from the body as opposed to inputs from the cerebral cortex. In a normal person these inputs may be in some sense balanced, but in the anorexic one or the other may tend to dominate.

Among the control parameters the normal factor is hunger, which in normal people governs the rhythmic cycle between eating and satiety. The splitting factor is the degree of abnormality of the anorexic's attitudes toward food; the abnormality steadily increases as her condition deteriorates. Diets become more severe, entire classes of foods are eliminated; carbohydrates are at first avoided and later actively feared.

The bias factor in the butterfly graph is loss of self-control, which can be measured by loss of weight. In the first phase of the disorder the anorexic's attitudes are already abnormal, but she has control of herself. As a result she is trapped on the bottom sheet of the behavior surface; all the time she is awake the limbic system remains in states corresponding to a fasting frame of mind, even when she is eating her minimal meals.

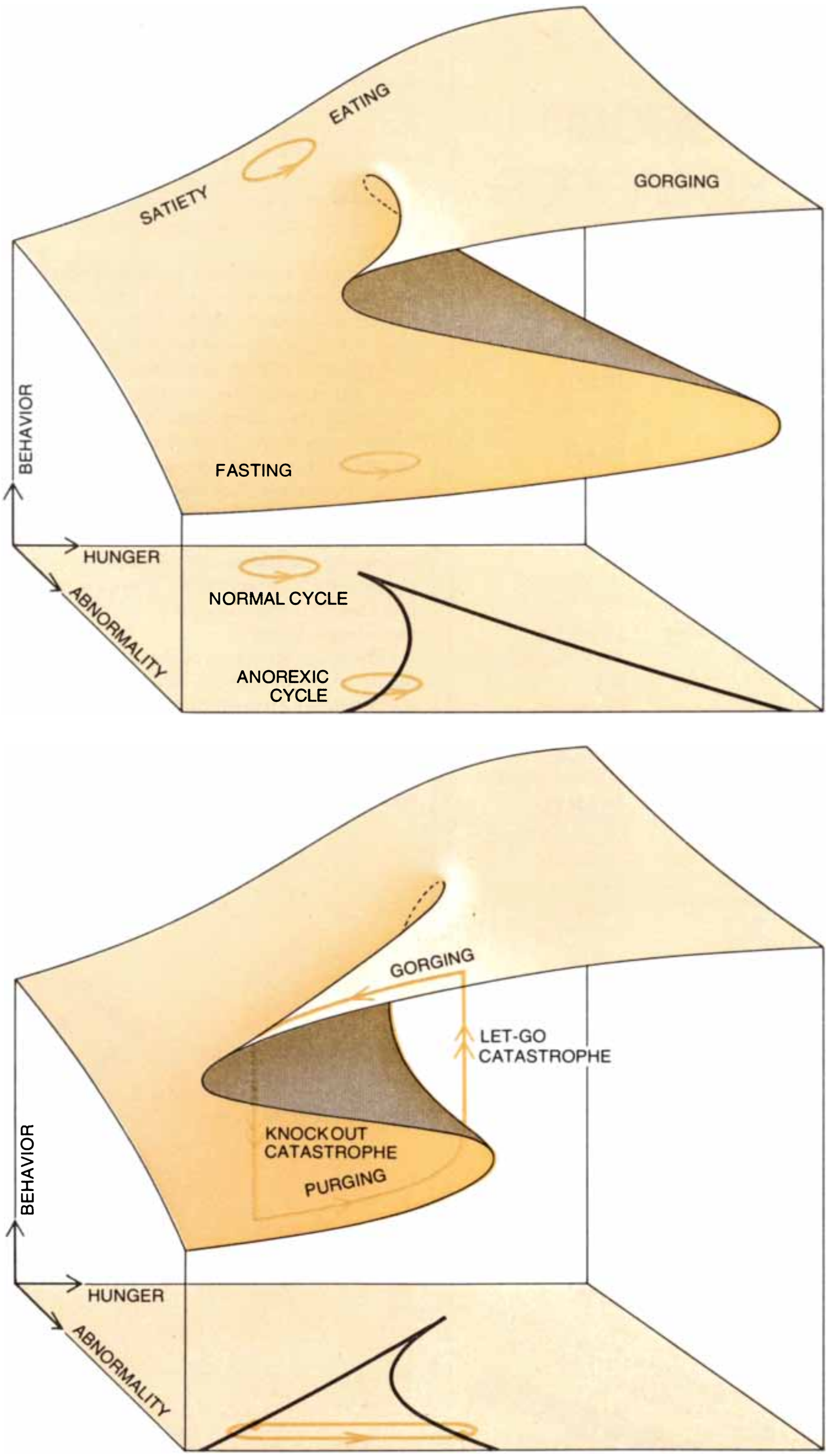
As the anorexic loses weight she also loses control, and the bias factor gradually increases. As a result the cusp swings to the left on the graph [*see illustration at right*]; if it moves far enough, the right-hand side of the cusp intersects the anorexic cycle, bringing the sudden onset of the second phase. Now the anorexic is no longer trapped in a cycle of constant fasting but is caught in a hysteresis cycle, jumping from the bottom sheet to the top one and back again. In the words of a typical anorexic, the catastrophic jump from fasting to gorging takes place when she "lets go" and watches helplessly as the "monster within her" devours food for several hours, sometimes vomiting as well. The catastrophic return to fasting comes when exhaustion, disgust and humiliation sweep over her, an experience that many anorexics call the "knockout."

The period of fasting that follows the "knockout" in the hysteresis cycle is different from the constant fasting of the first phase. It lies at a different position on the behavior axis and might better be called purging. The limbic state associated with the earlier fasting is dominated by inputs from the cerebral cortex and is directed toward forbidding food to enter. During gorging the limbic system is dominated by inputs from the body. The limbic state underlying purging is again dominated by cerebral inputs, but it has a bodily component directed toward ridding the body of contamination.

The trance therapy employed by Hevesi reassures the patient, reduces her insecurity and thereby enables her to regain access to normal behavior. Anorexics tend to sleep fitfully, and when they are awake, they experience naturally occurring trancelike periods; it is on these periods that the therapist builds. The trance may represent a third state of the limbic system, in the otherwise inaccessible zone between the gorging and the purging states. When the patient is fasting she views the outer world with anxiety and when she is gorging she is overwhelmed by that world, but during the trance she is isolated, her mind free both of food and of scheming to avoid food. It is only then that reassurance is possible.

Reassurance becomes the butterfly factor in the model. It creates the new sheet of the behavior surface, which lies between the other two sheets and which eventually gives access to the stable, normal region behind the cusp. Because therapy usually takes place during the fasting part of the cycle, entering the trance is a catastrophic jump from the bottom sheet onto the intermediate sheet. Coming out of the trance is another catastrophe, which can take the patient either to the bottom sheet or to the top one.

After about two weeks of therapy and in about the seventh session of trance the patient's abnormal attitudes usually break



ANOREXIA NERVOSA, a nervous disorder of adolescent girls and young women that involves obsessive fasting, can be described as a butterfly catastrophe. Two of the controlling factors are hunger and the abnormality of attitudes toward food. In normal people hunger leads to a cycle of behavior that oscillates between eating and satiety; in the anorexic person, with abnormal attitudes, the same hunger cycle leads to quite different behavior. In the first phase of the disease (*top*) the cycle is trapped on the bottom sheet of the behavior surface, and the anorexic remains constantly in a fasting frame of mind. The second phase (*bottom*) is induced by a change in a third controlling factor, self-control. As the patient loses control of herself over a period of two years or more, the bifurcation set is gradually skewed to the left until the hunger cycle crosses the right-hand side of the cusp. She then enters a hysteresis cycle: she fasts until hunger causes her to catastrophically "let go," then she gorges herself until, after a "knockout" catastrophe, she reverts to fasting and purges herself of what she perceives as contamination.

down catastrophically and the personality is fused into a complete whole again. When the patient awakens from this trance, she may speak of it as a "moment of rebirth," and she finds that she can eat again without fear of gorging. The trance has seemingly opened a pathway in the brain back to the more balanced limbic states, so that the patient regains access to normal behavior. Subsequent trance sessions reinforce the experience.

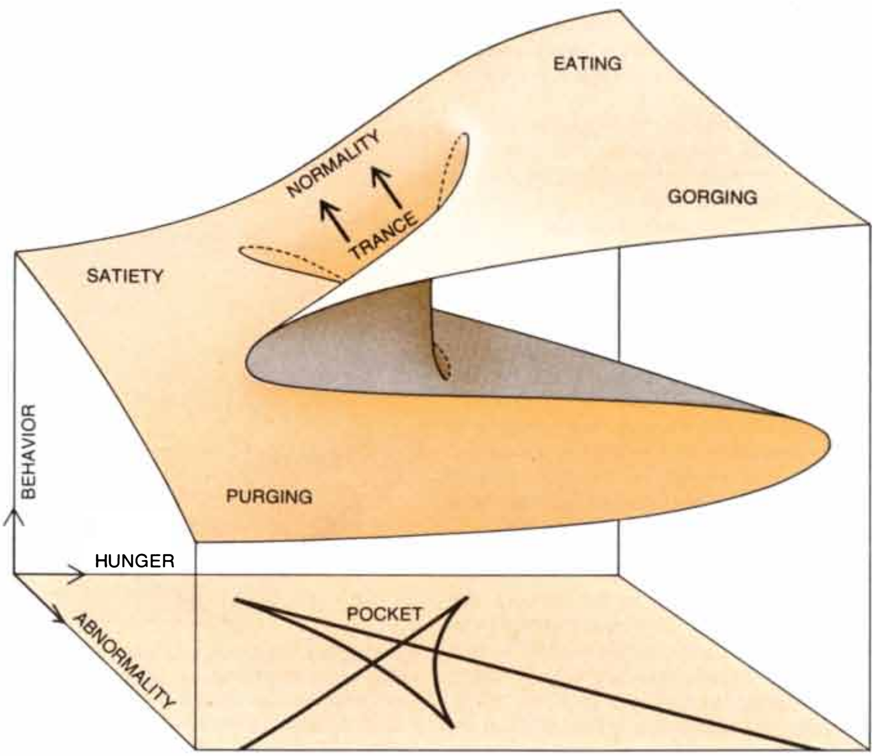
The model of anorexia presented here is incomplete in several respects. I have omitted an additional control factor, drowsiness, which governs the behavioral distinction between waking and sleeping and the associated catastrophes of falling asleep and waking up. As a consequence the path from trance to normality in the model is misleading in that it omits the catastrophe of awakening. I have also not discussed the other half of the model, which concerns personality as opposed to behavior and which explains the escalation of the disorder, its rigidity, the stability of the abnormal attitudes and the breakdown of that stability at the moment of cure.

One of the strengths of the catastrophe-theory model of anorexia is that it explains the patient's own description of herself. The seemingly incomprehensible terms in which some anorexics describe their illness turn out to be quite logical when viewed in the framework of the catastrophe surfaces. The advantage of a mathematical language in such applications is that it is psychologically neutral. It allows a coherent synthesis of observations that would otherwise appear to be disconnected.

The Future of Catastrophe Theory

Catastrophe theory is a young science: Thom published the first paper on it in 1968. So far its greatest impact has been on mathematics itself; in particular it has stimulated developments in other branches of mathematics that were required for the proofs of its theorems. The most important outstanding problems in the development of the theory concern the understanding and classification of generalized catastrophes and the more subtle catastrophes that arise when symmetry conditions are imposed. In addition there are problems associated with how catastrophe theory can be employed in conjunction with other mathematical methods and concepts, such as differential equations, feedback, noise, statistics and diffusion.

New applications of the theory are being explored in many fields. In physics and engineering, models have been developed for the propagation of shock waves, the minimum area of surfaces, nonlinear oscillations, scattering and elasticity. Michael V. Berry of the University of Bristol has recently employed the umbilic catastrophes



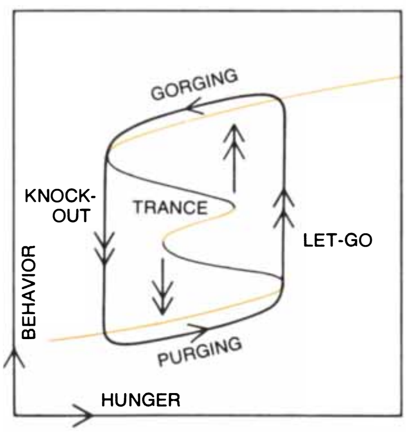
TREATMENT OF ANOREXIA relies on creating a third, intermediate mode of behavior. The new behavior is made possible by increasing the fourth control parameter of the butterfly catastrophe: reassurance. The effect of this fourth parameter is to create the pocket in the bifurcation set and thus to create the intermediate sheet in the behavior surface. In a system of therapy developed by J. Hevesi, a British psychotherapist, reassurance is encouraged by putting the patient in a trance. Initially the patient enters and leaves the trance through catastrophic

to predict new results in the physics of caustics and fluid flow, and he has confirmed those results by experiment.

Thom's own *Structural Stability and Morphogenesis,* inspired by the work of D'Arcy Wentworth Thompson and C. H. Waddington, was largely concerned with embryology, but as yet few biologists have pursued his ideas in the laboratory. I have constructed catastrophe models of the heartbeat, the propagation of nerve impulses and the formation of the gastrula and of somites in the embryo. Recent experiments conducted by J. Cooke of the Medical Research Council laboratories in London and by T. Elsdale of the Medical Research Council laboratories in Edinburgh appear to confirm some of my predictions.

Most of my own work, however, has been in the human sciences, as is suggested by the models described in this article. An increasing number of investigators are now suggesting models derived from catastrophe theory, and in the coming decade I look forward to seeing those models tested by experiment. Only then can we judge the true worth of the method.

Thom has employed the theory in an endeavor to understand how language is generated. It is an intriguing thought that the same mathematics may underlie not only the way the genetic code causes the embryo to unfold but also the way the printed word causes our imagination to unfold.



jumps from the intermediate sheet to the top and bottom sheets, as in the section at right. When the course of therapy has been successful, however, the patient makes a smooth transition from the intermediate sheet to the normal modes of behavior behind the pocket.

Opals

What makes these gems sparkle with flecks of pure spectral color when they are rotated in white light? The answer has been found in the optical behavior of regularly stacked tiny spheres of silica

by P. J. Darragh, A. J. Gaskin and J. V. Sanders

Most gemstones are large single crystals. Opal is not crystalline but is an amorphous form of silica. There are large natural deposits of "common opal," an opaque, milk white amorphous silica of no commercial value. In a few scattered deposits around the world, usually in association with weathered volcanic rocks, there is "precious opal," a mineral that can exhibit sparkling colors ranging across the entire visible spectrum, from violet to deep red. How an amorphous material can give rise to such vivid color displays was long a mystery to mineralogists and gemologists. Now the explanation is known.

The use of opal as a gemstone dates back at least to 400 B.C. The main source of precious opal from that time until late in the 19th century was mines in eastern Czechoslovakia, where commercial production continued until 1932. Some Mexican opal was brought back to Europe by explorers as early as 1520, but commercial development did not begin for another three centuries. The present mines near Querétaro and Magdalena lie near the center of a long geographical belt where precious opal is found in widely scattered localities. The belt stretches from Washington, Oregon and Idaho in the northwestern U.S. through Central America to Brazil (where a deposit is currently being mined).

Opal was first discovered in Australia in 1872. About 1890 the first of a number of extensive deposits was found at White Cliffs, some 500 miles north of Melbourne in New South Wales. Today the annual Australian production of precious opal is valued at about $30 million and dominates the world market. The Australian deposits are distinctive because the associated host rocks are not volcanic, as they are in other parts of the world, but sedimentary. The main opal fields are at Lightning Ridge, Coober Pedy and Andamooka. Here the opal fills assorted veins and cavities at levels within 40 meters of the surface. As we shall see, it is significant for such deposits that the environment is extremely arid.

Since opal is valued only as a gemstone, its color and beauty are of prime importance. In other precious stones the color is either the uniform body color seen in an emerald or a ruby or the fiery flashes seen in a faceted diamond. Pearls, which are not crystalline minerals but biological artifacts, exhibit a rainbowlike iridescence. The body colors of crystalline gemstones are created by the selective absorption of some of the wavelengths of white light as it passes through the mineral. The spectrum of the emergent colored light shows dark bands at wavelengths that are characteristic of the particular gem.

The "fire" of diamond and other materials with a high degree of color dispersion is produced by the facets of the stone, which act as prisms to spread the emergent light into a continuous sequence of spectral colors over a range of angles large enough to be resolved by the eye. The extent of the spreading depends not only on the optical dispersion of the material, which is a measure of the difference in the refractive index of the substance for different wavelengths of light, but also on the length of the light path through the stone. In the art of cutting a gem to ensure the optimum dispersion effects it is important to get the longest possible light path within the stone by arranging the facets so that the rays are internally reflected several times before they emerge. The larger the stone, the more dramatic the spectral separation, because of the longer light paths. The iridescent colors seen in pearls are the result of selective interference between light waves reflected from regularly spaced layers of calcium carbonate, which are separated by organic films.

The brightly colored beams that flash out of precious opal as the gem is rotated while being viewed in white light, preferably from a single small bright source, bear little obvious relation to the color displays of other gemstones. The fire does not depend on the size of the gem, and it is not the result of skillful faceting. Most opals are usually fashioned in the simple, rounded style known as cabochon (from *caboche,* an old French word for head).

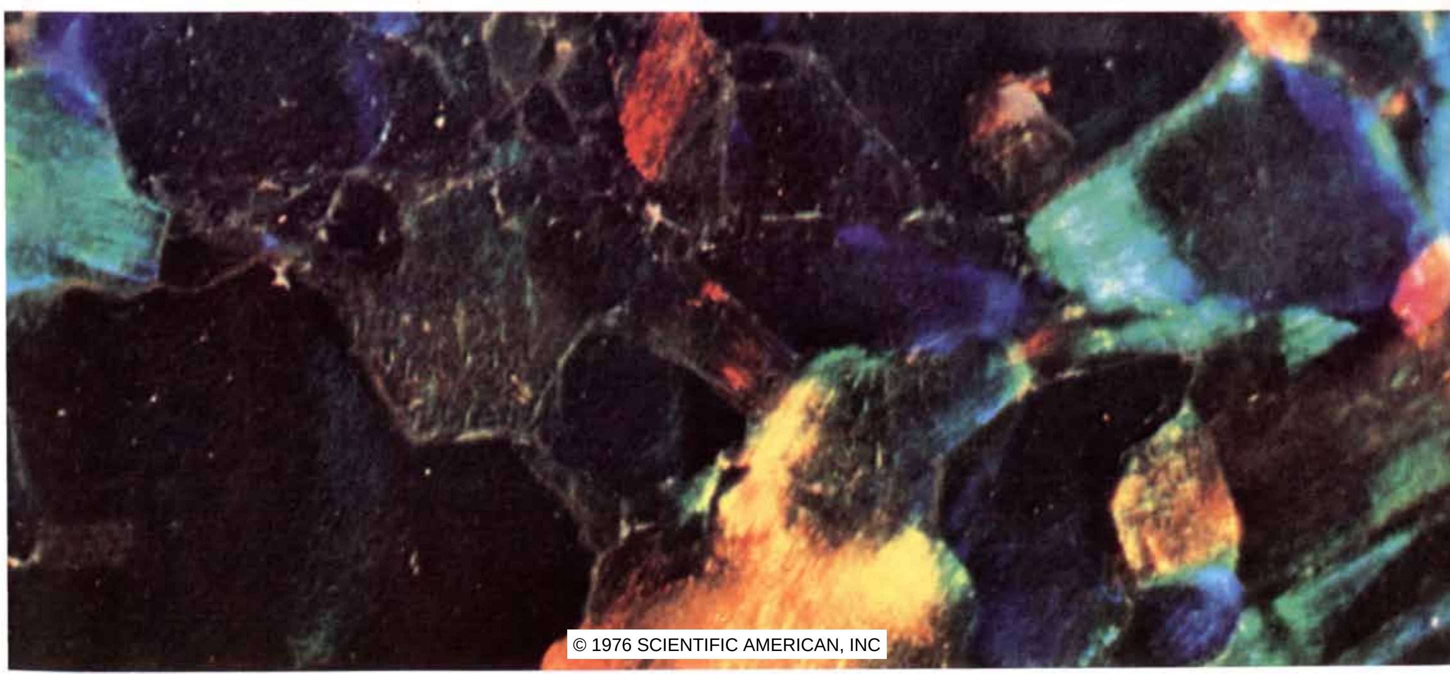
Opal is essentially an amorphous, hydrated form of silica that normally has the same optical properties throughout. The mechanism responsible for opal's unique fire was for a long time an enigma. Various explanations invoked interference effects from hypothetical thin internal films, but no evidence for the existence of such films could be found. It was recognized, however, that if there were systems of regularly spaced points or lines in the material, a mechanism capable of producing the pure colors of opal was diffraction.

There are three significant features in the typical appearance of opal that provide clues to the mechanism responsible for the color display. First, the color is associated with small grains, and throughout each grain the color is fairly uniform. Although the grains are irregular in shape, their size within any particular stone is more or less the same. In "pinfire" varieties of opal the grains are up to a millimeter across, but in the more typical varieties they are several millimeters across. Some stones have grains as large as a centimeter, in which case the entire gem may have just one or two large areas of uniform color. The character and value of a stone are greatly influenced by the size and arrangement of the grains, particularly if some attractive pattern is present.

The second significant feature of opal is the way the colors of the grains change as the orientation of the stone is changed with respect to the light source and the observer. This can readily be demonstrated by making several photographs of an opal with the camera held in a fixed position while the direction of the incident light is varied between exposures. The different grains show colors in some positions but not in others; moreover, the colors of the grains change [*see illustration on opposite page*]. The effect is known as the play of color.

The play of color is unique to each opal and is distinctly different from the rapid sequence of colors across the complete spectrum exhibited when diamond and other gems of high color dispersion are turned through a small angle. Most opals show

PLAY OF COLOR in an opal is shown in the photographs on the opposite page. All three pictures are close-ups of the same area of the same stone, with the camera in the same position. In each picture, however, the angle of illumination is different. Different grains are colored in each picture. From the middle picture to the bottom one the large grain at bottom center changes hue from red to yellow.



Your kind of Books...

COMPUTER POWER AND HUMAN REASON
From Judgment to Calculation

Joseph Weizenbaum

We do not know how to make a "wise" computer, and yet, as this book brilliantly documents, scientists, politicians, and ordinary citizens are today relinquishing to computers the responsibility for making genuinely human choices—choices that demand true wisdom. *Computer Power and Human Reason* is an account of the impact of technology on man's self-image and of his abandonment of judgment for calculation. It is a distinguished scientist's searching probe of the limits of computer power and of scientific rationality itself. It is above all an eloquent and moving defense of the sanctity of the human spirit. Joseph Weizenbaum is professor of computer science at MIT.

1976, 300pp., 13 illus., $9.95

A Book-of-the-Month Club Alternate Selection

FIELD PHOTOGRAPHY
Beginning and Advanced Techniques

Alfred A. Blaker

The former Director of the University of California Scientific Photographic Laboratory shares his many tricks of the trade in this clear, comprehensive, and beautifully illustrated book. *Field Photography* includes more than 70 original photographs (18 in full color), 85 drawings, and a separate, portable field supplement of technical data. This is sure to become the standard, indispensable reference for both professional and amateur nature photographers.

1976, 445 pp., 156 illus., $19.95

GEOLOGY ILLUSTRATED

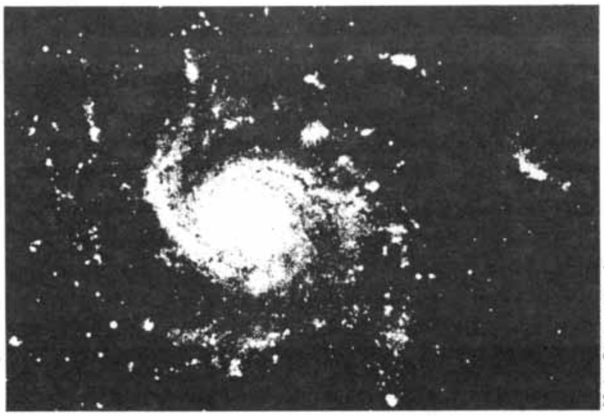
John S. Shelton

Drawings by Hal Shelton

Geologist/pilot Dr. John Shelton has been taking aerial photographs since 1939, and in this, his master work, he assembles nearly 400 of these photographs to produce a breathtaking view of the natural features of our home planet.

"Worth buying for the photos and book design alone, but you'll probably find yourself becoming interested in geology regardless of your original intentions."
—*The Whole Earth Catalog*

1966, 434 pp., 382 illus., $14.50



Hale Observatories

HIGHLIGHTS IN ASTRONOMY

Fred Hoyle

As his millions of readers will attest, Fred Hoyle has a flair for sharing the very latest astronomical discoveries with the general reader. Hoyle's latest book, *Highlights in Astronomy*, is a lavishly illustrated tour of the universe. The 149 illustrations include nearly 100 photographs, some of them never before published. This is quite simply the best popular description of what modern astronomy is all about.

1975, 179 pp., 149 illus., softcover, $5.95

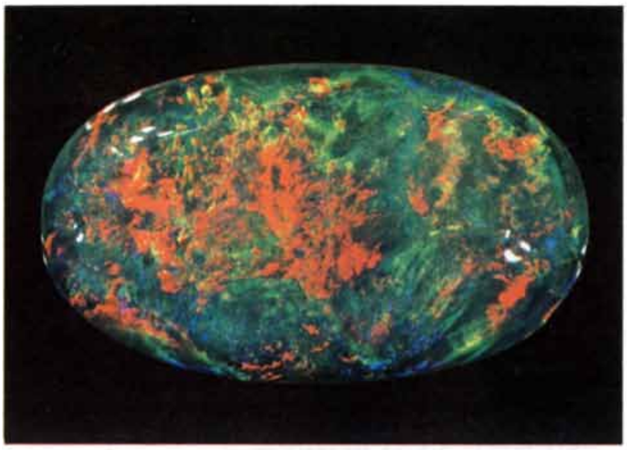
HELPLESSNESS
On Depression, Development, and Death

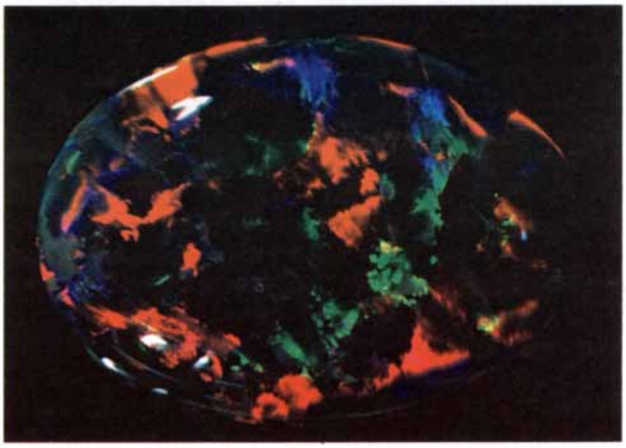
Martin E. P. Seligman

"A place to look for the problem when that hideous vague sense that 'something's wrong' begins to grow."
—*CoEvolution Quarterly* (of the *Whole Earth Catalog*)

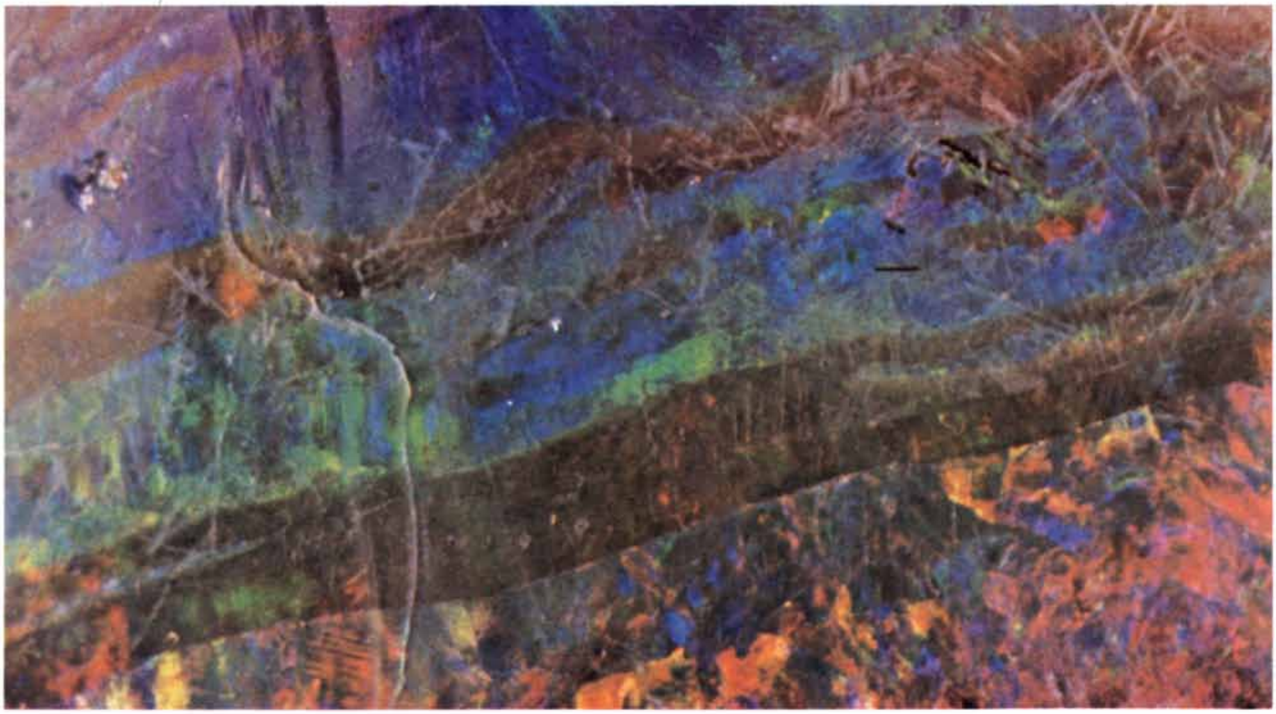
Helplessness is Dr. Martin Seligman's new theory of anxiety and depression, the result of a ten-year search for the causes and cures of these widespread disorders. Dr. Seligman's research suggests that depressed individuals can *learn* to exert greater control on their environments and thereby lessen their feelings of depression. Of special interest is Seligman's suggestion that early childhood experiences with control may effectively immunize a child against depression for life.

1975, 250 pp., 8 illus., softcover, $4.95





FOUR TYPES OF OPAL are shown in these photographs made by John Cubitto of the Gemological Institute of America. At top left is black opal, the most valuable form of opal; the stone is from Lightning Ridge in Australia. At top right is the commoner white opal; it is from Coober Pedy in Australia. At bottom left is Mexican opal, characterized by reddish brown color. At bottom right is synthetic opal.



SECTION OF RAW OPAL from Queensland in Australia is unusual in that from one angle its color ranges from violet at the top to red at the bottom. Electron microscopy of the material showed that the size of silica spheres of the opal increased smoothly from top to bottom.

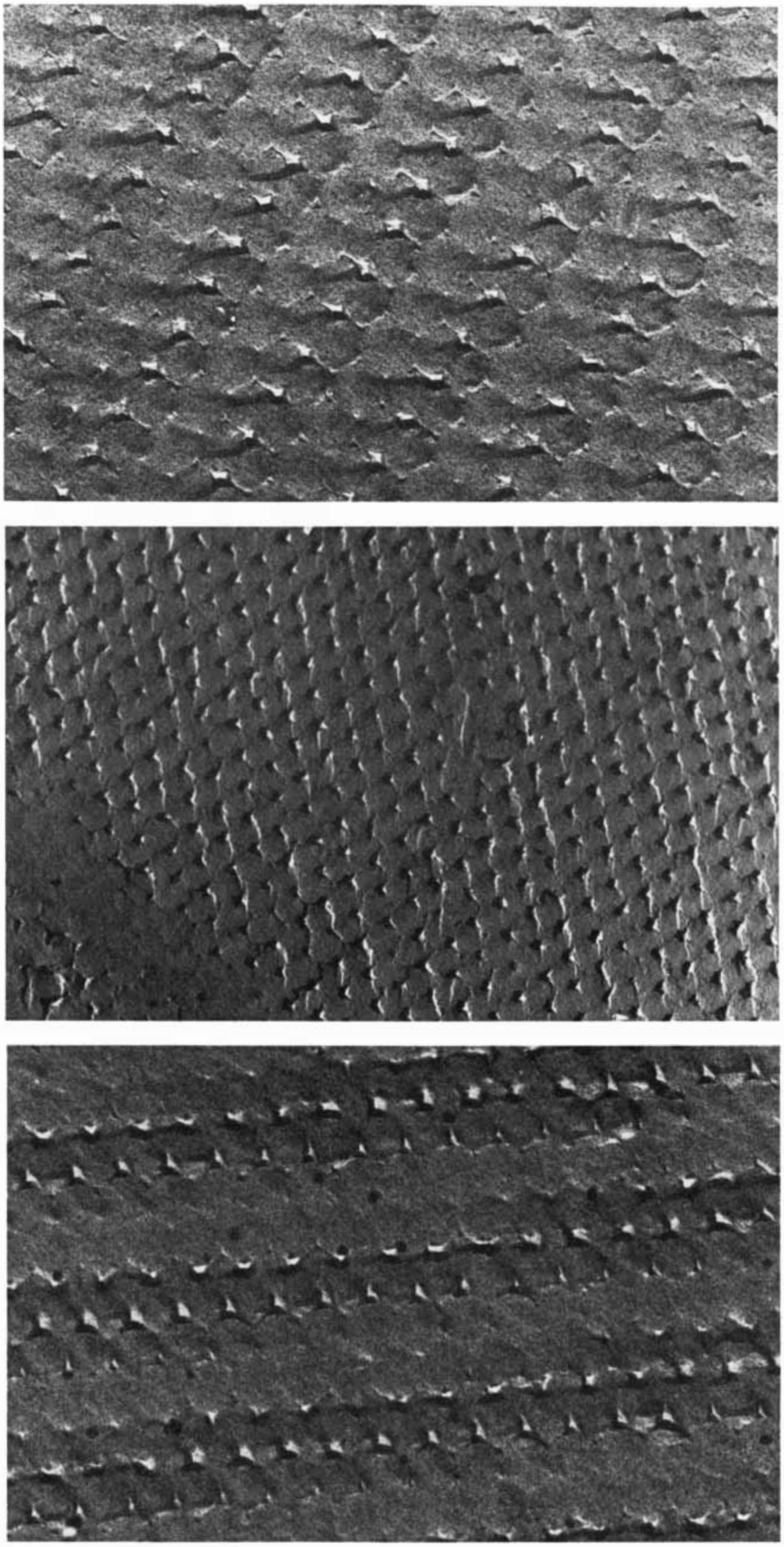
practically no color when they are viewed solely by transmitted light, but strong colors appear to be reflected from surfaces within the grains when an opal is viewed by direct reflection, with the light source immediately behind and slightly above the observer. When the stone is viewed in this way, it usually exhibits the fire of the longest wavelength it is capable of showing. In a good stone that wavelength would correspond to a deep primary red, and all the other colors of the visible spectrum would appear as the stone is rotated to other positions.

The third significant feature of opal fire is that generally the colors are spectrally pure, representing only a very narrow portion of the visible spectrum at any given viewing position for a single grain. Shades of color made up of mixtures of different wavelengths are rarely seen. When this feature is taken into consideration with the other two features, a mechanism based on some form of internal diffraction seems unavoidable. Although such explanations were suggested from time to time over the past century, no evidence for the existence of regular light-scattering structures was found in the purest forms of precious opal until about 10 years ago.

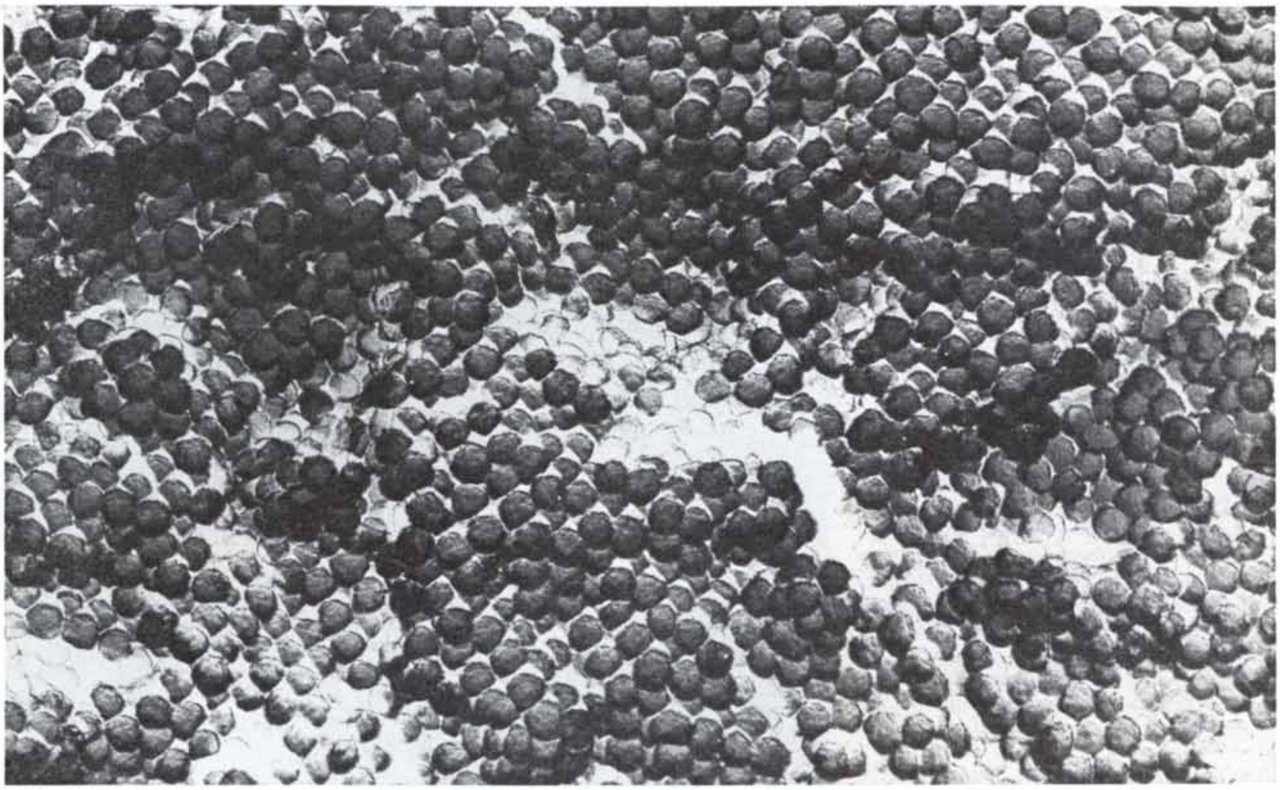
The evidence for periodic structures in opal was discovered by our group in the Commonwealth Scientific and Industrial Research Organization in Australia while we were trying to resolve a different question raised by one of our colleagues, E. Ralph Segnit. He asked whether opal contained voids capable of holding free water. In an effort to find out we applied the technique of transmission electron microscopy, which, as is well known, can reveal details several hundred times smaller than those that can be seen with the ordinary light microscope.

For our study we made thin-film replicas of the freshly fractured surfaces on chips of opal specimens. The cleaved surface is first coated in a vacuum with a thin layer of platinum, deposited at a shallow angle to reveal the surface topography by shadowing. The surface is coated evenly with a layer of carbon, which adheres to the platinum and forms a tough composite film. The opal chip is then dissolved away in hydrofluoric acid, and the film is viewed by the transmission of electrons in the electron microscope. The carbon layer is transparent to electrons, so that the different amounts of platinum deposited in different regions faithfully reveal the topography of the original opal surface.

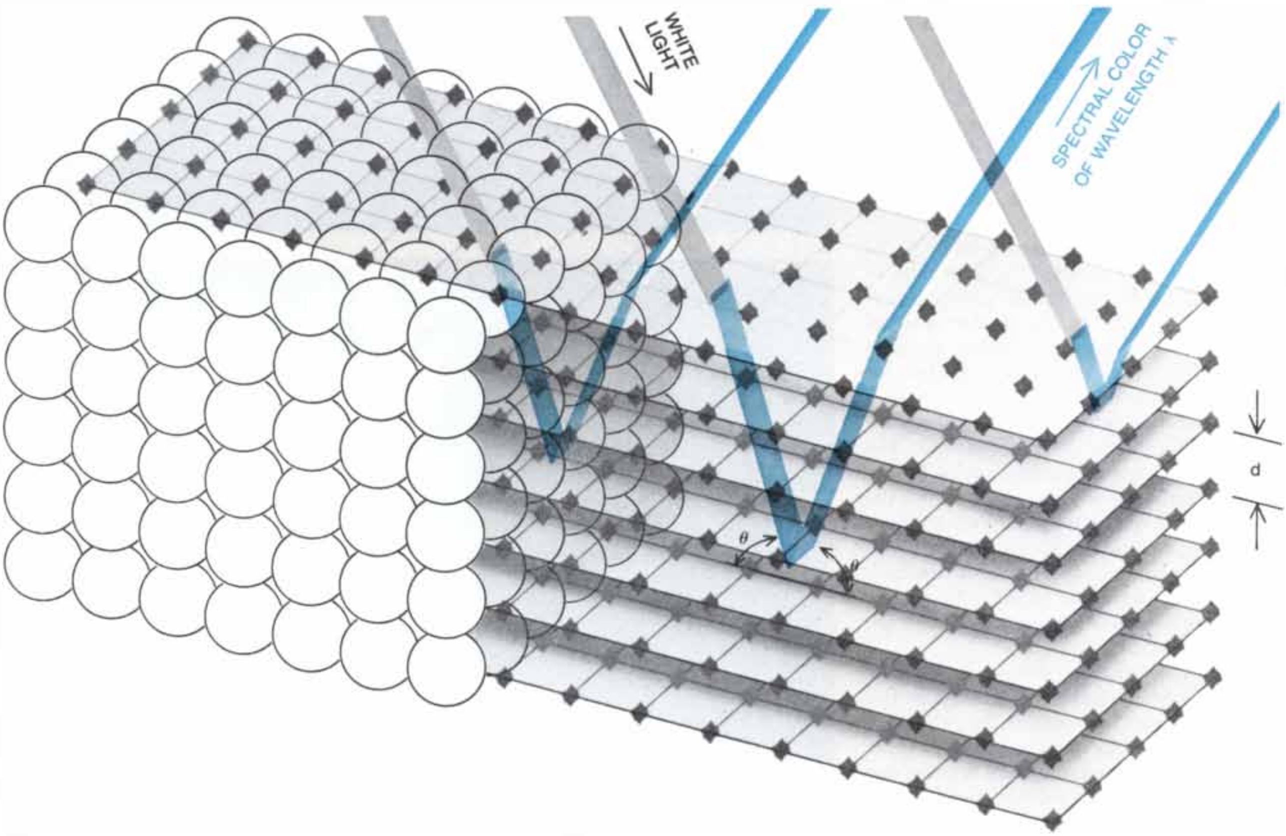
To our surprise and delight the electron micrographs revealed regular geometric patterns of tiny holes across the entire surface. Although the patterns and the shapes of the holes were found to vary between one fracture surface and another and between one opal grain and another, they could be seen on every fracture surface. It was quickly apparent that throughout the body of a precious opal there are arrays of holes that are sufficiently regular to act as three-di-



INTERNAL STRUCTURE OF PRECIOUS OPAL is shown in electron micrographs of replicas of three freshly fractured and platinum-shadowed surfaces of the same specimen. The replicas, magnified 30,000 diameters, exhibit different patterns because the fracture planes intercept the three-dimensional array of the opal structure at three different angles. Regularly spaced cusps are sections of voids, or holes, that exist between closely packed spheres of silica.



SURFACE OF PRECIOUS OPAL, after etching with hydrofluoric acid, is seen to consist of spherically shaped particles. The magnification is 19,000 diameters. In precious opal the amorphous silica spheres usually have diameters ranging from 150 to 300 nanometers.



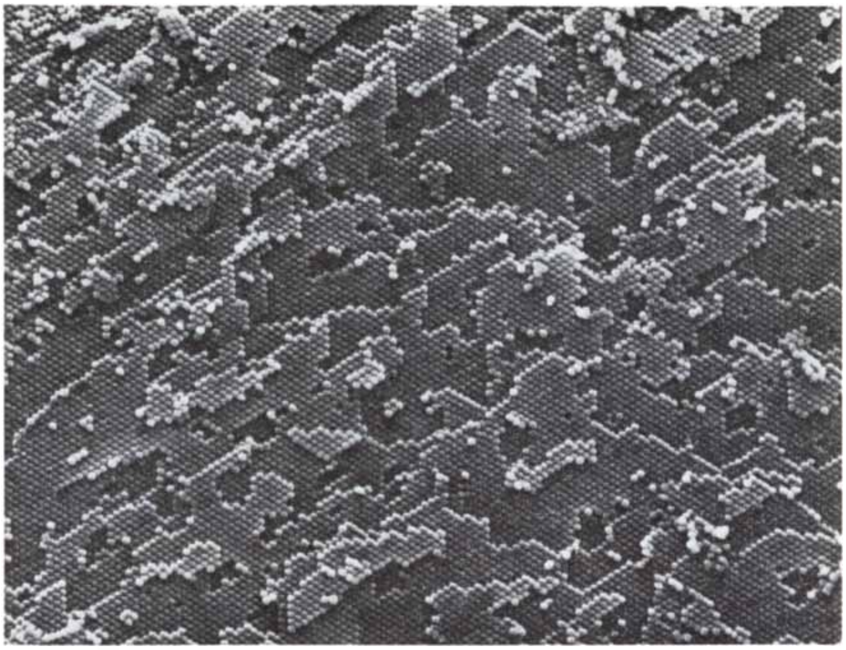
SPECTRAL COLORS OF PRECIOUS OPAL arise because light rays are diffracted by an orderly array of holes, which are separated by a spacing, *d*, that is much smaller than the wavelength of light. Coherent interference between light waves reinforces specific wavelengths extending in certain directions, as if a given wavelength were reflected from a set of parallel planes that passed through the scattering points. The wavelength of the diffracted spectral color is determined by the value *d* of the spacing and varies with the angle θ.

mensional diffraction gratings and hence to give rise to the stone's characteristic fire. Our excitement over this finding would have been somewhat dampened if we had known that at almost the same time J. Pense of the University of Mainz was giving a lecture describing his discovery of a regular fine structure in opal, also revealed by electron microscopy. He realized that the play of colors in opal was generated by diffraction from the regular arrays in the material, but apparently he did not pursue the matter further. We came on an abstract of his talk years later.

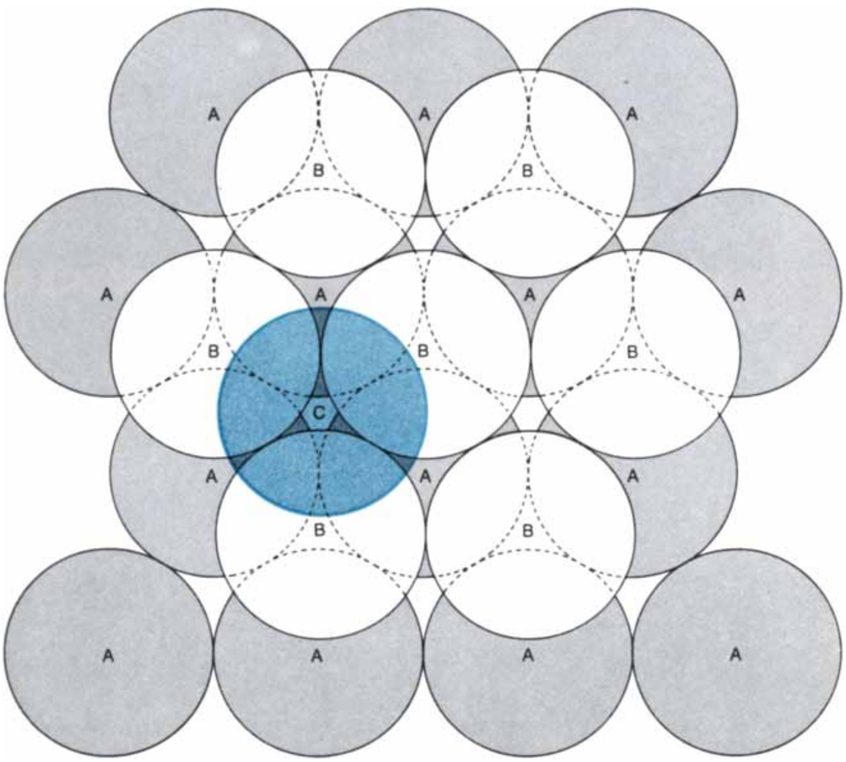
The next problem was to account for the existence of the arrays of holes. No explanation seemed satisfactory until one of us suggested that if the opal had formed as a collection of closely packed spherical particles of amorphous silica, in other words as an aggregation of giant colloidal sol particles, the holes could well represent the spaces between the packed spheres. The existence of such spheres was confirmed when we made electron micrographs of the surface of opals that had been etched with hydrofluoric acid.

We subsequently examined many specimens of gem opal in this way and found that all those exhibiting strong color displays contain spheres that are remarkably uniform in size and very regularly stacked, so that they have regular arrays of voids. The particular geometric pattern of the array and the shape of the voids seen in any fracture surface are largely incidental features determined by the angle at which the fracture intersects the array. Sphere diameters measured in normal varieties of opal range from 150 to 300 nanometers. In a few exceptional samples from volcanic localities, however, we have found spheres as much as 30 times larger (up to .01 millimeter in diameter).

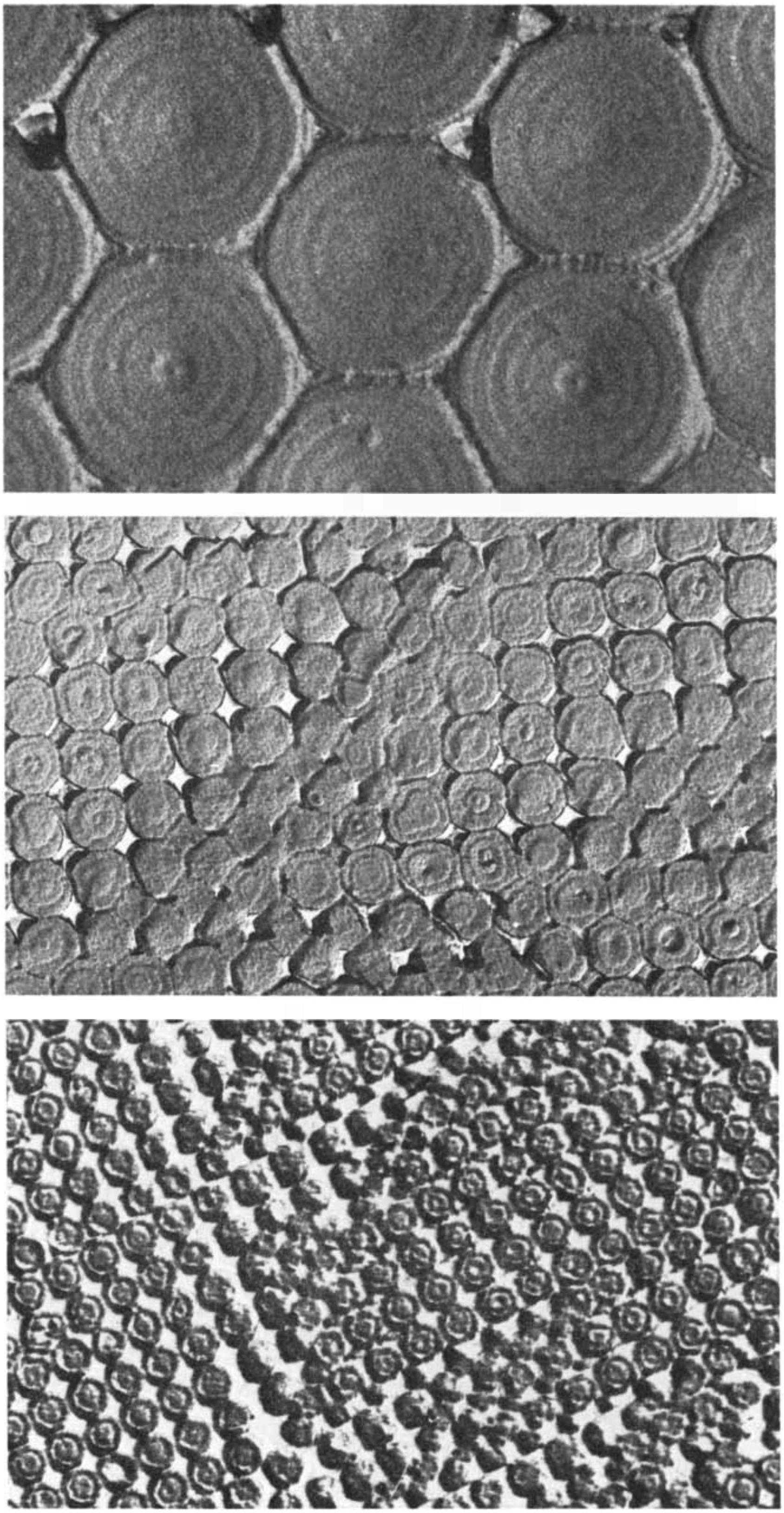
With the information from the electron microscope the mechanism that creates the color display of opal becomes clear. The body of the stone consists of transparent spherical particles of amorphous silica that are tightly packed together in good optical contact. The material is therefore substantially transparent and has the refractive index of amorphous silica, which lies between 1.435 and 1.455, depending on the degree of hydration. (The refractive index is a measure of the ratio of the velocity of light in a vacuum to the velocity of light in a transparent material medium, and hence it is a measure of the degree to which light is bent when it enters the material medium. The refractive index of water, for example, is 1.33 and that of diamond is 2.417.) At the interfaces of the voids, however, there is a reduction in the refractive index, because the voids appear to contain only air, water vapor or liquid water. (Thus the answer to Segnit's question was yes.) These tiny, regularly arranged discontinuities act as light-scattering points, forming a three-dimensional diffraction grating from which beams of a single wavelength are shined at specific angles of incidence and reflection when the



HEXAGONAL PACKING OF SILICA SPHERES is shown in this electron micrograph of an unusual opal sample in which the spheres are only weakly cemented together. As a result of this weak bonding, when the sample was fractured, the break left the spheres in adjacent layers intact rather than cleaving them, which is what happens in more typical samples. Also as a result of the weak bonding this opal was chalky in appearance and therefore of no commercial value.



ARRANGEMENT OF ADJACENT LAYERS of hexagonally packed spheres can follow either of two general systems. In both systems the first two layers, *A* and *B*, nest together in the same way, the only way possible for close-packing. The difference arises in the positioning of the third layer, where there are two possibilities. If the spheres in the third layer are centered directly over those in the first (*A*) layer, the sequence becomes simply *ABABAB*.... This is called hexagonal close-packing. In the alternative arrangement the spheres in the third layer (*C*) have no sphere directly below in layers *A* and *B*, thus yielding the sequence *ABCABC*.... This produces a face-centered-cubic structure. A layer that departs from either of these orderly arrangements introduces what is called a stacking fault. In normal opals the layers are generally stacked in a random mixture of the two patterns, with the cubic sequence slightly favored.



CROSS SECTIONS OF OPAL PARTICLES show they are built up from concentric shells of primary particles typically 20 to 50 nanometers in diameter. When an opal is fractured, the break usually passes through the silica particles rather than between them. By etching the freshly cleaved surface with hydrofluoric acid and making a platinum-shadowed replica of it the internal structure of the particles is revealed. These electron micrographs of replicas show particles enlarged 30,000 diameters. Unusually large particles in opal from Idaho (*top*) have six concentric shells. Normally (*middle, bottom*) two or three shells are seen in each particle.

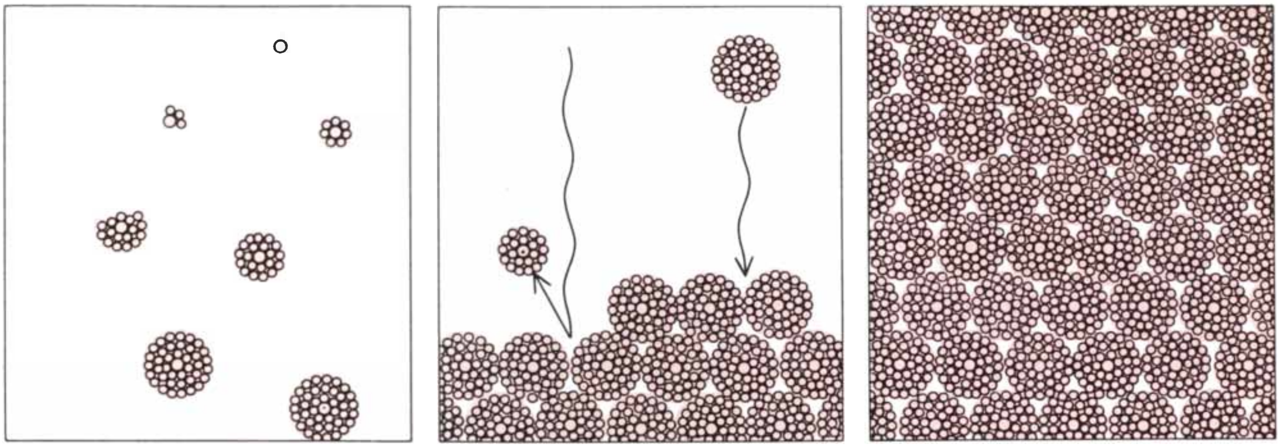
spacing between the voids is of the order of the wavelength.

The light-diffracting system of precious opal is in many ways a large-scale analogue of the well-known phenomenon of the diffraction of X rays by scattering from the regular arrays of atoms in a crystal. X rays have a wavelength of about a tenth of a nanometer, which is of the same order as the spacing between the atoms in a crystal lattice. Since the wavelengths of light range from 400 to 700 nanometers, the spacing between the voids in opal must approach those dimensions if diffraction is to take place. If we assume that the diffraction laws of X-ray crystallography apply, the relation between the spacing (d) of the planes of voids and the angle (θ) at which a diffracted wavelength (λ) can appear from opal is given by the Bragg equation: $n\lambda = 2d\mu\sin\theta$, where n is the "order" of the reflection and μ is the index of refraction. Since we are mainly interested in the first-order, or brightest, reflection, n has the value of 1. For μ we must insert the refractive index of opal, 1.45, because the actual process of diffraction takes place within the silica of the opal, where the wavelengths of light are reduced.

From the equation one can deduce that the range of the wavelengths of the colors seen in an opal will increase with the size of the spheres in the diffraction array and that colors of short wavelength will be seen only at low angles of incidence (low values of θ). For any given sphere size the wavelength diffracted will reach a maximum when θ is 90 degrees, in which case $\sin\theta$ equals 1 and the equation reduces to $\lambda = 2\mu d$, or $\lambda = 2.9d$.

These deductions are in accord with what is seen when different types of opal are selected by examination in white light and their appearance is correlated with the sphere sizes revealed by electron micrographs. Specimens containing only very small spheres, and therefore showing small values of spacings between voids, exhibit no visible colors. Violet light, with a wavelength of 400 nanometers, first appears when the spacing d equals 400 divided by 2.9, or 138 nanometers. Opals with larger values of d show a greater variety of colors ranging from violet and blue (at low viewing angles) to green, yellow or orange (at high angles). Red first appears when d equals 700 divided by 2.9, or 241 nanometers, and the material is viewed with the light source very near the eye, so that θ is as close to 90 degrees as possible. Opal cabochons that show flashes of deep red can exhibit all the other colors of the spectrum when they are viewed at suitable angles; they are therefore more valuable than stones whose maximum wavelengths are shorter than those of red.

A rough rule for estimating the radius of the spheres in a specimen of opal is to determine the longest wavelength of light that can be observed and divide by 5. Thus if one can see a red of 700 nanometers, the spheres in the opal must have a radius of about 140

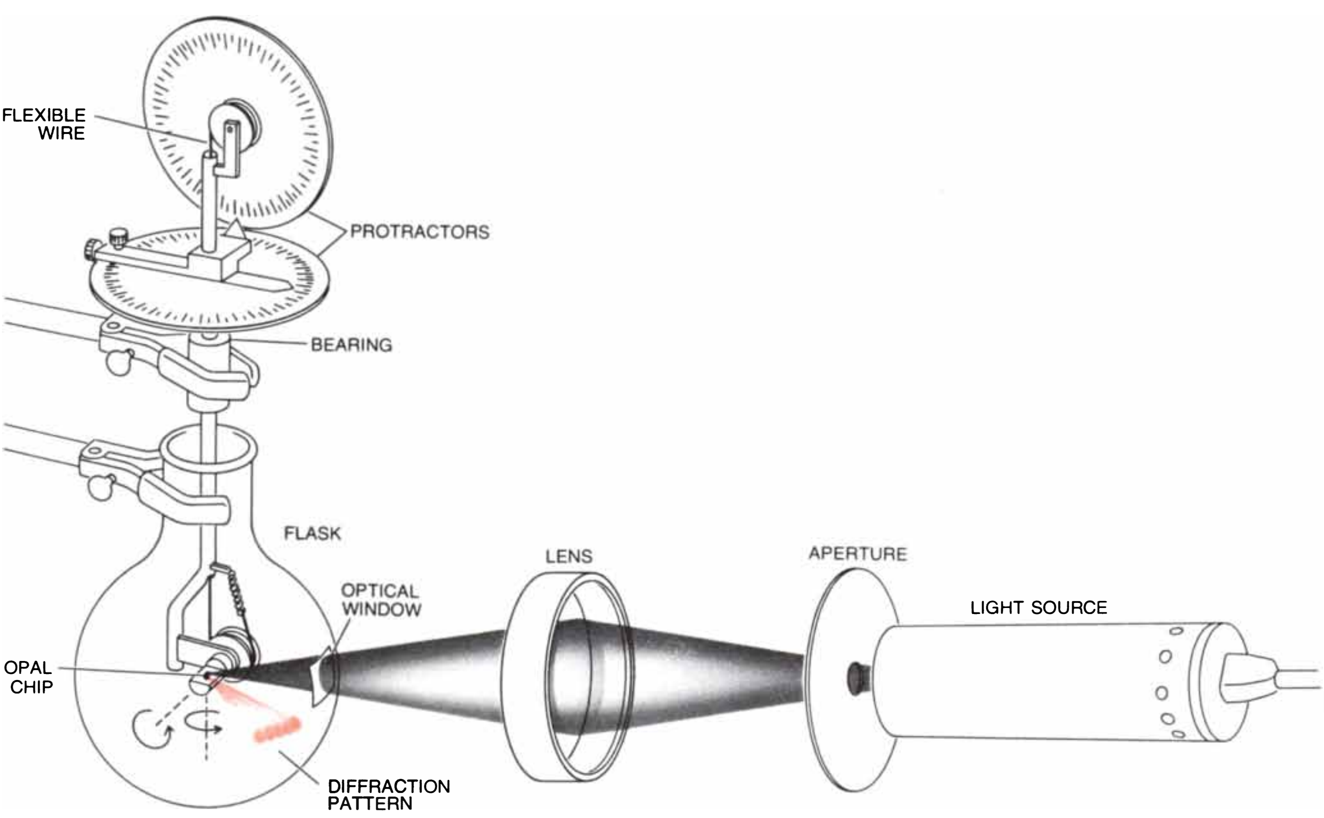


PROBABLE MECHANISM for the formation of the shell structure seen in opal particles is proposed by the authors. The evaporation of water from a solution of silica produces homogeneously sized particles, 20 to 50 nanometers in diameter, that cluster around a central particle (*left*). Spheres of uniform size slowly pack together in hexagonal arrays, while odd-sized particles are rejected (*middle*). The spheres become cemented together in a close-packed structure, leaving voids that act as a three-dimensional diffraction grating (*right*).

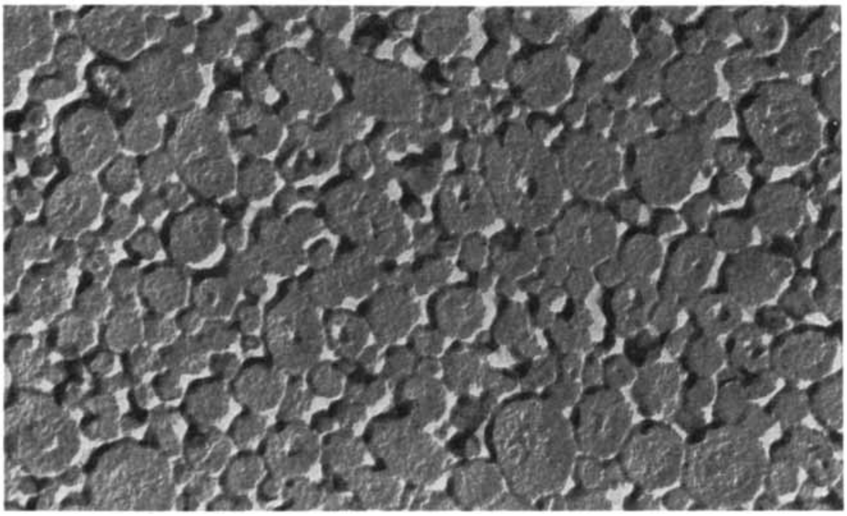
nanometers. This formula takes into account the geometrical relation between the size of the spheres and the spacing of the array of interstitial holes. It also assumes that all the spheres in any one specimen are the same size, which is generally true for precious opal. Some large slabs of opal, as they come from the mine, show obvious differences in sphere size between one side of the piece and the other, as is indicated by variations in the range of the wavelengths diffracted.

The diffraction effects are strongest in opals with distinct voids; the colors are less intense when the voids are smaller and when the amount of silica deposited between the spheres is larger. We now know that in the material called jelly opal the spheres are poorly defined. In Mexican opal and some of the opal from other volcanic deposits the filling of voids is so complete that discontinuities are no longer visible in electron micrographs of fracture surfaces; nevertheless, a periodic variation of structure can still be resolved by etching. We believe that here the refractive index of the secondary deposit of silica is sufficiently different from that of the spheres to produce diffraction effects.

Irregular arrays of voids give rise to a



APPARATUS FOR STUDYING DIFFRACTION in opals is similar to that used for recording the X-ray-diffraction pattern of crystals. The flask is coated externally to form a translucent screen except for a small window through which a collimated beam of white light can shine. The light strikes the opal and is diffracted backward at angles of up to 180 degrees, forming a pattern of streaks and spots on the outer surface of the flask. Colors of the longest wavelength the opal is capable of diffracting appear nearest the window. As the opal is rotated, colors of increasingly short wavelength appear in sequence. The orientation of the opal is read from the two protractors.



POTCH OPAL IS COMPOSED of particles of varying shapes and sizes, as shown in this electron micrograph of a fracture-surface replica, magnified 50,000 diameters. Because the irregular particles cannot form orderly arrays such opal is incapable of producing diffraction pattern.

general scattering of white light, causing an opal to have the milky appearance described by the term opalescent. Fine particles of light-absorbing material, such as carbon or oxides of iron and titanium, darken the stone. The most valuable form of opal, the black opal, contains these particles in combination with spheres of uniform size packed in highly regular arrays. The overall result is that intense diffraction colors are produced with high contrast because the undiffracted light is absorbed within the black body of the stone.

The essential optical properties of precious opal, as distinct from common opal that shows no fire, therefore depend on the existence in the stone of orderly arrays of tiny optical discontinuities, spaced precisely at some particular repeat distance in the range of 150 to 300 nanometers. If these arrays are to diffract light, the spheres not only must be all the same size and shape but also must be packed in a regular system, like the atoms in a crystal.

To study such packing systems we set up an optical analogue of the apparatus used in X-ray diffraction [*see bottom illustration on preceding page*]. When opals are placed in this apparatus, they give rise to complicated patterns of colored spots and streaks, but we have been able to isolate a basic type of pattern that is characteristic of most single grains of opal when they are rotated in the incident beam. The pattern consists of two spots of color that represent simply diffracted beams obeying Bragg's law, plus six parallel colored streaks that, in the language of the crystallographer, represent reciprocal lattice rod effects. The streaks reveal that there are stacking faults in the otherwise orderly packing of the spheres. Electron micrographs show that opals consist of layers in which the spheres are arrayed in hexagons. When such layers are stacked one above the other, the two simplest arrangements are "hexagonally close-packed" and "face-centered cubic" [*see bottom illustration on page 91*]. We interpret the observed diffraction patterns to mean that the layers of spheres in opal are most commonly stacked in a random sequence, so that they are neither hexagonally close-packed nor face-centered cubic but a mixture of the two arrangements. When the stacking changes from one regular arrangement to the other, it is said that there is a fault in the stacking.

In a normal Australian opal each grain has one set of these faults; moreover, neighboring grains have their faults in planes of different orientations. This arrangement of a high concentration of faults enhances the patterns of color in opals and helps to account for their shifting colors. The streaks of color seen in the diffraction pattern indicate that the various colors of the spectrum are produced at different angles. Hence an observer looking at a single grain would see the various colors in order when the stone was tilted in an appropriate way. Opals with large, unfaulted grains are rare and show rather uninteresting color effects, with flashes of fire appearing only at a few specific viewing angles.

A type of opal new to us has recently been mined near Spencer, Idaho. It is a star opal that shows streaks of color in symmetrically arranged angular patterns similar to the rays of a star sapphire. In sapphire the effect is created by oriented groups of needle-shaped inclusions; in opal it is produced by diffracted streaks from three sets of faults occurring in the three planes symmetrically arranged around the vertical axis of a close-packed structure and inclined at angles of 20 degrees to that axis.

Let us now consider how these precise arrangements of microspheres of amorphous silica have been produced in nature. One would expect the conditions required for the formation of great numbers of perfectly sized and regularly packed spheres to be so stringent that they would rarely be fulfilled. That expectation is borne out by the rarity of precious opal.

The small amounts of precious opal are accompanied by an enormous quantity of valueless "potch" opal, which looks like opal but shows no color and is generally discarded by the miners. Although potch opal is chemically similar to precious opal, electron microscopy shows that it usually consists of rounded particles of silica that are not well enough shaped, sized or ordered to form light-diffracting arrays.

It seems clear that opal is made when amorphous silica slowly settles out of a dilute water solution. It is also evident that there must be a general tendency for the silica to appear as rounded particles of colloidal dimensions. From field and laboratory evidence we have deduced that the particles form in stages by colloidal aggregation as a silica solution is concentrated by the progressive evaporation of water. The first particles to appear are roughly spherical "primary" aggregates 30 to 40 nanometers in diameter. These units do not continue to grow, but if they are left in suspension, they slowly cluster into larger spherical assemblies whose internal structure consists of concentric shells formed by the accretion of several layers of the primary particles around a central nucleus. If the nucleus is a reasonably spherical clump of primary particles, the shells can retain a high degree of sphericity under favorable circumstances of aggregation. The final size of the sphere is determined by the number of completed shells. The smallest spheres found in precious opal consist of a nucleus and one shell, yielding a diameter of about 150 nanometers. The largest spheres consist of a nucleus surrounded by three shells.

One can identify a number of critical factors necessary for the development of perfectly shaped spheres all with the same number of shells. The first requirement is a geological formation in which steady-state conditions can persist over long periods. In addition the deposition site must be free of mineral colloids, such as clay, and of dissolved salts, which would flocculate the primary particles into gel-like masses or cause the growing spheres to stick together prematurely. It is essential that the spheres persist as free units, held in suspension by Brownian motion, until they finally arrange themselves into an ordered accumulation of similarly sized spheres at the bottom of the system. The rate at which such tiny spheres settle is very low, even in pure water. If the viscosity of the medium is increased by the formation of a gel, the time required for the spheres to collect on the floor of the cavity holding the suspension becomes enormously long.

In the natural environment the essential requirements are met when an undisturbed space in a rock holds a clean solution of silica from which water is slowly removed over a period of thousands of years. In the Andamooka field in South Australia a typical situation for the deposition of opal is found in the free space between the pebbles of a bed of conglomerate underlain by an impervious layer of bentonite clay at a depth of 10 to 40 meters below an arid land surface. The rock overlying the conglomerate is a soft, permeable mixture of quartz and fragments of feldspar, which gradually weathered to kaolin and supplied silica in solution to the perched water-table level at which the opal is found.

The entire formation has dried out during the period of increasing aridity that has affected the region in recent geologic times, allowing the colloidal silica spheres to settle slowly out of solution and jiggle themselves into an orderly crystal-like array. Channels through the formation allowed the continual percolation of silica solution, which became concentrated by evaporation and deposited more silica onto the spheres, cementing them together and partly filling the voids between them until the entire array became impermeable. Further slow evaporation of water hardened the silica, turning the originally porous conglomerate into a hard, concretelike layer.

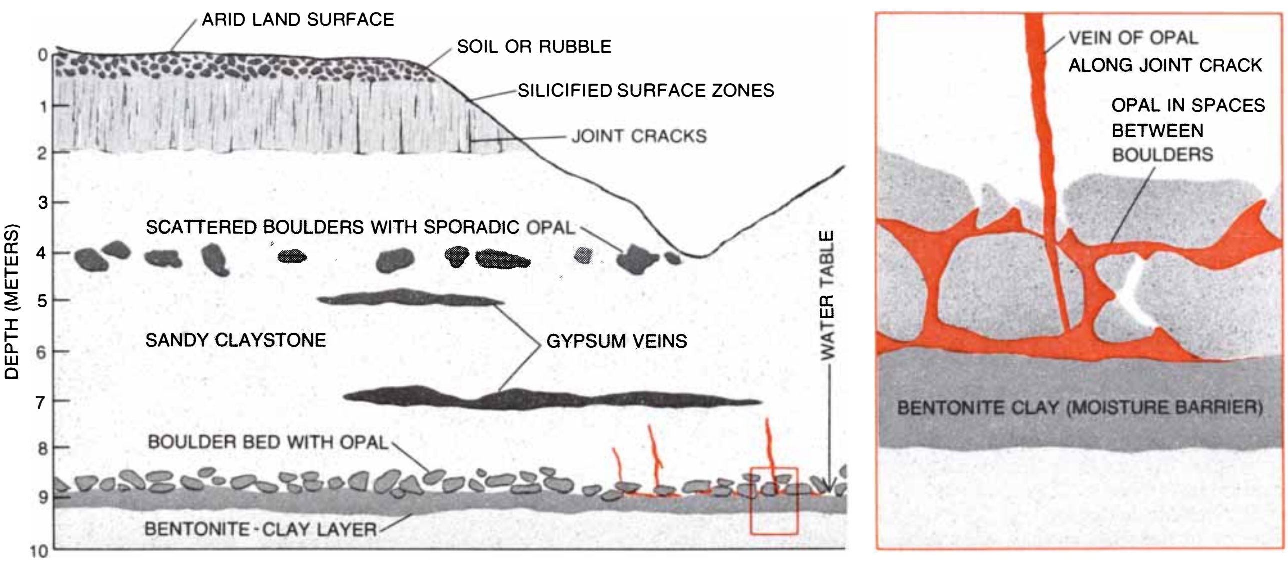
Other localities in Australia show similar patterns of occurrence, with the necessary free space being provided by cavities left when materials such as fossil shells, bones or crystals of soluble salts have been removed from the host rock by percolating groundwaters. Cavities of a different type are found in Queensland; they are created in the interior of gelled mixtures of iron oxide and silica that shrink on drying to leave hollow boulders, some of which are filled with "boulder opal."

In other parts of the world, where precious opal is typically found in volcanic rocks, gas bubbles were trapped and formed rounded holes when the lava solidified. Opal was deposited in some of the holes from solutions that extracted silica from the rock minerals while the lava was at temperatures higher than the normal ground temperatures. Much of this opal is different in appearance from varieties formed in sediments, probably because of its more rapid rate of deposition. The tiny silica spheres have no shell structure and are merged in close-packed arrays that show almost no trace of interstitial voids. As a result the material tends to be transparent and to lack a marked grain pattern. It exhibits only a diffuse band of color that moves through the stone as it is rotated. Most Mexican "fire opal" is of this type. (Here "fire" usually refers not to the play of colors but to the reddish body color of the stone.) On the other hand, opal from volcanic areas in Hungary, Idaho and Indonesia seems to have been formed under normal sedimentary conditions and shows well-defined grain patterns.

As far as we know, all precious opal is geologically young. It is not a substance that can withstand the effects of heat and pressure that almost always accompany long-term geological processes, particularly when the material is deep below the surface. Under such conditions opal progressively crystallizes into fine-grained masses of cristobalite or tridymite. Moreover, opal cannot survive exposure to the effects of surface weathering for any appreciable period. Under those conditions hydrated amorphous opaline silica tends to lose water and to become cracked and opaque. The results of the process can be seen in the dumps of the opal fields, where discarded potch, once hard and almost transparent, has become chalk white and crazed on exposure to the sun. Some kinds of precious opal can sustain extreme conditions of heat and dryness, but in general opal needs to be treated with care in order to reduce the risk of damaging it.

Once we knew the basic structure of precious opal we investigated the possibility that synthetic opal could be made. If the process of synthesis were a reasonably simple one, it obviously could have a major effect on the future value of the gem. Laboratory experiments showed that pure silica sols, made from aqueous solutions of sodium silicate by treatment with ion-exchange resins, slowly generated suitable microspheres if the sol was kept heated and stirred for weeks. The size of the spheres, however, varied greatly. Fractions of spheres of roughly similar size were separated by sedimentation in a centrifuge and allowed to settle quietly over a period of months. Eventually the particles accumulated in ordered layers that exhibited the typical play of colors of opal.

A later synthesis technique was based on the controlled hydrolysis of an emulsion of tetraethyl orthosilicate, a much quicker process and one allowing better control of sphere size. When the sedimented layers were dried and hardened by heating or other cementing procedures, they duplicated many of the characteristics of precious opal. We had difficulty, however, obtaining a silicified product as hard, as dense and as clear as natural opal. More recently Pierre Gilson, a Swiss expert in synthetic-gem production, has much improved the techniques of synthesis and has made some spectacular examples of both the milky and the black varieties of opal. The material has not been released commercially because it lacks stability and tends to crack on prolonged exposure to air. That is also a problem with natural opal from certain localities, and no solution has been found for it.



OPAL DEPOSIT at Andamooka in South Australia is depicted schematically in this cross-sectional drawing. A water table containing dissolved silica is supported by a highly impermeable layer of bentonite clay. The ground above is porous and the climate arid, which allows the water to evaporate at a steady rate. As the silica solution increases in concentration, colloidal spheres of silica slowly settle into cavities in the boulder bed above the clay base (*detail at right*). Eventually the spheres are jiggled into orderly arrays of opal.

The Analysis of Materials by X-Ray Absorption

Tiny wiggles in the characteristic X-ray-absorption "signature" of an atom embedded in a solid can now be interpreted to provide clues to the exact spatial arrangement of the neighboring atoms

by Edward A. Stern

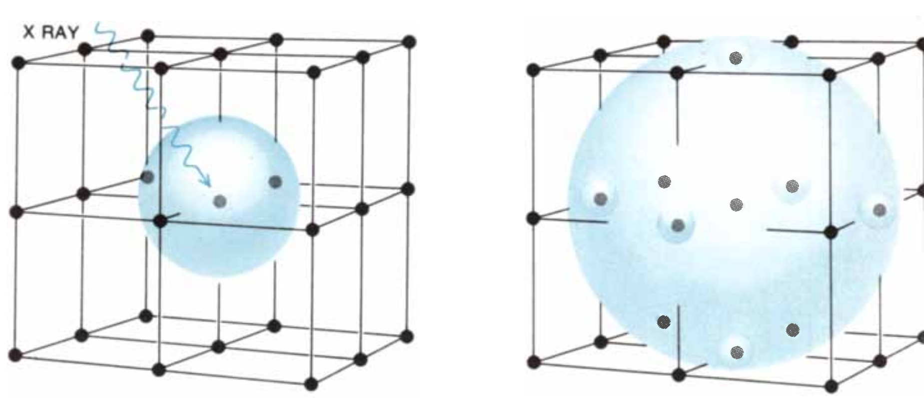
Our knowledge of the atomic structure of matter—the foundation on which much of modern physical and biological science is built—is derived primarily from X-ray crystallography: the versatile investigative method developed by W. H. Bragg and his son W. L. Bragg after the discovery by Max von Laue in 1912 that the very short waves of X rays are diffracted, or scattered, in a decipherable way by the repeating pattern of atoms in a crystal. Later it was found that electrons and neutrons also have wave properties that can be exploited to probe the geometric arrangement of atoms in various materials. The application of electron-beam and neutron-beam techniques has led to the determination of atomic structures inaccessible to the original X-ray method; nevertheless, these techniques are essentially adaptations of the same wave-diffraction principle.

This article concerns a completely different approach to the analysis of atomic structure, based not on the diffraction of X rays by an array of atoms but on the absorption of X rays by individual atoms in such an array. Specifically the new technique (which was first proposed by Farrel W. Lytle, Dale E. Sayers and me in 1971) measures how that absorption is affected by the atoms in the immediate neighborhood of the X-ray-absorbing atom. Although the existence of such an "extended fine structure" effect in X-ray absorption has been known for almost half a century, it is only in the past few years that the analytical potential of the phenomenon has come to be appreciated. Besides complementing the other crystallographic techniques, the analysis of materials by their characteristic X-ray-absorption "signature" has several decided advantages. For one thing, unlike the earlier methods, the new method can be applied equally well to materials with long-range order or to materials without such order. Furthermore, the local atomic environment can now be determined separately for each type of atom in a chemical compound.

In a sense the new technique is an extension of X-ray spectroscopy, the branch of physical analysis that relies on the absorption of X rays to disclose the allowed energy states of an isolated atom. Traditionally this line of research is quite distinct from X-ray crystallography; in fact, it was the gap between the two disciplines that was responsible for the long delay in the recognition of the value of the structural information contained in the X-ray-absorption signature of materials. This is not to say that no important experimental and theoretical work on the phenomenon of the extended X-ray-absorption fine structure was done before 1971. The basic mechanism underlying the technique of X-ray-absorption analysis was first studied in the 1930's by R. de L. Kronig and H. Petersen in Germany. Valuable contributions were also made by A. I. Kostarev and A. I. Kozlenkov in the U.S.S.R. and by T. Shiraiwa and his colleagues in Japan.

In a solid with long-range order, such as a crystal of a pure element, the atoms are all arrayed at fixed sites in a geometrically perfect three-dimensional lattice; no matter how far one moves from a given atom in any direction, the atoms are still likely to be found at fixed periodic intervals. In a solid without long-range order, on the other hand, there is no correlation between a given atomic site and another site some distance away. To put the matter differently, in an amorphous, or glassy, material an atom far from a given atomic site has an equal probability of being any distance from the original site.

Solids with and without long-range order yield inherently different X-ray-diffraction patterns [*see illustration on page 99*]. For materials with long-range order the standard X-ray-diffraction technique can determine to a high degree of accuracy the exact arrangement of the atoms in any local region of the material. For materials without long-range order the standard technique is



X-RAY-ABSORBING ATOM located in the midst of an array of atoms in any solid material can acquire enough energy from the impact of the X-ray photon to eject an electron from one of its inner shells of tightly bound electrons, the process known as the photoelectric effect. According to quantum theory, the energy state of the escaping photoelectron can be visualized as an outward-moving spherical wave (*color*) centered around the excited atom. This electron

not as satisfactory. To determine the local configuration of the atoms in the latter case it is first necessary to assume a model of the types of atoms present, their locations and their X-ray-scattering properties. One must then calculate the expected X-ray-diffraction pattern and compare it with the measured pattern. For simple materials composed of only one type of atom calculations of this kind yield a fairly accurate picture of the short-range order. For more complicated structures composed of two or more types of atom it is usually not possible to determine the short-range order unambiguously. The problem arises because the measured intensity of the scattered X rays is the sum of all the X rays scattered from each type of atom, and neither the measurements nor the calculations are accurate enough to distinguish between the different components of the diffraction pattern traceable to each type of atom. It is in precisely such a situation that our new X-ray technique can be expected to be most useful.

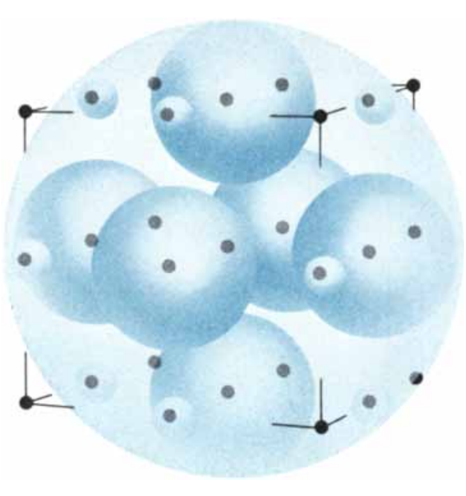
When X rays pass through matter, they are attenuated, owing to their interactions with the electrons in the matter. One of the mechanisms that absorb X rays is the photoelectric effect: the annihilation of an X-ray photon in a collision with an electron, accompanied by the transfer of all the photon's energy to the electron, which as a result is knocked loose from its atom. (According to quantum theory, all electromagnetic radiation, from low-frequency radio waves through the spectrum of visible light to high-frequency X rays, is composed of photons, the smallest possible units of the radiation. The energy of a photon is directly proportional to the frequency, and hence inversely proportional to the wavelength, of the radiation, the constant of proportionality being the fundamental physical quantity known as Planck's constant.)

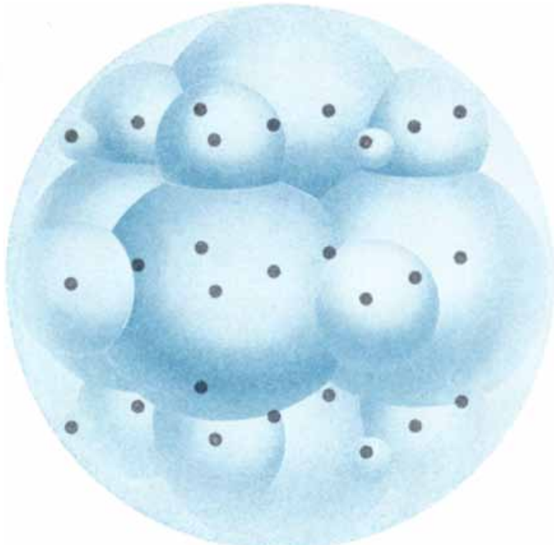
An excited electron, in order to be capable of absorbing the energy of an X-ray photon, must have an empty energy state where it can end up after gaining the photon's energy. The possible energy values that an electron can have in an atom are discrete for energies low enough for the electron to remain bound to the atom. For electrons that escape, however, all energy values are allowed. An electron that is excited enough by a photon to escape from an atom is called a photoelectron.

Consider an X-ray photon with an energy just high enough to knock one of the eight electrons out of the atom's *L* shell, the second most tightly held cloud of electrons in any fair-sized atom [*see bottom illustration on page 100*]. Such a photon does not have enough energy to liberate one of the two electrons present in the atom's *K* shell, the most tightly bound cloud of electrons. If the energy of the photon (that is, its frequency) were to be increased, however, a certain energy value would be reached where a *K*-shell electron could be pried loose from the atom. At that energy the measured attenuation of the X-ray beam would exhibit a sudden increase as the new mode of absorption is initiated. That abrupt increase is called the *K* edge of the absorption spectrum; similarly, the energy where the *L*-shell electrons can just be knocked free is called the *L* edge [*see top illustration on page 101*]. (In actuality three different *L* edges can usually be distinguished, but that detail need not be considered here, since in the following discussion the application of our technique will be described only with reference to absorption in the vicinity of the *K* edge of the X-ray-absorption spectrum. In general, however, the same considerations apply also to absorption near the *L* edges.)

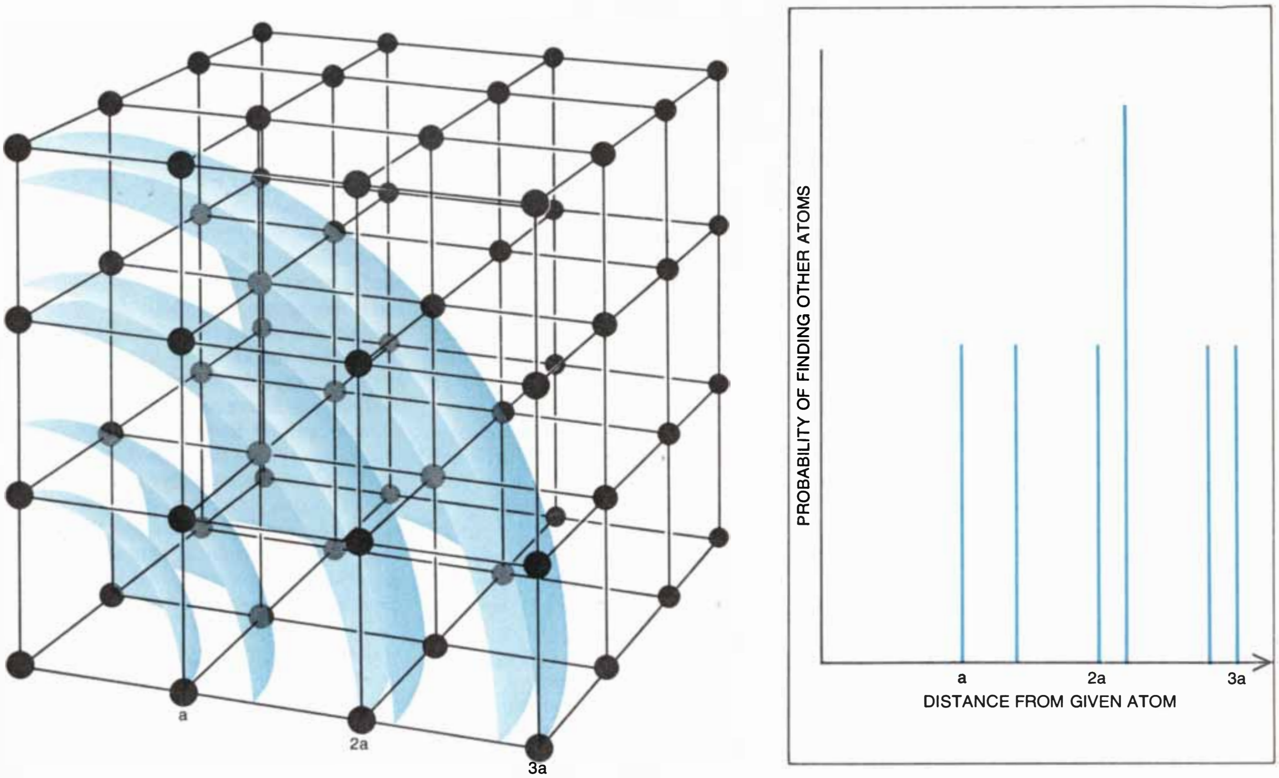
The attenuation of X rays as they pass through a material is expressed by a quantity called the absorption coefficient. For isolated atoms the variation of the absorption coefficient with the energy of the X-ray photon is monotonic, or smooth, except for the conspicuous absorption edges I have mentioned. For the atoms in a solid, however, the variation of the absorption coefficient at energies just above an absorption edge displays a complex fine structure [*see bottom illustration on page 101*]. The observed nonmonotonic variation above the *K* edge for copper in this example is clearly attributable to the atoms surrounding the X-ray-absorbing atom, since it appears in the X-ray-absorption spectrum only when other copper atoms are present.

To explain fully how this nonmonotonic variation is produced would require a highly mathematical excursion into quantum mechanics. The basic idea can nonetheless be illustrated quite simply. The probability that an X-ray photon will be absorbed by an electron depends on both the initial energy state of the electron (that is, its energy in the *K* shell of the atom) and the final energy state of the electron (its energy as a free photoelectron). Except for the very lightest atoms, such as those of hydrogen and helium, the initial energy state of the electron in the *K* shell is so embedded in the atom that it is not appreciably influenced by whether

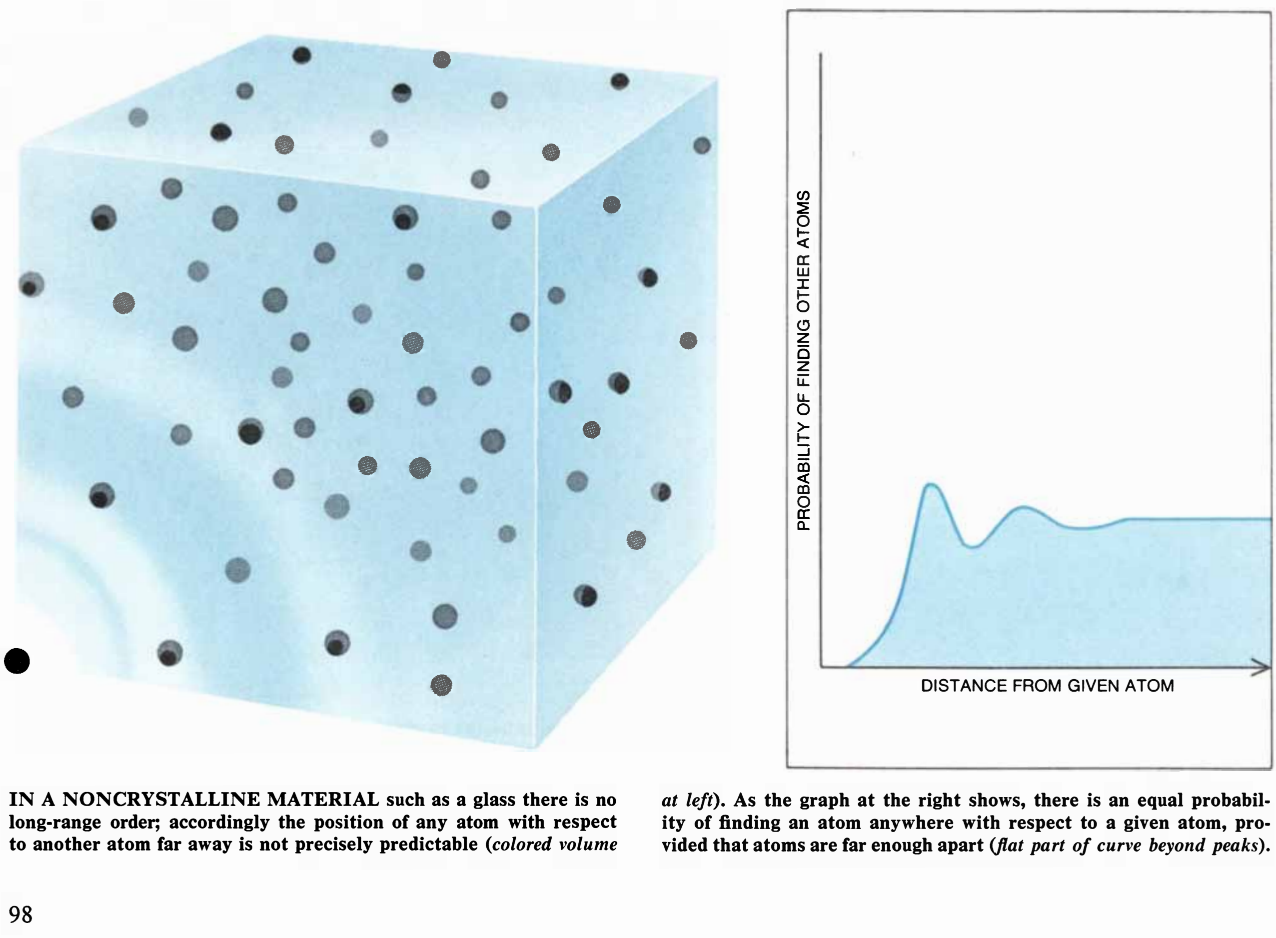




wave is scattered by successive ranks of neighboring atoms, and the new waves emanating from each scattering site are added to (or subtracted from) the initial outgoing wave. The interference of the outgoing and incoming waves near the center of the X-ray-absorbing atom affects the absorption of X rays by that atom in a complex way that is reflected in the fine structure of the X-ray-absorption spectrum. The structural information encoded there can in turn be deciphered to determine the locations, numbers and types of all the atoms that scattered the initial outgoing electron wave. The scattering from the first three "nearest neighbor" spheres of atoms is illustrated here.



IN A CRYSTALLINE MATERIAL that is distinguished by the presence of long-range order the position of any atom with respect to any other given atom is precise and predictable, even when the atoms are far apart (*colored surfaces at left*). As the graph at right shows, the probability of finding another atom at a fixed distance from any origin point consists of a series of discrete values (*vertical lines*).



IN A NONCRYSTALLINE MATERIAL such as a glass there is no long-range order; accordingly the position of any atom with respect to another atom far away is not precisely predictable (*colored volume at left*). As the graph at the right shows, there is an equal probability of finding an atom anywhere with respect to a given atom, provided that atoms are far enough apart (*flat part of curve beyond peaks*).

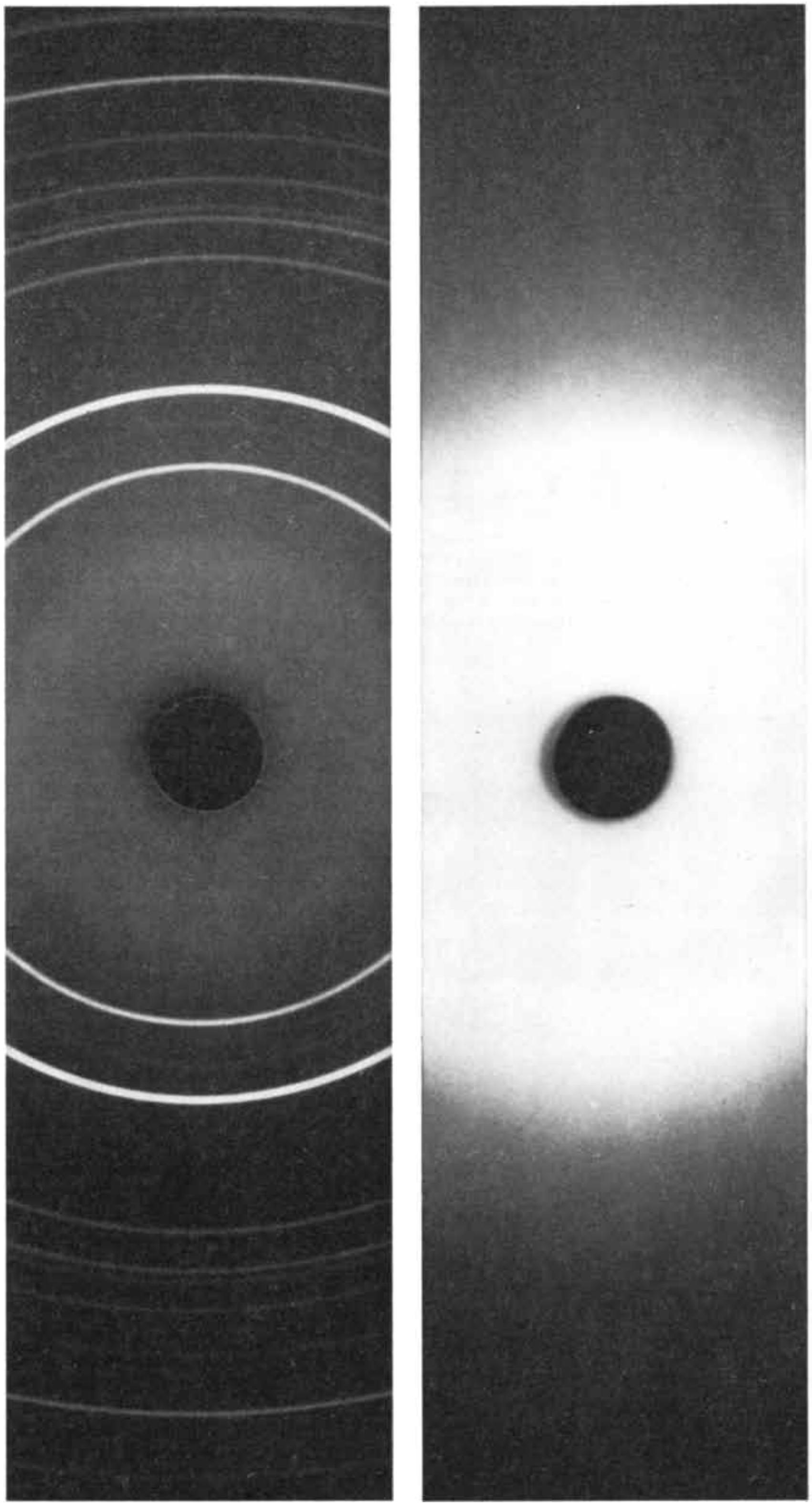
the atom is isolated or is located in a solid. Thus changes in the X-ray-absorption spectrum are caused only by changes in the final energy state of the electron.

For our purpose the pertinent final energy states are those of the photoelectron that escapes from the excited atom. In quantum theory the energy state of the escaping electron is represented as an outward-moving spherical wave centered around the X-ray-absorbing atom [*see illustration on pages 96 and 97*]. If the atom is isolated, that is all there is to the final state of the electron. If the atom is surrounded by other atoms, the electron wave will be scattered by those atoms, just as the outgoing circular ripple from a pebble thrown in a still pond will be scattered by reeds protruding from the water. In both cases the initial outgoing wave will have added to it the new waves radiating from each scattering site. It is the addition of such waves that gives rise to the fine structure in the X-ray-absorption spectrum.

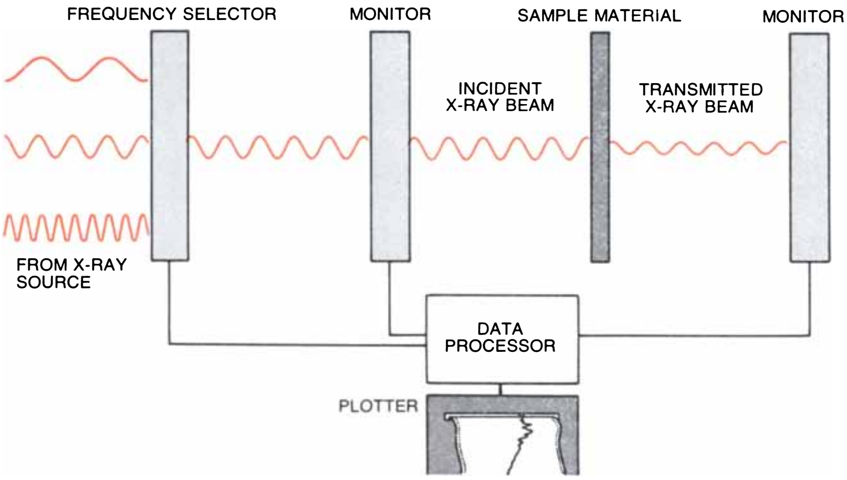
Whenever two or more waves interfere, the amplitude of the resultant wave is the sum of the amplitudes of the individual waves at any given point [*see top illustration on page 102*]. If the relative phases of the waves are opposite, one amplitude subtracts from the other. In extended X-ray absorption the incoming wave has the same wavelength as the outgoing wave but a smaller amplitude. Hence where the waves are exactly in phase the resultant amplitude is increased, and where they are exactly out of phase the resultant amplitude is decreased.

For *K*-edge X-ray absorption the important amplitude of the final state of the electron wave is the amplitude in the same region of space where the electron exists in its initial energy state, since the measured absorption originates only in this region. Typically the region consists of a very small spherical space around the center of the X-ray-absorbing atom. For example, the initial energy state of an electron in the *K* shell of an iron atom is confined to a spherical space whose radius is about an eightieth the size of the atom (that is, the space occupied by its outermost electrons). Roughly speaking, then, what we are interested in is the amplitude of the final state of the electron wave near the center of the X-ray-absorbing atom. If that amplitude is zero, there is no coupling between the final-state electron wave and the initial-state electron wave, and hence there is no absorption of X rays by the photoelectric effect. On the other hand, if the amplitude of the final-state electron wave is large near the atomic center, the probability of an X-ray photon's being absorbed by the excitation of a photoelectron is also large.

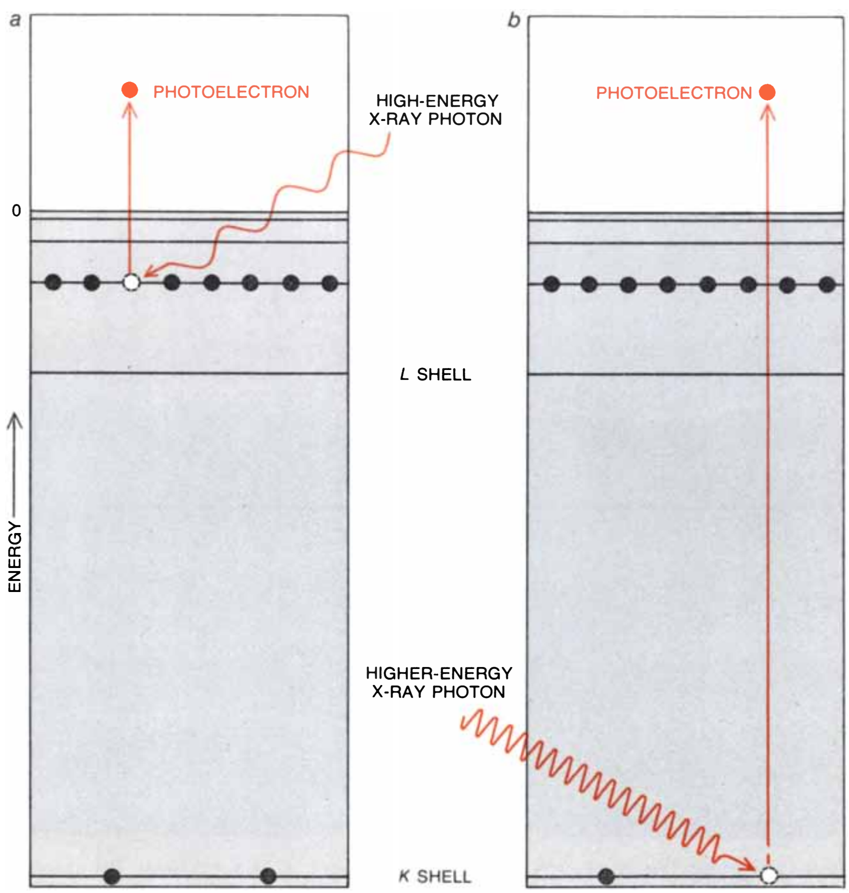
The amplitude of the final-state electron wave near the center of the atom is modified by the waves that are scattered by the surrounding atoms, since their amplitudes are added to or subtracted from the amplitude of the original outgoing electron wave. Depending on the relative phases of the outgo-



X-RAY-DIFFRACTION TECHNIQUE is used routinely to help determine the exact arrangement of atoms in materials that possess long-range order, but it is not as successful with materials that lack such order. In the X-ray-diffraction photograph at left, for example, the precise, orderly pattern of bright rings is a manifestation of the presence of long-range order in a sample consisting of powdered crystalline quartz (SiO_2). In the X-ray-diffraction photograph at right, in contrast, the fuzzy halo indicates the absence of long-range order in a sample consisting of powdered silica glass (also SiO_2). The structural significance of such diffuse patterns is much harder to interpret, particularly for materials composed of two or more types of atoms. Both photographs were made by Daniel E. Appleman of the Smithsonian Institution.



EXPERIMENTAL SETUP for measuring the X-ray-absorption characteristics of materials is depicted in this highly schematic diagram. A beam of X rays composed of many different wavelengths is first passed through a device that can be adjusted to select only X rays with a particular frequency, or energy. The attenuation of this X-ray beam in passing through the sample material is found by measuring the intensity, or amplitude, of both the incident beam and the transmitted beam. Recent experiments have been facilitated by harnessing the synchrotron X-radiation from the electron-positron storage ring at the Stanford Linear Accelerator Center.



DISCRETE ENERGY STATES at which electrons can remain bound to an atom are represented here by the horizontal black lines beginning at energies below zero. At energies above zero the electrons are free to escape from the atom. The particular atom shown has two electrons in its *K* shell, the most tightly bound cloud of electrons in any atom, and eight electrons in its *L* shell, the second most tightly bound electron cloud. In *a* an X-ray photon colliding with the atom has just enough energy to excite an electron from the *L* shell to a state where it can escape from the atom and become a photoelectron. In *b* a higher-frequency, and hence higher-energy, X-ray photon is able to knock an electron loose from the atom's innermost *K* shell.

ing and the incoming waves, the resultant amplitude could be greater than or less than that of the outgoing wave [*see bottom illustration on page 102*]. An increased amplitude would result in a larger X-ray-absorption coefficient; a decreased amplitude would yield a smaller X-ray-absorption coefficient. If the relative phases of the outgoing and the incoming waves of a photoelectron are changed by varying the energy of the X-ray photon, then a fluctuating absorption coefficient will result (which is of course what is observed).

Changing the frequency of the X-ray photon produces the required variation in phase between the outgoing and the incoming photoelectron waves, since (by the de Broglie relation) the wavelength of the photoelectron is inversely proportional to its momentum. Increasing the energy of the X-ray photon beyond its value at the absorption edge will increase both the energy and the momentum of the photoelectron and thus decrease its wavelength.

The oscillatory behavior observed in the fine structure of the extended X-ray-absorption spectrum provides a good demonstration of the wave nature of electrons, since only waves could account for such behavior. What is more important is that those tiny oscillations contain the key to the location of all the atoms that scattered the outgoing wave! It is simply a matter of deciphering the encoded information.

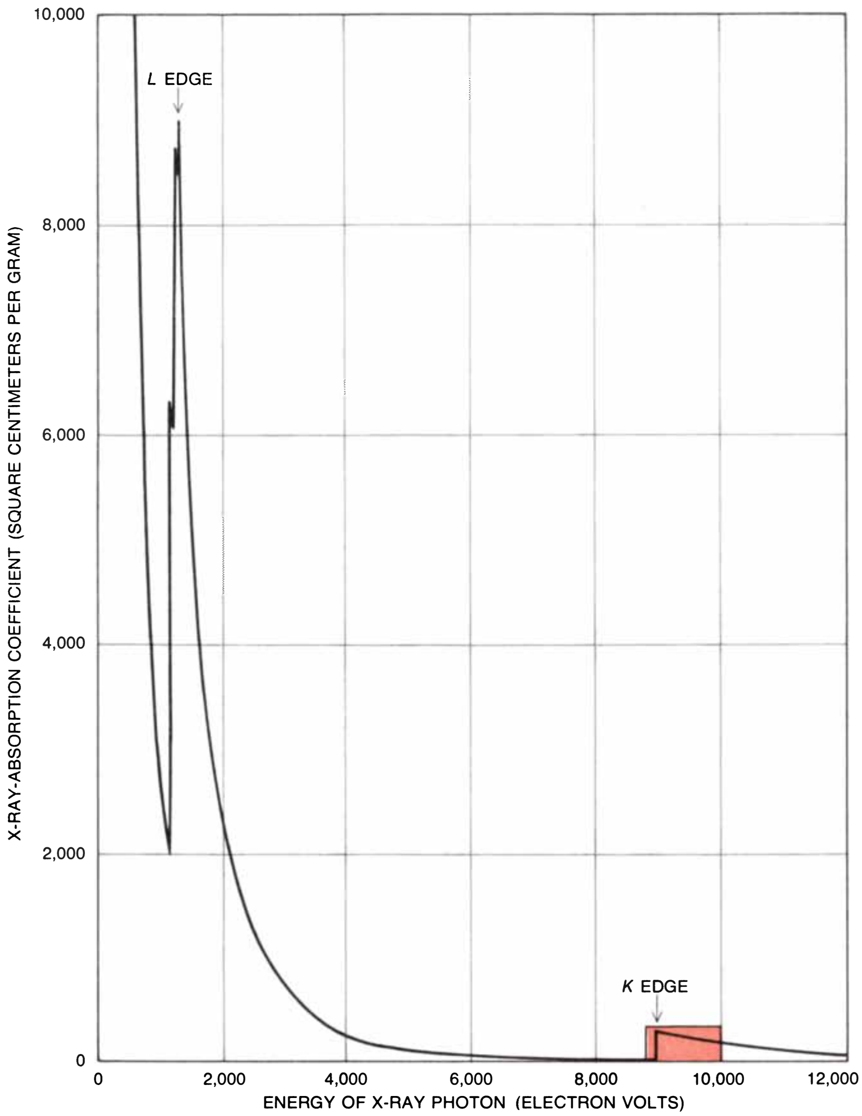
In certain respects analysis of the extended X-ray-absorption fine structure resembles the process of holography, since in both cases phase information is retained by recording the interference between an outgoing wave and a scattered wave. In holography the interference between the outgoing light wave from the monochromatic source and the scattered waves is recorded on a photographic film. In X-ray-absorption analysis the interference is recorded by measuring the variation of the X-ray absorption at a given point in space. In holography only one recording wavelength need be used because the interference pattern is recorded over a plane instead of at one point. To compensate for this limitation the wavelength of the electron in the X-ray-absorption technique must be varied by changing the energy of the X rays.

In both holography and the new X-ray technique the location of the scattering sites can be ascertained by processing the inference information. In holography the original object that caused the light scattering can be reconstructed by passing a laser beam through the developed film on which the interference pattern is recorded. In X-ray-absorption analysis the processing step involves the mathematical procedure known as Fourier transformation to relate the observed fine structure in the absorption spectrum to the momentum of the photoelectron. What one ends up with is a photoelectron-scattering profile plotted as a function of the radial distance from the excited atom [*see illustration on page 103*]. The peaks in such a curve correspond to the scattering of the electron wave from the several "nearest-neighbor spheres" of atoms surrounding the excited central atom. The size of the peaks is a function of the number of atoms in a given sphere, their distance from the central atom and the uniformity of their distances in a given sphere. The positions of the peaks indicate the positions of the surrounding atoms.

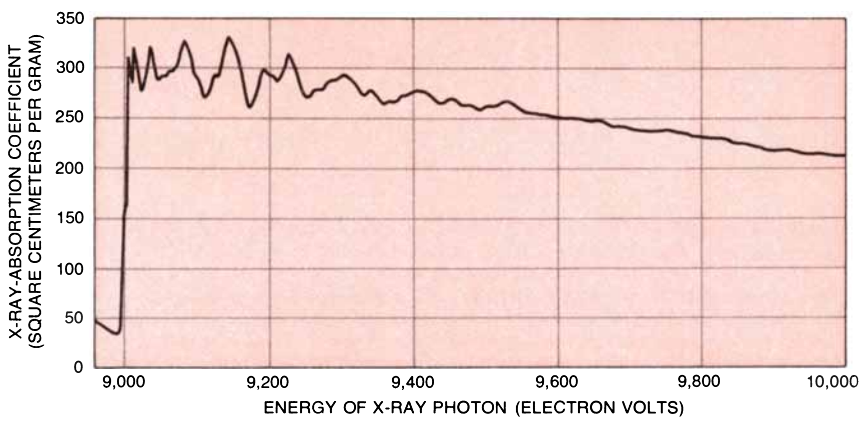
By appropriate analysis of this scattering profile one can in principle determine the locations, numbers and types of atoms in the first few neighboring spheres around a given atom. The information can be obtained for each type of atom separately by tuning the X-ray energy to coincide with the absorption edge of just one type of atom at a time. The resulting data and the required analysis are independent of whether or not the sample has long-range order. This factor is of great practical importance because it means that the X-ray-absorption technique can be calibrated on the basis of known structures, which are invariably those with long-range order, and the calibration can then be carried over to the study of unknown structures. Because of this calibration capability analysis of the extended X-ray-absorption fine structure can determine the short-range order of materials that lack long-range order more accurately than the standard diffraction techniques can, even when the material consists of only one type of atom.

The special characteristics of this new tool are useful not only because they complement the standard diffraction techniques but also because they open up for structural analysis entire classes of systems that could not be analyzed otherwise. Analysis of the extended X-ray-absorption fine structure holds great promise for the study of noncrystalline condensed matter composed of many different types of atoms, particularly when the environment around one of the types of atoms is of special interest. Some important examples of such systems are biological molecules, catalysts and glass compounds.

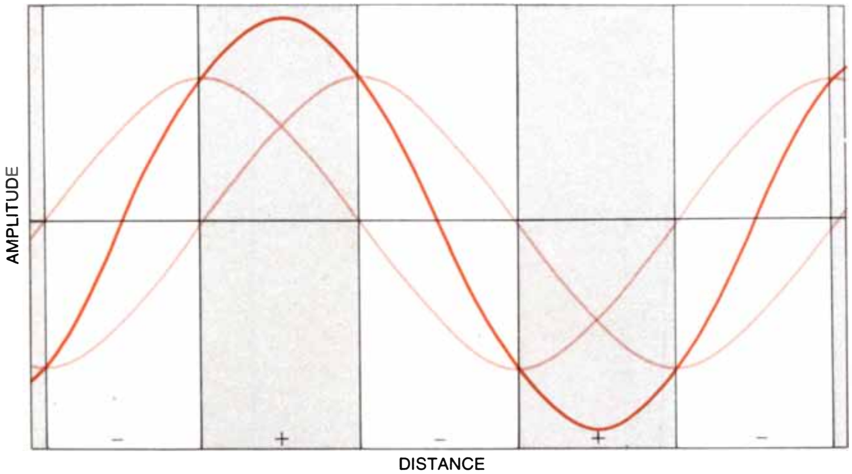
The glass most carefully studied by X-ray-absorption analysis is germania glass (GeO_2), a compound closely related to silica (SiO_2), the major constituent of ordinary glass. By comparing germania glass with the crystalline form of germanium dioxide Sayers, Lytle and I were able to determine the disorder in the first two atomic spheres surrounding the germanium atoms with greater accuracy than was attainable with standard diffraction techniques. Our results made it possible for the first time to determine which of two competing models for glasses is correct. It was found that the random-bond model is valid and not the microcrystalline model. The incorrect microcrystalline model assumed that the glass is made of very small crystallites some 30 angstroms in diameter. The random-bond model assumes that germania glass is composed of many small tetrahedrons, each with a germanium atom at the center surrounded by four oxygen atoms. These tetrahedrons are not appreciably distorted in the change



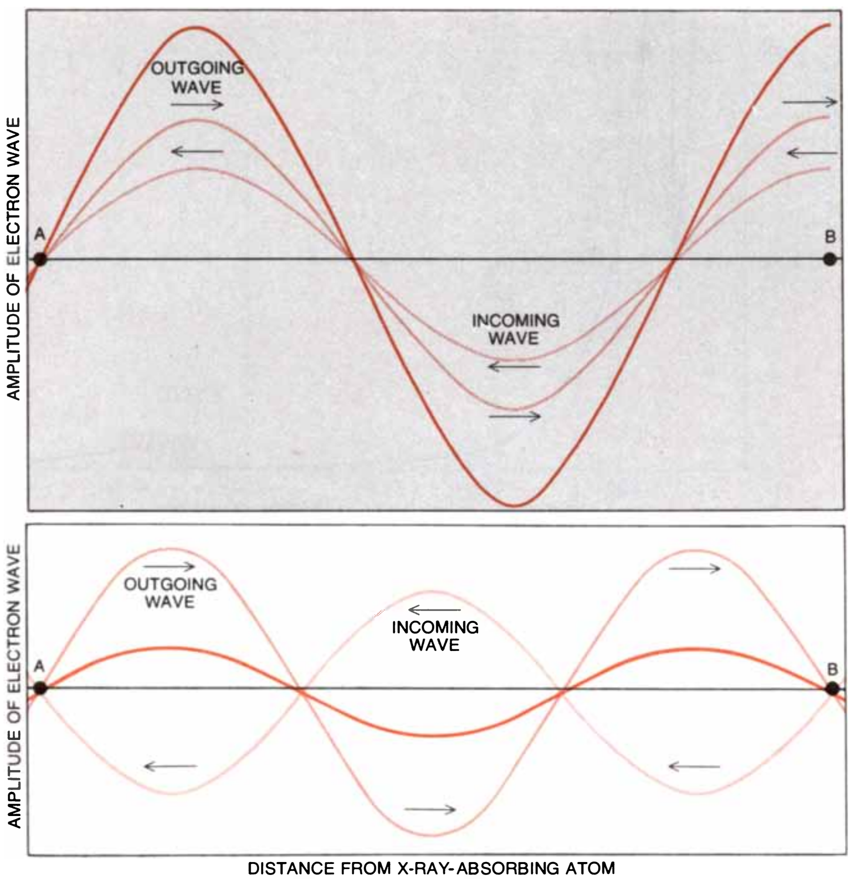
ABRUPT INCREASES, called absorption edges, appear in the measured attenuation of an X-ray beam as each new mode of absorption is initiated. For example, in this graph of the X-ray-absorption coefficient of a sample of copper as a function of the energy of the incident beam of X rays the conspicuous *L* and *K* edges (*arrows*) correspond to the energies at which an X-ray photon can just begin to free an electron from copper atom's *L* shell and *K* shell respectively.



COMPLEX FINE STRUCTURE is revealed at energies just beyond the *K* edge for copper in this enlarged graph of the variation of absorption coefficient with X-ray-photon energy. The wiggles in the curve are attributable to the electron-scattering effect of the atoms surrounding the X-ray-absorbing atom. The analytical technique devised by the author and his colleagues is concerned with such wiggles in the vicinity of the absorption edges of various materials.



INTERFERENCE of any two waves yields a resultant wave (*heavy colored line*) whose amplitude is the sum of the amplitudes of the individual waves at any point. Where waves are in phase (*gray areas*) amplitudes add; where waves are out of phase (*white areas*) amplitudes subtract.



AMPLITUDE OF RESULTANT ELECTRON WAVE in the immediate vicinity of the X-ray-absorbing atom depends on the relative phases of the initial outgoing wave and the incoming scattered wave. The relative phases of the two waves depend only on their wavelength, which in turn can be varied by varying the frequency of the absorbed X-ray photon. This wavelength dependence is illustrated for two different electron wavelengths here. Point *A* and point *B* are assumed to be the positions of the X-ray-absorbing atom and the nearest electron-scattering atom respectively. In the upper diagram the wavelength of the initial outgoing wave from point *A* is such that at point *B* the scattered wave is reflected back exactly in phase with the outgoing wave. (In the scattering process the amplitudes of the reflected waves are somewhat reduced as shown.) The outgoing wave and the incoming wave are in phase at point *A*, and hence they add to produce an increased amplitude in the vicinity of the X-ray-absorbing atom. In the lower diagram the wavelength of the outgoing wave from point *A* is such that the phase of the incoming wave is exactly opposite to that of the outgoing wave. Hence the two electron waves subtract to produce a resultant wave with a reduced amplitude. The X-ray-absorption coefficient of the atom at *A* is dependent on the amplitude of the resultant electron wave in its vicinity.

from the crystalline form to the glass form, and hence they can be considered the basic building blocks of the compound. The orientation of the tetrahedrons with respect to one another is distorted in the glassy form, leading to the destruction of long-range crystalline order.

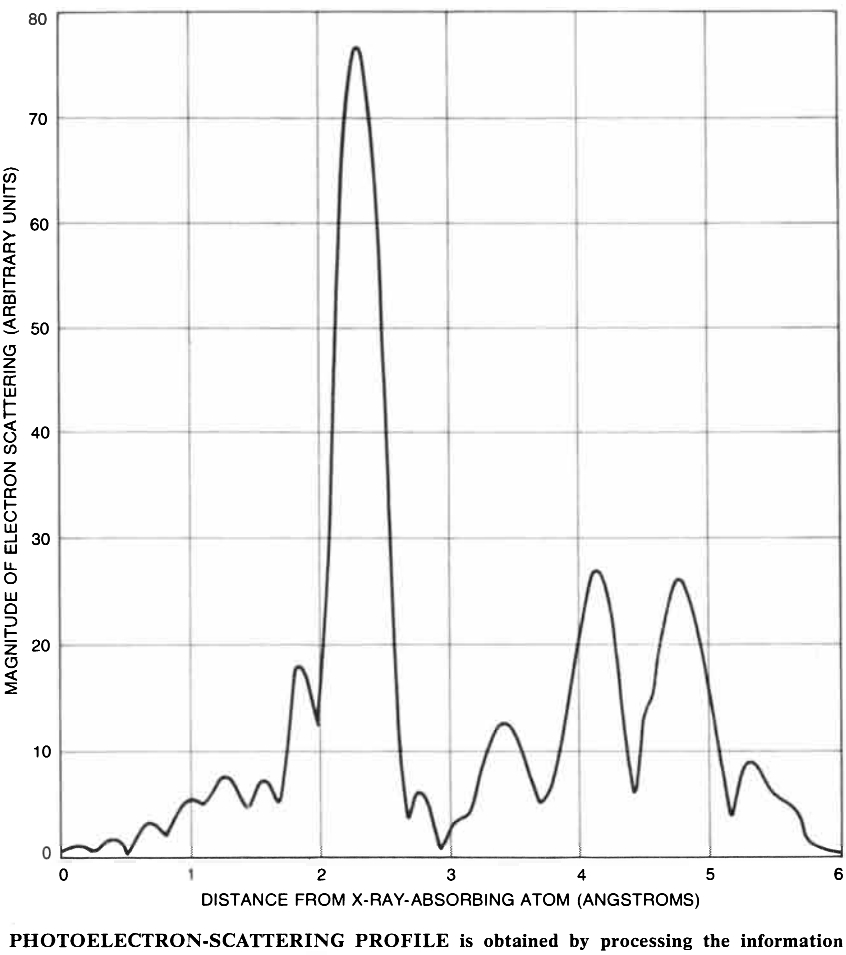
Lytle has been applying the X-ray-absorption technique to the study of catalysts, substances that accelerate chemical reactions without being used up themselves. Catalysts are of great commercial importance, playing a major role in producing goods worth more than $100 billion a year. The chemical industry, for example, is dependent on catalysts for converting crude oil into commercially useful products such as gasoline and plastics. Nature also uses catalysts to control the chemical reactions of life.

In spite of the importance of catalysts our knowledge of their basic mode of operation is so limited as to have almost no predictive value. The tailoring of a catalyst to perform a desired function is currently a matter of trial and error. If catalysis is to be more a science than an art, it will be necessary to determine both the atomic structure of catalysts and how that structure changes during catalysis. These phenomena are almost all unknown, since commercial catalysts are usually made in a form where standard diffraction techniques yield only limited information. Some progress has been made with idealized systems, but they have the limitation that they can be studied only under ultrahigh-vacuum conditions, instead of in the normal-pressure atmospheres that surround real catalysts. Our technique has the advantage of enabling one to study catalysts during actual reactions in the presence of actual contaminants. Analysis of the extended X-ray-absorption fine structure of certain catalysts has already produced enough results to prove its ability to determine both the active site of catalysis and the local atomic environment. The application of this technique to study catalysis has hardly begun.

Perhaps the most important application of the new X-ray method will turn out to be in the study of biological molecules. Much of the richness and intricacy of living processes is at present beyond our grasp. Although a great deal of progress has been made in our understanding of some biological processes, only a little has been understood on a fundamental atomic level. It is therefore not surprising that much effort is being exerted to apply the new mode of analysis to various simple but interesting biological systems. In addition to our group at the University of Washington, workers at Bell Laboratories, Stanford University and the University of California's Lawrence Berkeley Laboratory are all busy studying biological systems. The new technique appears to be best suited for analyzing problems such as the oxygenation of hemoglobin and the active site in photosynthesis, where the reactions of greatest interest take place in the vicinity of a heavy atom. Analysis of the extended X-ray-absorption fine structure is capable of determining the environment of the heavy atom and the changes that take place in that environment as the reactions proceed.

The recognition of the usefulness of this new tool as a way of measuring atomic structure was fortunately soon followed by the availability of a new high-intensity X-ray source of continuously variable wavelength. In the late spring of 1974 the synchrotron radiation from the electron-positron storage ring at the Stanford Linear Accelerator Center was harnessed to a suitable test facility, and the first high-intensity measurements of the extended X-ray-absorption fine structure of materials were made. The intensity of this new source at the Stanford Synchrotron Radiation Project is 10,000 times higher than standard X-ray sources, making measurements feasible that could not be considered before. Measurements made with standard X-ray tubes, with which it takes weeks to accumulate the necessary data, can now be made in half an hour at the Stanford facility; moreover, the intensity of this X-ray source makes it feasible for the first time to measure extremely dilute trace elements.

The heavy-atom studies of biological systems would be impractical without the Stanford facility. Trace atoms on the order of a few parts per million have had their extended X-ray-absorption spectra measured at Stanford by means of fluorescent techniques developed by workers at the Lawrence Berkeley Laboratory and Bell Laboratories. The chemical form in which the trace atom is present can be readily determined by its extended X-ray-absorption fine structure, a feature that makes this approach particularly suited for the study of pollutants. The fluorescent techniques provide an indirect way of applying the new X-ray tool to measure much smaller amounts of a given atom than would be possible by direct X-ray absorption. There is no doubt that the addition of the fluorescent techniques will make possible the application of X-ray-absorption analysis to a host of scientifically and technologically interesting systems.



PHOTOELECTRON-SCATTERING PROFILE is obtained by processing the information contained in the fine structure of the measured X-ray-absorption spectrum with the aid of the mathematical technique known as Fourier analysis. The electron scattering is plotted here as a function of the radial distance from the central excited atom. The positions and sizes of the peaks in the curve can be interpreted to determine the locations, numbers and types of the atoms in the region immediately surrounding X-ray-absorbing atom. In the case of a compound material comparable structural information can be obtained for each type of atom separately by tuning X-ray energy to coincide with absorption edge of just one type of atom at a time.

Galileo and the First Mechanical Computing Device

It was the sector, which could be employed to solve mechanically a number of mathematical problems. Galileo devised it to attack an insoluble problem and then perceived its value for simpler ones

by Stillman Drake

People have resorted to many kinds of mechanical devices to avoid the trouble of doing arithmetical calculations. The oldest of them is the abacus, with its sets of beads strung on parallel wires; by pushing beads back and forth according to set rules one can get the same result as one does when working with written numbers. One now pushes buttons on a small electronic calculator according to set rules. Since mathematicians long ago found ways to reduce geometry and physics to numbers, every practical problem involving mathematics can now be solved without any tedious computations on paper.

Up to a few years ago people used the slower mechanical calculating machine, pushing buttons that set in motion numbered wheels. Half a century earlier, before the electric motor was applied, similar machines were operated by turning a crank and shifting a carriage. The basic mechanism goes back to the 17th century, when Blaise Pascal put digits around successive wheels that he connected to provide for carry-over. In the same century Gottfried Wilhelm von Leibniz introduced the movable carriage to speed up multiplication. Credit for the basic idea of a calculating machine goes back before Pascal to John Napier, who carved numbers on movable rods of bone rather than on wheels.

Napier is deservedly more famous for his invention of logarithms in 1611. That invention makes him ultimately responsible for the basic idea behind the slide rule, the chief alternative device for avoiding arithmetical calculations in modern times. It soon occurred to Edmund Gunter of England to lay out Napier's logarithms along a line and measure along it with an ordinary ruler. "Gunter's line" gave quick approximations without the nuisance of adding and subtracting numbers taken from tables. Slide rules quickly followed, first in circular shape and then in the more familiar slip-stick form.

Now the small electronic calculator is rapidly converting the slide rule and the mechanical calculating machine into museum pieces. Yet it is less than a century since the slide rule and the mechanical calculator themselves converted into a museum piece another mechanical calculating device that had been invented a bit earlier and that competed with them successfully until motors were put on calculating machines and the slide rule came into wide use as a result of modern technical education. The earlier device was the sector. The form in which it first excited wide interest was invented about 1597 by Galileo.

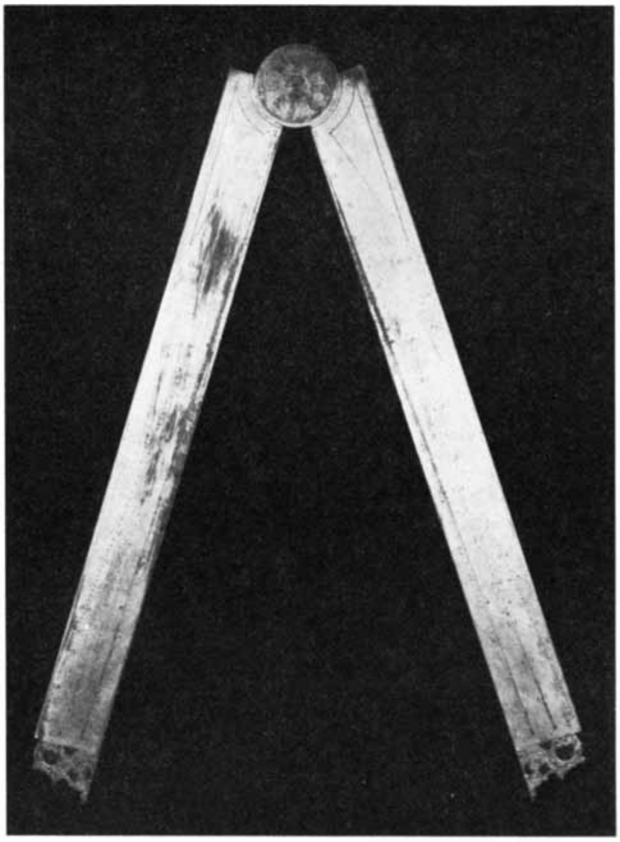
In its basic design the sector consists of two arms joined at one end by a pivot. The arms are of equal length (from four to 12 inches) and bear identical numerical scales. A geometrical problem that can be worked out with such an instrument calls for inscribing a polygon of equal sides in a circle of a given diameter. One sets a point at the free end of each arm of the sector so that the separation of the two points equals the diameter of the circle. Then, supposing that a pentagon is the figure to be inscribed in the circle, one measures the distance from the 5 on one arm of the sector to the 5 on the other arm. That distance is the length of each side of the pentagon [*see illustration on page 107*]. Such a problem, however, lacks the feature of calculation that was added later.

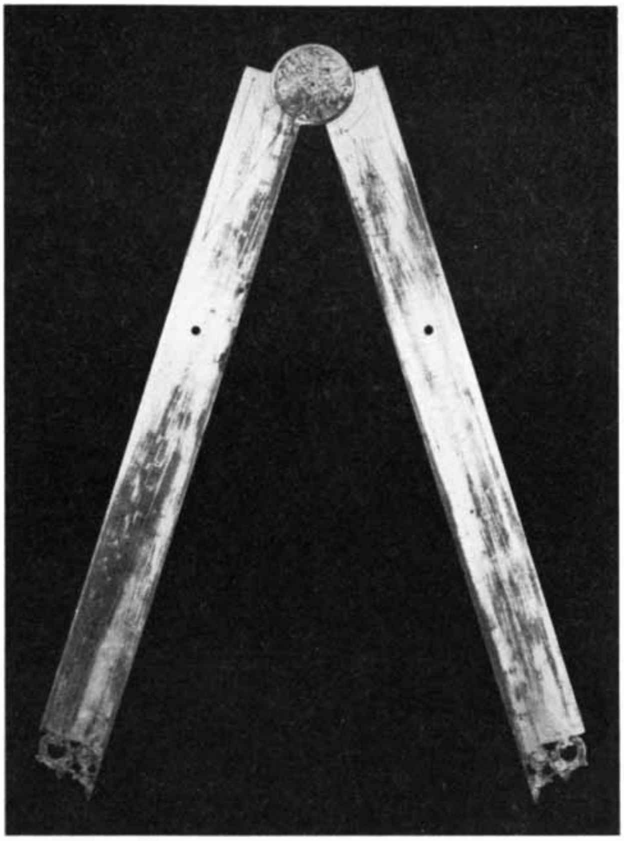
One advantage of the sector was that it could be easily understood and used by people of little education. Another was that for many common practical problems the user did not even need to think in terms of numbers. Indeed, one of the influences that led Galileo to conceive of the sector as a universal calculator for all practical purposes was a problem he confronted that was beyond the mathematics of his time.

The sector of old was a handsome instrument, usually made of brass or silver. At least 1,000 such instruments survive in museums and private collections. During the past century a less ornate version of the sector became part of the standard kit of the carpenter, the craftsman and the draftsman. These instruments, made of wood or bone in mass production, can still be encountered in commerce. Nevertheless, few people today have even heard of the sector, and the story of its invention has remained more the subject of guesswork and controversy than of serious historical investigation.

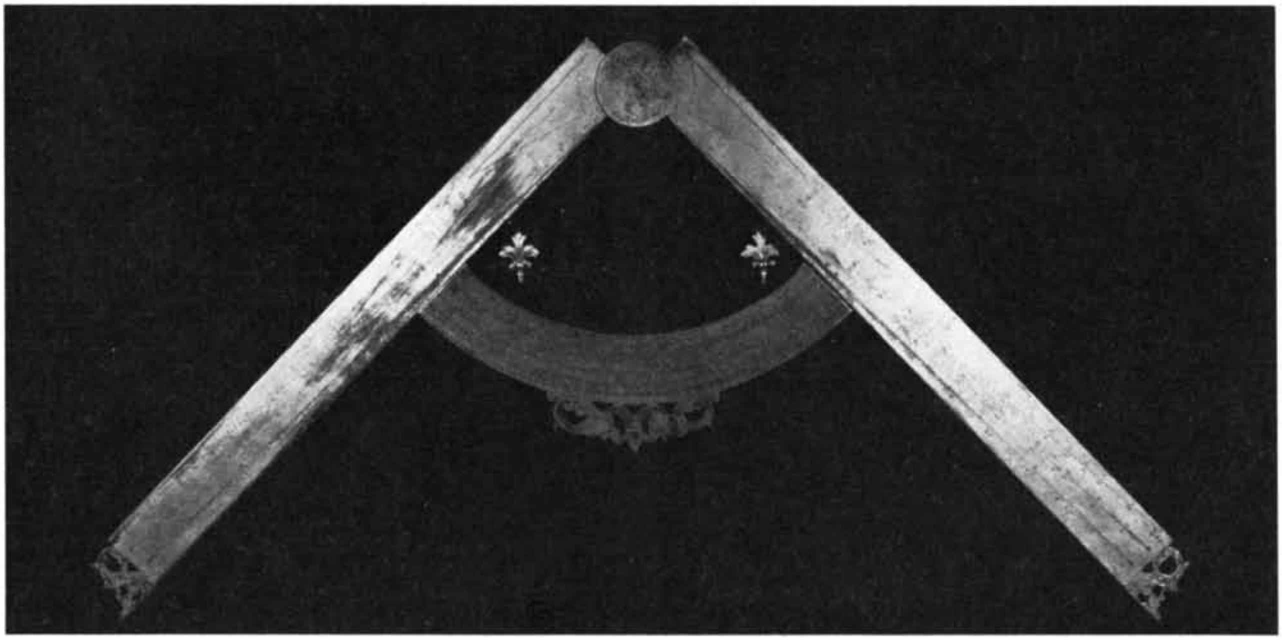
The story I shall tell is based on an examination of manuscript instructions that Galileo wrote before 1606, when he first printed a revised version of them, and on my inspection of many early sectors. My interest was recently reawakened when Anahid Iskian of New York, an expert on rare drawings and prints, called to my attention (and enabled me to acquire) a manuscript copy of Galileo's instructions made in 1605. The instructions differed from those in Galileo's book, *Operations of the Geometric and Military Compass,* published in 1606. Previously I had found a different manuscript version, dating from about 1599, in the Rocco-Watson collection at the California Institute of Technology. These two manuscripts, together with the five other known versions (all in the Ambrosiana Library in Milan and dating from 1597 to 1600), have made it possible to reconstruct the evolution of Galileo's instrument. Its history turns out to be quite different from what was supposed.

The use of a sector as a mechanical calculating device was first described in print by Thomas Hood. His account, written in English, was published in 1598. At the time Hood knew nothing of an Italian sector, and in 1597 Galileo had not heard of the English one, which probably had been in service for some years before Hood's book appeared. Independent simultaneous discoveries and inventions are not uncommon in science and technology. What is unusual here is that the two inventors had quite different backgrounds, Galileo being a professor of mathematics and Hood a practical scientist. In arriving at the sector they had different starting points and different objectives.

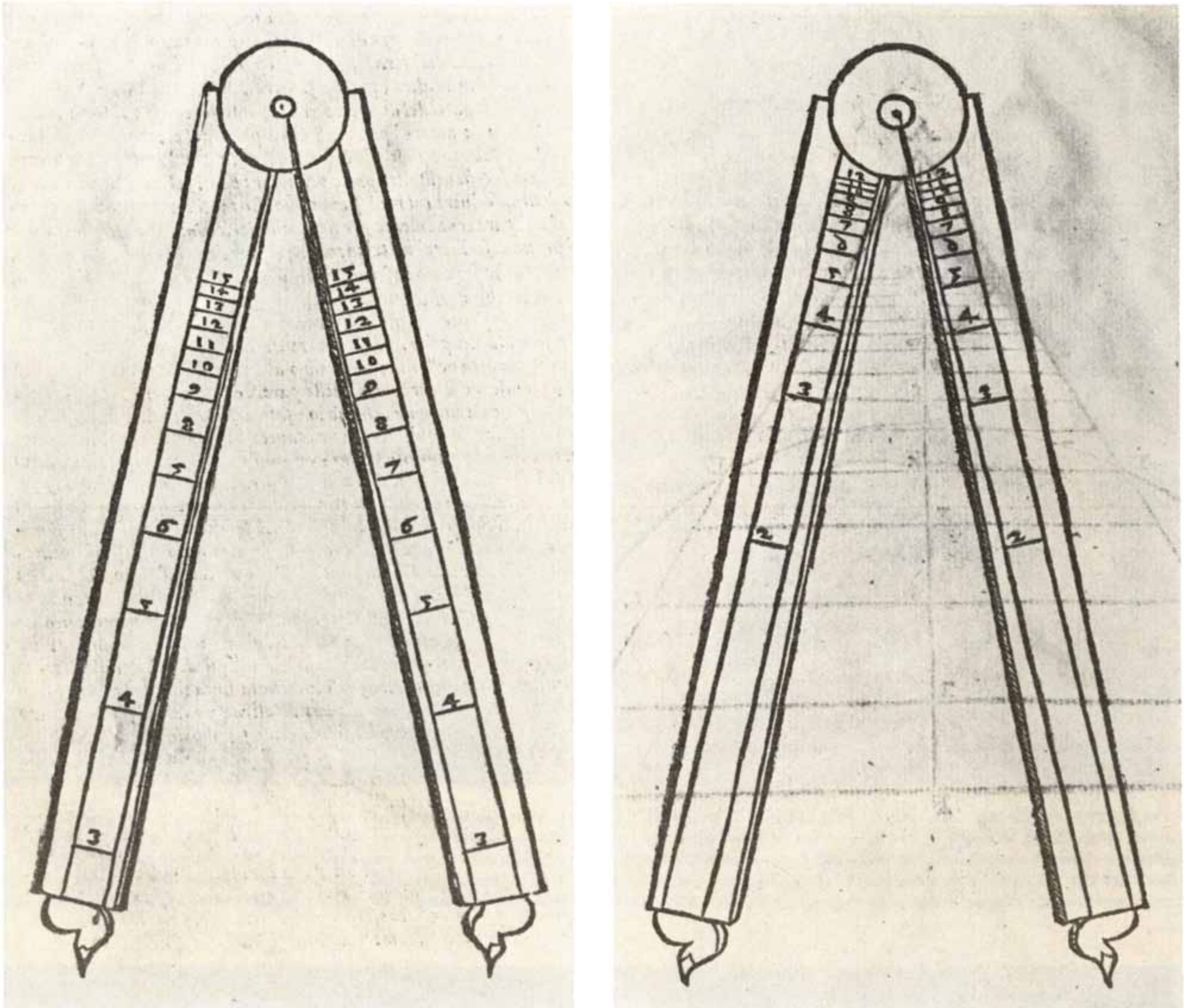




GALILEO'S "GEOMETRIC AND MILITARY COMPASS" is shown in the form in which it was made from 1598 onward. Its two sides are shown; the numerals are indistinct because of wear and the fact that the instrument appears at less than full size. This instrument, which was probably made in Florence, is from the author's collection. Galileo designed his compass, which later became known as the sector, to solve a problem he called "making the caliber." The problem was to find the appropriate charge for an artillery weapon of a given bore, calibrated according to material the ball was made of. Scales of instrument are shown in illustration on pages 112 and 113.



COMPASS WITH QUADRANT is portrayed. The quadrant made the instrument useful for astronomical sightings and for surveying. Galileo designed the quadrant to be detachable so that the entire instrument could be easily carried. A scale that reads from zero at each end to 100 in the center was also his idea; in modern terms each division is 1 percent of grade. With the quadrant the geometric and military compass could be employed in the determination of heights, distances and slopes, all of which were important military problems.



FORERUNNER OF SECTOR was devised by Galileo's friend the Marchese Guidobaldo del Monte. These two drawings appeared in a book that was printed in Venice in 1598. Ink from the reverse page shows through in each case; the faint lines in the drawing at right are therefore extraneous. On one side of the instrument (*left*) were scales that gave the sides of regular polygons to be inscribed in a circle with a diameter equal to the distance by which the two points at the bottom of the instrument were separated. With the scales on the other side (*right*) the user could divide into various equal parts a line of a length equal to the separation of the endpoints of the instrument.

Hood's sector bore three scales. Galileo's 1597 model had seven, only one of which (for constructing regular polygons) was on Hood's. Galileo eliminated this scale from his sector a year later, when he first included a scale for obtaining ordinary ratios, although that was the simplest scale of all and had from the outset played the main role on Hood's sector.

The accessories of Hood's sector suggest that it originated as a surveying instrument. They included pairs of removable sights, a plumb line and a graduated quadrant attached to one arm. The scale of equal divisions on each arm was probably provided originally for mapping to any given scale. The variable opening, by altering the measurements between corresponding points along the two arms, was a simple mechanical aid in the solution of all problems of proportionality.

A noncalculating type of sector was illustrated in a book on instruments printed in Venice in 1598. The instrument had been devised not long before by the Marchese Guidobaldo del Monte, Galileo's friend of many years, as a simple and inexpensive aid in two common problems of drafting and design. One problem was dividing a circle into a given number of equal arcs or constructing a regular polygon in a circle. The other was dividing a straight line of given length into an exact number of equal parts.

Guidobaldo's sector was a consolidation and an improvement of two drafting instruments that had been in use in Italy since the 1560's. One was the ordinary proportional compass, which is still in use today; it had points at both ends and a movable pivot. The other was the reduction compass, with a fixed pivot, two fixed points and two sliding points at right angles to the arms. Later models were made so that all four points met the paper at right angles. All such instruments were expensive and required frequent resetting of the movable parts. Guidobaldo made a simple hinge the only movable part of his sector and provided permanent scales that gave direct readings for the number of parts of a circle or a line.

The origin of Galileo's "military compass" has been conjecturally related in various ways to the proportional compass, the reduction compass and Guidobaldo's sector. It has always been supposed that Galileo took over a calculating device already in service and added to it more complex scales. One trouble with this supposition is that no calculating sector has been found that was in use in Italy before Galileo devised his "military compass" in 1597. Another objection can now be added: Galileo's sector was already quite complex before it bore the simplest scale of all. Let us start with its earlier history.

The ancestors of Galileo's calculating sector were two instruments of a quite different kind, both invented 60 years earlier for military applications by Niccolò Tartaglia, a practical mathematician. They were combined into one by Galileo, with improvements, before he thought of the idea of mechanical calculation.

In 1537 Tartaglia published in Venice a book, *The New Science,* in which mathematics was applied to artillery practice. The book introduced a gunner's elevation gauge consisting of a kind of carpenter's square that had one long leg, which was placed in the mouth of the cannon, and a fixed quadrant arc divided into 12 equal parts called points [*see top illustration on next page*]. A plumb line that was hung from the vertex indicated the elevation of the cannon, so that a dead-level shot was called "point-blank" and a shot made at 45 degrees was said to be "at six points." The device was quickly adopted throughout Europe.

Tartaglia also discussed the determination of the height and distance of targets by sighting and triangulation. For this purpose he invented a second instrument, which was also based on the square. Between his time and Galileo's several men proposed more appropriate and convenient instruments for triangulation in the field.

Many of Galileo's students at Padua from 1592 on were young noblemen destined for military careers. He tutored them privately in military architecture and fortification, and that work led him to improve on and consolidate Tartaglia's two instruments. First he remarked that it is not without peril to stand in front of a cannon, exposed to enemy fire, while adjusting the elevation. For that reason and others it would be best to gauge the elevation near the breech, which could be done by putting the ends of the arms of the instrument on top of the barrel and reading points of elevation from the center of the quadrant rather than from one end. Galileo made the arms of the instrument equal in length. Since a cannon thickens toward the breech, a compensation had to be provided by somewhat lengthening the forward arm of the instrument. For this adjustment Galileo fitted his instrument with a "movable foot" mounted on a cursor and held in place by a setscrew.

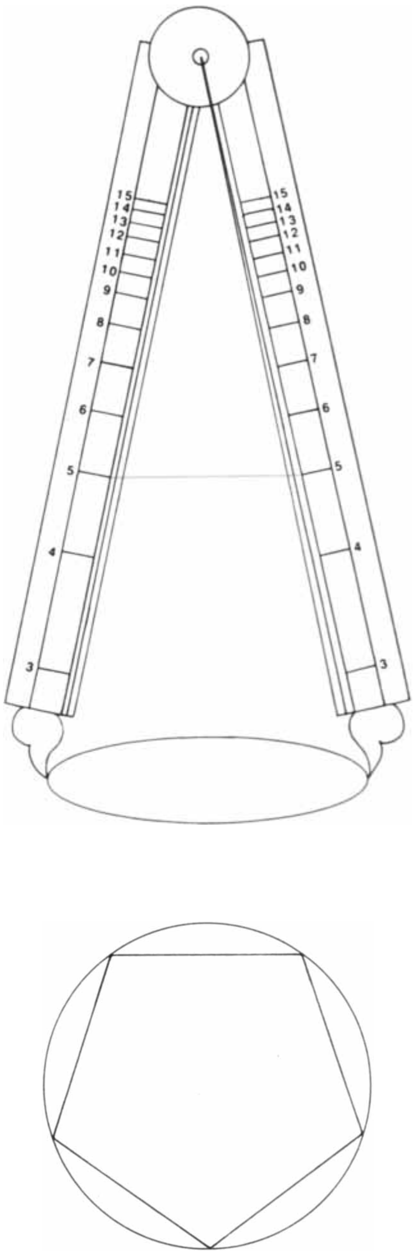
Next Galileo added graduations of the quadrant so that they extended for 90 degrees. The change made the instrument useful for astronomical sightings during long marches. He also added a clinometer scale reading in units of 100 percent of gradient; it enabled military architects to determine the slope of escarpments. That scale, giving the units of vertical drop per unit of horizontal advance, in turn suggested a simplification in the determination of height and distance by sighting. Galileo divided his quadrant into 200 equal parts, reading from zero at each end to 100 in the center, that is, at 45 degrees. Because such units are in modern terms 1 percent of grade, the scale dispensed with certain common calculations and made others a matter of simple mental arithmetic. The resulting instrument not only eliminated the need for two separate instruments for gunners but also was of value for civilian mappers and surveyors.

It was probably in 1595 that Galileo wrote a brief untitled treatise on the uses of the combined instrument. The last part, which concerns triangulation, was copied at the end of one early manuscript of his first instructions (1597) for the use of the sector. Later additions to this appendix on triangulation show that it was composed before Galileo had perfected his calculating sector.

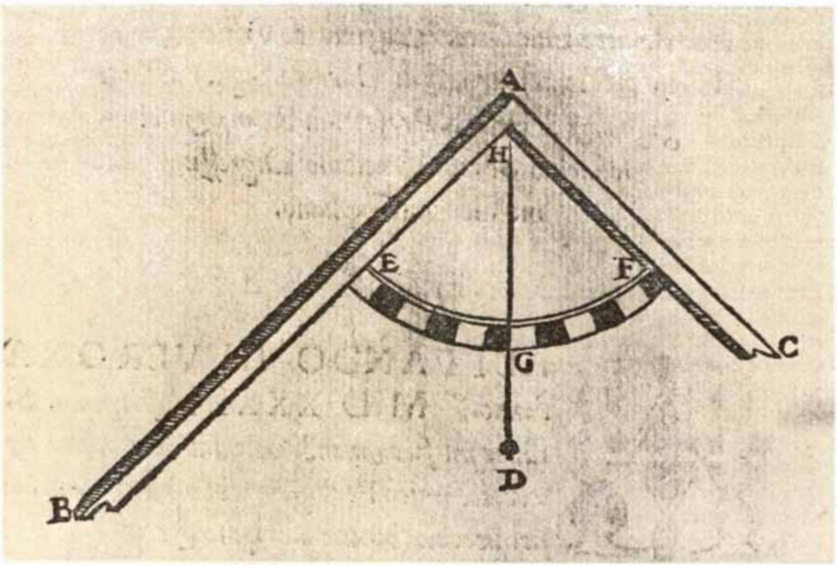
The stage that Galileo's instrument had reached in 1595 or 1596 is illustrated in a drawing made by a German student and inserted in the recently discovered manuscript [*see illustration on page 110*]. The quadrant was drawn exactly the same size as that of Galileo's own sector, which is now preserved in Florence. The cursor and the setscrew for the "movable foot" are also shown, along with a bracket and a universal joint to mount the instrument for surveying work. This is the only known drawing of these accessories, but Galileo's account books show that he manufactured to order a *nocella* (universal joint) to fit the instrument on a tripod.

Tartaglia's quadrant made an integral part of his elevation gauge. As long as the instrument was used only by gunners its large size and awkward shape did not matter. When Galileo incorporated the scale for triangulation, however, making the instrument useful also for surveyors, he detached the quadrant and hinged the arms so that the instrument could be carried more easily. This modification automatically created a sector. It was a natural step then to mark near its inside edges the two scales of Guidobaldo's sector, since both were helpful to mappers as well as to military architects.

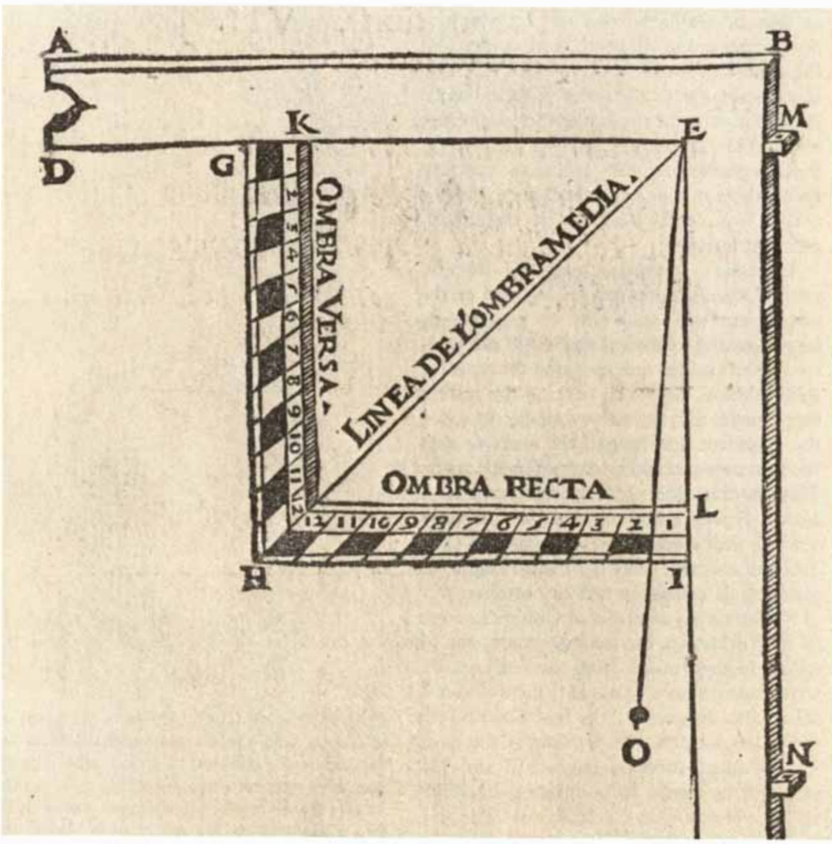
No surviving example of Galileo's sector of 1597 is known, but the instrument can be easily reconstructed from his earliest instructions on how to use it. I have drawn a schematic diagram of the instrument with the scales labeled; the drawing is the basis for the illustration on pages 112 and 113 and will be useful in comparing Galileo's earlier version with his later one. The two scales employed by Guidobaldo (*G-1* for constructing a regular polygon and *G-2* for obtaining equal sections of a line) were on the 1597 instrument but not on later models. Galileo added a scale, which I have designated *V,* that was even more useful than *G-1* to military architects because it facilitated the construction of a given regular polygon on a line of given length. Fortifications often had parts of regular polygons in their design, and the length of one side was likely to be dictated by some feature of the terrain or by part of an older fort. Be-



POLYGON IN CIRCLE shows a typical use of Guidobaldo's instrument and the later sector. Here the problem is to inscribe a pentagon with sides of equal length in a circle that has as its diameter the distance between the two endpoints of the instrument. With that distance established (*top*) one measures the distance (*color*) between the 5 on one scale and the 5 on the opposite scale. That distance is the length of one side of the desired pentagon. The inscribed polygon (*bottom*) also divides the circle into equal arcs of 72 degrees.



GUNNER'S GAUGE invented early in the 16th century by the Italian mathematician Niccolò Tartaglia was one of two instruments modified by Galileo to make his military compass. Tartaglia's gauge is shown here as it was illustrated in a book published in 1537. The longer leg (*left*) was put in the mouth of the cannon; the elevation of the gun could then be read in "points" on the quadrant by means of the plumb line. A level gun was said to be at "point-blank," marked *F*.



TRIANGULATION INSTRUMENT, the second of Tartaglia's inventions adapted by Galileo, is shown as it appeared in another illustration in Tartaglia's book *The New Science*. The instrument's function was to assist artillerymen in determining the height and distance of targets by sighting and triangulation. Galileo made his military compass by combining the two instruments devised by Tartaglia, changing the square to a quadrant and adding a clinometer.

tween *V* and *G-1* Galileo placed another line (*VI*), which he called the "tetragonic." This scale gives directly, for any regular polygon, the side of any other such polygon having an equal area. The same lines served for giving the approximate quadrature of any circle and for comparing areas in square measure. (By the reduction of any linear figure to triangles its total area could be equated to a square.)

Scales *V* and *VI* mark the emergence of a specialized mechanical computing device, still confined to basically geometric comparisons. Galileo's next step was to solve mechanically an important practical problem in artillery for which no solution had been known. It involved arithmetic, geometry and physics. Even when he had done so, however, his sector was still incapable of solving simple rule-of-three proportionality problems encountered in daily life by ordinary men. The idea of a single instrument for all purposes had not yet occurred to him. We shall see how the first general mechanical calculator evolved in his hands.

As Galileo remarked, the same names were often applied in different places to weights and measures when the quantities described were quite different. A captain of artillery must know how to charge a gun of any bore for a ball of any material without relying on anything but his own precise knowledge of the proper charge for a specific bore and a ball of known material. The knowledge was necessary because captains were often called suddenly to foreign places and because when enemy guns were captured, the captains needed to know how to turn them against their former owners. Wasteful and dangerous errors could result from mistaken units of measure even when charging data were marked on unfamiliar guns. The only sure safeguard against burst cannon, killed or injured gunners and wasted ranging shots was to be able to swiftly solve in the field the problem Galileo called "making the caliber."

The solution of the problem required scales *I* and *II*, which respectively gave the relative volumes of equal weights of various metals and stones and the spherical volume relations corresponding to equal increments of radius. With their aid even a gunner without mathematical training could solve any problem of calibration in a few seconds. In 1597 algebra had not yet been applied to geometry, let alone to physics, so that Galileo himself could not have written a practical formula for the problem. Even if he could have, it would have been of no use to gunners or even to most captains because of their limited mathematical knowledge.

Galileo's mechanical solution of this problem inspired him to exploit the sector for other problems. Its priority in his mind is clear from the fact that in every version of his instructions before 1600 the first problem taken up was calibration. In 1600, however, he rewrote his instructions to begin with a different scale (*IV*), which provided matching equal linear divisions, as if on two rulers. This scale had not even appeared on the 1597 model of the sector; it was added about a year later. By 1600 Galileo had found so many uses for the sector that thereafter the problem of calibration was not explained until his 20th chapter.

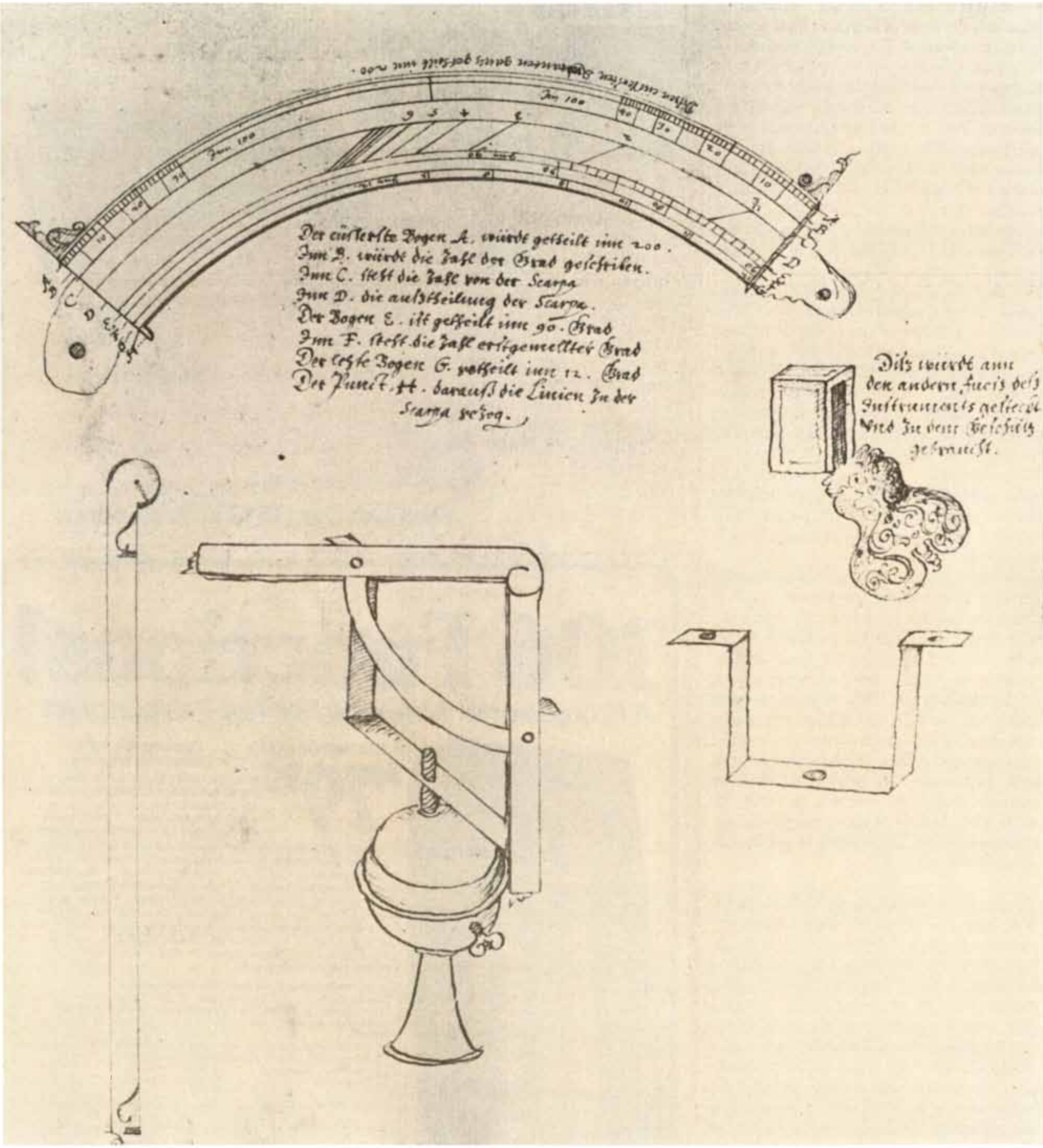
The manuscripts show how mistaken historians have been in relying on common sense and the form of Galileo's final printed instruction manual to reveal his methods of invention. Common sense suggested that he started with something simple (the scale of equal divisions as on Hood's sector) and then saw how it could be applied to more complex problems, just as he later arranged his instructions. Actually he had previously devised a mechanical means of solving a problem that was not susceptible to the mathematics of his time: the calibration problem, which was a function of two independent variables. Only later did it occur to him that simple proportionality problems could also be solved mechanically. Why should a professor of mathematics be concerned about problems that gave him and his students no trouble? In time, however, Galileo came to care about such problems for the benefit of ordinary people, who could not do square roots and even had trouble with multiplication and division. Mathematically untrained gunners had been enabled to do precision work; now, for the first time, mathematically untrained civilians could learn even such occupations as surveying.

Insertions made in the appendix on triangulation after 1600 reflect the evolution of Galileo's interest in dealing mechanically with simple mathematical problems. Originally (in 1595 or 1596) he gave detailed arithmetical calculations for triangulations. In the manuscript of 1605 and in his book each such example was followed by a passage showing how the same answer could be approximated quickly on the sector without using arithmetic. He introduced the new passages with such phrases as "But for those who cannot manage arithmetical calculations . . . " or "If we wish to avoid tedious calculations. . . ."

Evidence from affidavits filed in 1607, from the dating of manuscripts preserved in Milan and from the form and content of the two other manuscripts of Galileo's instructions indicates that the final model of his sector was designed within a year or so of the 1597 instrument. He made two fundamental additions and some minor changes. One of the major additions resulted from the fact that Galileo regarded his instrument as being imperfect as long as it did not enable him to determine the area of every figure bounded by straight lines and circular arcs in any combination. For this problem he devised a scale that would give the area of any segment of a circle. It is the scale I have designated as *VII*. It requires two sets of numbers, one on each side of the line. Hence for legibility it had to be placed near the outer edges, where scale *V* had been. Since *V* was directly useful in fortification, Galileo abandoned *G-1* and put *V* in its place, adding an instruction that showed how the inscribed polygons of *G-1* could easily be obtained from *V* instead.

The other major addition in 1598 replaced scale *G-2* with the simple scale (*IV*) of equal divisions, starting from the pivot. Identical rulers hinged at a common point make possible the immediate solution of all problems of proportionality, which is equivalent to solving all linear equations. Together with scales *III* and *II*, which in effect extended this capability to certain quadratic and cubic equations, scale *IV* turned Galileo's final sector into an instrument that could solve mechanically many algebraic problems. In fact, it conquered all the practical mathematical problems of the time.

By 1606, when Galileo printed his book



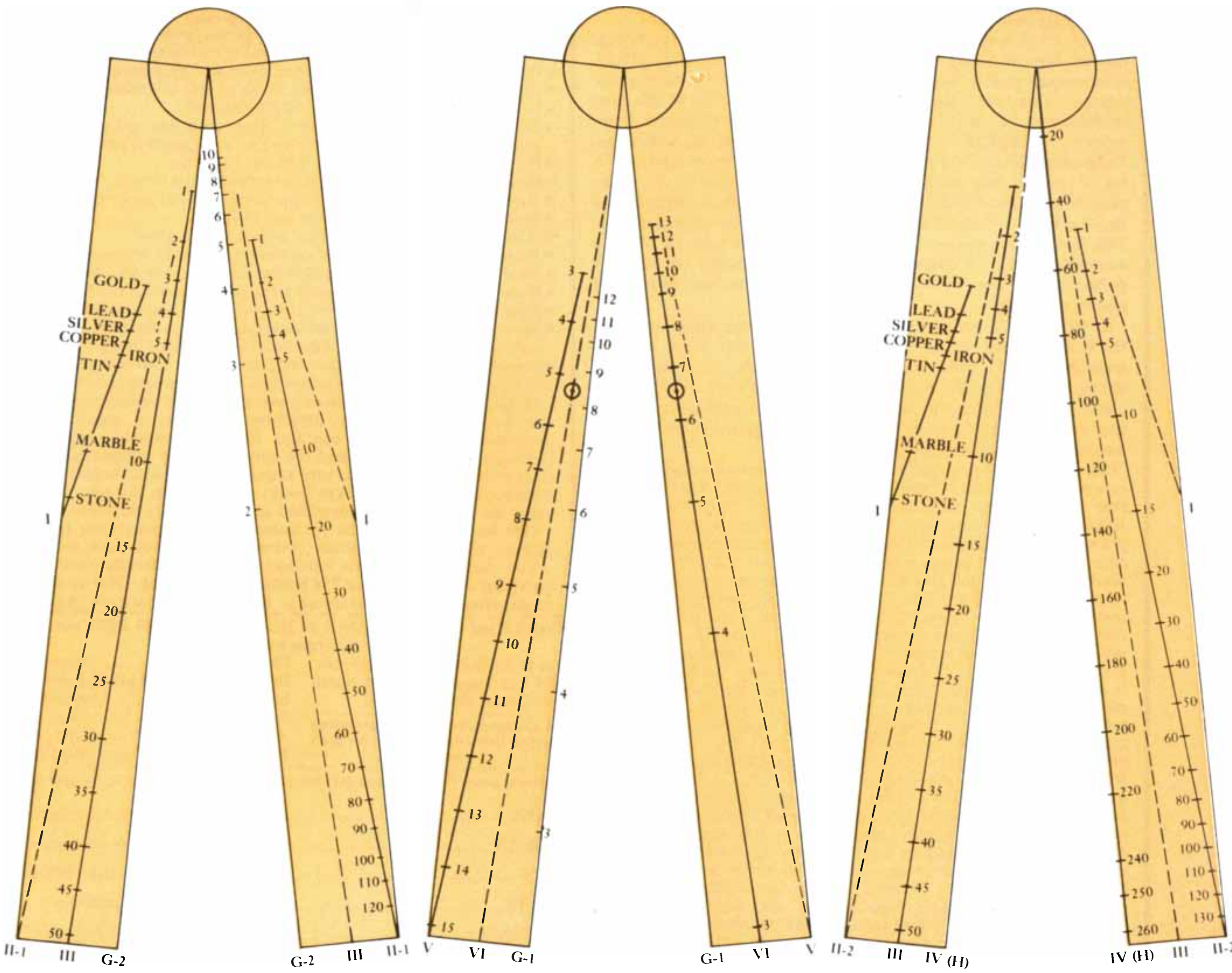
STUDENT'S DRAWING of accessories for Galileo's military compass was found in a 17th-century manuscript recently acquired by the author. The manuscript contained Galileo's instructions on the use of the sector. The drawing was probably made by one of Galileo's many German students before sector scales were added. At the top is the quadrant. Below it to the right is the cursor, which was fitted to one arm of the compass to carry a "movable foot" that Galileo designed to compensate for the taper of a cannon so that the gun's elevation could be read at the breech instead of the mouth. Below that are a bracket that mounted the instrument on a universal joint (*bottom center*), whose conical base could fit on a tripod, and (*bottom left*) the arms of the instrument in the folded position in which the compass was carried.

on the "geometric and military compass," about 100 of the instruments had been sold to students or given to friends and dignitaries in Italy and abroad. Galileo's account books show that at least 20 were in Germany, Austria, France and Poland. His book was printed in Italian for the benefit of the general reader. Within months a Paduan student plagiarized the book in Latin and implied that Galileo had taken the invention from him. Since Galileo had dedicated his book to Prince Cosimo de' Medici, with whom he was seeking employment, this was a serious matter, and he took legal action against the plagiarist. A proper Latin translation was published in Germany in 1613, by which time sectors were in use all over Europe.

A rash of other claims appeared in other languages and other countries up to the 1630's. Meanwhile sectors were made with a great variety of combinations of scales. For many years I have collected copies of the books and examples of the instruments. Apart from Hood's independent invention of a calculating sector of less general applicability, only one other claim appears to me to have any merit. That is the assertion in 1610 by the Belgian mathematician Michel Coignet, who corresponded with both Guidobaldo and Galileo in the 1580's, that he had had since that time his "pantometric scale," which consisted of graduated lines similar to those on the later sector. The lines were engraved on a metal plate without moving parts. Coignet's instrument was used, together with an ordinary pair of dividers, to solve similar problems by construction and measurement, not by mechanical computation as Galileo's was.

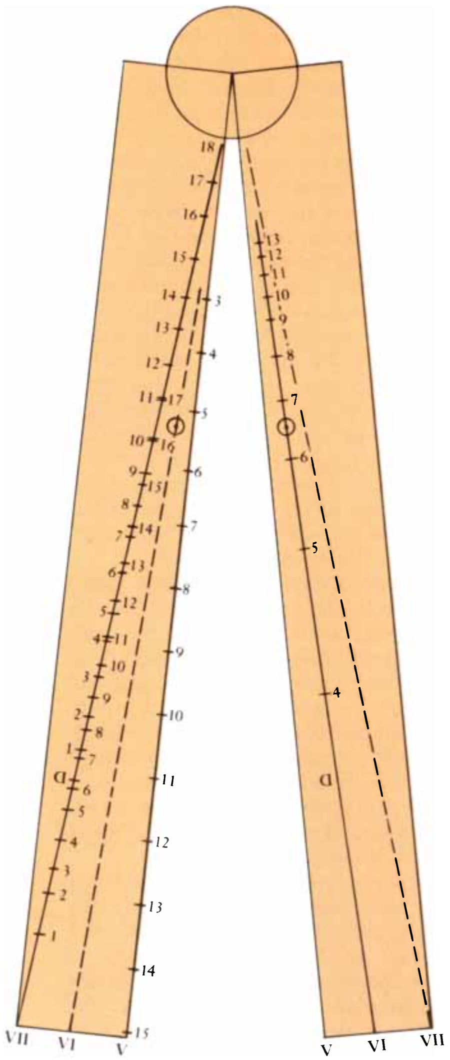
Apparently in 1612 Coignet transferred his pantometric scales to a pair of sectors. This modification was the basis, after his death in 1623, for a claim by his French editor that Coignet had invented the sector in the 1580's. Since Coignet himself made no protest against the books of Galileo and other rivals, it seems doubtful that he meant to claim the calculating sector when he asserted his priority in the "pantometric rule." Coignet's pantometric instrument in its original form was in fact more like a set of portable tables than a mechanical computing device, at least in the usual sense of the term.

Mechanical computers characteristically do several things: they abbreviate lengthy calculations, enable trained mathematicians to solve problems otherwise be-



GALILEO'S SCALES are portrayed schematically as they appeared on the 1597 model of his sector (*left*) and on the sector as it was made after 1598 (*right*). Each scale was inscribed on both the left and the right arm of the sector; here the numerals appear only once and the representation of the scale on the other arm is indicated for interior scales by broken lines. Numerals at bottom that identify the scales are the author's. Scales *I* and *II* dealt with the artilleryman's problem of "making the caliber." *I* gave the volumes for equal weights of various metals and stones. *II* provided a means of obtaining cube roots, *III* of obtaining square roots. *IV*, which first appeared in 1598, gave equal divisions of linear measure; the *H* signifies that a sector devised independently by Thomas Hood had a similar scale.

yond their powers and put at the disposal of the mathematically untrained the power of calculating methods otherwise unavailable to them. It is significant that Galileo's sector did all three things from its inception, as did the later slide rule. Hood's sector and the calculating machine that evolved in the 17th century from the work of Pascal and Leibniz seem to me to be somewhat different in that they were limited to problems that could already be solved by older methods (although less conveniently). Galileo attacked mechanically a problem he could not solve in any other way. In the course of doing so he came to see that mechanical means could be made available for solving all the practical mathematical problems of the day, much as our practical problems in mathematics are now solved by electronic devices.



V was for constructing a given regular polygon on a line of given length. *VI* gives for any polygon the side of any other polygon of equal area. *G-1* and *G-2* are like Guidobaldo's scales: *G-1* is for fitting a regular polygon in a circle, *G-2* for obtaining equal sections of a line.

The Nile Crocodile

Found both in the Nile and in most other African rivers and lakes, it has a remarkably large repertory of social behavior that includes parental protection of the young

by Anthony C. Pooley and Carl Gans

The largest of the living reptiles belong to the order Crocodilia. There are some 20 members of the order; for most people the best-known are the alligator of the New World and the Nile crocodile of the Old World. The Nile crocodile (*Crocodylus niloticus*) is now scarce in the lower Nile, but it is common in the upper waters of the river and in the other rivers and lakes of tropical and southern Africa, from the Sudan and Somalia southward well into South Africa and up the west coast of Africa as far as Senegal. Its tolerance of salt water has allowed it to occupy the coastal waters, and it has colonized not only the continent's offshore islands but also the great island of Madagascar, 250 miles out in the Indian Ocean. The same tolerance once extended the Nile crocodile's range along the Mediterranean coast as far west as Tunis and as far north as Syria.

The extent of this range, which overlaps the ranges of three smaller African crocodiles (the sharp-nosed *Crocodylus cataphractus* and two species of a dwarf genus, *Osteolaemus*), is an indication of the outstanding biological success of the Nile crocodile. Until a decade or so ago, however, this success was difficult to explain on the basis of the animal's recorded natural history. The consensus of incidental observations made by hunters, missionaries and an occasional naturalist was that the Nile crocodile was a lethargic hulk that spent most of its life basking in the sun, now and again rousing itself long enough to attack an unwary animal at the water's edge. Today, thanks to the systematic observation of individual crocodiles and of crocodile populations over periods of months and even years, we know that the animal is a remarkably efficient predator with a life cycle that includes substantial social components. The Nile crocodile often hunts cooperatively and has been seen to divide its prey. The adults are at least seasonally monogamous, and they protect their young by means of a complex repertory of behavior. The animals' adaptations to their water's-edge environment suggest why it is that crocodilians are found in all tropical and subtropical regions and how it is that the Nile crocodile has come to occupy such a large territory.

The adaptations that account for the success of the Nile crocodile are structural, physiological and behavioral. To deal with structure and physiology first, *C. niloticus* is among the largest of the crocodilians. Adults weighing 1,000 kilograms have been reported. The maximum recorded length, from the end of the snout to the tip of the tail, is five meters. Among reptiles only a few sea turtles and the estuarine crocodile, *C. porosus,* are heavier than the Nile crocodile, and only a few snakes and the estuarine crocodile are longer.

The most impressive part of the crocodile is its head. Between half and two-thirds of its length consists of the animal's elongated jaws. They bear a single row of stout, conical and slightly curved teeth that do not meet but interdigitate when the jaws are closed. The teeth of the upper jaw generally lie slightly to the outside of the teeth of the lower jaw, so that they are visible when the jaws close; the lower teeth fit into sockets in the upper jaw. In crocodiles, but not in alligators or caimans, an enlarged fourth lower tooth fits into a notch in the upper jaw rather than into a socket; it therefore remains visible when the jaws are closed and is a distinctive feature of the crocodile's appearance.

The crocodile's tongue is wide; it can be lifted from the floor of the mouth and its motion can be delicately controlled, but it cannot be protruded. The crocodile's mouth can be completely isolated from its pharynx by the gular fold, a tissue that can rise from the floor of the mouth to overlap a bony fold on the roof of the mouth. A crocodile is therefore able to keep its mouth open in the water without flooding its lungs and its alimentary canal. It can also breathe when its mouth is filled with water because its valved nostrils, which are on a raised area at the extreme tip of its snout, open into long tubes that lead not to its oral cavity but beyond the gular fold to a cup-shaped cavity in the roof of the pharynx [*see bottom illustration on page 122*]. The valves of the nostrils prevent flooding of the nasal tubes when the animal is submerged.

The crocodile is unique among reptiles in having a series of sensitive pressure receptors between its teeth and its jaws. Pressure receptors of this kind are found in mammals, including man; they gauge the intensity of our bite. If they provide crocodiles with the same kind of information, it could help to explain a number of the animals' remarkable patterns of behavior.

The eyes of a crocodile have bony eyelids; after the lids are closed they can be retracted, thereby streamlining the animal's head. When the eyes are open, they provide excellent fields of vision both to the sides and to the front, where an overlap of the fields may result in binocular depth perception. Immediately behind the eyes flaps of skin cover the oblong ear openings. Both anatomical and physiological studies suggest that crocodilians in general have superior sound perception in the range from about 100 to 4,000 hertz (cycles per second). The crocodile's sense of smell is also advanced. Apparently it can detect odors by flushing air through its nasal tubes; each sniff carries a fresh sample of air into the olfactory chamber adjacent to the olfactory lobe of the brain. The crocodilian brain is more complex than that of any other reptile.

The crocodile's limbs, trunk and viscera also show evidence of adaptation to life in a water's-edge environment. The forelimbs are somewhat longer than the hind ones, but both are stubby. The feet are long and bent, and the toes are tipped with sturdy claws; the toes of the hind feet are webbed. The thick tail, which accounts for some 40 percent of the adult crocodile's length, is almost rectangular in cross section. Its angularity is emphasized by two rows of scales that run along its top edges.

The skeleton of the trunk is powerfully constructed; the partly cartilaginous ribs are long, each consisting of a back, side and belly segment. The vertebrae are solid, and an extra set of cartilaginous abdominal ribs provides reinforcement for the belly. The shoulder girdle retains some flexibility with respect to the trunk; the limbs are attached low along the sides of the trunk and have a marked freedom of forward and backward movement, although vertical movement is restricted.

Most of the space within the bony rib cage is filled by a pair of complexly chambered lungs. The liver is to the rear of the lungs, where it is held in place by a radial array of connective tissue. The stomach and intestines occupy the visceral cavity to the rear of the liver. These organs can stretch well beyond their normal dimensions to accommodate large meals. The skin of the crocodile incorporates a geometrical arrangement of horny plates. Many of the plates have a bony core, and the plates that cover the crocodile's head are actually fused to the skull.

In the past decade a number of investigations of crocodilian physiology have been conducted, for example studies of how the animals digest their food, how they breathe and how their circulatory system is adapted to their amphibious way of life. All such research is useful, but it should be noted that the experimental animals have generally been immature crocodilians, ranging in age from hatchlings to specimens little more than three years old. The life span of any crocodilian is at least 25 years and probably exceeds 50 years, so that making projections on the basis of such studies is the equivalent of basing a physiology of man on tests conducted exclusively with infants.

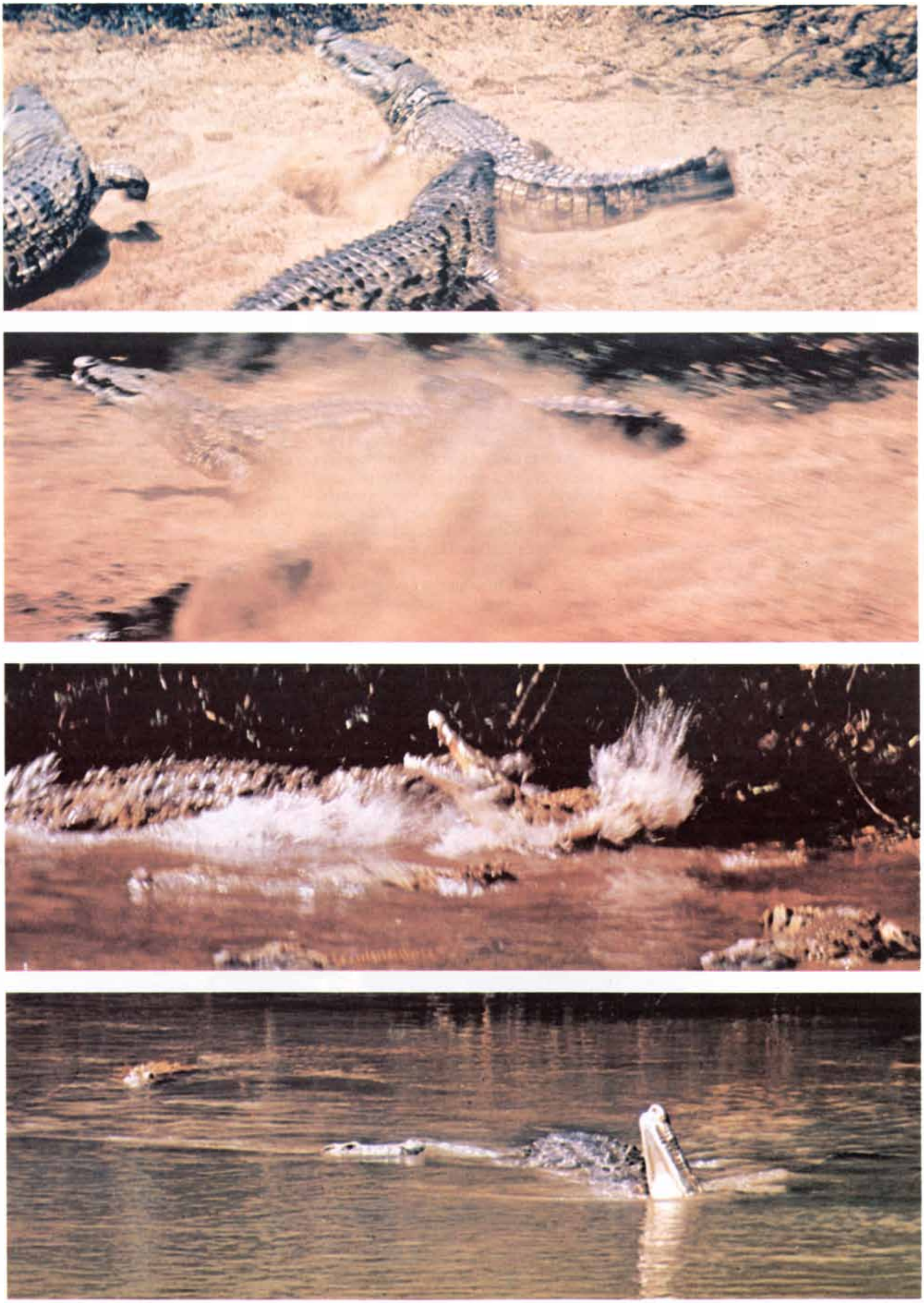
The studies of crocodilian digestion suggest that the passage of food through the alimentary system is rapid. The food is crushed and torn into large chunks that the animal gulps down. The chunks are further reduced by a kind of milling or churning: stones contained in the animal's stomach are set in motion by regular contractions of the strong stomach muscles. Secretions of hydrochloric acid maintain the digestive fluid at a *p*H of about 1.5, which is more acid than the optimum for digestive activity. The tissue fluids present in the crocodile's food, however, dilute the digestive fluid in contact with the food to the optimum *p*H (from 2 to 2.2). The estimated elapsed time from ingestion to elimination is somewhat more than 72 hours.

Even when food is constantly available, the animals' intake is variable. Temperature is one factor that affects feeding; adult crocodilians will not eat when their body temperature falls below a fixed threshold. Juvenile crocodilians are not as sensitive to lowered body temperature and continue to eat when adults will not. Adults will also refuse food when their temperature is at or above the threshold value but the barometric pressure is dropping. This behavior is apparently related to the fact that a falling barometer usually signals a one- or two-day period of lowered air temperature. Cooler weather retards the crocodile's rate of digestion, which is a matter of some significance in the northern and southern extremes of the crocodilian range.

Crocodilian breathing is a complex phenomenon, as might be expected in an animal that moves frequently from land to water and back again. To the rear of the gular fold the floor of the pharynx is normally raised so that the larynx neatly meets the cup-shaped cavity where the long nasal tubes terminate. The cycle of breathing is intermittent: an exhalation, an inhalation and then a protracted pause. The valved nostrils at the end of the snout dilate rhythmically in time with the cycle. In exhalation the space available to the lungs is decreased as the animal's ribs fold slightly and contractions of the abdominal and intercostal muscles shift the liver forward. In inhalation the contraction of the longitudinal diaphragmatic muscles, which connect the liver to the pelvis, pulls the liver backward. The lung volume is increased, and the air is aspirated much as it is when the piston of a syringe is pulled.

The differences between breathing on land and breathing in the water are marked. On land the crocodilian's rib cage is distended as the weight of the animal displaces the flexible ribs outward. In the water hydrostatic pressure tends to fold the ribs inward. The pressure of the surrounding water increases the muscular effort required for inhalation; electromyograms of the inhalatory muscles record a rise in effort as the depth of immersion increases. The crocodilian's inhalations therefore become shorter with depth. At the same time the activity of the exhalatory muscles decreases

MALE CROCODILE holds in its mouth a crocodile egg that is ready to hatch. The male, one of a parental pair, was being tested to see whether its reaction at hatching time was as positive as the reaction of the female parent. The male gently rolled the egg back and forth between its tongue and its palate until the hatchling broke out of the shell. Once free, the hatchling swam ashore to join its crèche-mates.



in proportion to the depth. It is as if the animal were recovering some of the energy it has to expend by inhaling against pressure.

The crocodilian's sniff is entirely unrelated to its breathing. The air is flushed through the nasal tubes by a lifting and dropping of the floor of the pharynx as the gular fold and the glottis remain closed. Electroencephalograms show that changes of air in the olfactory chamber stimulate the olfactory lobe of the animal's brain.

The crocodile's heart is analogous to a mammal's. Both direct measurements and cinefluoroscopy confirm that its four chambers can pump blood from the body to the lungs and from the lungs to the body without mixing. A bypass equivalent to one in the heart of the human fetus, however, connects the two ventricles, and it confers a unique advantage. When the crocodile is submerged and its lungs have given up their oxygen, most of the blood bypasses the lungs, so that the heart works primarily to circulate blood through the body.

A behavioral adaptation that contributes to the success of the Nile crocodile is basking. Like other reptiles, crocodilians conserve energy because of their low metabolic rate. The Nile crocodile conserves additional energy when its body temperature is raised by basking. This direct use of solar energy enables large crocodiles to survive even when food is scarce. (At any one time more than 30 percent of the crocodiles in many populations have an empty stomach.) When food is seasonally abundant, the animal can quickly recoup and go on to store the energy needed for growth and reproduction.

These specializations are probably shared by all crocodilians. They provide the basis for the animals' success as predators, a success so outstanding that to judge by the fossil record the order of crocodilians has undergone no significant structural alteration since the end of the Mesozoic era, some 60 million years ago. In considering the animals' predatory adaptations let us return to the Nile crocodile.

The most striking characteristic of crocodilians is their large size. Even the dwarf crocodiles of Africa are more than 1.6 meters long, and the average length for adults of all crocodilian species must be more than three meters. The newly hatched Nile crocodile weighs no more than 125 grams; the female parent weighs 300 to 500 kilograms. Thus the weight gain from immaturity to adulthood is 2,400- to 4,000-fold. The food supply required to achieve this gain might present problems if it were not for the fact that the ecological niche occupied by the Nile crocodile is unusually broad. For example, during its life span it takes prey that is progressively larger; the size of the food objects increases roughly in proportion to the predator's increase in size. The crocodile's predatory behavior undergoes a similar development.

RITUAL CHASE establishes the dominance of a senior male crocodile over a potential challenger. In the top photograph on the opposite page a challenged senior male leaves the water in pursuit of the challenger. The junior male takes flight (*second from top*); the pursuit drives several basking crocodiles off the bank into the water (*third from top*). The junior male now signals its submission by raising its head and exposing its throat (*bottom*). The senior may take one of the junior male's legs in its mouth but will not bite it.

Numerous analyses of the contents of the Nile crocodile's stomach and observations of the capture of prey indicate that the animals are not sit-and-wait hunters but highly active and versatile ones. Juvenile crocodiles necessarily subsist mainly on small prey: insects, snails, frogs and fish fry. Adults can successfully subdue large animals; they routinely capture drinking antelopes, and they can seize and drown Cape buffaloes as heavy as themselves. At the same time the adults will feed opportunistically on lesser prey: frogs, crabs and small fish. Subadults have a similar feeding behavior. As they become larger they seek out the larger prey, but they do not lose the ability to deal with the smaller.

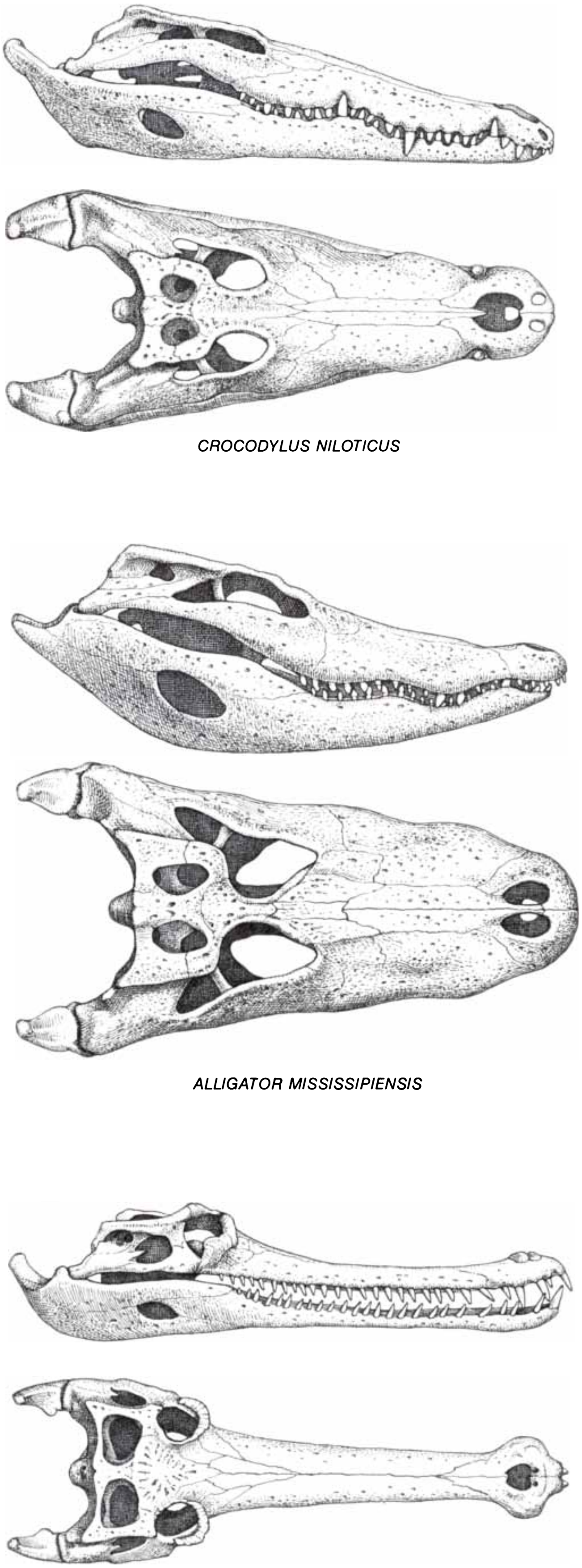
The Nile crocodile probably got its false sit-and-wait reputation from its hiding behavior. Lying in the water, it can breathe, smell, see and hear while only its nostrils and the top of its head are visible above the surface. It regularly enhances this natural concealment by lying next to a stand of reeds or drifting alongside a floating object. Moving in this manner, or swimming underwater, the crocodile can catch fish with a sudden sideways snap of its jaws. All crocodilians have a flattened snout that offers little resistance to lateral movement, and a sideways sweep is mechanically efficient in that most of the crocodile's body remains motionless while most of the mass of its head, the only moving part, is located near the center of rotation. Even so, the slenderer the snout of a species is, the more effective its sideways snap is and the more likely it is that there is a high proportion of fish in its diet. For example, whereas the Nile crocodile eats fish in addition to other prey, its slender-snouted relative *C. cataphractus* feeds mainly on fish and crabs.

The crocodile can also attack by lunging entirely out of the water and onto the land. Swimming crocodilians generally propel themselves by sculling with their tail, keeping their limbs folded against their sides to maximize streamlining. When the crocodile lunges to meet its prey head on, it has usually got close by swimming underwater, perhaps bringing its head to the surface once or twice to check the prey's location. The final lunge may carry the attacker several times its own length. The acceleration imparted by the powerful tail is combined with a simultaneous forward swing of the hind legs as the crocodile touches bottom. The toes and the feet dig into the bank and the powerful legs lever the body upward; if the bank is steep, the crocodile appears to vault straight out of the water. If the prey is still out of reach, the hind-leg stride may be repeated, and the crocodile's head may lower to hook over the top of the bank and support its body while its legs swing forward for another stride. Juvenile crocodiles have been seen to lunge when they were catching insects. An adult Nile crocodile can also dash across dry land for several yards, moving swiftly in a running stance with its trunk held above the ground.

There are numerous variations on both the land the and water patterns of hunting behavior. For example, a Nile crocodile can stun a small antelope at the water's edge with the same sideways head motion it uses to catch fish. It can also employ its tail for hunting purposes. If there is a nest of weaverbirds in the reeds along the riverbank, the crocodile can bend the reeds down with its tail and flip the nestlings into the water, where they are easily snapped up. In another tail-hunting tactic the crocodile swims slowly parallel to a riverbank with its tail bent toward the bank. The scales at the top of the tail riffle the water slightly, and small fish in the shallows move along ahead of the disturbance. When the crocodile turns its head around to the bank, the fish are trapped and are seized by a sideways sweep of its open jaws. The Nile crocodile is so agile that a fast reverse sweep will even intercept fish that are trying to escape by jumping over its back.

Larger prey animals are crushed by a series of bites along their length. If the prey is too large to be swallowed whole, the crocodile reduces it primarily by jerking and twisting motions. Such manipulations are inertial, with each flip of the crocodile's head moving the food object closer to the gullet; small prey is oriented headfirst to make it easier to swallow. In an alternative tactic the crocodile leaves some unwanted part of the prey outside its mouth and jerks the rest of the tissue away from it; the Nile crocodile uses this maneuver to break off the bony head of an armored catfish. In its hunting behavior the crocodile also makes good use of its adaptation to both land and water. Prey that is caught on land is often killed by drowning and then dismembered in the water; prey that is caught in the water is often flipped into the air, a medium that offers less resistance to jerking than the water does.

The hypothesis that crocodilians have a socially advanced feeding system is supported by observations of Nile crocodiles apparently cooperating in predation. The most commonly observed form of cooperative behavior is one where certain kinds of prey are reduced to pieces of a size convenient for swallowing. The twisting maneuver, in which the crocodile seizes some part of the prey in its jaws and then rolls over repeatedly until the seized part is torn away, obviously does not work with animals that are not large enough: when the crocodile rotates, so does the prey. In such circumstances crocodiles have been observed to



move the carcass toward another crocodile. The second crocodile bites the carcass and holds it while the first crocodile rotates, or perhaps both rotate in opposite directions. Each crocodile eats what it tears off without any hostility toward the other.

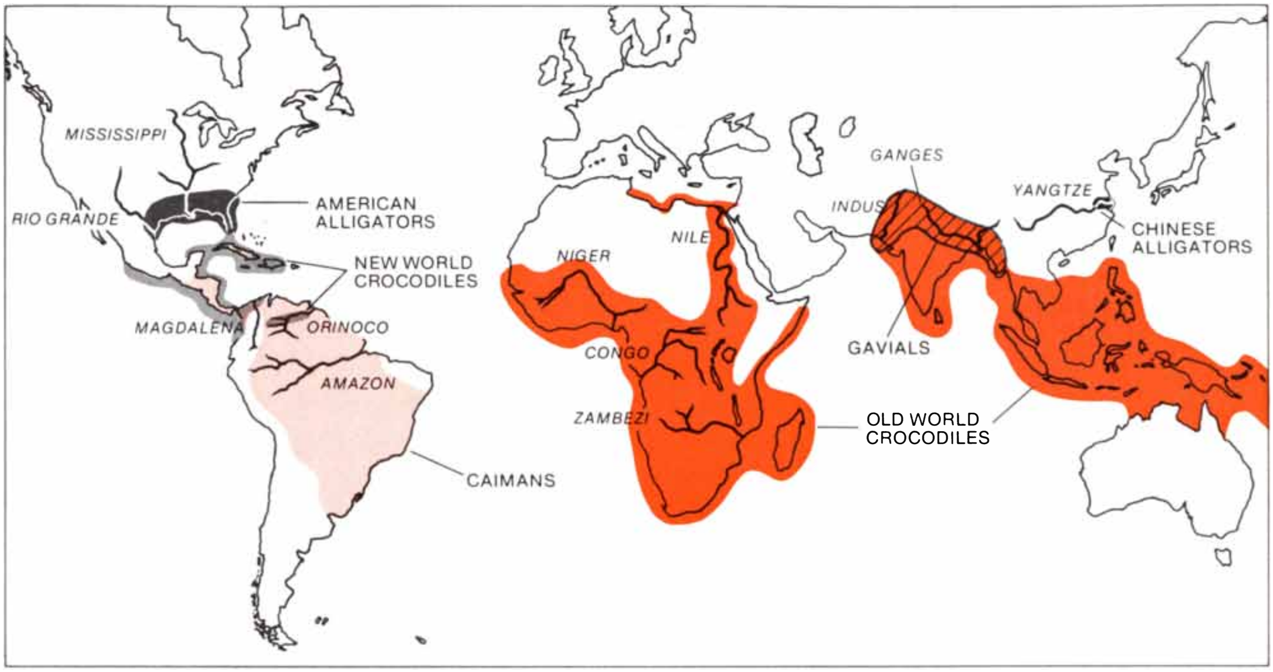
One of us (Pooley) has witnessed an even more important example of cooperation: two crocodiles walking overland side by side, carrying the carcass of a nyala antelope well off the ground between them. It was not clear whether one or both of the crocodiles had killed the antelope or whether it was carrion. (Crocodiles have been seen to sniff out carrion on land and drag it into the water.) Observations of this kind suggest a high level of nervous integration in the Nile crocodile.

Another example of cooperation may be seen in the early spring, when rivers rise and the water flows into channels leading to pans, or natural depressions, along the river. Subadult crocodiles often form a semicircle where a channel enters a pan, facing the inrushing water and snapping up the fish that emerge from the river. Each crocodile stays in place and there is no fighting over the prey. Any shift in position, of course, would leave a gap in the crocodiles' ranks through which the fish could escape, so that what might be a momentary advantage for one crocodile would be a net loss for the group.

Recent observations have helped to clarify our understanding of the Nile crocodile's pattern of reproduction. In southern Africa the animals evidently reach sexual maturity when they are 12 to 15 years old; they are then between two and three meters long and weigh between 70 and 100 kilograms. It appears that a dominance hierarchy is established among the mature males of a population early in the mating season. The crocodile's aggressive display is a complex sequence that includes its blowing bubbles out of its open mouth (the exhalation bypasses the nostrils and emerges through the gular fold), arching its neck, raising its tail and lashing the water. The bubbling display is accompanied by a grunt or growl. The animal may also partially submerge its head and blow water through its nostrils.

Dominance is established when two mature males display to each other. Whichever

EIGHT GENERA of living crocodilians fall into three broad divisions, illustrated by three skulls shown from the top and from the side: *Crocodylus niloticus* (*top*), *Alligator mississipiensis* (*middle*) and *Gavialis gangeticus* (*bottom*). The three genera of caimans, here grouped (*color*) with the genus *Alligator*, may actually belong in a separate category. Crocodiles have an oversized fourth lower tooth that remains in view even when the animal's mouth is closed because the tooth fits into a prominent gap in the upper row of teeth. This distinguishes crocodiles from alligators, whose oversized lower tooth is concealed behind an overlapping row of the upper teeth.



RANGE OF NILE CROCODILE (*center*) is compared with that of other crocodilian species. New World species of crocodiles are found from Cuba and southern Florida, where they overlap the range of the American alligator (*dark gray*), to northern South America, where they overlap the range of the caimans. Old World species of crocodiles are found in Africa and from the Indus River southeastward to New Guinea and northern Australia; their range in Asia overlaps the range of the gavial. A small population of alligators is also found in China. The Nile crocodile no longer ranges as far along the Mediterranean coast as is shown here and has become rare in the lower Nile.

animal proves to be the subdominant one eventually turns and swims off at high speed; the dominant male, usually the larger of the two, pursues, threatening to bite [*see illustration on page 116*]. If the subdominant animal slows in its flight, it raises its head almost vertically to expose its throat. The pursuing male may then seize one of the subdominant animal's limbs in its jaws, but it will not bite. Interactions of this kind result in the displacement of subdominant males from the territory occupied by mated pairs.

We are beginning to obtain evidence that paired Nile crocodiles stay together for some time. The species appears to be monogamous at least during the sexually active part of its annual reproductive cycle. The male and the female join in an elaborate courtship ritual that includes a mutual lifting of heads and rubbing of jaws. The ritual is also marked by a wide gaping of the jaws, but there is no biting. Within two or three days following the courtship ritual the animals copulate at least twice while they are in the water.

Crocodilians, like most reptiles, lay eggs. Nile crocodiles lay theirs some five months after fertilization. The clutch varies from 16 eggs to more than 80. The larger the female is, the larger are both the number of eggs and the weight of the individual egg (from 85 to 125 grams). When egg-laying time approaches, the female selects a nesting site. After the site is first used it becomes a permanent one; the female returns to it year after year, resting at it after mating and defending it against the approach of other females until egg-laying time.

When the nesting site is being used for the first time, the female occupies it for a few days. Then one night she digs a hole with her hind legs, deposits her eggs and buries them under 30 to 45 centimeters of soil. She now remains on the nest or in its immediate vicinity. She defends a fairly large territory around the nest by charging at intruders.

The incubation period lasts 84 to 90 days. During this period the female evidently does not feed at all, and she has become quite inactive by the time the young begin to hatch. The male parent also remains in the general vicinity of the nest, although he does not approach the nest itself and goes off at intervals to feed. One student of the Nile crocodile, M. L. Modha of the Kenya Game Department, observed an entire breeding season on the shores of Lake Rudolf; his observation blind overlooked an area that included several nesting sites. He reported that the largest males in the local population patrolled the beach in the vicinity of the nesting sites and that no subdominant crocodiles were to be seen there.

At hatching time the young crocodiles begin to call from within the egg. The sound they make is loud enough to penetrate the overlying soil and to be heard as much as 20 meters from the nest. On hearing the sound the female moves to the nest and starts to excavate it, working with her forelimbs and scraping and biting with her jaws. When the nest is opened, a remarkable event in the crocodile's life cycle takes place: the female picks up the hatchlings one by one until they are all in her mouth and then carries them down to the water.

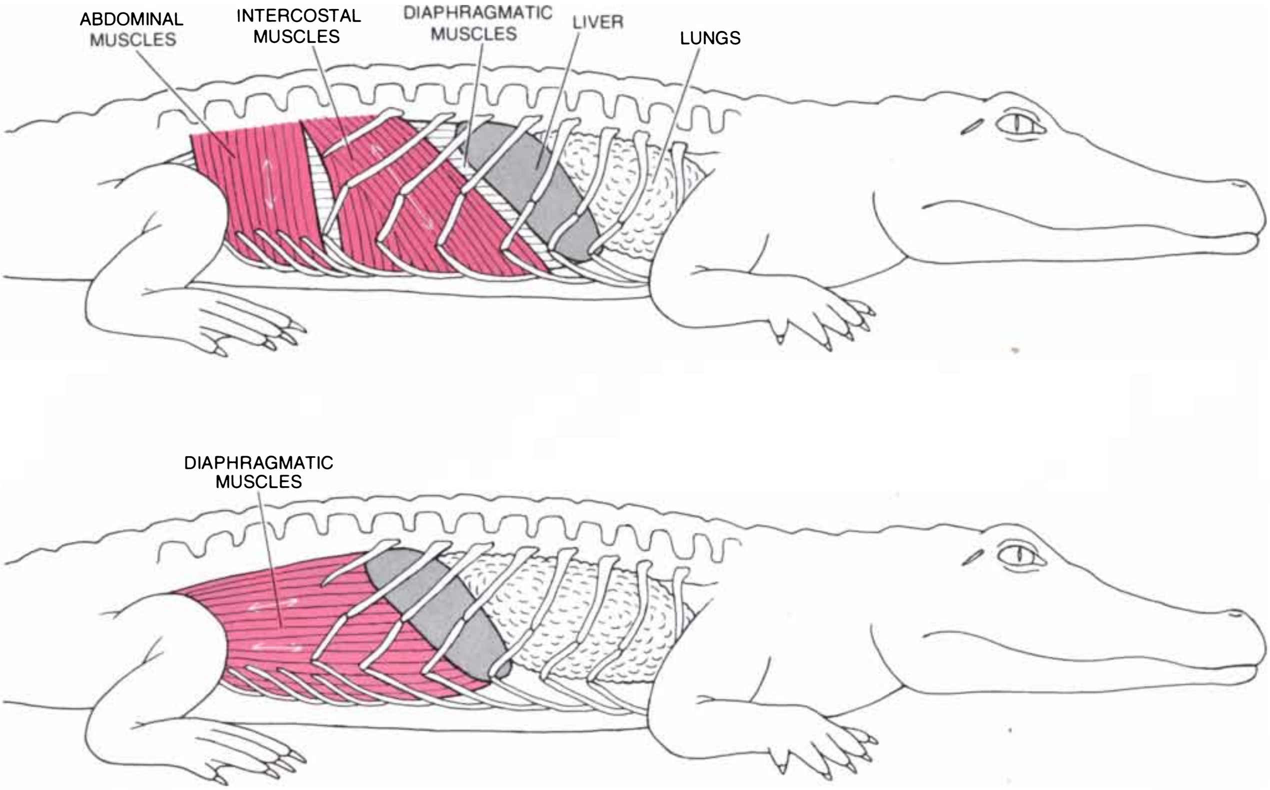
To do this the female rotates her head 45 to 90 degrees, gently picks up one of her tiny offspring with her teeth and flips it into the back of her mouth. To make room for all the hatchlings she depresses her tongue and the entire floor of her mouth, forming a pouch. As the young crocodiles enter the female's mouth their call changes to a softer chirping. Those hatchlings that wander away from the nest while the female is busy collecting the others soon begin to emit distress calls. Some turn to approach the female; others move around at random until she locates them by their calls and puts them in her mouth.

When the entire brood has been collected, the female enters the water and releases the young by opening her mouth and swinging her head from side to side in the shallows. Washed clean of sand from the nest, the young swim ashore and mill about at the water's edge, emitting pulsating chirps. This chorus elicits a vocal response from adult and subadult crocodiles in the vicinity. The parental male now approaches the female and is greeted by a low warble.

The hatchlings do not disperse for some six to eight weeks. Both parents remain near the crèche during that time, defending their young against other crocodiles and against predators in general.

Experiments conducted by one of us

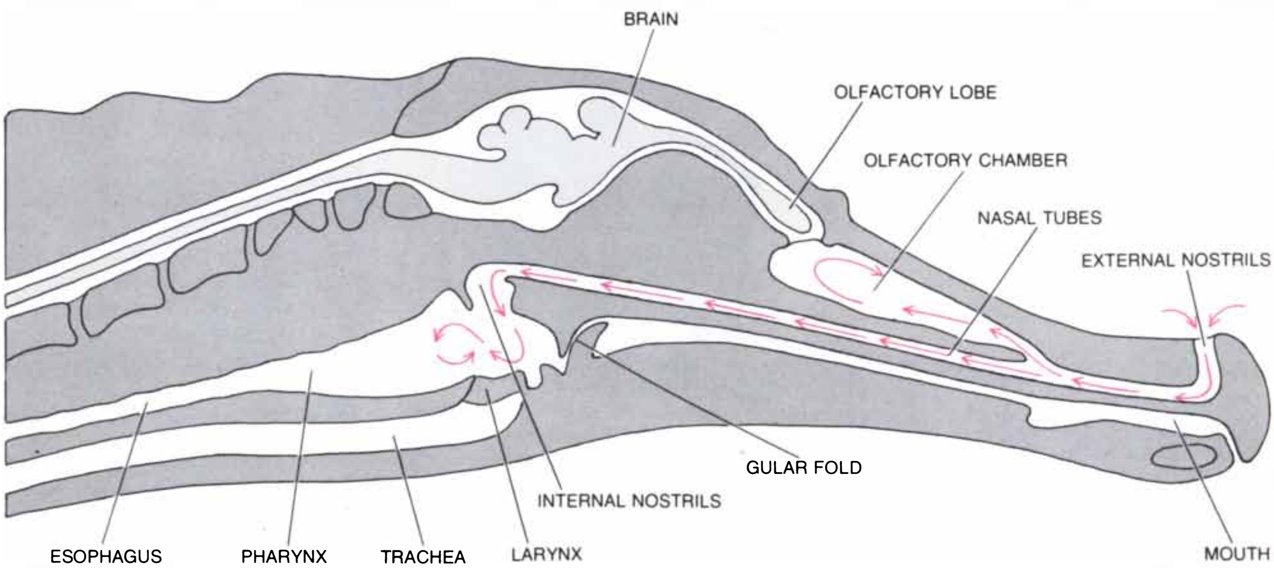

Did you miss reading any of these issues?
SCIENTIFIC AMERICAN
February 1974
SCIENTIFIC AMERICAN
SCIENTIFIC AMERICAN
January 1974
CIENTIFIC
MERICAN
July 1973
SCIENTIFIC AMERICAN
March 1974
June 1973
Our subscribers didn't.
Use the postpaid envelope in this issue to enter your subscription — and be secure in the knowledge that you will not miss any further issues of SCIENTIFIC AMERICAN.



CROCODILE'S BREATHING **is assisted by the movement of its liver, which lies just behind its lungs. In exhaling (*top*), space for the lungs is decreased as the ribs fold slightly and contraction of certain muscles (*color*) shifts liver forward. The inhalation that follows (*bottom*) begins when contraction of other muscles (*color*) shifts the liver backward; the air is drawn in as if it were pulled by a syringe piston.**

(Pooley) show that if the parental male has the opportunity, he too will collect the hatchlings in his mouth and deposit them in the water. The male will even break eggs to free the hatchlings. This he does by taking the egg in his mouth and rolling it back and forth between his tongue and palate. If one remembers that the weight ratio of the adult to the hatchling can be as much as 4,000 to one, both the male's grasping action and his palpation of the egg are demonstrations of a spectacular oral sensitivity and muscular control. The same jaws that can crush the femur of a Cape buffalo can pick up an egg without harming the little crocodile inside.

If one of the juveniles in an aggregation is disturbed or encounters any other kind of trouble while it is still a member of the crèche, it emits a long, loud, high-pitched distress call; the other hatchlings may join in, chorusing for as long as 30 seconds.



SNIFFING, an action that is independent of respiration, is made possible by the crocodile's ability to shut both its glottis and its gular fold, thereby making a closed system out of the nostrils and pharynx. By dropping the floor of its pharynx the crocodile is able to draw a sample of fresh air (*color*) into the olfactory chamber. Encephalograms show that the change of air that is produced by a sniff stimulates the olfactory lobe of the crocodile's brain; the mechanism may account for the crocodile's ready location of carrion in its vicinity.

Simultaneously the hatchlings take shelter, fleeing to the water or hiding under vegetation. All the adult crocodiles in the vicinity respond to the distress call by moving immediately toward the sound, even leaving the water to do so. The subadult crocodiles, however, do not react. The adults respond to distress calls even after the crèche disperses, although their reaction is less lively.

At the end of the crèche phase the young crocodiles gradually disperse, seeking out pools and streams that are not inhabited by subadults or adults. Their earlier positive response to the presence of an adult crocodile is reversed; when they encounter a crocodile larger than themselves, they turn and move away. Young crocodiles soon use their claws and teeth to dig a tunnel in the riverbank; the burrowing is sometimes a communal enterprise, with several young sharing the work. The tunnels are up to three meters long, and the animals use them as a shelter until they reach the age of five years or so. The tunnels afford the juveniles protection against various predators, including adult crocodiles, but it may be that the warmth they provide is equally important. During the winter months in southern Africa the air temperature in the tunnels can be eight to 10 degrees Celsius above the temperature of the river water.

There is almost no information on the life span of the Nile crocodile in its natural environment. This is hardly surprising; it is only in the past few years that systematic observations of crocodilians in the wild have been undertaken. We can confirm the impression that the young and the subadults have the highest mortality rate. Relatively few live to reach sexual maturity, let alone maximum growth. Predators that take hatchling and juvenile crocodiles include crabs and large fishes, other reptiles, herons and storks and such mammals as the mongoose and the hyena. A crocodile that survives to sexual maturity may play host to a few parasites but rarely encounters predators.

It seems clear that the 12-week crèche period, when not only the parental pair but also other adults behave protectively toward the hatchling crocodiles, must greatly reduce predation at a time when the young animals are at their most vulnerable. At the same time simple arithmetic—the 20- to 40-year reproductive span of the adult female crocodile multiplied by the annual production of 16 to 80 eggs—provides persuasive evidence that the annual loss of immature crocodiles to predators is a substantial one. The same arithmetic also suggests the crocodile's potential for the invasion of new habitats and for the renewal of populations after catastrophes.

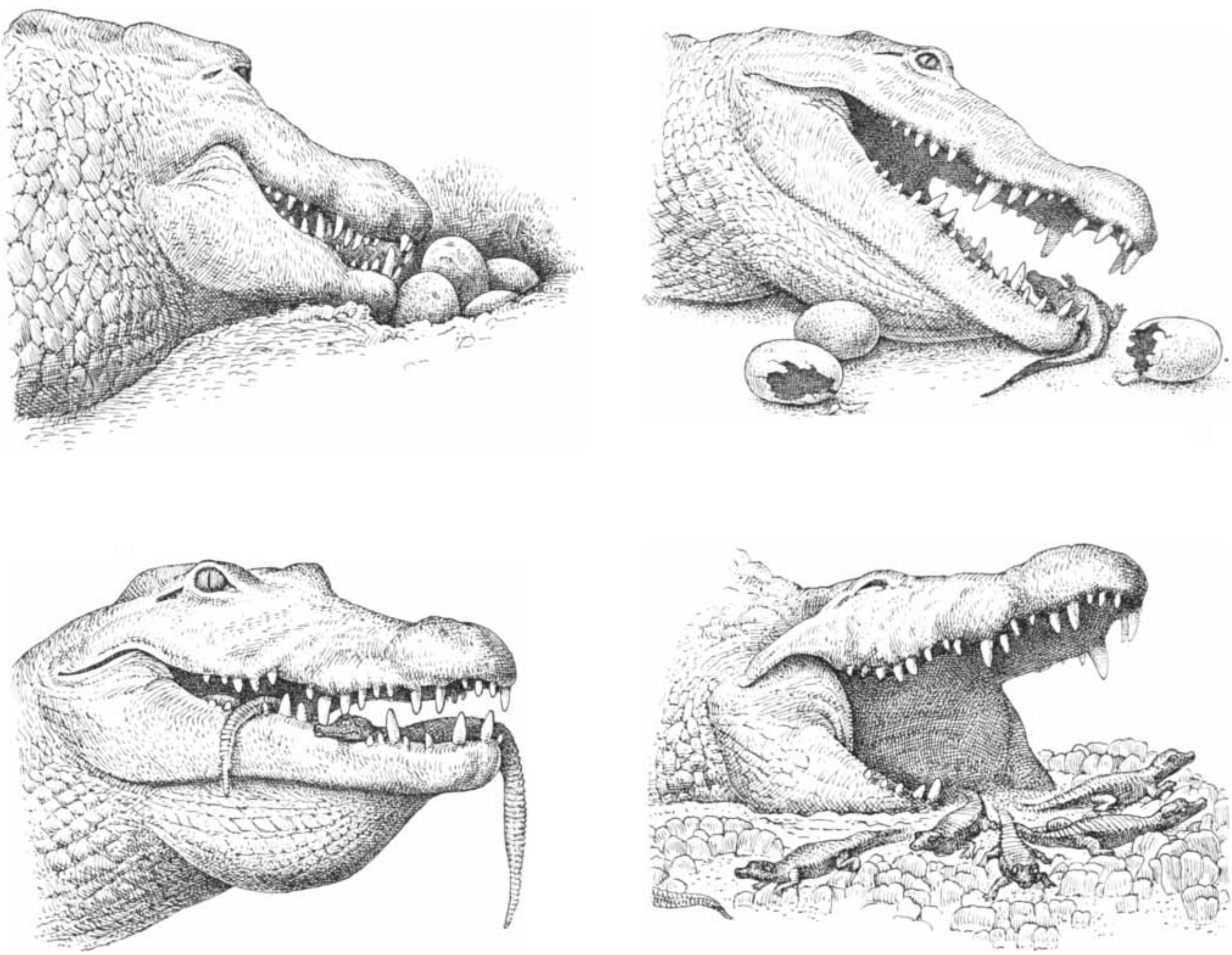
Bearing all this in mind with respect both to the Nile crocodile and to crocodilians in general, one can briefly summarize the nature of the Nile crocodile's successful adaptation. Any food chain has a bottom level occupied by animals that feed on plants, the primary producers. Above the herbivores are primary predators and above them are secondary and tertiary predators. The crocodile, as a tertiary predator, can feed on any of the animals below it in the chain, and in addition it can play the role of a scavenger. Moreover, as it grows larger it can add larger prey to its diet without losing the ability to secure small prey.

The crocodile's keen senses and advanced nervous system, which enable it to track live prey and to locate carrion, also facilitate cooperation with other crocodiles in the capture and dismemberment of prey. Its low rate of metabolism makes it possible for the female to endure the long fast while her eggs are incubating and also enables an entire population to survive when energy sources are limited. The crocodile's ability to maintain its optimum body temperature by basking means that a larger share of its total energy input can be allocated to growth and reproduction. The net effect is that crocodiles extract enough energy from a variety of waterside environments to maintain large populations of large animals.

The crocodile's large size confers its own selective advantage. The mature crocodile combines formidable power and speed with remarkable agility; the larger the crocodile is, the less likely it is to be the victim of predation. At the same time the larger males have the advantage in dominance confrontations, and the larger females produce more and larger eggs and are probably able to defend their offspring more effectively. That ability in turn increases the hatchlings' chances for survival. In this connection, it was once the custom to assess an animal's "reproductive effort" on the basis of the gross weight of each litter of its offspring. The application of this standard to the Nile crocodile would neglect its behavioral investment in the next generation: the female's three-month fast while she is guarding the nest. The weight of prey the female forgoes during her abstinence far exceeds the 10-kilogram mass of the largest clutch of eggs.

The burrowing habit of juvenile crocodiles, together with their ability to feed at body temperatures below the normal adult threshold for ingestion, are mechanisms that protect the young animals and accelerate their passage through the hazardous small-size stage. The same is true of the postcrèche dispersal patterns of the young crocodiles. The dominance hierarchies among the adults also do much to explain how crocodile populations manage to live with little overt conflict in locales where the area available for basking sites may be limited.

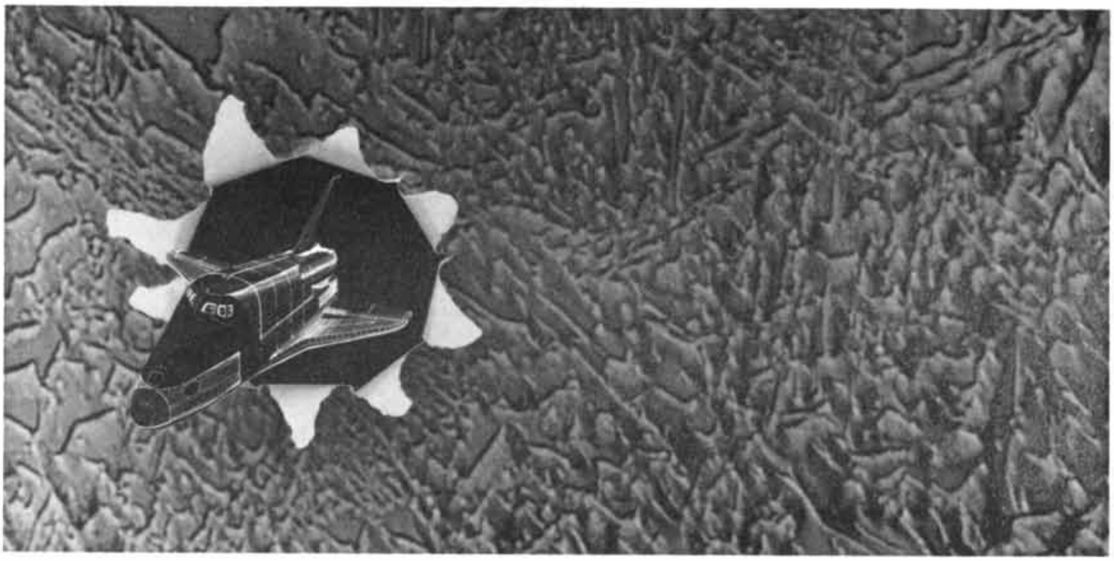
The Nile crocodile has adjusted successfully to a wide variety of African water's-edge habitats. It is nonetheless in danger of extinction, not because it has failed to find its place in nature but because it is the prey of human hide-hunters. One might not want to have a crocodile in every water hole, but it will be a sad day when a few stuffed or pickled specimens are all that remain of these splendid animals.



MATERNAL ASSISTANCE to hatchlings, shown in this series of drawings based on photographs made in the field, begins when the nest-guarding female is attracted to the nest site by the cries of the hatchlings. She first removes the soil that covers the eggs (*top left*). Next, as the nestlings break from their shells, the female turns her head to one side and, forming her mouth into a kind of pouch by depressing her tongue, gently gathers them in (*top right*). When all the hatchlings are in her mouth (*bottom left*), she moves from the nest site to shallow water. There she opens her mouth (*bottom right*) and deposits the hatchlings in the water by swinging her head from side to side. After the little crocodiles swim ashore they congregate; both parents guard the crèche during the weeks before the young disperse.

Conversation Pieces

Planning for Breakthroughs in Materials Technology



What can the Space Shuttle and Spacelab programs really be expected to do for technology and, ultimately, for mankind? Well, let's ignore the gee-whiz stuff. Let's stick to practical probabilities, beginning with medical technology.

Manufacturers of serums, vaccines, antibiotics, and other biomedical products, strive constantly for purity. But, on earth, it's very difficult to separate different kinds of living cells. It only takes a tiny minority of unwanted cells to contaminate an almost perfectly purified culture. Minute differences in the electrical charge on each type of cell could be used to achieve significantly better separation if it weren't for gravity. So, suspend your media in zero-g, apply an electrical field and various cells can be withdrawn with the most delicate precision.

Or, suppose you're a metallurgist, interested in alloys; earth's gravity tends to separate the components of many melts as they cool and harden. In zero-g, mixtures tend to stay uniformly mixed. The same goes for cool mixtures of fluids that differ in density. Immiscible on earth, they're easily kept homogeneous in zero-g.

Crystal-growers face similar problems. The benefits that zero-g processing can bring to makers of semiconductors alone are considerable. As for that curious noncrystal, glass, the prospects for optical technology are exciting, to say the least.

One of the greatest attractions of space manufacturing is containerless processing. On earth, even vessels that seem perfectly clean can actually contaminate their contents by reacting at the high temperatures that are essential to many processes. In zero-g, you can contain melts in electrostatic, magnetic, or acoustic fields; power requirements are low yet contaminants are quite easy to keep out.

The combination of zero-g and vacuum, that's available in orbital flight, is expected to facilitate developments in materials technology that range from difficult to impossible on earth. But the breakthroughs, whether they're surprising or reasonably predictable, aren't going to come automatically. It's going to take very careful planning.

At TRW, we have a team of systems engineers working the cost, schedule, and technical tradeoffs right now. They're supported by biologists, chemists, and physicists, who cut their teeth on difficult processing problems. We're working closely with NASA and other government agencies and we're teamed on specific projects with Beckman Instruments, Owens-Illinois, and U.S. Steel. By starting so early and proceeding with care, we hope to help develop new materials that will benefit everybody.

TRW®
SYSTEMS GROUP

Marketing Communications, E2/9043
One Space Park Redondo Beach, California 90278

MATHEMATICAL GAMES

Snarks, Boojums and other conjectures related to the four-color-map theorem

by Martin Gardner

"...they sought it with care;
They pursued it with forks and hope."
—LEWIS CARROLL,
The Hunting of the Snark

In the mail that comes from readers, which I open joyously, proofs of the famous four-color-map conjecture now outnumber trisections of the angle by about three to one. I return all such documents at once (if postage is supplied) without reading them. There are three reasons: (1) I am not qualified to evaluate such proofs; (2) if I were, it would often take hours, sometimes days, to discover where a proof had gone astray, and (3) when a four-color-prover is told about his error, he sometimes refuses to accept the verdict. That can lead to pointless correspondence, intrusive telephone calls and even unannounced visits.

In its simplest form the four-color conjecture is that four colors are sufficient to color any map so that no two areas of the same color meet. At the moment the most knowledgeable mathematician who is convinced he has proved the conjecture is Joseph Miller Thomas. The proof was published privately as a pamphlet in 1969, and it is now incorporated in Thomas' hardcover volume, *A Primer on Roots,* which can be obtained from the author. (His address is 60 Slocum Street, Philadelphia, Pa. 19119.) I should add that Thomas is a former editor of *Duke Mathematical Journal* and the author of several excellent books on modern algebra. His proof was sharply criticized in *Mathematical Reviews* by Frank R. Bernhart, an authority on the four-color conjecture. At a mathematical conference in 1975, when Bernhart gave a talk titled "How Not to Prove the Four-Color Conjecture," Thomas rose to his feet to defend himself. I am told that the scene was fraught with emotion and potential fisticuffs. Bernhart has a high opinion of some of the ideas in Thomas' proof, but it is exceeded by Thomas' low opinion of Bernhart.

A number of distinguished mathematicians, H. S. M. Coxeter for one, believe the four-color theorem is false. "If I may be so bold," Coxeter wrote in 1959, "I would guess that a map requiring five colors may be possible, but that the simplest such map has so many faces (maybe hundreds or thousands) that nobody... would have the patience to make all the necessary tests." A section on the theorem in Coxeter's classic *Introduction to Geometry* is headed by the following quotation from *Through the Looking-Glass:*

"I doubt it," said the Carpenter,
And wept a bitter tear.

If a countermap exists, it must have more than 51 regions. This new lower bound considerably improves Oystein Ore's 1968 proof of impossibility for maps with 39 or fewer regions. The new figure was established by Walter Stromquist (see "The Four-Color Theorem for Small Maps," *Journal of Combinatorial Theory,* Series B, Vol. 19, pages 256–268; December, 1975) on the basis of new reducible configurations discovered by Bernhart and others.

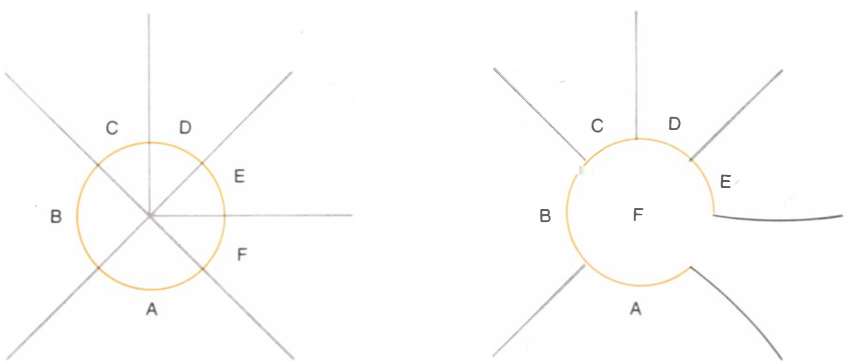
Many people who devise four-color proofs confuse the conjecture with a much simpler theorem that is easy to prove: No more than four regions can be drawn on the plane so that every pair share a common border segment. Even without such confusion it is tempting to think that this fact leads to a proof that all maps on a plane (or a sphere, which is topologically the same problem) can be colored with four colors so that no two regions of the same color share a border. Unfortunately it is possible (as far as anyone knows) for a map of more than 51 regions to be structured so that it cannot be four-colored. Any attempt to four-color such a map will inevitably result in a spot where two colors come together. Of course, one could always eliminate the trouble spot by altering colors, but then one would turn up somewhere else.

Dozens of other conjectures, seemingly unrelated to maps, are equivalent to the four-color theorem in the sense that if you settle any one of them, you settle the four-color problem. Is it always possible to slice corners from a convex polyhedron until every face is a polygon with a number of sides that is a multiple of 3? If you can do that, the four-color conjecture is true! For instance, by truncating four corners of a cube in such a way that no two corners are diagonally opposite, you can produce a solid with four triangular faces and six hexagonal ones. If you can find a convex polyhedron that cannot be properly transformed by truncation, you will have found a solid whose skeleton will produce a map disproving the four-color theorem.

A much easier way to search for a countermap is to look for a graph (a set of points called vertexes joined by lines called edges) that has the following properties:

1. It is connected. (It is all in one piece.)
2. It is planar. (It can be drawn on the plane with no edge intersections.)
3. It has no bridge (or isthmus). A bridge is an edge such that if it is removed, the graph falls apart into two disconnected pieces.
4. It is trivalent. (Three edges meet at every vertex.)
5. It is not three-colorable. (The edges cannot be colored with three colors, one to an edge, so that all three colors meet at every vertex.)

To explain this more fully, let us go back to a paper published in 1880 by Peter G. Tait, a mathematical physicist at the University of Edinburgh. Tait showed how easily any map can be transformed into a trivalent map with the same coloring properties. If a vertex has more than three edges, draw a small circle around it. Erase what is inside, and also erase one of the circle's arcs [*see illustration at left*]. The vertex of n edges is replaced by an extension of one region, now surrounded by $n - 2$ trivalent vertexes. It is obvious that any coloring of the regions will color the original map. We do not have to worry about a vertex with



Producing a trivalent map

only two edges because it is just a spot on a border and can be removed. In brief, any map can be changed to a network of trivalent forking lines, producing a trivalent map, and if the trivalent map can be four-colored, so can the original map. In addition Tait was able to prove that if the regions of a planar trivalent map can be four-colored, the edges of its graph can be three-colored, and vice versa.

The equivalence of the two colorings is evident from the following procedures. Assume that the regions of a trivalent map are colored with *A, B, C* and *D*. Label each edge with a letter that is the "sum" of the regions on each side, using the following addition table:

$$A + B = B$$
$$A + C = C$$
$$A + D = D$$
$$B + C = D$$
$$B + D = C$$
$$C + D = B$$

The result is a three-coloring of the edges. To go from edge-coloring to region-coloring assume that the edges of a trivalent graph are colored with *B, C* and *D*. Label any region *A*. From *A* take any path that visits each region just once. When you cross an edge, label the new region with the "sum" of the edge and the last region visited. Use the same addition table as before if the two letters differ, and call the new region *A* if they are the same. The result is a four-coloring of the regions.

Tait believed all trivalent graphs are three-colorable (and therefore all maps are four-colorable), but he overlooked two kinds of trivalent graph that are not three-colorable. One class consists of trivalent maps with bridges. Three simple graphs of this type are shown in the top illustration on this page. The two loops in the first graph make it obviously uncolorable, and it is almost as trivially obvious that the other two graphs also cannot be three-colored. Such graphs cannot, of course, be those of any legitimate map because the bridge would divide the outside region—a connected region if the map is on a sphere—from itself. (When any map is four-colored on the plane, the "outside" must always be treated as a region.) The bridge would be an absurd border: if you crossed it, you would still be in the same region.

The other class of uncolorable trivalent maps Tait missed are all nonplanar (impossible to draw on the plane without at least one intersecting edge). The simplest example, known as the Petersen graph, is shown in the upper illustration at the right. The form on the left is the one usually seen in textbooks. Rufus Isaacs of Johns Hopkins University, an applied mathematician noted for his work on game theory, prefers the one on the right. It is drawn with fewer strokes, and all the vertexes except one are on the outside, where they can be used for "hooking" the graph to other graphs in a manner to be explained below. By tracing each graph along its edges, taking the spots in order, you can easily verify that the two are topologically identical. The inner star is colored to make the isomorphism more obvious.

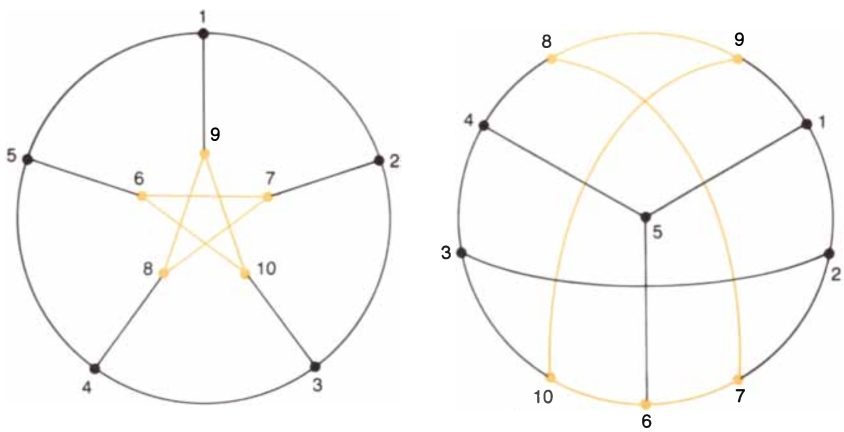
Trivalent maps with bridges

What can we conclude from all of this? If an uncolorable trivalent graph exists that is equivalent to a map disproving the four-color theorem, it must be trivalent, planar and bridge-free. No such graph has been found, and the reader would be ill-advised to search for one. On the other hand, as Isaacs recently discovered, the search for nonplanar graphs that cannot be "Tait-colored" (three-colored) can be a delightful pastime. Isaacs' book *Differential Games* (Robert E. Krieger, 1975) has provided material for several of my columns. The results of his search for uncolorable trivalent graphs are reported in his fascinating paper "Infinite Families of Nontrivial Trivalent Graphs Which Are Not Tait Colorable," in *The American Mathematical Monthly*, Vol. 82, No. 3, pages 221–239; March, 1975.

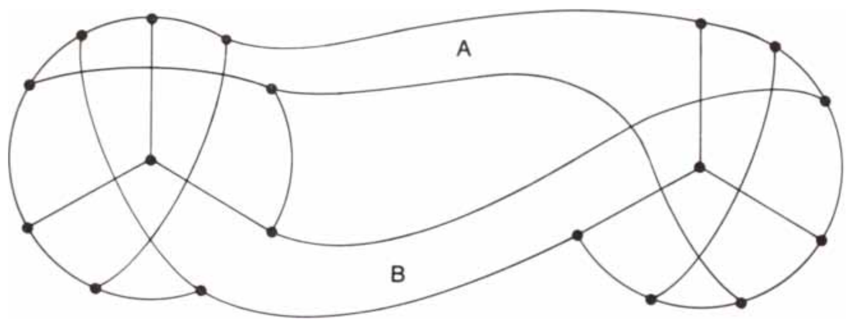
By nontrivial Isaacs means primarily a graph without a bridge. It is so easy to attach a bridge to any graph and render it uncolorable that we can ignore all graphs with bridges and concentrate on trivalent graphs without them. Isaacs also considers trivial any graph that has a "digon" or "triangle" (also to be explained below) because these are trivial features that can be added to or removed from any uncolorable graph without altering its uncolorable property.

To avoid the constant use of "nontrivial uncolorable trivalent" it would be helpful to have a single term for it. My first thought was the acronym NUT, suggesting that such graphs, like buried nuts, are hard to find. Indeed, as Isaacs puts it, anyone who searches for them "will be vividly impressed with the maddening difficulty of finding" a single one. NUT, however, implies that the search is a bit nutty when in fact it is serious mathematical business. The problem of defining and classifying all nontrivial uncolorable trivalent graphs is as worthy of being tackled as proving the four-color-map theorem. If it is ever solved, and one can prove that all such graphs are planar, the four-color-map problem will also be solved.

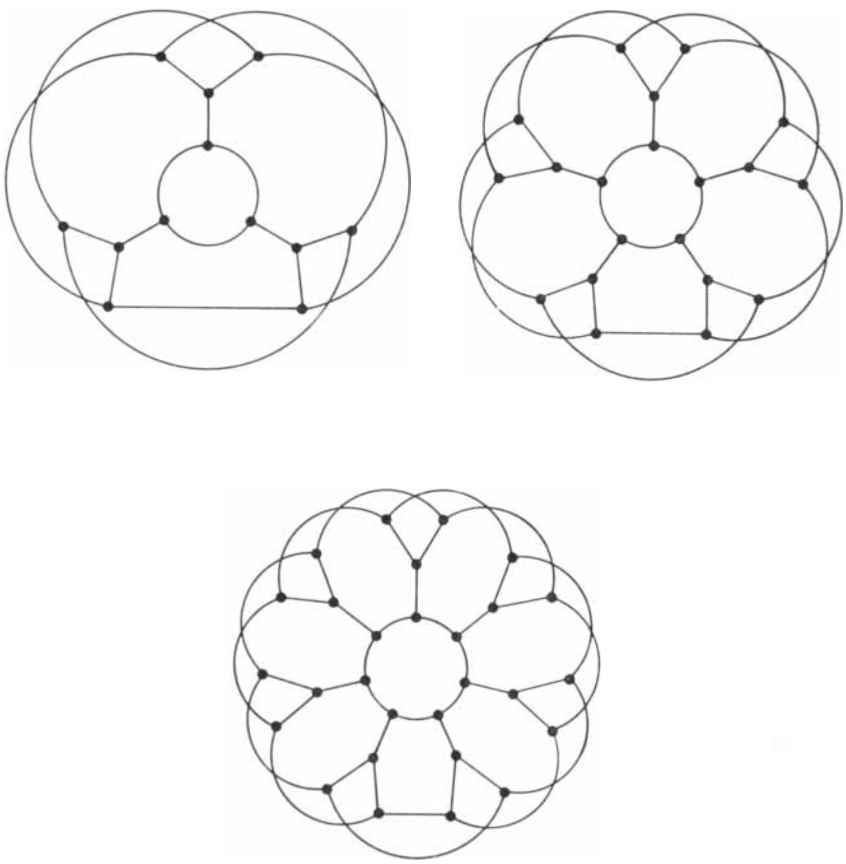
I propose calling nontrivial uncolorable trivalent graphs Snarks. A trivalent graph is a network of forking paths, and the person who tries to prove that it is uncolorable is certainly pursuing "with forks and hope"



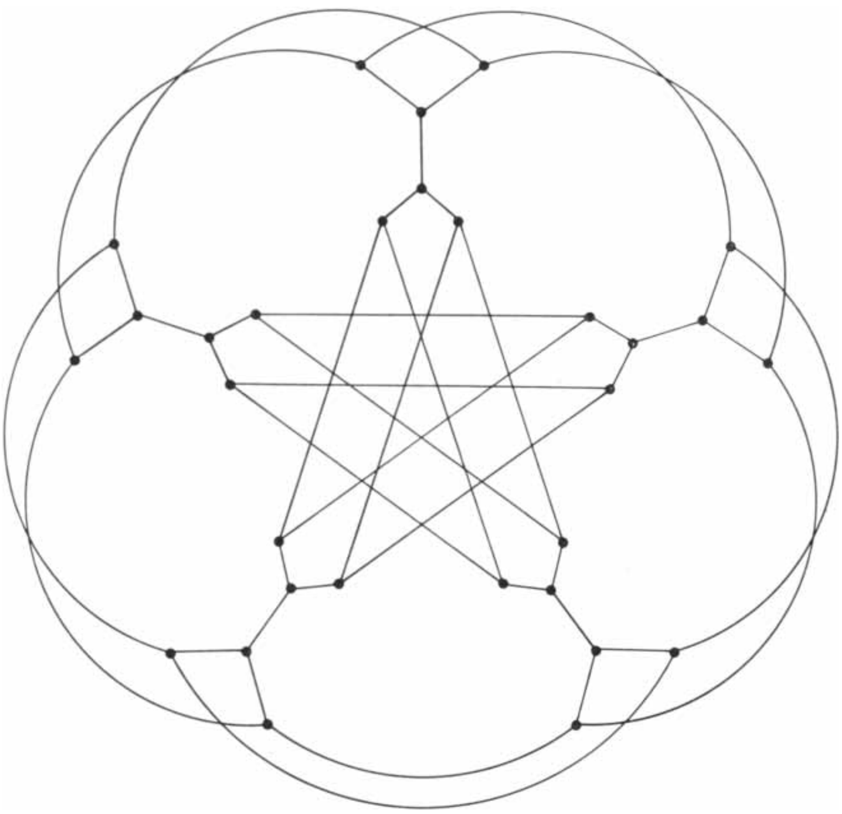
Two ways to draw a Petersen graph



How two Petersen graphs join to make a Blanuša graph



First three members of an infinite genus of Snarks



The double-star Snark, a wild variety

like the mad Snark-hunting crew in Lewis Carroll's immortal nonsense ballad. We know that Snarks are difficult to find, and that there is an exceedingly rare and dangerous variety called a Boojum. In our terminology the Boojum is none other than the planar Snark: the trivalent graph that explodes the four-color conjecture by providing a countermap. If anyone discovers a Boojum, he and the graph are instantly translated into hyperspace. This may explain why the four-color problem remains open. Was Judge Crater an amateur mathematician?

The Petersen graph, published in 1891, not only is the smallest possible Snark but also is (as W. T. Tutte, a great expert on graph theory, has shown) the only Snark with as few as 10 spots on its hide. It is hard to believe, but more than half a century passed before the second Snark (with 18 spots) was discovered by Danilo Blanuša, who published it in 1946. Two years later Blanche Descartes published a 210-spot Snark. Not until 1973 was the fourth Snark (50 spots) published by G. Szekeres.

The main outcome of Isaacs' Snark expedition was the discovery of two infinite sets of Snarks. One set includes the graphs of Blanuša, Descartes and Szekeres. Isaacs calls them BDS graphs to honor the three mathematicians on whose work he based his own. The graphs are formed by hooking together graphs previously known to be uncolorable, and also by hooking on other arbitrary graphs. Blanuša was not aware of it, but his graph can be obtained by joining two Petersen graphs in the manner shown in the bottom illustration on the preceding page. Remove any two nonadjacent edges from the Petersen graph on the left, then remove the three adjacent edges and their two vertexes shown missing on the Petersen graph on the right. Hook the two graphs together as indicated. The pairs of edges at *A* can be crossed or not, and the same is true of the pairat *B*. Moreover, at *A* or *B* (or both) one can hook in arbitrary graphs without destroying colorability. Any two uncolorable graphs can be joined in this way, with or without additional arbitrary graphs, to produce an infinity of Snarks. Szekeres' graph is formed by hooking together five Petersen graphs. Descartes's graph is a combination of Petersen graphs with inserted nonagons. Readers interested in these constructions can consult Isaacs' paper for details.

The second infinite set of Snarks found by Isaacs is shown in the top illustration at the left. The first graph is the Petersen graph with a trivial "triangle" of three spots substituted for the central vertex. By increasing the number of large petals in the series 3, 5, 7, 9, ... one obtains an infinite set of flower Snarks with spots in the series 12, 20, 28, 36, ... All trivalent graphs, by the way, necessarily have an even number of spots. If the number is $2n$, the number of edges is $3n$, and if the graph is three-colorable, there are n edges of each color.

Isaacs also discovered one Snark (30 spots) of a wild variety that does not belong to either the BDS or the flower sets. He calls it the double-star [*see bottom illustration on opposite page*]. Of course, the double-star Snark, as well as any flower graph, can be hooked to BDS graphs, or the flower graphs can be hooked to one another. The combinatorial possibilities are endless, and complex BDS graphs can be drawn in such a way that it is extremely difficult to sort out their component parts.

To introduce readers to the excitement of Snark hunting four simple trivalent graphs that are three-colorable are presented in the illustration on this page. Readers are invited to see how quickly they can three-color any one graph or all four graphs. The bottom graph is drawn in canonical form. Obviously any graph can be similarly distorted so that all its vertexes lie on a straight line. After some practice at three-coloring the reader may want to try the more difficult task of proving that the Petersen graph or any of the other Snarks is uncolorable. To do this you must test all possible ways of three-coloring, and that can be time-consuming, particularly if you are not using an efficient procedure.

The following backtrack algorithm, which is not given in Isaacs' paper, is the one he has found most useful, both for three-coloring trivalent graphs and for proving their uncolorability. I shall describe it in pencil-and-paper terms [*see upper illustration at left on next page*], although it can be made even more efficient by working with extremely large graphs and small numbered counters.

1. Draw a large picture of the graph in ink. Let 1, 2 and 3 stand for the colors. All the labeling should be done with a soft lead pencil because there may be many erasures.

2. Pick any vertex and label its three forking paths 1, 2 and 3. It does not matter how you distribute the numbers. They merely stand for different colors, so that there is no loss of generality by permuting them.

3. Move to any adjacent vertex. Its two unlabeled edges can be labeled in two ways. In our example the upper edge must be 1 or 3. Label it 1, putting a bar over the numeral to indicate that it is a free choice. Add subscript 1 to show that it is your first free choice. We call this the step number.

4. Label all edges that are determined by the first free choice. In our example there is only one. It gets a 3. Add the same step number (subscript 1). No bar goes over the 3 because the labeling is forced.

5. Move to another adjacent vertex where there is a free choice. As before, the first decision gets a bar over it, but now the subscript is 2. This shows that it is your second free choice.

6. Continue in this way until the entire graph is labeled or you encounter a contradiction—a forced choice that puts two edges of the same color at a vertex. When that occurs, at step n, erase all labels having subscript n. It is wise to leave the erasure of the barred label to the last.

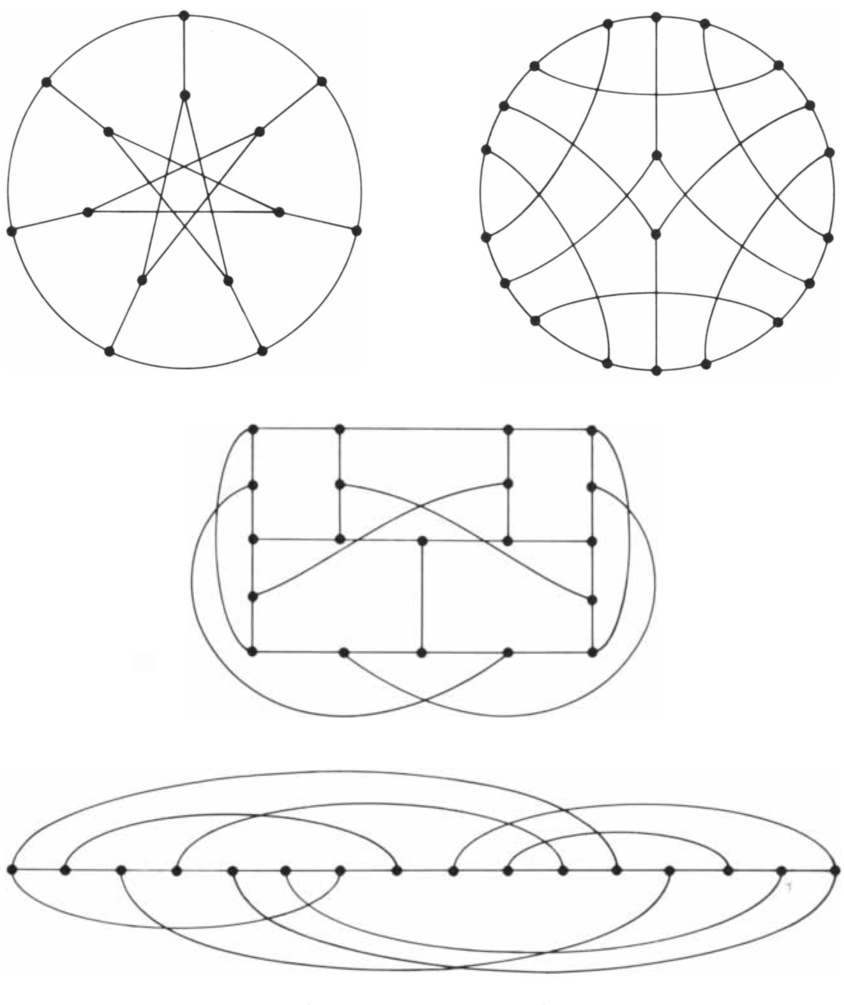
7. Make the alternative choice at step n. This time, however, no bar is put over the label. Why? Because it is no longer a free choice. It was forced by the contradiction that resulted from the previous choice at step n. It gets the subscript $n - 1$. This shows that you were forced to backtrack one step. In other words, the new step has now become part of the previous step, so that it gets the previous step's number.

8. Keep repeating the procedure. If the graph is colorable, you will eventually color it. If it is not, you will keep encountering contradictions that force backtracking. Subscripts will become fewer only to increase again. If there is a contradiction after all edges have acquired labels with no bars, then the graph is uncolorable. You have found a Snark.

"The task is simpler," Isaacs writes in a letter, "the more omniscient we are in foreseeing the consequences of each step." After a while numerous dodges will occur to the experienced Snark hunter. The lower illustration at the left on the next page shows some useful coloring tips. For example, a digon (a two-sided polygon) on a path can be skipped because clearly the same color is forced on both sides. Similarly, a triangle can be treated like a single vertex because the three edges leading to it (as you can easily prove) must have three colors. Squares can be simplified by remembering that either the four edges leading to it are all the same color or two adjacent edges are one color and the other two are another. Pentagons are simplified by remembering that, of the five leading edges, three adjacent edges must be the same, the other two colors being on the remaining pair.

It is good, Isaacs advises, to hunt for forks of the kind depicted in the last figure in the lower illustration at the left on the next page. If two edges are the same color, as is shown, the edge marked with the arrow must also be the same color. Finding forced labels of these kinds can eliminate unnecessary backtracking.

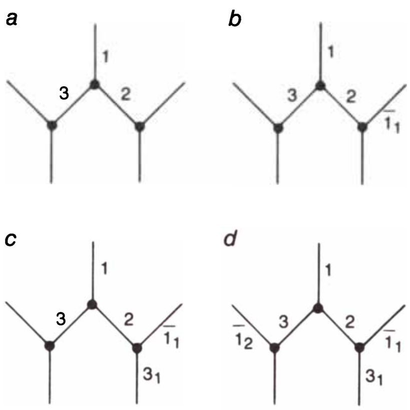
Every known Snark, Isaacs tells us, contains at least one Petersen graph. This means that by erasing certain edges and removing spots from edges that remain, you are left with a structure topologically identical with the Petersen graph. It does not mean that the Petersen graph is a subgraph. Subgraphs have to correspond, point for point, edge for edge, to a portion of a graph. Although subgraphs are contained within graphs, not all contained graphs are subgraphs. Indeed, the Petersen graph cannot



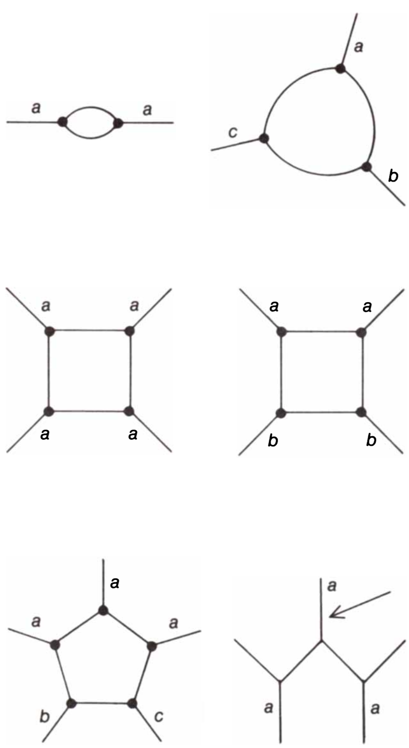
Three-colorable trivalent graphs

be a subgraph of a trivalent graph, because if you add an edge to any vertex of the Petersen graph, it raises the order of the vertex to 4.

The top illustration on this page shows one way to remove five edges (shown as broken lines) and 10 spots from the flower Snark at top right in the top illustration on page 128 to leave a Petersen graph. The graph is numbered and colored to correspond with the graph in the second illustration from the bottom on page 127. The presence of this graph inside the Snark proves that it is not planar, because no planar graph can contain a nonplanar one. There is a famous theorem that states all nonplanar graphs (not necessarily trivalent ones) must contain either the complete graph for five points or the six-point "utilities" graph (see



Rufus Isaacs' coloring procedure



Useful coloring tips

this department for April, 1964). Along similar lines Tutte has conjectured that all Snarks contain Petersen graphs. If that is true, the four-color theorem is also true and Boojums do not exist.

In our Carrollian terminology Petersen graphs are the "bathing machines" that Snarks are so fond of, and that Carroll says every Snark

"... constantly carries about,
And believes that they add to the beauty
of scenes—
A sentiment open to doubt."

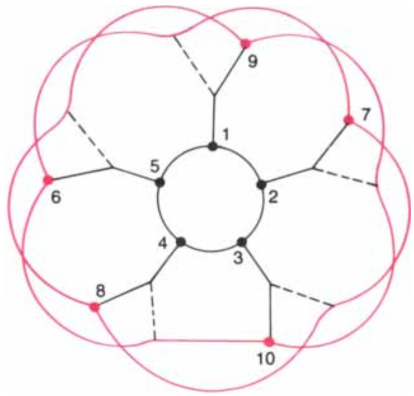
Finding bathing machines inside Snarks is not always easy. You might try looking for one in the other two flower Snarks and inside the double-star. Perhaps you will become sufficiently intrigued to embark on a Snark-hunting expedition of your own. Try drawing and testing a variety of trivalent graphs. Your ability to color them will improve rapidly with practice, and you will become impressed with how difficult it is to find genuine Snarks. If you bag any with more than 10 spots (a Petersen graph) and fewer than 18 (a Blanuša graph), I should like to see their picture. No Snarks are known with spots numbering between 10 and 18. The first flower Snark (12 spots) does not count because the triangle at its center makes it a trivial variation of the Petersen graph.

Hold on a minute! I have just finished sketching a fantastic trivalent graph with 50 projections that stick out like feathers. There are no intersections. It just might be a Boo....

Here are some comments on letters about the five short problems of last December. Many readers pointed out that the numeral 1 is commonly called *eins* in German, not *ein* as it appeared in one of the cryptarithms. William C. Giessen was the first to add that if this change is made, the cryptarithm still has a unique solution: $1329 + 1329 + 1329 + 1329 = 5316$.

Ransom's paradox, in which the "least different" symbol becomes the "most different," was explicitly recognized by Banesh Hoffmann in his devastating book *The Tyranny of Testing* (1964), page 19.

Stephen I. Warshaw of the Lawrence Livermore Laboratory of the University of California generalized the cake-cutting problem. The procedure given for the square cake, he showed, also applies to any cake with a polygonal base each side of which is tangent to the same circle. This includes all triangles and rhombuses, all regular polygons (as well as the limiting case of the circle) and all polygons of unequal sides that meet the proviso. Even more surprising, Warshaw found that this generalization led to the solving of a sticky problem in his work: "How to find a simple way to divide a discrete mesh cell into parts of given proportions in computer simulation studies involving hydrodynamic motions. The solution, once seen, is 'obvious.'"



Bathing machine carried by a flower Snark

Many readers could not comprehend why the third-man chess problem does not have many other solutions, such as a white pawn in place of the white bishop. The reason is that none of these positions rules out the possibility that it is White's move, and if White can move, the position fails to meet the demand that neither side can move. For instance, assume that the black king is on the QB1 cell. It is checked by an advancing white pawn. The black king moves to the corner. Now it is White's move. The given solution is the only one for which it can be shown by retrograde analysis that it must be Black's move, proving that the position is a stalemate for Black.

Concerning Lewis Carroll's poem, G. A. Crum, Ross Eckler, Phyllis and Margaret Gottlieb, Robert Johnson and David Shulman cited *editor* as an anagram for "to ride"—certainly as likely as *rioted* to be what Carroll had in mind. Both Shulman and Johnson added *palters, psalter* and *platers* as anagrams for "let's rap." Among proper names, which Carroll surely meant to exclude, readers mentioned *Wagner, Rodney, Yeats* and *Yates,* and of course Carroll would not have intended such words as *triode* and *goofer.*

In February I reported the finding of two perfect magic cubes of order 7. Before the report appeared readers too numerous to mention succeeded in constructing order-7 cubes. Other readers sent magic cubes, along with construction procedures, for order 9 and higher, as well as magic cubes in dimensions higher than three.

From time to time games that were first introduced in this department have been featured in science fiction. Sprouts, for example, is played on several occasions in Piers Anthony's *Macroscope* (Avon, 1972), and now pentominoes, with due acknowledgment to their creator, Solomon W. Golomb, are mentioned throughout Arthur C. Clarke's new novel *Imperial Earth* (Harcourt, 1976). The 12 tantalizing pieces are an integral part of the plot, symbolizing on one level the combinatorial richness of the DNA molecule and on a higher level the combinatorial possibilities of Everything.

BOOKS

The importance of honey; early concepts of probability, and animal architecture

by Philip Morrison

HONEY: A COMPREHENSIVE SURVEY, edited by Eva Crane. Crane, Russak & Company, Inc. ($42.50). Stolen sweets are held to be the best, and honey is twice-stolen: "taken by the bee from the flower... and then stolen from the bees by their keepers." The bee pays in part with the service of pollination, which its visit generally ensures; the keeper too may pay, compensating his bees with shelter and with surrogate sugar. These relations are nothing new, since honey-storing social bees have been around since the Miocene, for 10 or 20 million years, and chimpanzees and baboons (not to mention black bears and honey badgers) have long fed greedily on that stored sweetness. In this thick, well-illustrated book you can see representatives of bee husbandry from painted cave wall (late Paleolithic) through the ancient empires to today.

One pre-World-War-II ethnographer found Paraguayan Indians whose staple diet was honey robbed from wild bees, but for the most part this volume, the joint product of more than a dozen specialists, deals with honey from the domesticated hived species: the European honeybee *Apis mellifera* and its close relative in the Far East, *Apis cerana*. The stingless tropical bees of the New World produce a much smaller surplus, an order of magnitude less than that of today's honeybee. Nevertheless, Aztec and Maya beekeepers were a skillful and productive group; "the most important medicinal and tonic beverage" for the Aztecs was a preparation of powdered cacao, vanilla, pepper and honey. (Keeping stingless bees, by the way, is no painless idyll: they can bite or secrete a fluid that burns the skin.)

The key to domestication is the honeybee's habit (shared by the European and the Asian species) of building combs within dark enclosures. By offering bees hives made of logs, pots or baskets to rival the hollow tree or cliff-wall cavity of nature, Neolithic people found they could bring the bees home rather than search the forest for bee trees. Sweetness was certainly part of the spur but not all of it. It is pretty clear that it was the fermentation of honey sugars by natural yeasts (when the honey was diluted in water) that was the major attraction to early users. Mead may be the oldest alcoholic drink, predating the fermented fruits of vine or grain, certainly among the peoples who speak Indo-European languages. Mead was the favorite alcoholic drink of England and other northern European lands until the end of medieval times. Similar drinks were found in pre-Columbian America, and to this day four-fifths of the honey used in Africa south of the prohibitions of Islam goes to make a long drink, honey beer, "the best of all beers"—and the more honey and the less commercial sucrose, the better. (Benedictine and Drambuie simply continue a high tradition.)

The honey market in a place such as Addis Ababa is filled with goatskin sacks, each holding 50 kilograms or so of broken honeycombs, roughly harvested from crude hives. The modern winning of honey involves a more subtle hive management that need not much disturb the bees. The capping invention was made in the 1850's by a Philadelphia clergyman, L. L. Langstroth. It amounts to the provision of light, interchangeable frames on which the bees can build and fill their combs. The bees are then fully controlled; they can be sustained by replacing their surplus with sugar in seasons of dearth, or induced to collect prodigies of nectar during the brief, intense subarctic flowering season or the two-month flow of *Eucalyptus platypus*. A colony patrols a region up to a mile in radius (about as far as the sound of a church bell carries), but some truck-borne apiarists migrate with their bees 1,000 miles to follow the blossoming. There are 50 million honeybee colonies in the world, so that the world honeybee population is about 500 bees per person. Reckon for a year about 10 kilograms of surplus honey per colony, with several times that much for the bees' own use. The worker honeybee offers a paradigm of fuel economy: she gets some seven million miles per gallon of high-test.

The source of most honey is the dilute sap of higher plants, usually from the special nectar glands of flowers, sometimes from cut stems (as of sugarcane) or in the secretions of sucking insects. In the sap itself the sugar is chiefly sucrose; in the nectar the simpler sugars are common, although the composition differs strongly from species to species. Honey in the comb is thoroughly hydrolyzed, and it contains mainly the simple invert sugars fructose and glucose. The ripening of honey involves a chain of transfers from bee to bee, during which most of the enzymatic action takes place. The bees concentrate the watery nectar by repeatedly exuding and taking back their loads of nectar in tiny drops, the film losing water in the warm, dry nest. At about 20 percent water content the honey is ripe and is sealed into the comb.

The comb honey, fragile and beautiful, is increasingly popular; there is much to say about the task of shipping it from hive to home. Most table honey (which is the predominant use of the sweet—tobacco, graham crackers, bread or *baklava* notwithstanding) is centrifugally spun out of the cut comb, strained, rapidly heated and cooled in order to kill yeasts and remove the nuclei that would speed crystallization, blended and packed. Finely granulated honey is the most common form except in the U.S. and Australia; in its preparation a large inoculum of fine crystals is added. The starter crystals grow, but they are so numerous that no one of them can grow large, and the desired product results.

The largest producers of honey are the U.S. and the U.S.S.R., wide lands with many modern frame-hive beekeepers. Europe is saturated; its overall yield is about that of the U.S. or the U.S.S.R., but there are seven times as many bee colonies per square kilometer in Europe. It is to Europe (particularly to Germany) that the honey of world trade flows, although much also goes to Japan, where the advent of the American breakfast has created a nation of new buyers of mild white honey. Argentina, and recently China, supply that demand. Mexico, the fastest-growing of all exporters, sends out a yellow honey from yellow blossoms of a composite in Cuernavaca and also one of the world's most aromatic honeys, drawn from the greenish flowers of the shrub the Maya called *dzidzilché*. The New World has accepted the honeybee. The People's Republic of China now exports in appreciable amounts a honey that is the take of *A. cerana*. Not much is known about the small differences in the honey of the two domesticated species.

The microscopy of pollen, Latvian terms for wax and mead (much like the Sanskrit), the half-lives of enzymes, long lists of flowering plants and their nectar-yield potential, the viscosity of honey by flower source—what is not included in this comprehensive work? Bees and beekeeping, to be sure, are here only in the context of honey; they are quite another story. The editor has produced a remarkably readable yet authoritative and technical work; she has traveled from Zululand to the Peace River to accomplish it.

There is a chapter on the biomedical properties of honey. Its first draft was written by the late Professor M. H. Haydak, who himself subsisted for months on milk and honey alone. With a vitamin supplement that diet can supply the sole food for adult men. Haydak and the expert colleagues who completed his draft endorse King Solomon's advice: "Eat honey, my son, for it is good." They see it entirely as a source of sugars that are rapidly absorbed without digestive effort. (Its many other constituents are present in such small amounts compared with those in a normal diet that it is hard to regard honey as useful for anything but its sugars.) It is bactericidal and safe to use as a wound dressing (if not quite as good as leaving the wound alone, without any dressing at all). There is no support here for a honey mystique, although the universal sucrose craze is no friend to public health.

The endpapers display a 17th-century woodcut. It shows the bees, their diverse flowers—rose and thistle—and their hive, a scholar and a cleric eating a piece of a honeycomb. The legend reads: "All plants yield honey as you see / To the Industrious Chymick Bee." Under the rubric "Everything is to be found in books" the scholar gathers knowledge from a hundred flowers blooming. The bee-devoted authors of this volume have happily depicted themselves!

THE EMERGENCE OF PROBABILITY: A PHILOSOPHICAL STUDY OF EARLY IDEAS ABOUT PROBABILITY, INDUCTION AND STATISTICAL INFERENCE, by Ian Hacking. Cambridge University Press ($15.95). The forest king Rtuparna amazes his charioteer, a gambler-king in exile, by examining a single twig and concluding that the great tree bears just 2,095 fruits. His companion confirms the result after an all-night count, one by one. Rtuparna expected no other outcome, because, he reminds his friend,

I of dice possess the science
and in numbers thus am skilled.

So runs a passage of the *Mahábarata,* the Sanskrit epic completed before the fall of Rome. This subtle anticipation of probability, with its remarkably modern implication that dicing and sampling are related, is unmatched in the West. Gaming, to be sure, is old enough. The author himself has used "exquisitely balanced" dice from ancient Egypt "which the guards kindly allowed me to roll for a long afternoon."

So runs the opening chapter of this compact, felicitous book, an essay at once meticulous enough about the evidence to be historical and sufficiently probing about the concepts it studies to be philosophical. Its presentation is enlivened with candor and is disarmingly explicit about both methods and assumptions. The concept whose origins it teases out of a subtle tissue of thought is of course probability, that modern touchstone of science and philosophy. For Professor Hacking the concept is inescapably dual, one aspect aleatory, relating to games and combinatories, and the other epistemological, "dedicated to assessing reasonable degrees of belief" even where no statistical trials are anywhere in sight.

The powerful godmother, duality, was there at birth. By 1660 the modern conception was in place; 50 years earlier it did not exist, at least as a feature of learned public discourse. By common consent the history of mathematics—since Leibniz himself—has recognized in Pascal that "austere Jansenist" who founded the theory to answer the practical questions of sporting gentlemen. So much is granted, but the theme of this discussion is built on the fact that the same Pascal, in two sheets scratched with writing in all directions, full of erasures and afterthoughts, proposed an awesome wager. Headed *Infini-rien,* the famous statement argues that one ought rationally to believe in God, since the expectation of the pious from eternal salvation exceeds that of the damned libertine, however small the chance that God exists. Even two generations later Voltaire himself is a puzzled witness to the originality of Pascal's method, whatever the force of his argument. "This article seems a little indecent and puerile: the idea of a game, and of loss and gain, does not befit the gravity of the subject." Today we fully accept games as models of decision under conditions of uncertainty. Probability was born dual, and dual it lives.

Hacking starts with St. Thomas, who held the view that real knowledge was of necessity from first principles. It was within the "low" empirical sciences—alchemy, medicine, mining—that novelty arose. They could not work so grandly and they could not rely on authoritative witness or plausible discourse; their evidence took the form of natural signs. Hobbes knew that. Writing before Pascal, he said: "If the signs hit twenty times for one missing, a man may lay a wager of twenty to one of the event; but may not conclude it for a truth."

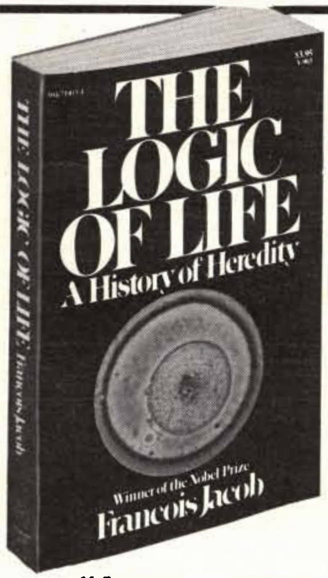
The roles of Leibniz and Huygens are better known and here become central. Leibniz is shown to be close to the origin not merely of the combinatorial mathematics of the aleatory world but even more plainly of a "new kind of logic." For Professor Hacking, Leibniz' views are about as good as any metaphysics since his time, even though Leibniz was prisoner to the high belief that true knowledge must be essentially demonstrative. He hoped, as many do today, for a statistical logic that could grant not certainty but only the decision as to what conclusion was most probable in the light of the information.

We encounter at some length Jacques Bernoulli; the probability concept emerged fully in the posthumous publication of his *Ars conjectandi* in 1713. The central limit theorem is there, well proved, but what it means remains to this day less clear. Already probability had its "mathematical profundity, its unbounded practical applications, its squirming duality and its constant invitation for philosophizing." Now we briefly encounter the moderns, Jerzy Neyman, Bruno de Finetti, Rudolf Carnap, L. J. Savage, J. M. Keynes and the rest. Every school still claims Bernoulli; from its birth this most powerful of evidential doctrines has had a fruitful inner conflict.

The historical stage is crowded. There is the "gambling scholar" Girolamo Cardano, anticipating by a century the frequency concept in a work that was not published until after Pascal. There are Dutch actuaries pricing state annuities at a losing rate, there is a coach to Poitou making fellow travelers of gamblers and noble scholars, there are Prussian law courts and London bills of mortality. The later chapters tend to the philosophical: Hume's "bombshell" of a root-and-branch attack on induction, the role of God's design in the stability of such frequencies as the male-female birth ratio, Berkeley on the inadequacy of a science that found not necessary connections but mere lawlike regularities. The general reader will not always grasp these niceties of metaphysics, but the argument is fairly put and rich matter for reflection. It seems that Newton's triumph without any mechanical hypotheses paradoxically threatened the ancient role of First Cause, and itself entailed that Herr Gott who does throw dice. Another such study, we can hope, will elucidate the role of the later physicists of atomic chance, from Boltzmann and Gibbs to Born and Heisenberg; there are allusions to Professor Hacking's interest in those matters.

The bibliography is up to date for both

A 17th-century woodcut from **Honey: A Comprehensive Survey**



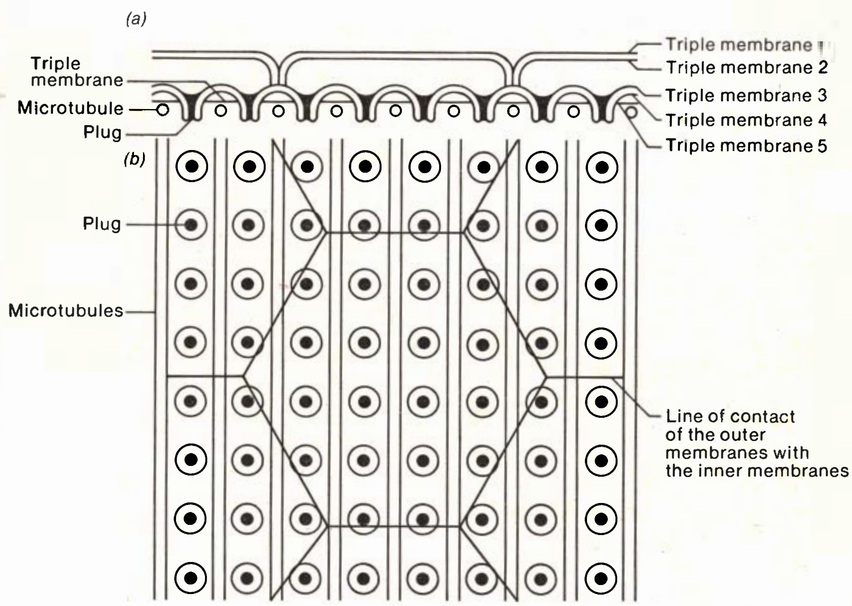
books and periodicals, with an unusually honest apology for its scamping of Russian material; moreover, it "tries to include all published works on probability between 1654 and 1700," helpfully annotated.

STRUCTURAL MATERIALS IN ANIMALS, by C. H. Brown. John Wiley & Sons ($32.50). Airframe, ship's hull or house, the structure supports and protects every complex internal system against external forces. So it is with animals: every organism is built as a mechanical structure, and the chemical-physiological subtleties of information transfer or free-energy use are all maintained, and sometimes constrained, by that easily overlooked essential. On the eve of World War II the work of the late Mark Pryor on the hardening of the cockroach cuticle opened a renaissance in this domain of biology, a generation of work mapped here for students and all biologists. The book is a zoologist's survey, organized systematically, taking animals phylum by phylum and class by class, which makes it convenient for reference but exacts the price of much reduced appeal to the general reader. Yet the detailed material is fascinating.

As everywhere in biology, a few inventions are used over and over. Human engineers rely on wood and steel, nuts and bolts, frames and skins in most of their construction. Composite materials are the rule in living structures. Long fibers of polymers provide tension elements, sometimes aligned for directed stresses, as in tendon, and sometimes oriented at random for strength in all directions. The long chitin chain, common to fungi and invertebrates but never found in vertebrates, is an amine-substituted form of cellulose, the plants' glucose polymer. The ropy protein collagen is found in nearly all multicellular animals from sponge to mammal. Even the soft jellyfish depend on it, although their "surprisingly strong" internal substance has only 1 percent of organic matter in a saltwater gel! Hair, feather, nail and horn are keratin, another protein-fiber helix. Keratin, however, remains intracellular; it does not form a continuous network in those strong structures, whose good tensile properties depend on the protein mortar between the keratinized bricks.

The compression-resistant matrix of such materials is also diverse. In the arthropod cuticle the chitin fibers are bound to a protein matrix; in cartilage (found very widely) the matrix is one of a number of mucopolysaccharides, crudely analogous to chitin but not organized as fibers. There are special spring materials in legs, wings, hearts and lungs, low-loss and very elastic (some withstand threefold elongation): protein rubbers of high quality, not yet understood in microdetail. Inorganic crystals are of course the inexpensive hardening device of backbone and clamshell. (Lion whiskers contain calcium phosphate, just as lobster shells contain calcite.) The hardening of insect cuticles is best secured not by adding inorganics but by specialized covalent cross-linking: the protein chains of the outer layers (not the chitin) are connected by a linking ring, usually a quinone. This action is called tanning, since it is what happens in the commercial preparation of leather, when the collagen fibers of fresh animal skins are linked for increased mechanical and chemical stability.

Such quinone-linked proteins, called sclerotins, are usually colored brown or black, accounting for the hard beetle look we all know. Special developmental biochemistry does this trick so well that entirely organic sclerotins can be hard enough to scratch calcite. As for chemical stability, the beachcomber will certify how often he encounters crab shells and dogfish egg cases, both surfaced by quinone-tanned pro-



The surface structure of a dinoflagellate, from **Structural Materials in Animals** teins. They do not seem to be biodegradable by any enzymatic reaction, although they are soon reduced by abrasion to small particles that survive only microscopically or that gradually oxidize through inorganic processes.

Form by form, there is a lot we do not know. D'Arcy Thompson held that "a hollow shell is admirable for small animals but Nature does not and cannot make use of it for large." Yet the Carboniferous period saw "centipedes" five feet long running on dry land, and certain fossil marine arthropods were up to nine feet in length. Whatever has kept arthropods small, "it has not been the mechanical limitations of the exocuticle." (Nor was it the diffusion limit of oxygen transport, since insects do pump air through their tracheae by abdominal movement.) The hollow form puts the strength outside; if the femur were organized in this way, it would be seven times stronger. The stiff material outside then becomes vulnerable to impact load, which may be a significant consideration.

The vertebrates may have grown larger simply because they could use a large size better, with more control built in. Then the invertebrates, with less computer power (mostly mere servo controls), headed for those niches open to smaller sizes. The largest invertebrates, the cephalopods, have indeed sunk the shell deep into the body, developed a cartilaginous endoskeleton and for complex behavior now "demand serious comparison" with vertebrates. This convergence of design suggests that internal armature, and not a hollow shell, is the right design for the thinking animal—until one recalls the hollow protective cranium!

This book incorporates the literature very widely up to about 1970; later references are listed, but they do not enter the detailed explanatory text.

The Quest for Ulysses, by W. B. Stanford and J. V. Luce. Praeger Publishers ($17.50). Works of Poussin and Turner and Chagall are reproduced here in a book as close to art history as a scientific reviewer can presume to go. So, however, are the products of goldsmiths of Minoan and of Mycenaean time, of a great muralist of Imperial Rome and of the painter of a red-figured Attic vase, together with a photograph of the "cave of the Nymphs" on Ithaca. That cave fits in many ways the Homeric description of the secret place to which the Wanderer returned to stow his treasures amid sheets of stalactite, "great looms of stone where the Nymphs weave marvelous fabrics of sea-purple," as the epic tells us.

In this deeply instructive and beautiful book, two classics professors from Trinity College, Dublin—the city that in this century most of all gave another home to Ulysses—carry off a bold and moving feat. They seek out Ulysses twice, before Homer and after. First they trace his origin. He is the provincial "Achaean prince of Western Greece [who] won renown among the commanders in the Trojan War." Five hundred

years later, out of the embellishments of folklore and the tales of seamen and colonists, the great poet Homer "created a complex and subtle character," fusing fact and legend with the imagination of an artist hardly matched over the millenniums to make a personality still alive in modern image. He remains *polytropos,* the man of many turns, as Homer says in the very first line of the epic. Then the authors follow him as image and idea through Ovid and Dante to Matisse and Kazantzakis.

Stanford the scholar speaks first of the old traditions, the gloss on Homer. In western Greece the hero is called Olixes or Ulixes, which has no obvious etymology in Greek, but Homer always uses the eastern Odysseus, probably meaning "child of wrath." The name is dual, like the man. The western name probably goes back to some pre-Greek trickster of folklore; the eastern one, with its aristocratic form and solemn import, is the work of the poets.

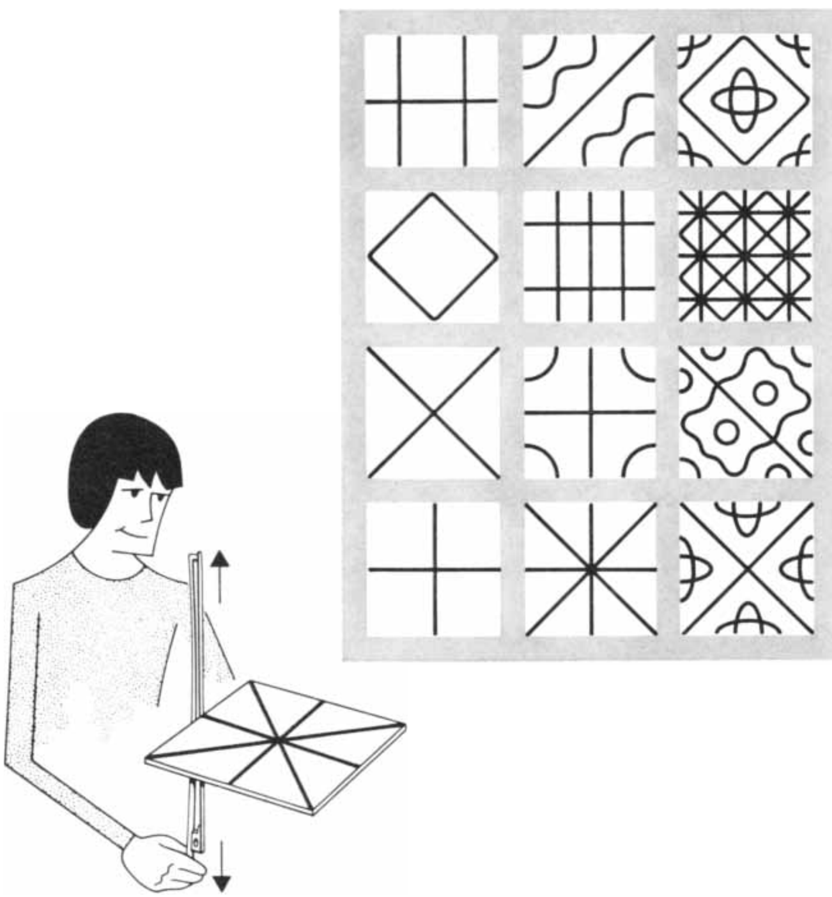
Luce, the more archaeological partner, tells of the find of what seems really to be Nestor's palace in Pylos, of texts in Linear *B,* of the island Ithaca and its present scenes, of fiery Stromboli in the Lipari Islands. He steers a persuasive middle course along the voyages of the Ithacan. No less a personage than Eratosthenes once held: "You will find the course of Odysseus' wanderings when you find the cobbler who sewed up the bag of the winds." One can, like old Strabo, be less skeptical: "These traditions must not be pressed to yield accurate information, but neither are they to be dismissed as baseless and ungrounded." For Luce the travels mirror real Mycenaean Bronze Age voyages, from Troy on the coast of Anatolia westward beyond the Liparis to the west coast of Italy. Circe and the Cyclops Polyphemus are not themselves to be found on any maps, yet the culture of the giants is described factually as being pastoral. The marvelous tale of the Cyclops mixes the universal folklore motive of an awful ogre outwitted by cunning with a true picture of an upland cave-dwelling people of Italy in the same period. We know of such a place, where metal tools were lacking but where one hoard of bronze swords has been excavated. Was it the loot from an Achaean ship? Does the story commemorate some actual ship that once put in to the beach, to be overwhelmed by rude, strong tribesmen of the caves? The fabric is equally intricate in most of the voyages.

The duality is strong; it runs to us through the art of all Europe, ancient, medieval and modern. Home-seeker and home-deserter, rogue and hero, liar and philosopher, Odysseus, as Kazantzakis says, "goes on ceaselessly, his neck stretched forward like the leader of birds migrating."

THE FLYING CIRCUS OF PHYSICS, by Jearl Walker. John Wiley & Sons ($5.50). "What I mainly wanted to show here is that physics is not something that has to be done in a physics building." So writes the author, a recent Ph.D. and a member of the faculty at Cleveland State University. He has accomplished his aim, one already acceptable to readers of the wonderful series of columns "The Amateur Scientist," which were displayed for many years in this magazine by their impresario and proponent, the late C. L. Stong. Red Stong would have liked this rich but light-hearted book. It is not as demanding as most of the clever instruments of Red's delight, but it is a beginning toward them. Boxed among many sketches, some whimsical drawings (by Ron Markman) and even cartoons from the comic strip "B.C." stand about 600 questioning paragraphs, grouped into seven chapters by topic. Most of them are questions for verbal—or calculational—reply; a few of them induce simple experiments, and many will generate amusement. They range from the familiar "What causes the ocean roar that you hear in a seashell?" to the incredible "Should an earthquake produce lightning discharges?" Answers are not given; in their place there is a splendid keyed list of references, more than 1,600 of them, where an answer or at least a serious conjecture can usually be found. Sometimes the very title of the reference is a hint; for example, the paper cited for the second question above is titled "The Piezoelectric Theory of Earthquake Lightning." Believe it or not. No physics teacher—amateur or pro—can afford to miss this book, and it should be a stimulus to libraries for a long time to come; indeed, its one requirement is a library. The references include the scientific weeklies, *Scientific American,* teaching journals in chemistry and physics, admirable treatises and texts in many branches of physical science, many major professional journals, works of Enrico Fermi and G. I. Taylor and Lord Kelvin and David Hilbert and a thousand other less renowned but fascinated physicists, biologists, earth scientists and puzzlers.

Falling chimneys, singing wineglasses, winds whistling through the open ground floor of an M.I.T. building, maps of ice fields seen in the Arctic sky, fixed-point theorems in coffee-stirring, a magnetized dollar bill, yo-yos and plumes of shrimps in streams serve as a random sample of this circus of a hundred rings. (Don't open it if you cannot tolerate deferred resolution.)

Why can pure reds be found only in the vertical portions of rainbows? "Fraser points out that this simple feature... escaped notice until his paper of 1972. As another example of modern work, it has only been recently that photographs of the infrared rainbow have been taken, thus allowing man to see for the first time what has ... hung in the sky for millions of years."



Chladni figures **(upper right)** *obtained by bowing a plate, from* **The Flying Circus of Physics**

BIBLIOGRAPHY

Readers interested in further reading on the subjects covered by articles in this issue may find the lists below helpful.

THE SCIENCE-TEXTBOOK CONTROVERSIES

The Ecology of the Public Schools: An Inquiry into Community Control. Leonard J. Fein. Pegasus, 1971.

Religious Movements in Contemporary America. Edited by Irving I. Zaretsky and Mark P. Leone. Princeton University Press, 1974.

The Monkey Laws and the Public Schools: A Second Consumption? Frederic S. Le Clercq in *Vanderbilt Law Review,* Vol. 27, No. 2, pages 209–242; March, 1974.

Science or Scripture: The Politics of Equal Time. Dorothy Nelkin in *Science and Its Public: The Changing Relationship,* edited by Gerald Holton and William Blanpied. D. Reidel Publishing Company, 1976.

THE SENSING OF CHEMICALS BY BACTERIA

Chemoreceptors in Bacteria. Julius Adler in *Science,* Vol. 166, No. 3913, pages 1588–1597; December 26, 1969.

Chemotaxis as a Model for Sensory Systems. Daniel E. Koshland, Jr., in *Federation of European Biochemical Societies Letters,* Vol. 40, Supplement, pages S3–S9; March 23, 1974.

Chemotaxis in Bacteria. Julius Adler in *Annual Review of Biochemistry,* Vol. 44, pages 341–356; 1975.

Chemotaxis in Bacteria. Howard C. Berg in *Annual Review of Biophysics and Bioengineering,* Vol. 4, pages 119–136; 1975.

SUBJECTIVE CONTOURS

Stereopsis and Anomalous Contour. R. B. Lawson and W. L. Gulick in *Vision Research,* Vol. 7, pages 271–297; March, 1967.

Selective History of Theories of Visual Perception: 1650–1950. Nicholas Pastore. Oxford University Press, 1971.

Untersuchungen zum Problem virtueller Konturen in der visuellen Wahrnehmung. Michael Stadler and Jürgen Dieker in *Zeitschrift für experimentelle und angewandte Psychologie,* Vol. 19, No. 2, pages 325–350; 1972.

Subjective Contours and Apparent Depth. Stanley Coren in *Psychological Review,* Vol. 79, No. 4, pages 359–367; July, 1972.

Cognitive Contours. R. L. Gregory in *Nature,* Vol. 238, No. 5358, pages 51–52; July 7, 1972.

Contours without Gradients or Cognitive Contours? Gaetano Kanizsa in *Italian Journal of Psychology,* Vol. 1, No. 1, pages 93–112; April, 1974.

CATASTROPHE THEORY

Primary and Secondary Waves in Developmental Biology. E. C. Zeeman in *Lectures on Mathematics in the Life Sciences, Vol. 7: Some Mathematical Questions in Biology,* edited by Simon A. Levin. American Mathematical Society, 1974.

Catastrophe Theory: Its Present State and Future Perspectives. R. Thom and E. C. Zeeman in *Dynamical Systems—Warwick 1974: Proceedings of a Symposium Held at the University of Warwick 1973/74,* edited by Anthony Manning. Springer-Verlag, 1975.

Structural Stability and Morphogenesis. René Thom, translated by D. H. Fowler. W. A. Benjamin, Inc., 1975.

The Classification of Elementary Catastrophes of Codimension ≤ 5. D. J. A. Trotman and E. C. Zeeman in *Symposium on Catastrophe Theory; Seattle 1976.* Springer-Verlag, in press.

OPALS

The Opal Book. Frank Leechman. Ure Smith, 1961.

The Book of Opals. Wilfred C. Eyles. Charles E. Tuttle Company, Inc., 1964.

The Nature and Origin of Opal. P. J. Darragh and A. J. Gaskin in *Australian Gemmologist,* No. 66, pages 5–9; 1966.

Diffraction of Light by Opals. J. V. Sanders in *Acta Crystallographica, Section A,* Vol. A24, Part 4, pages 427–434; July 1, 1968.

U.S. Patent 3,497,367: Opaline Materials and Method of Preparation. Peter J. Darragh and Arthur J. Gaskin. U.S. Government Patent Office, February 24, 1970.

Australian Precious Opal. Archie Kalokerinos. Thomas Nelson & Sons (Canada) Ltd., 1971.

Notes on Synthetic Precious Opal. P. J. Darragh and J. L. Perdrix in *The Journal of Gemmology and Proceedings of the Gemmological Association of Great Britain,* Vol. 14, No. 5, pages 215–223; January, 1975.

THE ANALYSIS OF MATERIALS BY X-RAY ABSORPTION

New Technique for Investigating Noncrystalline Structures: Fourier Analysis of the Extended X-Ray-Absorption Fine Structure. Dale E. Sayers, Edward A. Stern and Farrel W. Lytle in *Physical Review Letters,* Vol. 27, No. 18, pages 1204–1207; November 1, 1971.

X-Ray Absorption Spectra. Leonid V. Azaroff and Douglas M. Pease in *X-Ray Spectroscopy,* edited by Leonid V. Azaroff. McGraw-Hill Book Company, 1974.

Theory of the Extended X-Ray-Absorption Fine Structure. Edward A. Stern in *Physical Review B,* Vol. 10, No. 8, pages 3027–3037; October 15, 1974.

Determination of Atomic Structure in Non-Crystalline Biological Systems. Edward A. Stern, Dale E. Sayers and Farrel W. Lytle in *Collective Phenomena and the Applications of Physics to Other Fields of Science, Prepared for Delivery at a Seminar, Moscow, USSR, 1–5 July, 1974,* edited by Norman A. Chigier and Edward A. Stern. Brain Research Publications, Inc., 1975.

New Method to Measure Structural Disorder: Application to GeO_2 Glass. D. E. Sayers, E. A. Stern and F. W. Lytle in *Physical Review Letters,* Vol. 35, No. 9, pages 584–587; September 1, 1975.

GALILEO AND THE FIRST MECHANICAL COMPUTING DEVICE

The Instruments of Galileo. Silvio A. Bedini in *Galileo: Man of Science,* edited by Ernan McMullin. Basic Books, Inc., 1967.

Scientific Instruments of the Seventeenth and Eighteenth Centuries. Maurice Dumas, translated and edited by Mary Holbrook. Praeger Publishers, 1972.

THE NILE CROCODILE

The Nile Crocodile, *Crocodilus niloticus:* Notes on the Incubation Period and Growth Rate of Juveniles. A. C. Pooley in *The Lammergeyer,* Vol. 2, No. 1, pages 1–55; February, 1962.

Diving Bradycardia and Withdrawal Bradycardia in *Caiman crocodilus.* Abbot S. Gaunt and Carl Gans in *Nature,* Vol. 223, No. 5202, pages 207–208; July 12, 1969.

How Does a Baby Crocodile Get to Water? A. C. Pooley in *African Wildlife,* Vol. 28, No. 4, pages 8–11; Summer, 1974.

MATHEMATICAL GAMES

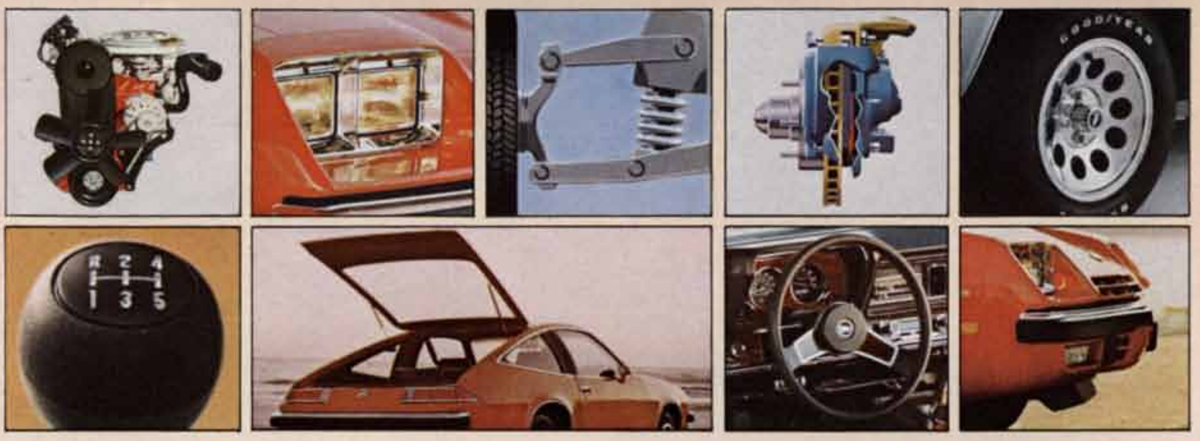
The Four-Color Problem, 1840–1890. H. S. M. Coxeter in *The Mathematics Teacher,* Vol. 52, No. 4, pages 283–289; April, 1959.

The Four-Color Map Theorem in *Martin Gardner's New Mathematical Diversions from Scientific American.* Martin Gardner. Simon and Schuster, 1966.

Thirteen Colorful Variations on Guthrie's Four-Color Conjecture. Thomas L. Saaty in *The American Mathematical Monthly,* Vol. 79, No. 1, pages 2–43; January, 1972.

MONZA 2+2

A small car and then some.



Monza 2+2 offers all the standard equipment and forward-looking options you'll ever need to make it all the road car you'll ever want. *Affordably*.

Monza offers a variety of engines and transmissions. Engines range from the Dura-Built 2.3 Litre 4-cylinder to a GT-class 5-litre V8. There's a choice of three manual transmissions including a smooth-shifting 5-speed gearbox. Turbo Hydra-matic transmission is available with all Monza engines.

Every Monza engine has High Energy Ignition, a system that aids starting, engine performance and plug life.

Outside, advanced styling and sleek aerodynamics let a Monza 2+2 turn heads as easily as it turns corners. New square headlights are set into a resilient front end that bounces back from minor bumps, and there's an easy-opening hatchback.

Underneath, each wheel has computer-selected coil springs. Front disc brakes self-adjust and the torque-arm rear suspension smooths the ride. All this is standard equipment that makes Monza more exciting.

Available options that will help to make your Monza 2+2 even more of a driver's car include: power-assist brakes, variable-ratio power steering, stiffer sport suspension, steel-belted radial tires, Positraction rear end and Spyder equipment. Available aluminum wheels, rear spoiler and front air dam add to Monza 2+2's competition look.

The value-adding equipment list also includes Four-Season air conditioning, AM/FM radios, stereo tape decks, a choice of luxury interiors and much more.

All this tells you that your Monza 2+2 can unquestionably be a lot more small car. Just how much more is up to you.

Chevrolet

DEWAR'S® PROFILES

(Pronounced Do-ers "White Label")

BLENDED SCOTCH WHISKY • 86.8 PROOF • ©SCHENLEY IMPORTS CO., N.Y., N.Y.

PAUL SZEP

HOME: Scituate, Massachusetts

AGE: 33

PROFESSION: Nationally syndicated editorial cartoonist

HOBBIES: Lithography, golf, classical music

MOST MEMORABLE BOOK: Low's Autobiography

LAST ACCOMPLISHMENT: Won the 1974 Pulitzer Prize for editorial cartoons.

QUOTE: "The satirist is an extremely important person in a society that takes itself far too seriously and is incapable of laughing at its own frailties and insignificance."

PROFILE: Relaxed and easygoing, but very sensitive to social injustice. In his work, he reduces complex events to what concerns him most, their effect on ordinary people.

SCOTCH: Dewar's "White Label"®

Authentic. There are more than a thousand ways to blend whiskies in Scotland, but few are authentic enough for Dewar's "White Label." The quality standards established in 1846 have never varied. Into each drop go only the finest whiskies from the Highlands, the Lowlands, the Hebrides.

Dewar's never varies.