SCIENTIFIC AMERICAN

NONVISUAL LIGHT RECEPTION

ONE DOLLAR

March 1972

The Gould Early Warning System.

It tells the nurse when a newborn baby has trouble breathing.

Newborn babies need all the attention they can get. Especially those few who show some evidence of possible respiratory problems. If mucus stuffs up a baby's nose, it can cause suffocation. And if the doctor can have an accurate and continuous record of how the infant is breathing, this information can be very helpful in making a diagnosis and prescribing the necessary treatment.

Now a baby's breath can be monitored.

The Proud Inventors at Gould have developed an Early Warning System. It constantly monitors and records a baby's delicate breath, and sounds an alarm should breathing stop. So a nurse can quickly revive him and clear the problem.

Right now, biomedical equipment like this is a growing part of Gould's business. And it's important for two reasons.

Every hospital has patients that need monitoring.

As the use of this equipment expands, so will the size of our market. Our niche in the patient monitoring business is creating sophisticated instrumentation equipment that's easier to use and more economical. So more medium and small hospitals can take advantage of it.

Our work in biomedical equipment is also important because it lets our Proud Inventors be involved on the threshold of medical breakthroughs. These opportunities to help meet socially relevant challenges add a dimension of satisfaction and motivation that carries through in their work on all our product lines.

You'll be seeing the Gould name on a lot of new things.

Besides our biomedical equipment, we've also invented a low cost rechargeable battery for small appliances, and a more efficient heating element for forced air systems, and a battery power system for lawn and garden vehicles.

Now some of the things our Proud Inventors are working on are an energy absorbing bumper system for your car and a solution to automotive emissions. You'll be hearing more from us. Gould Inc., 8550 West Bryn Mawr Avenue, Chicago, Illinois 60631.

GOULD

the proud inventors

Once maps were made by hand.

But why today?

Once, a man told another of what he'd seen and that man drew a map that all others could follow.

All of that was done by hand. That was then.

Today, a man takes a picture from an airplane of what he sees. And a second man prepares a manuscript from these photos. And then, this manuscript is transferred to film.

And then—incredibly—all of the lines that will make up the map (the rivers, the mountains, roads and streets) are *scribed* onto a negative master. By hand.

Finally, a swivel knife is used to cut outlines of specified areas. By hand. In the seventies of the twentieth century.

Someone doesn't trust someone.

We, CalComp, have told cartographers that our 745 flatbed plotter will scribe lines equal to the tolerances and standards of the most skilled mapmaker's hand.

Cartographers have told us that they tried plotters once. And the lines were not accurate. And they were uneven. And wiggly.

The CalComp 745 Plotter is accurate to a rate of plus or minus .001 inch. The lines it scribes or cuts are smooth and even. (Their step size is only .0001 inch.)

And our 745 plots at a speed of 4.2 inches per second. No hand alive can do that accurately.

If you make maps by hand, call us. We'll help you get from here to there.

Write us at California Computer Products, Inc., Dept. SM-M2-72 , 2411 West La Palma Avenue, Anaheim, California 92801. Or call (714) 821-2011.

CALCOMP

Established 1845 **SCIENTIFIC AMERICAN** March 1972 Volume 226 Number 3

ARTICLES

DEPARTMENTS

PUBLISHED MONTHLY BY SCIENTIFIC AMERICAN, INC., 415 MADISON AVENUE, NEW YORK, N.Y. 10017. SECOND-CLASS POSTAGE PAID AT NEW YORK, N.Y., AND AT ADDITIONAL MAILING OFFICES. AUTHORIZED AS SECOND-CLASS MAIL BY THE POST OFFICE DEPARTMENT, OTTAWA, CANADA, AND FOR PAYMENT OF POSTAGE IN CASH. SUBSCRIPTION RATE: $10 PER YEAR. EUROPEAN SUBSCRIPTION COPIES DISTRIBUTED C/O INTERNATIONAL DISTRIBUTORS, VLAGSTRAAT 128, BORGERHOUT 2200, BELGIUM.

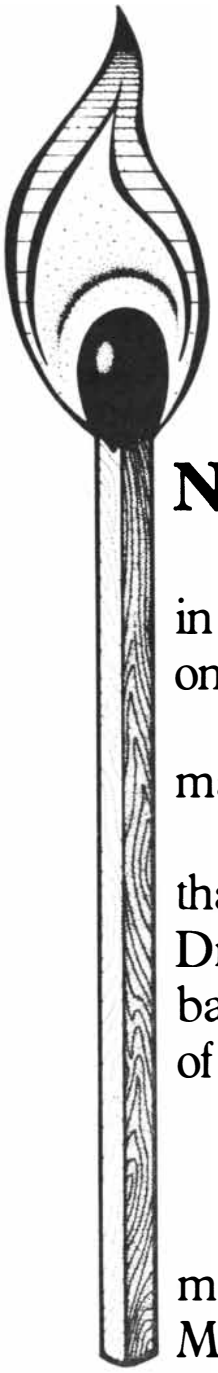
THE COVER

The photograph on the cover reproduces an experiment familiar to most children. By placing a hand over a lighted flashlight in a darkened room they discover that flesh is surprisingly transparent to red light. The demonstration is pertinent to the experiments reported by Michael Menaker in the article titled "Nonvisual Light Reception" (*page 22*). Menaker and others have found that light passing through the skull of birds, quite independently of the visual system, coordinates a number of physiological rhythms with the length of the day. To make the cover picture Ben Rose improved on the simple flashlight experiment by carefully masking the subject's hand so that no light leaked around the edges of the fingers. Illumination was provided by an 800-watt-second flash lamp. Exposure was one thousandth of a second at *f*/5.6 on Ektachrome X (ASA rating 64).

THE ILLUSTRATIONS

Cover photograph by Ben Rose

Page	Source
16–17	Jerome Kuhl
18–19	Lorelle A. Raboni
20	Jerome Kuhl
23	Eric Mose (*top*), Graphic Presentation Services, Inc. (*bottom*)
24	Michael Menaker, University of Texas at Austin
25	Eric Mose
26–29	Graphic Presentation Services, Inc.
31–33	National Aeronautics and Space Administration
34–38	Bunji Tagawa
41	National Aeronautics and Space Administration
46	Ben Rose
49–50	Jim Egleson
51	Research-Cottrell, Inc.
52	Jim Egleson
53	Ransburg Electro-Coating Corporation
54–57	Jim Egleson
58	Duncan C. Blanchard, State University of New York at Albany
61–67	Tom Prentiss
68–69	Robin Ingle
70	Tom Prentiss
75	Paul Weller
76–77	Alan D. Iselin (*top*), Eric Mose (*bottom*)
78–81	Alan D. Iselin
85	John D. Hanlon for *Sports Illustrated*
86–91	Allen Beechel
92	Paul Weller, from the New York Public Library
94–105	Dan Todd
108–112	Alan D. Iselin
114–119	Roger Hayward



Now what starts fires, stops them.

Phosphorus used to seem like magic, back in Chemistry I. Expose it to air and it catches on fire, remember?

Now we've made phosphorus seem like magic again. We've turned it into a fire-retardant.

It's a Monsanto product called Phos-Chek,® that's used in fighting brush and forest fires. Dropped from a plane or copter, it produces a barrier that puts out the fire, then does a couple of nearly unbelievable things.

It helps keep the area safe from further fire.

It fertilizes the ground for new seedlings.

Phos-Chek® fire-retardant is just one of the many kinds of products coming from the new Monsanto Company. In areas as diverse as swine genetics and electronics, fashion fibers and steel, environmental engineering and plastic packaging. Ideas just seem to catch fire at **Monsanto: the science company.**

FOR MORE INFORMATION ABOUT MONSANTO'S FIRE RETARDANT PRODUCTS, WRITE BOX G5EBF13, MONSANTO COMPANY, 800 N. LINDBERGH BLVD., ST. LOUIS, MO. 63166

Some things are changing for the better.

Many people know us as an instrument manufacturer: we make more than 2,000 products for measurement, test and analysis. Others know us as a computer company: more than 10,000 own our programmable calculators and computers. We prefer to think that our business is to serve measurement, analysis and computation needs . . . in science, industry, medicine and education. This is the rationale behind every new instrument, computer or system that we tell you about in these ads. This month:

When the HP 9600 rolls through your door, your real-time and data acquisition tasks become a lot easier to perform. This new systems family's long suit is the efficient and economic handling of multitudes of analog and digital information, simultaneously.

A sensor-based system that makes real sense.

There's a growing demand in industry and research laboratories for sensor-based computer systems that handle great quantities of analog and digital information. Systems built from programmable instruments usually are too expensive; people pay for equipment features that they don't need. Yet the alternative has been a piecemeal approach – break down the customer's problem into several parts and use separate "mini-systems" to solve each part independently.

Now there's a third choice – Hewlett-Packard's new family of compact data acquisition and control systems for cost-effective automation in industry and research. A 9600 Series system monitors, collects, and processes information from sensor-based sources. It then can generate reports, control power supplies, alert operators, drive graphic displays and plotters, and produce control signals for closed loop operations. Although you can't be everywhere at once – supervising and trouble-shooting – our system can.

Two new subsystems within the 9600, one analog and one digital, now do the things a number of programmable instruments used to do. These instrument functions are contained on plug-in cards. Instead of adding individual instruments, you merely slip in an inexpensive printed cirucuit board.

The 9600 data acquisition systems are modular. Start with a minimum low-cost system to control a single test or experiment, and expand with your growing needs.

The full story on the 9600 System family is yours for the asking.

Nothing can outperform this new digital GC-even at twice the price.

Because the gas chromatograph (GC) is essentially a tool for qualitative and quantitative chemical analysis, its value ultimately depends on how well it does this job. Over the years, many new models have been introduced that perform more accurately than previous

This would be an unusual case – using a battery-powered counter to check out the frequency of a mountain rescue-team's radio equipment – but it illustrates that HP's portable instruments can go anywhere service is needed.

instruments – at a price. The truly amazing thing about the new HP 5700 GC is this: it produces more accurate and precise retention time (qualitative) and peak area (quantitative) data than any GC ever built. Yet it costs about half as much as top-of-the-line GCs of comparable quality.

A new bulletin on the 5700 fully documents this perhaps startling claim. Until you have a chance to study this data consider this: one of the first 5700s off the production line was used "as is" to make two series of replicate analytical runs, one series before and one after an overnight shutdown. The sample used in both series contained seven components, out to C_{17}.

The results speak for themselves. In terms of repeat accuracy, the mean retention time of each of the seven components differed less than 0.01 minute after the overnight shutdown; the normalized area % varied only within ±0.001%. In terms of precision, the standard deviations of the replicate retention time measurements fell within 0.0175, both before and after the overnight shutdown; the standard deviations of the area % data were all within 0.0038. No other GC, regardless of price, can do better.

For a fully documented proof of performance as well as a factual description of this new all-digital, computer-compatible automatic GC, write for Bulletin 5700.

Portable instruments go where the problem is.

Capital equipment such as mobile or remote communications systems and million dollar computers have at least two things in common. They are electronically complex, and they can't be taken into a service center when they need repair. Today's traveling field service engineer must have laboratory quality equipment that will go where he goes.

HP's portable instruments enable service engineers to diagnose and repair this equipment on the spot, reducing expensive downtime. Our portable scopes are small enough to fit under an airliner seat, and, at 24 pounds, are light enough to be carried up antenna masts and into other hard-to-reach places. An HP electronic counter can be held in one hand – it takes only seconds to snap on a function module that provides the specific measuring capability needed. Then there's our multi-function meter – a high performance, instant-reading voltmeter and ohmmeter rolled into one.

And the length of HP's portable measuring capability isn't limited by the distance to the nearest wall socket. Most of our portable instruments feature their own accessory battery pack. Many can run off ordinary car, plane or boat batteries as well as a standard power line. And all of them deliver HP precision in a rugged, portable package.

Ask for the full story on portable instruments that go where the problem is. Write Hewlett-Packard, 1502 Page Mill Road, Palo Alto, California 94304; Europe: 1217 Meyrin-Geneva, Switzerland.

00202

HEWLETT hp PACKARD

LETTERS

Sirs:

Like Jean Lipman-Blumen ["How Ideology Shapes Women's Lives"; SCIENTIFIC AMERICAN, January], I have done a lot of empirical work on women, and have observed the traditional and contemporary ideologies she identifies. The striking thing about the holders of these ideologies, however, is that they *do* the same things while professing very different attitudes.

Her sample was biased from the start, because she chose her interviewees from the "wives of graduate students in the Boston area." That is, she did not choose individuals; she chose the role "wife of graduate student" to interview. Then she reported on what they told her about their ideas and aspirations. She did not report on what they actually did in their daily lives.

It has been my observation that the woman who hopes to complete her Ph.D. and the woman who is content with an A.B. are equally likely to be found at the dishpan or the clerical job. What they profess to believe about the role of women often does not really influence their actions. This situation is now beginning to change because of the women's liberation movement; it would therefore be interesting to revisit these same women four years later and see if their ideology has been converted into action.

E. E. C. CROMLEY

New York University
New York

Sirs:

I am encouraged by Mr. (or Ms.) Cromley's report of "observing" the traditional and contemporary ideologies discussed in my article. I welcome replication of my findings; however, I think it is appropriate to sound a note of methodological caution. Cromley doubtless would agree that unless sex-role ideology was measured in comparable ways in both studies, we would not have "tapped" the same ideological dimensions.

The criticism that I do not indicate the similarities in behavior of these two ideologically distinguishable groups is somewhat curious on several counts. First, since educational aspiration (that is, intended or planned behavior) varies according to sex-role ideology, I could hardly claim that these data suggest that my respondents "do the same things while professing very different attitudes." Moreover, the data reported on religious conversion rates for the two groups indicate variation along another significant dimension of adult behavior.

Second, although I share Cromley's concern with what these women "did in their daily lives," my basic research questions with regard to the sex-role ideology data were formulated with a slightly different focus (although I consider religious conversion and making plans for future education part of one's daily life in a very real and important sense). I was primarily interested in the impact of sex-role ideology on a very specific kind of behavior—educational-planning behavior—that has serious implications for one's entire life pattern. After establishing this relationship, my research design involved focusing on the *antecedent* factors associated with this predictor of planned adult behavior.

The question raised about sample selection is always of great concern to social scientists, in view of the inherent problems associated with making any statistical inferences about a parent population from a finite sample. As Cromley notes, the original sample consisted of graduate students' wives. Since both contemporary and traditional ideologies were found adequately distributed among these women, there is no reason to believe the role of graduate student's wife introduces an unacceptable bias. Although there was a preponderance of women with the contemporary ideology, the contemporary-traditional distribution is probably not unlike the distribution of these ideological positions among other married college women.

The sample was further limited to women who had no more than an A.B. degree. This was done in order to have a consistent baseline from which to measure the original dependent variable: educational aspiration. Although these methodological decisions clearly mean that my findings are not necessarily applicable to the universal set of women, I, along with most serious social scientists, only expect the findings to be generalizable to a specified subset of individuals. I did not maintain that the results were applicable beyond the population of married women who have attended college, within the age limits of the sample. Comparison of this sample with a national sample of college women on a wide range of variables, however, showed marked similarities.

The U.S. Department of Labor reports that in the year the data were collected there were 11,268,000 women in the U.S. who had completed one to four years of college. The majority of these women were married. If one prefers a more conservative position, generalizing only to the wives of graduate students, we note that at that time there were 664,000 civilian males enrolled in graduate schools on a full-time basis, not to mention the larger number of part-time male graduate students. There were an additional 2.2 million men who already had received a master's, doctor's or other professional degree, less than 10 percent of whom were single. I report these figures merely to suggest that whether one wishes to generalize the reported findings to the larger population of college-educated wives or simply to those whose husbands are attending or have attended graduate school, within comparable age ranges, the population to which the data may be referred represent substantial groups about whom such information would be meaningful.

JEAN LIPMAN-BLUMEN

Department of Sociology
Stanford University
Stanford, Calif.

Scientific American, March, 1972; Vol. 226, No. 3. Published monthly by Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017; Gerard Piel, president; Dennis Flanagan, vice-president; Donald H. Miller, Jr., vice-president and secretary; George S. Conn, treasurer.

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 Advertising Manager, SCIENTIFIC AMERICAN, 415 Madison Avenue, New York, N.Y. 10017.

Offprint correspondence and orders should be addressed to W. H. Freeman and Company, 660 Market Street, San Francisco, Calif. 94104. For each offprint ordered please enclose 25 cents.

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

Please enter change of address ☐
Please enter subscription ☐
☐ 1 year, $10 ☐ 2 years, $18 ☐ 3 years, $25

NAME

NEW ADDRESS

OLD ADDRESS

There are two cars built in Sweden. This is the one with front wheel drive for a firmer grip on snow and ice.

When we built the first Saab, we built it with front wheel drive for Swedish winters.

From that one to today's Saab 99E, we've seen no reason to change.

Because, with the weight of the engine over the drive wheels, you get a better grip on snow.

And with the engine pulling you around curves instead of pushing you, you're less likely to skid on ice.

(In a rear wheel drive car, the rear wheels have a tendency to keep going straight even as the front wheels start to turn.)

But front wheel drive isn't the only thing that makes us different from the other car built in Sweden.

We have rack and pinion steering for quicker response.

Impact-absorbing bumpers that reduce your collision insurance rates 15% at Allstate.*

And "roll cage" construction, the kind that soon, by law, may be required on all cars.

As standard equipment, we also have radial tires and 4-wheel disc brakes.

And while we were making the Saab 99E a good-handling and safe winter car, we also made it a comfortable winter car.

With things like separate heating controls for the back seat. And the world's only electrically-heated driver's seat to keep you warm until the rest of the car warms up.

But the best way to tell the difference between the two cars built in Sweden is to drive both. Once you've done that, we think you'll want to keep driving the Saab 99E.

SAAB 99E

Before you buy theirs, drive ours.

*Allstate discount available in most states. For the name and address of the dealer nearest you, call 800-243-6000. In Connecticut, call 1-800-882-6500.

50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

MARCH, 1922: "Dr. E. P. Lewis of the University of California at Berkeley gives the following statement of the picture of atomic structure due to the work of Lorentz, Zeeman, Rutherford and Bohr: 'There is a central nucleus that may be a hydrogen or helium atom or a combination of these with binding electrons, and with an excess number of positive charges equal to the atomic number as defined by Rutherford and Moseley. Around this nucleus circulate one or more electrons in circular or elliptical orbits that have radii several thousand times greater than the diameter of the nucleus. These electrons control chemical valency and emit or absorb ordinary light waves as they pass from an outer to an inner orbit or conversely. As shown by the work of Rutherford, the nuclei of the heavy radioactive elements disintegrate spontaneously, emitting alpha particles (helium atoms), beta rays (similar to cathode rays) and gamma rays, which are very short X-rays. Rutherford has recently also shown that the energy of an alpha particle from one of the disintegration products of radium is sufficient to disintegrate the atoms of oxygen or nitrogen with which it collides, showing that they too are built up of smaller parts, one of these parts being certainly hydrogen and another probably helium. With an increasing knowledge of atomic and spectroscopic phenomena the Bohr hypothesis seems to present a more and more satisfactory picture of atomic structure. Yet there are underlying this hypothesis assumptions for which we have no explanation."

"Niagara marks the point where nature has found an outlet for the waters of Lake Erie, but man has found it convenient to order the matter otherwise. Niagara is not to become dry, but with the completion in late December of the Queenston-Chippawa power canal much of the water that has taken a 162-foot plunge will go over the cliff at a point where the head available for power generation will be 305 feet. Dredges have for nearly four years been cutting a channel through earth and rock in a wide swing around the falls. They have taken out 13,000,000 cubic yards of earth and 4,000,000 of rock. The hydroelectric plant's full power development in contemplation amounts to 500,000 horsepower. The turbines and generators are larger than any in use elsewhere."

"It takes an aerial view to give New Yorkers a comprehensive view of the town they live in, to make them comprehend fully the extent to which the trees have given way to the buildings, and the buildings to other buildings. It is only recently that they have been able to look down, to get a bird's-eye view of the entire layout. In the past six months the first complete aerial-mosaic map of Manhattan Island has been assembled, and photographic prints are being made on a scale that brings out a wealth of interesting detail and opens up a vast source of information that will be useful in further expansion of the city's commerce, industry and home building."

SCIENTIFIC AMERICAN

MARCH, 1872: "The history of gas light in brief is as follows: In 1792 Wm. Murdock of England lighted his own dwelling with gas; in 1803 a machine shop and in 1805 a cotton factory were similarly lighted. Murdock began to lecture upon the subject, but not until 1810 could a company get a charter for the manufacture of gas. In 1813 Westminster bridge was lighted, and in 1815 Guildhall. Still there was great opposition even from scientific men, and there were also great difficulties for want of machinery to make and use the gas. Gun barrels screwed together were used to convey it from place to place. Finally, however, every obstacle was surmounted, and now there is not a city of any size in the civilized world that is not lighted by gas."

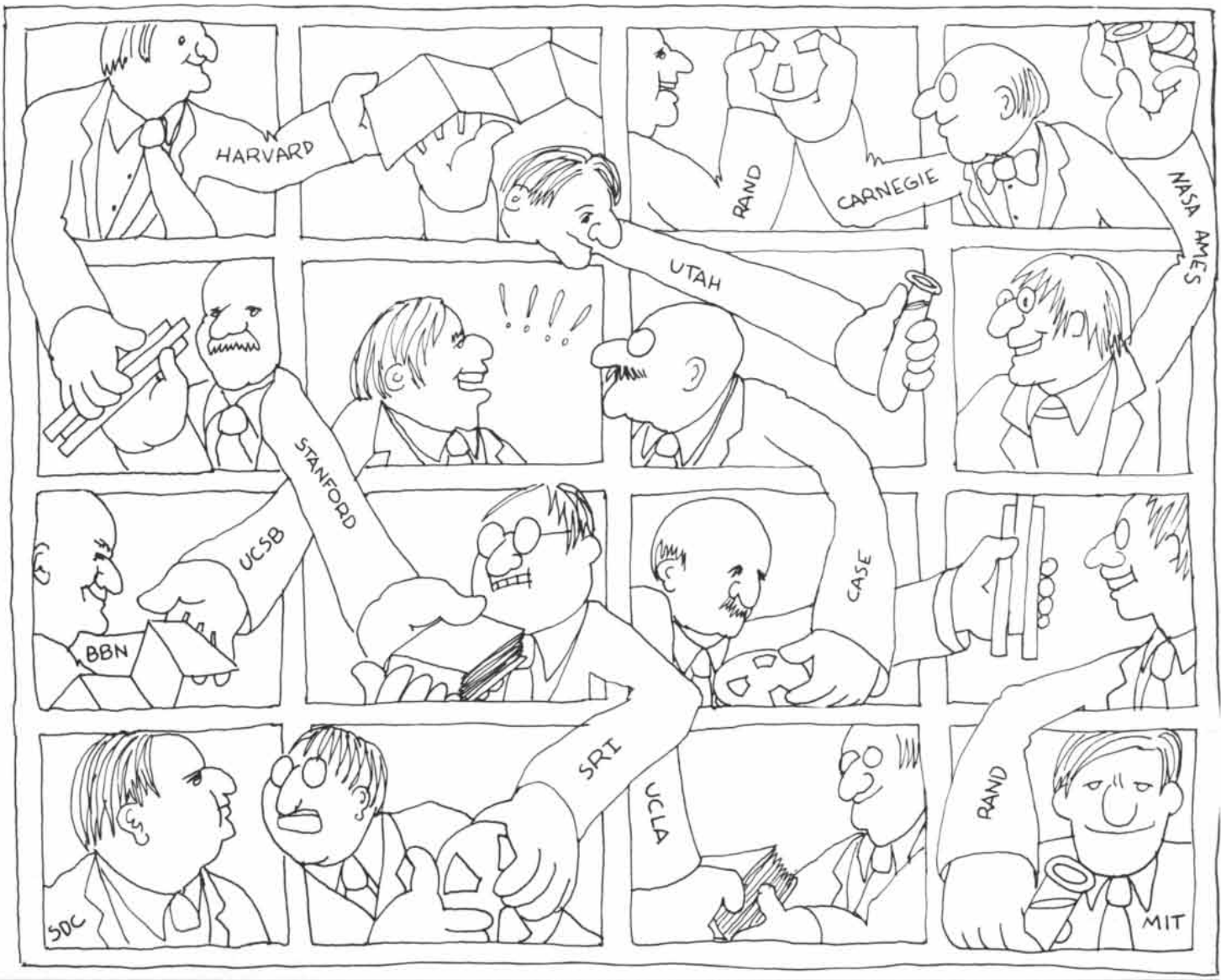
"The Museum of Natural History in New York is rapidly becoming an attractive and important institution. It is open to the public free and is daily visited by thousands. Since its first opening and reception last year many valuable acquisitions have been received, the more valuable being the collection of Prince Maximilian of Germany, which contains a vast number of specimens. A fine specimen of ichthyosaurus has been added. More than 14,000 specimens of birds, besides several hundreds not mounted, are in course of rapid preparation for public inspection and study. Other additions are constantly received."

"Dr. Colton recently lectured in Brooklyn, giving some practical illustrations of the peculiar effects of nitrous oxide, or laughing gas, which is composed of a mixture of two parts of nitrogen and one part of oxygen. Abroad it had been used for some time to light large buildings, such as the opera houses in Paris and Vienna. Since 1844 Dr. Colton has given the gas to 55,923 persons for dental operations, and none of them has felt the worse for it. They had removed 19 teeth from a Brooklyn lady that morning, and she never felt the slightest pain; indeed, she was astonished when she awoke to find that her teeth were out."

"M. Faye recently read to the French Academy two papers in which he summarized all the theories that have as yet been given to the world on the nature of comets. He attacked, with some satire, Sir William Thomson's view, given to the British Association at its meeting for 1871, that the comet's tail still remains an insoluble mystery. M. Faye asserts that the tails of comets are the effect of the repulsion of the sun, and he has supported this view by argument and experiment, stating that all white-hot bodies exert a repulsive force on extremely rarefied matter. We may reasonably expect that the spectroscope will enlighten us, as it has done already on many other phenomena scarcely believed to be within man's powers of explanation and hypothesis."

"The painful effects of poisoning by lead are not by any means confined to painters, white-lead manufacturers and others whose trades bring them in constant contact with this deleterious metal. There are some persons whose obstinacy allows them to use it in cosmetics and hair washes in spite of the warnings of the medical profession, and the evil is augmented by the fact that such preparations may be used for years with impunity, and the palsy, paralysis and other effects do not appear till the entire system is thoroughly impregnated."

"Recent letters from Professor Agassiz report the safe arrival of his exploring ship at Pernambuco, Brazil. His explorations of the sea bottom are continued with undiminished zeal. He has made many discoveries of fossil and live animals, sponges, etc., the mere nomenclature whereof is enough to break the uninitiated jaw."



ARPA has a network of Supercomputers.

There are two dozen huge computer systems in the Advanced Research Projects Agency network.

Over half of them are DECsystem-10s. Our Supercomputer. MIT has two. So does Utah. Then there's Harvard, BBN, Carnegie, Case, SRI, Stanford and Rand.

Which should give you some idea of how popular our DECsystem-10 really is.

In the ARPA network, DECsystem-10's are doing state-of-the-art research into weather forecasting, econometric studies, resource management, computer sciences, and much more. Everyone shares their computer and expertise with everyone else. Everyone comes out ahead.

Additional DECsystem-10's will be a part of ARPA's mammoth ILLIAC IV number crunching complex now being installed at NASA-Ames Research Center in California. They'll handle all communications while at the same time managing up to a trillion bits of file storage. Once ILLIAC IV is in gear, ARPA members will be able to do in hours jobs so big that they wouldn't even attempt to do them now.

ARPA is one of the biggest brain trusts ever assembled. If half of its members have a DECsystem-10, you really ought to know about it.

Write for the literature that explains why 62 of the leading universities and research institutions in the country have selected DECsystem-10's. (Hint: It does computation *and* timesharing at half the cost of other systems—without sacrificing throughput.)

DECsystem-10 Scientific Group. Digital Equipment Corporation, 146 Main St., Maynard, Mass. 01754. (617) 897-5111.

digital

THE AUTHORS

G. O. KERMODE ("Food Additives") is head of the food-standards program conducted jointly by the Food and Agriculture Organization and World Health Organization and also head of the former organization's Codex Alimentarius secretariat in Rome. Born in the United Kingdom, he studied economics at the University of Liverpool and in 1956 entered the British civil service. He served in the Inland Revenue Department and in the Ministry of Agriculture, Fisheries and Food; he also was secretary to a number of government committees and private secretary to three agriculture parliamentary secretaries. Before taking up his present work he was head of the Fisheries Branch and the Food Standards Branch in the Ministry of Agriculture, Fisheries and Food.

MICHAEL MENAKER ("Nonvisual Light Reception") is associate professor of zoology at the University of Texas at Austin. He notes that he grew up in New York City "directly across the street from, and much of the time actually in, the American Museum of Natural History." Menaker was graduated from Swarthmore College in 1955 and obtained his Ph.D. from Princeton University in 1960. He writes: "In biology my broad interests are in attempting to understand the ways in which organisms make use of the informational content of the environment. I have worked on circadian rhythms and the physiology of hibernation in bats. In addition to the work reported here I have an active interest in pineal physiology. I teach an introductory course in organismic biology. I value this contact with undergraduates, particularly because of the opportunity it provides to transmit and illustrate my conviction that science is properly viewed as an important aspect of culture and not as a source of grist for the technological mill."

ROBERT S. DIETZ ("Geosynclines, Mountains and Continent-building") is a marine geologist with the National Oceanic and Atmospheric Administration, working at the Atlantic Oceanographic and Meteorological Laboratories in Miami. His degrees are from the University of Illinois, where he received his Ph.D. in 1941. He has written extensively on geosynclines, plate tectonics, sea-floor spreading, continental drift, marine mineral resources and deep-research vehicles. "As a departure from investigating the ocean floor," he writes, "I occasionally take trips to various parts of the world to study geologic scars of ancient meteoritic or cometary impacts. These are not circular holes in the ground but usually complex disrupted domes revealing evidence of intense shock." For his research in geotectonics Dietz recently received the Walter H. Bucher medal of the American Geophysical Union and the gold medal of the U.S. Department of Commerce.

A. D. MOORE ("Electrostatics") is professor emeritus of electrical engineering at the University of Michigan, where he taught for 47½ years until he retired at the beginning of 1964. He continues to be active in his profession, maintaining his experimental laboratory at the university, traveling extensively (in a station wagon loaded with apparatus) to give a lecture-demonstration on electrostatics, founding (in 1970) the Electrostatics Society of America and writing books (most recently *Electrostatics* in 1968 and *Invention, Discovery and Creativity* in 1969). He is now editing a book, *Electrostatics and Its Applications*. Outside his profession he served as a member of the City Council of Ann Arbor from 1940 to 1957 (he writes that it was the "longest continuous service of an elected city official in the history of the city"), has been and continues to be toastmaster at "innumerable banquets" and at one time "concentrated on three-cushion billiards, with a high run of 11; knew the world's greats, Willie Hoppe and Charles C. Peterson, and played with them in demonstration games." Moore was graduated from the Carnegie Institute of Technology in 1915 and obtained his master's degree at Michigan in 1922.

BJÖRN KURTÉN ("The Cave Bear") is a lecturer in paleontology at the University of Helsinki, where he obtained his Ph.D. in 1954. He has written on such subjects as fossil carnivores, dating of early man, late Tertiary and Quaternary stratigraphy, population dynamics, evolutionary theory and paleobiogeography. This is the year of publication of three of his books: *The Ice Age, The Age of Mammals* and *Not from the Apes*.

JEROME KAGAN ("Do Infants Think?") is professor of human development at Harvard University. He was graduated from Rutgers University in 1950 and took his Ph.D. at Yale University in 1954. Before going to Harvard he was at the Fels Research Institute for seven years, serving from 1959 to 1964 as chairman of the department of psychology. He writes that he is "deeply involved in field research on isolated Indian populations in Guatemala assessing biological and social factors controlling cognitive development." Kagan's most recent books, published last year, are *Change and Continuity in Infancy* and *Understanding Children*.

RODOLFO MARGARIA ("The Sources of Muscular Energy") is professor of physiology at the medical school of the University of Milan. "I was born in the Alps," he writes, "more precisely on the Matterhorn—not quite on top of it, but in the nearest little town at its feet. Being a mountaineer directed my research activity toward muscular exercise and high altitude. Having become an expert on high altitude, I was moved to study high-altitude flying, and so I shifted somewhat to aviation medicine, and with the advent of space flight my interest was shifted higher up to space medicine." In recent years his main field of study has been the biomechanics of human locomotion at reduced gravity, such as exists on the moon; he describes this work as "a kind of sublimation of the actual practice of mountaineering, which I had to drop because of old age." Margaria obtained his M.D. at the University of Turin in 1924.

CURTIS WILSON ("How Did Kepler Discover His First Two Laws?") is professor of history at the University of California at San Diego. He did his undergraduate work at the Berkeley and Los Angeles campuses of the university, graduating in 1945. "In the middle of my junior year," he writes, "feeling cramped for room in which to speculate and not finding that I could see through a microscope what I was supposed to see, I deserted a major in biology for history," a field in which he obtained his Ph.D. at Columbia University in 1952. From 1948 until 1964 he was a tutor at St. John's College in Maryland, serving also from 1958 to 1962 as dean of the college. "Along with a few other tutors I gave much of my effort to The Laboratory, as we called it: St. John's attempt to cope with the problem of including the grammar of modern science in a liberal education. Our endeavors, which were sometimes fairly successful and at other times failed miserably, provided an opportunity for which I remain grateful: the chance to try to reconstruct, using some logic, imagination and history, the steps and processes whereby conceptual changes have occurred in the sciences."

This Superb Silver Medal

Yours for just $1.00

To introduce a new kind of investment in art and in precious metals.

OVER 3 ½ TIMES MORE SILVER THAN IN A NEW UNITED STATES SILVER DOLLAR
Shown larger than 40mm actual size (to show detail).

Share the Pride of "America's Natural Legacy" - 36 Exquisite Medallions...Minted to the fine Jewelry Standards of Wittnauer With the Assistance of the Sierra Club

Yes, you may have this incredibly beautiful Art Medallion in either 24 Karat Gold-Plated Sterling Silver, or Solid Sterling Silver for just $1.00...less than the cost of the precious metal alone. Why?

We introduce you to a most spectacular opportunity.

36 exquisite medallions depicting American Wildlife can be yours in this extraordinary and historic series. Each medal has been sculptured by one of America's leading artists and forged into rich, high relief medallions. Magnificent, lasting tributes to our rich heritage.

Imagine what a set of this magnitude can mean to you and your family. What unforgettable lessons in our natural heritage can be learned and shared. Imagine also, the possible financial growth that lies in this rare combination of fine sculpting and precious metals...forged into exquisite medallions to the uncompromising standards of the world famous and highly honored Wittnauer Jewelers.

To further enhance their value, every coin in your series will bear your individual and personal serial number—engraved on its edge, alongside the Hallmark.

Daring Repurchase Guarantee

The Longines Symphonette Society considers this set so fine, and of such strong growth potential, that it promises to buy back any completed set, for the full purchase price, any time to December 31, 1974, if you are not satisfied either in its progress...or in any other way. Moreover, this repurchase guarantee is backed by a $25,000 bond from an internationally famous insurance company.

Take The First Medal For Just $1.00

Each medallion in your personally numbered series is a radiant work of art. Each depicts one of America's native animals in its natural habitat. Unlike ordinary coins, the high relief and hand-crafted finish give elegance and realism to the medallions. The Bald Eagle hovers majestically over a nest of chicks...the Plains Bison glances from across a meadow and you actually feel the presence of its massive dignity, now captured forever in solid sterling silver, or 24-Karat Gold-on-Sterling. It is a collection without equal...carefully designed to become your proudest possession today, a cherished heirloom that will be passed on to future generations.

WE CANNOT GUARANTEE YOUR RESERVATION UNLESS WE HAVE YOUR ORDER BY JUNE 30, 1972

ALSO RECEIVE COMPLETE INVESTOR'S KIT—FREE


Read This Daring Guarantee

1. We Guarantee To Maintain Your Low Subscriber price (just $10* for the solid sterling silver and $15* for gold on sterling) NO MATTER HOW THE PRICE OF SILVER MAY SOAR IN THE MONTHS AHEAD.
2. We guarantee that each medal will contain no less than 525 grains of pure sterling silver—MORE THAN 3½ TIMES THE SILVER IN A NEW U.S. SILVER DOLLAR!
3. We Further Guarantee to repurchase any complete set of 36 medals until December 31, 1974. A bold promise, backed by a $25,000 bond from an internationally famous insurance company.

*Includes 50¢ shipping and handling.

- **Lovely American Wood Frames** to display your collection. We even include a personalized plaque with your name and series number.
- **Monthly Newsletters** keep you up to date on the progress of your collection and give you news of major events in the coin and medal world...events that could affect the value of your series.
- **Luxurious Membership Certificate**...suitable for framing and a tribute to your good taste.
- **Beautiful and Informative Booklets,** prepared under the supervision of the Sierra Club, give you enlightening information about the history and habitats of each of the 36 animals.

FILL IN AND MAIL THIS COUPON TODAY

America's Natural Legacy

c/o Longines Symphonette Society Box 1972, Larchmont, N.Y. 10538

Please reserve my AMERICA'S NATURAL LEGACY subscription. I will receive the first two medals in the series for FREE 10-DAY EXAMINATION and, even if I decide to return the second medal and not begin my subscription, I MAY CHOOSE TO KEEP THE AMERICAN BALD EAGLE MEDAL FOR ONLY $1. I understand that I will be billed on a regular basis for each consecutive monthly shipment of two medals produced for my account.

Please Check One—Reserve my subscription in:
☐ 24 Karat gold—electroplated on sterling ☐ solid sterling

☐ Mr.
☐ Mrs.
☐ Miss PLEASE PRINT
Address

City

State Zip 151AZ 152AZ 131

The $15 price for each gold-plated sterling silver medallion and the $10 price for each sterling silver medallion includes 50¢ shipping, handling and insurance. Add sales tax where applicable.

AN EXTRA SERVICE FOR YOU:
(Please check if you wish to charge your purchases.)
☐ American Express ☐ Diners Club
☐ BankAmericard ☐ Master Charge

My credit card number is:

BankAmericard Expiration Date: Master Charge Bank #

GCA helped get it there.

A decade and a half ago the miniature computer was unknown. Integrated circuits that could handle multiple computer functions on a single silicon chip did not exist. And one of the main reasons why miniaturization was impossible was the inability to make precise reductions of circuit artwork . . . the microimages that were the first photomasks.

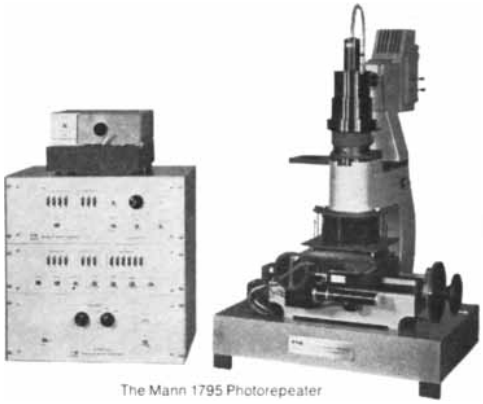
Then GCA's David W. Mann Company introduced the first commercially available Photorepeater®. It combined special optics with precision step-and-repeat motions into a microimaging system which fast became the industry standard. As IC technology improved, GCA photomasking systems provided the reliable, precise means necessary to meet manufacturing demands for more circuit functions on smaller chips.

The rest is design and packaging history.

Through the Mann product line, GCA provides total photomasking systems which handle smaller and smaller circuit geometries and enable computer people to build more computer power into less space.

If you could use a fast-turnaround, high yield photomask production capability backed by the most experience in the industry, turn to GCA. We've made the progress. We have the systems. And the information. Write GCA Corporation, Burlington Road, Bedford, Mass. 01730.

GCA CORPORATION



The Mann 1795 Photorepeater

Established 1845 SCIENTIFIC AMERICAN March 1972 Volume 226 Number 3

Food Additives

Perhaps as many as 2,500 substances are currently being added to foods for flavoring, coloring, preservation and other purposes. How are the necessity and safety of these substances determined?

by G. O. Kermode

Men have added nonfood substances to their food throughout recorded history, but in recent decades they have become concerned about such practices because of the large number of substances and motivations that have become involved. The questions at issue for any food additive are whether or not it is necessary and, if so, whether or not it is safe. For many years the United Nations (through the Food and Agriculture Organization and the World Health Organization) and many governments have kept watch on additives with these questions in view. The questions must be faced whenever a new additive is proposed; sometimes, as the recent case of cyclamate additives in the U.S. showed, they must be reconsidered when new evidence puts the safety of an old additive in doubt.

The distinction between food ingredients and food additives is somewhat imprecise. Sugar, being a natural product, is usually regarded as an ingredient, whereas saccharin, being an artificial sweetener, is likely to come under the heading of an additive. Perhaps the best method of classification is by function. Additives are employed for such purposes as enhancing flavor, improving color, extending shelf life and protecting the nutritional value of a food. They are, in short, valuable but not always essential items in the manufacture of food products.

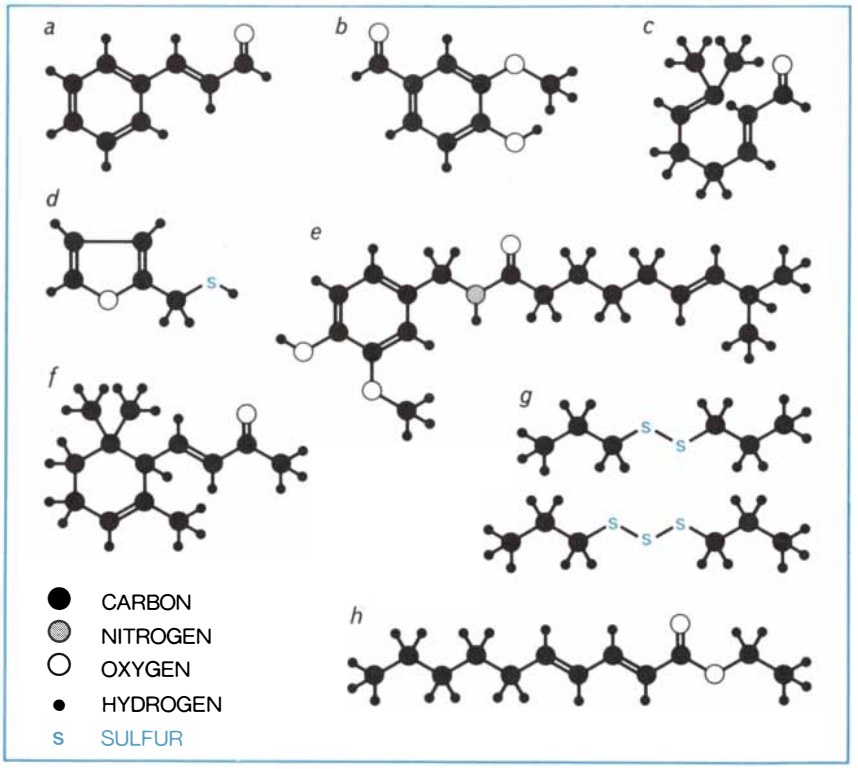
Whatever one's views on additives may be, it is true that without additives many food products could not be offered for sale in their present form. This is exemplified in particular by the many convenience foods that have become popular in North America and in western Europe. Moreover, if food production is to increase enough to keep pace with population growth and the effort to improve nutrition generally in undernourished areas, chemicals that are not normally part of food will inevitably play an increasingly important role.

From the earliest times foods were preserved with incidental additives that resulted from cooking. Food was also preserved extensively in ancient times by heating, drying, salting, pickling, fermenting and smoking. Food colors were used in ancient Egypt. In China kerosene was burned to ripen bananas and peas; the reason the method succeeded, although the Chinese did not know it, was that the combustion produced the ripening agents ethylene and propylene. Flavoring and seasoning were arts in many ancient civilizations, with the result that spices and condiments were important items in commerce.

Additives have not invariably been employed with beneficial aims. The adulteration of food, in order to pass an inferior product off as a good one, is as old as trade. Expensive items such as tea, coffee, sugar, spices and essential oils were often adulterated. Common adulterants included coloring substances and burned or roasted vegetable material, which was mixed with flour. Bread, beer, and wine were widely adulterated.

Eventually such practices led the authorities of the time and place to try to suppress them. The earliest food laws were often designed to control the more obvious forms of adulteration and fraud. In addition to these efforts the merchant guilds tried to protect the genuineness and the reputation of their products. The means at hand for testing foods were limited; checking the appearance, taste and smell of a food was about all one could do. The basis of knowledge making possible the national food laws that are common today was not established until about the middle of the 19th century. In the latter part of the century pure-food laws were enacted in country after country to control the composition of food and regulate the use of additives.

These developments coincided with a number of discoveries, mainly in organic chemistry, that led to the production of several of the important food additives now in use. For example, discoveries that resulted in the development of aniline and the coal-tar dyes led eventually to many of the synthetic colors now added to food. The active principles of odor and flavor were isolated from vegetables and other organic materials, leading first to alcoholic solutions of those materials as flavors and later to synthetic flavors, some of which did not appear to be present in natural edible aromatic substances used to flavor foods and some of which proved to have more flavoring



TYPICAL FLAVORING COMPOUNDS have chemical structures shown here. Cinnamaldehyde (*a*) supplies a cinnamon flavor; vanillin (*b*), a vanilla flavor; citral (*c*), lemon; furfuryl mercaptan (*d*), coffee; capsaicin (*e*), the pungent ingredient of red pepper; alpha-ionone (*f*), a principal component of strawberry and raspberry flavors; propyl disulfide (*g*), onion, and ethyl *trans*-2,*cis*-4-decadienoate (*h*), pear. Some 1,400 flavorings are in use.

power than the analogous natural flavors. By 1900 the flavorings in use were nearly all artificial and, except for vanilla, lemon, orange, peppermint and wintergreen, were being made with synthetic substances.

More than 40 functions now served by food additives can be listed. In this discussion, however, I shall group additives broadly in five categories: flavors, colors, preservatives, texture agents and a miscellaneous group.

Flavors constitute the largest class of food additives; estimates of the number of natural and synthetic flavors available range from 1,100 to 1,400. It is probably fair to say that flavors pose the largest regulatory task, not only because there are so many of them but also because of insufficient toxicological data, rapid changes in the field and many other factors. In general little is known about the toxicological aspects of flavors. Part of the problem is that many natural flavors have been used for centuries, and fully evaluating them all for safety would be an immense task. It is often argued that doing so would divert a large part of the effort that is needed to investigate the safety of more important and potentially more dangerous additives.

Over the past 30 years the use of flavorings has grown tremendously, paralleling the expansion in new types of food, new food-processing techniques and new methods of distribution. Governments have approached the question of controlling flavors from various directions. Some publish lists of permitted and prohibited flavors; some have a short list of prohibited flavors, many of which are natural, and others allow flavorings (both natural and synthetic) that are found only in the aromatic oils of edible plants.

Related to flavors are the additives known as flavor-enhancers. The commonest of them is monosodium glutamate (MSG), which is the monosodium salt of glutamic acid, one of the amino acids. A good deal of research is under way to find other flavor-enhancers, particularly in the group of substances known as the 5′-nucleotides. Similar work is being done on enhancers for fruit flavors. In recent years the use of maltol, which can intensify or modify the flavor of preserves, desserts, fruit, soft drinks and foods generally high in carbohydrates, has expanded greatly.

Manufacturers are also doing considerable research to find flavors that are cheaper or more effective than existing flavoring agents and flavor-enhancers. It is probably in this field that the greatest need for new additives will arise in the future, particularly for additives that can be put in simulated food products to imitate the complex flavor properties of traditional foods. At present the most widely sold simulated products are meat substitutes made from spun soybean proteins or proteins from other vegetables. With the addition of flavors, colors, vitamins, emulsifiers, acidifying agents and preservatives these proteins are sold as "vegetable steaks," "soya chicken breast" and "vegetable bacon" or are included in compounded products that normally have meat as a major ingredient. Other simulated foods are substitutes for dairy products. Flavored drinks, made so as to simulate the properties of genuine fruit juices, are also on the list. Because additive flavorings are expected to play such an expanding role in these products, they represent the field where ways of protecting the consumer's interest will need close attention, particularly with regard to designation and labeling of simulated foods.

Colors are put in food mainly to give it an appetizing appearance, on the tested assumption that the way food looks has an effect on its palatability. Foods are also colored to enhance the appreciation of flavor. Many people have become accustomed to the standardized color of a food product and would not accept the product if the color were substantially changed, even though nothing else had been done to the food. One need think only of blue or red butter to recognize the importance of accepted colors.

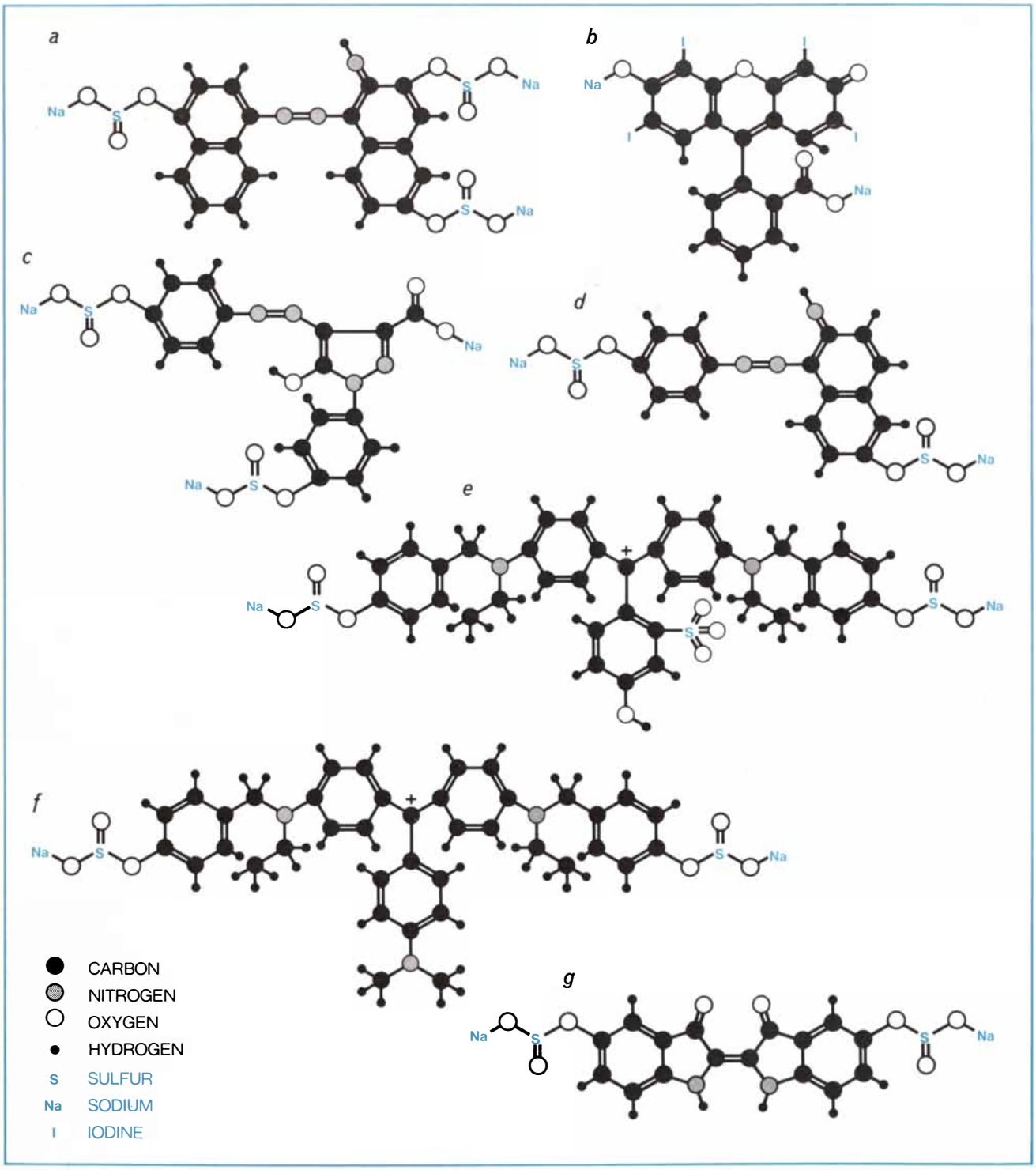
Much research has gone into food coloring. The colors most used in the food industry are synthetic dyestuffs. They are notably pure. Since they also have strong coloring power, little coloring is needed to achieve the desired result in a food product.

The manufacturer needs a color that not only produces the desired appearance but also will remain stable under certain conditions of manufacture, storage and cooking. Color put into candy, cakes and biscuits must be stable both to high temperature and to the action of carbon dioxide. Other colors must be able to withstand high processing temperatures and the action of acids.

Color regulation, like flavor regulation, varies from country to country. Many countries have fairly short lists of permitted food colors. The regulations specify purity and identity for the per-

mitted colors and also restrict the number of foods to which color can be added. Since most of the lists are based on the toxicological evaluation of the dyes, one might expect a reasonable degree of uniformity among the lists. It is not so, however, and therefore one of the most troublesome problems facing a food manufacturer who wants to export his products is the need to vary the color according to the different regulations of the importing countries.

The World Health Organization has evaluated more than 140 kinds of coloring matter, declaring a number to be unsafe and publishing a fairly short list of colors deemed to be safe. In some countries the food industry manages quite well with a choice of no more than a dozen dyes. Other countries allow more dyes. The difference is illustrated by a



CERTAIN COLORS employed as food additives are portrayed according to their chemical structure. The colors have both numerical and descriptive names: (*a*) red 2, amaranth; (*b*) red 3, erythrosine; (*c*) yellow 5, tartrazine; (*d*) yellow 6, sunset yellow; (*e*) green 3, fast green; (*f*) violet 1, benzylviolet, and (*g*) blue 2, indigotine. The characteristic ring structure evident in the seven diagrams is more likely than an aliphatic, or open-chain, structure to produce color because of the way it absorbs and reflects light.

BOTANICAL SOURCES of four natural additives are depicted. The annatto (*a*) is a tropical tree, *Bixa orellana*, that produces a yellowish-red coloring agent made from the pulp around the seeds. Natural vanilla extract comes from the pods of several species of orchid, chiefly *Vanilla planifolia* (*b*). Sap from the papaya tree (*c*), *Carica papaya*, is the source of the enzyme papain, which is

problem that the British will face when their country becomes a member of the European Economic Community: kippers will no longer be golden and sausages will no longer be "nicely pink" unless the Community's list of permitted colors is extended.

Preservative additives are one means of deterring food spoilage caused by microorganisms. The seriousness of spoilage is shown by the World Health Organization's estimate that about 20 percent of the world's food supply is lost in this way. Indeed, shortages of food in many parts of the world could be alleviated with the wider use of preservatives.

Spoilage can be prevented or retarded not only with additives but also with physical and biological processes such as heating, refrigeration, drying, freezing, souring, fermenting and curing. Some of these processes, however, achieve only partial preservation. Additives therefore have a role in prolonging a food's keeping qualities.

A number of different types of preservative have to be employed, depending on the kind of food, the method of manufacture, the way the food is packaged and stored and the nature of the microorganisms involved. Baked goods, for example, go stale rapidly. Once made, they are often exposed to mold spores that become active in warm weather or high humidity. In bread the spores produce a condition called "rope." Sodium diacetate, acetic acid, lactic acid, monocalcium phosphate, sodium propionate and calcium propionate are all effective in preventing rope. Sorbic acid and its salts have many uses, such as preventing mold in cheese, syrup and confections containing fruit or sugar. Benzoic acid and sodium benzoate serve as preservatives in margarine, fruit-juice concentrates, juices and pickled vegetables. Sulfur dioxide is widely used to inhibit mold and discoloration in wine, fruit pulps, fruit-juice concentrates, fruit drinks requiring dilution and dried fruits and vegetables.

Sulfur dioxide is giving rise to concern in a number of countries where the average wine consumption is so high that the people who drink a good deal of wine are in danger of exceeding the acceptable average daily intake of sulfur dioxide. The acceptable daily level is 1.5 milligrams per kilogram of body weight, which means about 100 milligrams a day or a half-liter of wine containing 200 parts per million of sulfur dioxide. Studies on sulfite in the rat found that .1 percent in the diet inhibited the growth rate, probably because sulfite destroys vitamin B_1. The significance of this finding in man, whose diet does not consist exclusively of sulfited food as in the experiments with rats, is questionable; nonetheless, more work is needed to dispel the uncertainty about the toxicity of sulfur dioxide and sulfites.

As a result of such uncertainties serious attempts to find alternatives to preserving foods with chemicals are under way. Among the recent advances is the development of antibiotics as antimicrobial additives. Antibiotics commonly have a more transitory effect than the traditional preservatives and are more selective. These advantages are significant when antibiotics are directed against known food pathogens and when their action is required only during the manufacturing stage. Antibiotics can be said to have a major disadvantage, however, in that by changing the normal spoilage pattern of certain foods they may result in unfamiliar forms of spoilage that consumers cannot recognize.

A number of countries have permitted such antibiotics as tetracyclines, nystatin, nisin and pimaricin as direct or indirect additives to chilled or raw fish, meat, poultry, cheese and bananas. The applications are strictly limited. Many other countries, although they recognize the efficacy of antibiotic additives, have taken the view that it would be unwise to approve them widely for food, since the antibiotics are important in medicine and their liberal use in food might produce resistant strains of pathogens that could affect humans.

Another development that has attracted interest as an alternative way of protecting food is the experimental work wherein ionizing radiation is employed to destroy the microorganisms and insects that cause food spoilage. An advantage of irradiation is that it produces little or no rise in the temperature of the food during treatment. A disadvantage at present is the possibility that irradiation will have an effect on the food and leave residues. For example, it is possible for the extranuclear structure of atoms to be excited under the influence of ionizing radiation. If the atoms are constituents of a molecule, the molecule as a whole may be excited, which may lead to rupture of one or more chemical

d

employed as a meat tenderizer. One source of wintergreen flavor is the leaves of the evergreen plant *Gaultheria procumbens* (*d*).

bonds, giving rise to free radicals. The free radicals may be capable of starting chemical chain reactions.

Nonetheless, work with irradiation has advanced to a point where in a small number of countries the sale of certain irradiated foodstuffs is now allowed. Irradiation raises the possibility that perishable foodstuffs could be more widely distributed in a fresh or nearly fresh condition. It is likely to take several years, however, for the irradiation of food to become a widespread practice because of the effort that must still be devoted to developing procedures for testing irradiated foods. The International Atomic Energy Agency, in conjunction with the Food and Agriculture Organization and the World Health Organization, has indicated a number of possible treatments by ionizing radiation to achieve long-term preservation, without refrigeration or chemical preservatives, of perishable foods and also to prevent food poisoning by destroying microorganisms such as salmonella.

Many traditional preservatives—notably salt, vinegar and sugar—still play an important role in homes and factories. It can be argued that recent improvements in food processing, coupled with improved standards of hygiene, should reduce the need for chemical preservatives. On the other hand, developments in making ready-to-use foods and the widespread resort to prepackaging have tended to increase the need for preservatives.

Related to preservatives are antioxidants, which are added to fatty foods primarily to prevent rancidity. Typical products containing these additives are margarine, cooking oils, biscuits, potato chips, cereals, salted nuts, soup mixes and precooked meals containing fish, poultry or meat. Certain foods, such as virgin olive oil, contain their own natural antioxidants in the form of tocopherols and therefore do not need the addition of antioxidants. If such foods are heated in a manufacturing process, however, they tend to lose their natural antioxidants, which should be restored if the product is to have a reasonable shelf life.

The most widely added antioxidants are butylated hydroxyanisole, butylated hydroxytoluene, propyl, octyl and dodecyl gallates and natural or synthetic tocopherols singly or in combination. Certain acids (ascorbic, citric and phosphoric) combined with antioxidants increase the antioxidant effect.

Preventing rancidity is not the only problem with a number of foods. The growing practice of using transparent wrapping for food presents its own problems by exposing the product to light and increasing the likelihood of discoloration. Ascorbic and isoascorbic acid have proved effective in preventing discoloration in certain fruit juices, soft drinks, canned vegetables, frozen fruits and cooked cured meat such as ham.

Often more than one antioxidant is put into a food, producing a synergistic action that allows more effective control of the product. Many countries authorize several antioxidants as food additives, but a number of others will allow only the so-called natural antioxidants, such as ascorbic acid (vitamin C) and the tocopherols (vitamin E). Much research is in progress to find compounds that are more potent than the present antioxidants. The search is particularly keen for antioxidants that are less likely than the present ones to impart odor, flavor or color to foods. Another quest is for antioxidants with a required solubility in both water and oil. With the development of simulated foods the search for antioxidants that are more effective in extending shelf life will gain further impetus.

In the class of texture agents I have included emulsifiers, stabilizers and thickening agents. In terms of quantity consumed they probably constitute the largest class of additives, being employed extensively in preparing bread, pastry, ice cream, frozen desserts, whipped products, margarine, candy and certain soft drinks and milk products. Many of the newer convenience foods have only become practicable as a result of the development of new and improved emulsifiers and stabilizers. Among other things, the texture agents permit oil to be dispersed in water, produce a smooth and even texture and supply the desired body and consistency of many food products.

The first emulsifiers were few in number and were either natural substances such as gums, alginates and soaps or synthetic substances of fairly simple composition. Their action was often variable. Progress in chemical synthesis has now made available a large number of new texture agents with characteristics suitable for almost any requirement. Among the most common emulsifiers and stabilizers, aside from the natural ones, are stearyl tartrate, complete glycerol esters, partial glycerol esters, partial polyglycerol esters, propylene glycol esters, monostearin sodium sulfoacetate, sorbitan esters of fatty acids and their polyoxyethylene derivatives, cellulose ethers and sodium carboxymethyl cellulose. Thickeners include natural products such as agar, alginates, celluloses, starches, vegetable gums, dextrins and pectin; modified celluloses such as methyl cellulose, and starches modified by bleaching, oxidizing and phosphating.

The miscellaneous group of additives is so numerous that I can only indicate a few of the functions they serve. Acids, alkalis, buffers and neutralizing agents are added to many processed foods where the degree of acidity or alkalinity is important; manufacturers of baked goods, soft drinks, chocolate and processed cheese employ these additives extensively. The baking industry also makes heavy use of bleaching and maturing agents, which render flour whiter and bring it to maturity sooner. Sequestrants are added to food to bind trace metals and thus prevent any oxidative activity the metals in an ionized state might have on the food; in shortening, for example, unsequestered metals could catalyze processes leading to rancidity. Humectants, which are hygroscopic, offset changes in the humidity of the environment to which food is exposed, so that a desired level of moisture can be maintained in a food product such as shredded coconut. Anticaking agents keep many salts and powders free-flowing. Glazing agents make certain food surfaces shiny and in some cases protect the product from spoiling. Firming and crisping agents prevent flaccidity in processed fruits and vegetables and also aid the coagulation of certain cheeses. Release agents help food to separate from surfaces it touches during manufacture or transport. Foaming agents coupled with propellants make whipped

ANTICAKING AGENTS
- Aluminum calcium silicate
- Calcium silicate
- Magnesium silicate
- Sodium aluminosilicate
- Sodium calcium aluminosilicate
- Tricalcium silicate

CHEMICAL PRESERVATIVES
- Ascorbic acid
- Ascorbyl palmitate
- Benzoic acid
- Butylated hydroxyanisole
- Butylated hydroxytoluene
- Calcium ascorbate
- Calcium propionate
- Calcium sorbate
- Caprylic acid
- Dilauryl thiodipropionate
- Erythorbic acid
- Gum guaiac
- Methylparaben
- Potassium bisulfite
- Potassium metabisulfite
- Potassium sorbate
- Propionic acid
- Propyl gallate
- Propylparaben
- Sodium ascorbate
- Sodium benzoate
- Sodium bisulfite
- Sodium metabisulfite
- Sodium propionate
- Sodium sorbate
- Sodium sulfite
- Sorbic acid
- Stannous chloride
- Sulfur dioxide
- Thiodipropionic acid
- Tocopherols

EMULSIFYING AGENTS
- Cholic acid
- Desoxycholic acid
- Diacetyl tartaric acid esters of mono- and diglycerides
- Glycocholic acid
- Mono- and diglycerides
- Monosodium phosphate derivatives of above
- Propylene glycol
- Ox bile extract
- Taurocholic acid

NUTRIENTS AND DIETARY SUPPLEMENTS
- Alanine
- Arginine
- Ascorbic acid
- Aspartic acid
- Biotin
- Calcium carbonate
- Calcium citrate
- Calcium glycerophosphate
- Calcium oxide
- Calcium pantothenate
- Calcium phosphate
- Calcium pyrophosphate
- Calcium sulfate
- Carotene
- Choline bitartrate
- Choline chloride
- Copper gluconate
- Cuprous iodide
- Cysteine
- Cystine
- Ferric phosphate
- Ferric pyrophosphate
- Ferric sodium pyrophosphate
- Ferrous gluconate
- Ferrous lactate
- Ferrous sulfate
- Glycine
- Histidine
- Inositol
- Iron, reduced
- Isoleucine
- Leucine
- Linoleic acid
- Lysine
- Magnesium oxide
- Magnesium phosphate
- Magnesium sulfate
- Manganese chloride
- Manganese citrate
- Manganese gluconate
- Manganese glycerophosphate
- Manganese hypophosphite
- Manganese sulfate
- Manganous oxide
- Mannitol
- Methionine
- Methionine hydroxy analogue
- Niacin
- Niacinamide
- D-pantothenyl alcohol
- Phenylalanine
- Potassium chloride
- Potassium glycerophosphate
- Potassium iodide
- Proline
- Pyridoxine hydrochloride
- Riboflavin
- Riboflavin-5-phosphate
- Serine
- Sodium pantothenate
- Sodium phosphate
- Sorbitol
- Thiamine hydrochloride
- Thiamine mononitrate
- Threonine
- Tocopherols
- Tocopherol acetate
- Tryptophane
- Tyrosine
- Valine
- Vitamin A
- Vitamin A acetate
- Vitamin A palmitate
- Vitamin B_{12}
- Vitamin D_2
- Vitamin D_3
- Zinc sulfate
- Zinc gluconate
- Zinc chloride
- Zinc oxide
- Zinc stearate

SEQUESTRANTS
- Calcium acetate
- Calcium chloride
- Calcium citrate
- Calcium diacetate
- Calcium gluconate
- Calcium hexametaphosphate
- Calcium phosphate, monobasic
- Calcium phytate
- Citric acid
- Dipotassium phosphate
- Disodium phosphate
- Isopropyl citrate
- Monoisopropyl citrate
- Potassium citrate
- Sodium acid phosphate
- Sodium citrate
- Sodium diacetate
- Sodium gluconate
- Sodium hexametaphosphate
- Sodium metaphosphate
- Sodium phosphate
- Sodium potassium tartrate
- Sodium pyrophosphate
- Sodium pyrophosphate, tetra
- Sodium tartrate
- Sodium thiosulfate
- Sodium tripolyphosphate
- Stearyl citrate
- Tartaric acid

STABILIZERS
- Acacia (gum arabic)
- Agar-agar
- Ammonium alginate
- Calcium alginate
- Carob bean gum
- Chondrus extract
- Ghatti gum
- Guar gum
- Potassium alginate
- Sodium alginate
- Sterculia (or karaya) gum
- Tragacanth

MISCELLANEOUS ADDITIVES
- Acetic acid
- Adipic acid
- Aluminum ammonium sulfate
- Aluminum potassium sulfate
- Aluminum sodium sulfate
- Aluminum sulfate
- Ammonium bicarbonate
- Ammonium carbonate
- Ammonium hydroxide
- Ammonium phosphate
- Ammonium sulfate
- Beeswax
- Bentonite
- Butane
- Caffeine
- Calcium carbonate
- Calcium chloride
- Calcium citrate
- Calcium gluconate
- Calcium hydroxide
- Calcium lactate
- Calcium oxide
- Calcium phosphate
- Caramel
- Carbon dioxide
- Carnauba wax
- Citric acid
- Dextrans
- Ethyl formate
- Glutamic acid
- Glutamic acid hydrochloride
- Glycerin
- Glyceryl monostearate
- Helium
- Hydrochloric acid
- Hydrogen peroxide
- Lactic acid
- Lecithin
- Magnesium carbonate
- Magnesium hydroxide
- Magnesium oxide
- Magnesium stearate
- Malic acid
- Methylcellulose
- Monoammonium glutamate
- Monopotassium glutamate
- Nitrogen
- Nitrous oxide
- Papain
- Phosphoric acid
- Potassium acid tartrate
- Potassium bicarbonate
- Potassium carbonate
- Potassium citrate
- Potassium hydroxide
- Potassium sulfate
- Propane
- Propylene glycol
- Rennet
- Silica aerogel
- Sodium acetate
- Sodium acid pyrophosphate
- Sodium aluminum phosphate
- Sodium bicarbonate
- Sodium carbonate
- Sodium citrate
- Sodium carboxy-methylcellulose
- Sodium caseinate
- Sodium citrate
- Sodium hydroxide
- Sodium pectinate
- Sodium phosphate
- Sodium potassium tartrate
- Sodium sesquicarbonate
- Sodium tripolyphosphate
- Succinic acid
- Sulfuric acid
- Tartaric acid
- Triacetin
- Triethyl citrate

SYNTHETIC FLAVORING SUBSTANCES
- Acetaldehyde
- Acetoin
- Aconitic acid
- Anethole
- Benzaldehyde
- *N*-butyric acid
- *d*- or *l*-carvone
- Cinnamaldehyde
- Citral
- Decanal
- Diacetyl
- Ethyl acetate
- Ethyl butyrate
- Ethyl vanillin
- Eugenol
- Geraniol
- Geranyl acetate
- Glycerol tributyrate
- Limonene
- Linalool
- Linalyl acetate
- 1-malic acid
- Methyl anthranilate
- 3-Methyl-3-phenyl glycidic acid ethyl ester
- Piperonal
- Vanillin

GROUP OF ADDITIVES included in the U.S. Food and Drug Administration's list of additives "generally recognized as safe" is given, except for large groups of natural flavors and oils. To be on this list an additive must have been in use before 1958 and have met certain specifications of safety. Additives brought into use since 1958 must be approved individually. Occasionally substances are removed from the list by the FDA in the light of new evidence; recent examples include the cyclamate sweeteners and saccharin.

toppings come out of their containers as a foam, whereas foam inhibitors have an opposite role, where a tendency to foam, as with pineapple juice, makes filling a container difficult. Clarifying agents remove small particles of minerals from liquids such as vinegar, which might otherwise turn cloudy. Solvents serve as carriers for flavors, colors and other additives, and solvent extraction is the method whereby oil is obtained from oilseeds, coffee is decaffeinated and a number of instant beverages are prepared.

Additives have become a public issue because of recurrent episodes bringing into question the safety of certain additives that have been used for some time. Cyclamates were tested extensively in the U.S. before they were put on the market as artificial sweeteners, but in 1969 it was reported that large doses had caused bladder cancer in rats. The U.S. Government ordered cyclamates off the market. Subsequently it was reported that rats fed with cyclamate and saccharin at a sixth of the dose that led to the original ban also developed bladder cancer. As a result saccharin is now being critically reviewed in the U.S. and other countries. Sodium nitrite, which fixes a red color in frankfurters, sausages and hams, is currently under review in many countries because of the possibility that it may form a cancer-producing agent during digestion and storage. Laboratory evidence has linked monosodium glutamate with the "Chinese restaurant syndrome" (more precisely Kwok's disease), a tightening of the muscles of the face and neck, occasionally accompanied by headache, nausea and giddiness, experienced by some people who have eaten in restaurants where monosodium glutamate has been used in large amounts. Many countries have therefore restricted the use of monosodium glutamate or required its presence in food to be prominently stated on the label.

Food additives, unlike the chemicals put in pesticide preparations, are not designed to be toxic, and most of them would have to be ingested in large single doses to produce acute toxic symptoms. Many additives by nature are of extremely low potential toxicity. It is therefore difficult to determine their possible hazards to man, even after exhaustive testing. It is probably true to say that there will always be an area of doubt concerning the possible effects of ingesting small amounts of additives over the course of a lifetime. One cannot be fully sure of the safety of an additive until it has been consumed by people of all ages in specified amounts over a long period of time and has been shown conclusively, by careful toxicological examination, to have no harmful effects.

Since humans cannot be used for testing by exposing them to unknown chemicals for a substantial period of time, tests are made on rats and other animals such as mice and dogs. Test animals are fed quantities of the additive that far exceed the amount likely to be found in food. Tests are made both for short periods and over the animal's lifetime and are often continued into succeeding generations. Any change in growth, body function, tissue and reproduction is reported, as is the incidence of tumors.

The largest dose that appears to produce no effects in animals is taken, and a safety factor reducing that dose by about 100 is applied in most countries in order to arrive at an acceptable dose for humans. The "acceptable daily intake" thus calculated is the daily intake that for an entire lifetime appears to be without appreciable risk on the basis of all known facts at the time. It is expressed in terms of milligrams of the additive per kilogram of body weight. One must then calculate how much of the additive a person might be expected to ingest in a day from all dietary sources and compare this figure with the acceptable daily intake in order to decide whether the applications of the additive should be permitted and whether the specific tolerances or maximum limits required for it by good manufacturing practices in individual foods are safe to the health of the consumer.

Many national authorities publish information on the tests they require for proposed additives. International guidelines have been published by the Joint Expert Committee on Food Additives of the Food and Agriculture Organization and the World Health Organization. They require a comprehensive series of tests on laboratory animals, including short- and long-term studies covering acute toxicity, metabolism of the additive and carcinogenic effects, among others.

Assessing safety on the basis of toxicological tests calls for expert judgment of all the evidence available. The judgment may have to be modified in the light of further experiments and experience with the additive in food for humans. The tests required to obtain approval of an additive can cost upward of $100,000.

It is therefore the hope of the government and the food industry in many countries that the work of such international bodies as the Codex Alimentarius Commission of the Food and Agriculture Organization and the World Health Organization will lead to a greater exchange of toxicological data and to the evaluation of the safety of more additives as soon as possible. The commission has published a list of six general principles on the use of food additives. The first is that the use of an additive is justified only when it has the purpose of maintaining a food's nutritional quality, enhancing its keeping quality or stability, making the food attractive, providing aid in processing, packing, transporting or storing food or providing essential components for foods for special diets, and that an additive is not justified if the proposed level of use constitutes a hazard to the consumer's health, if the additive causes a substantial reduction in the nutritive value of a food, if it disguises faulty quality or the use of processing and handling techniques that are not allowed, if it deceives the customer or if the desired effect can be obtained by other manufacturing processes that are economically and technologically satisfactory.

The second principle is that the amount of additive should not exceed the level reasonably required to achieve the desired effect under good manufacturing practice. The third principle calls for additives to conform with an approved standard of purity; the fourth holds that all additives, in use or proposed, should be subjected to adequate toxicological evaluation and that permitted additives should be kept under observation for possible deleterious effects; the fifth states that approval of an additive should be limited to specific foods for specific purposes and under specific conditions, and the sixth relates to the use of additives in foods consumed mainly by special groups in the community. In this case the intake of the food by the group should be taken into account before authorizing the use of the additive.

Food additives have become part of everyday life and undoubtedly will play an increasing role with advances in food technology. The prospect is not necessarily bad, because properly used additives can bring the consumer significant benefits. Moreover, provided that in each case sound justification for the additive is demonstrated, that government and manufacturers exercise the utmost care to ensure that the additive entails no appreciable risks to health and that clear labeling informs the consumer of the nature and composition of the product he is buying, consumers should be reasonably assured as to the safety of officially authorized food additives.

NONVISUAL LIGHT RECEPTION

In the brains of some vertebrates are unidentified organs that respond to daylight. Experiments with blinded birds demonstrate that these organs control biological rhythms

by Michael Menaker

Pattern vision, which we regard as being synonymous with seeing, is so important for our conscious perception of the world that we tend to assume that all effects of light are mediated by the retina of the eye and the associated areas of the brain. It may seem surprising, therefore, that the behavior of many animals with image-forming eyes is regulated by light in ways that do not depend on the perception of a visual pattern but only on the organism's ability to distinguish between light and dark.

The behavior that is regulated in this nonvisual manner is of two quite distinct kinds: day-to-day and year-to-year. For example, in most animals the 24-hour cycle of alternating light and darkness serves to synchronize certain physiological activities so that the rhythms of these activities also follow a cycle that is exactly 24 hours in length. Such rhythms are driven by a "biological clock," which, although it is precise in its own way, does not keep a 24-hour schedule. In the absence of cues from the environment, such as the natural daily cycle of light and darkness, the biological clock has a period close to, but not exactly, 24 hours. The rhythms the clock controls are therefore called circadian, from the Latin words for "approximately" and "day." When environmental cues are absent, the circadian cycle is "free-running."

In nature the free-running rhythms are forced by the cycle of night and day to maintain a length of cycle that exactly matches the natural 24-hour cycle. The process that synchronizes circadian rhythms in this precise way is called entrainment.

Many animals also respond to changes in the length of day and night that have a seasonal cycle. For example, the length of the day can serve as a reliable cue for the regulation and timing of reproductive cycles and other physiological processes that must be correlated with different times of the year. This response to seasonal changes in the length of the day is somewhat loosely called photoperiodism. The fact that plants as well as animals exhibit both photoperiodism and entrainment with respect to the daily cycle of light and darkness demonstrates that an organized eye is not essential for coupling an organism to the light cycles of the environment. Nevertheless, the effort to demonstrate a nonretinal photoreceptor in animals, particularly the higher vertebrates, has long been marked by controversy.

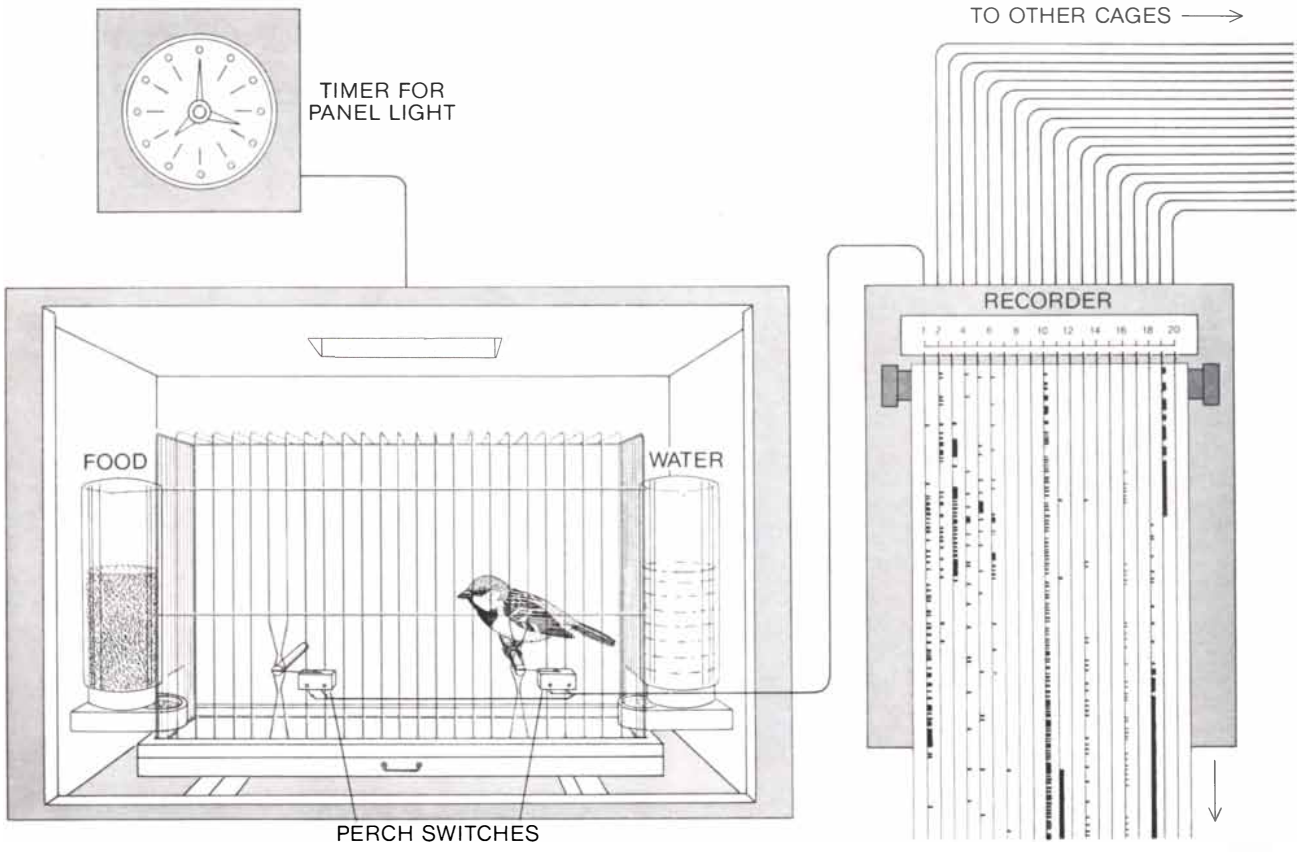
It is well established that among fishes, amphibians and reptiles the pineal organ, a small structure embedded in the top of the brain, and such associated structures as the parietal "eye" are sensitive to light. The function of the pineal organ in the life of the lower vertebrates remains uncertain, and it has not been shown that the structure acts as a light detector in either birds or mammals. In 1935 Jacques Benoit, who was then working at the University of Strasbourg Medical School, reported finding a light-sensitive area in the brain of Pekin ducks; direct illumination of the duck's head stimulates the growth of the duck's testes. Testis growth in ducks is a normal response to long days. Benoit's work was the first indication that complex image-forming eyes, in animals that possess them, are not necessarily the exclusive mediator of photoperiodism. Although Benoit and his colleagues have persevered in their study of extraretinal photoreception and have uncovered many fascinating phenomena, the photodetecting structure itself has not been identified.

Until 1967, when we began studying extraretinal photoreception in house sparrows at the University of Texas, Benoit's experiments had not been repeated outside his laboratory. Although ducks have some advantages as experimental animals, they also present certain technical difficulties. They are not only relatively large but also they (and many other birds) will not eat if they are kept in the dark or if they are blind. Because ducks are inconveniently large and because Benoit had to feed his ducks by hand the number he could maintain and study at any one time was limited. The difficulty of working with such animals may have discouraged other physiologists from repeating Benoit's experiments.

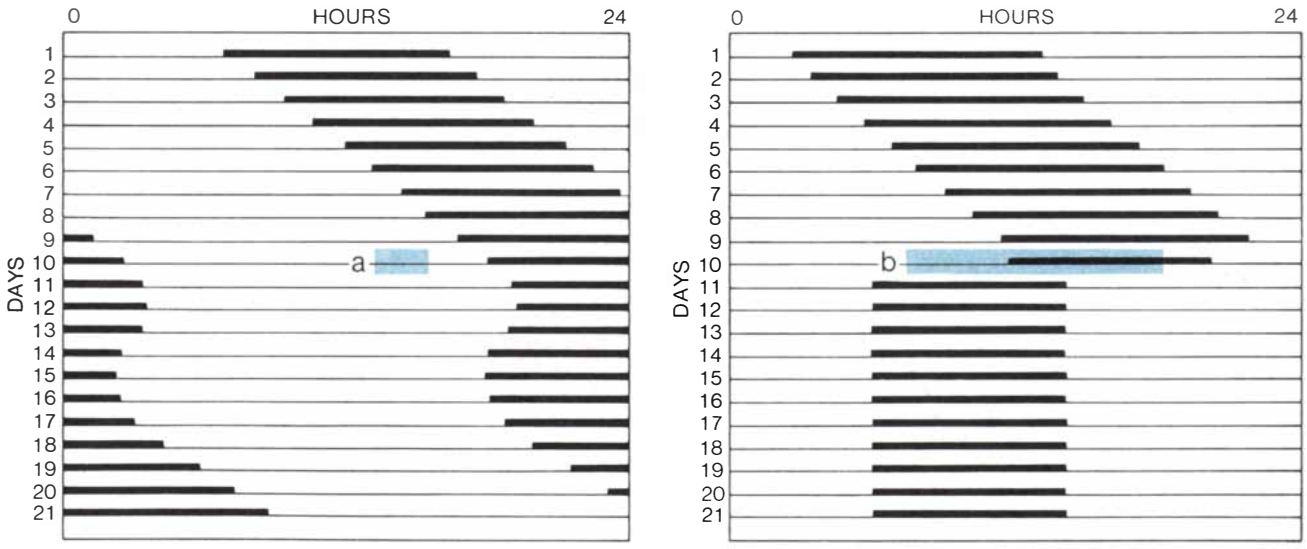
My students and I did not set out to repeat Benoit's experiments when we first began working with house sparrows. Our original intention was to investigate photoperiodism: the mechanism that enables birds with normal vision to synchronize their reproductive activities with the changing seasons. Knowing little about birds, we picked the house sparrow (*Passer domesticus*) primarily because of its toughness, adaptability and continuous availability. The house sparrow turned out to be a fortunate choice for quite different reasons.

Our experimental plan called for us to measure both a circadian rhythm and a photoperiodic response simultaneously in the same bird. Although there was a literature on photoperiodism in sparrows, there was no previous work on circadian rhythms. We therefore began to collect baseline data by recording the sparrow's locomotor activity when the ratio of light to darkness was varied artificially.

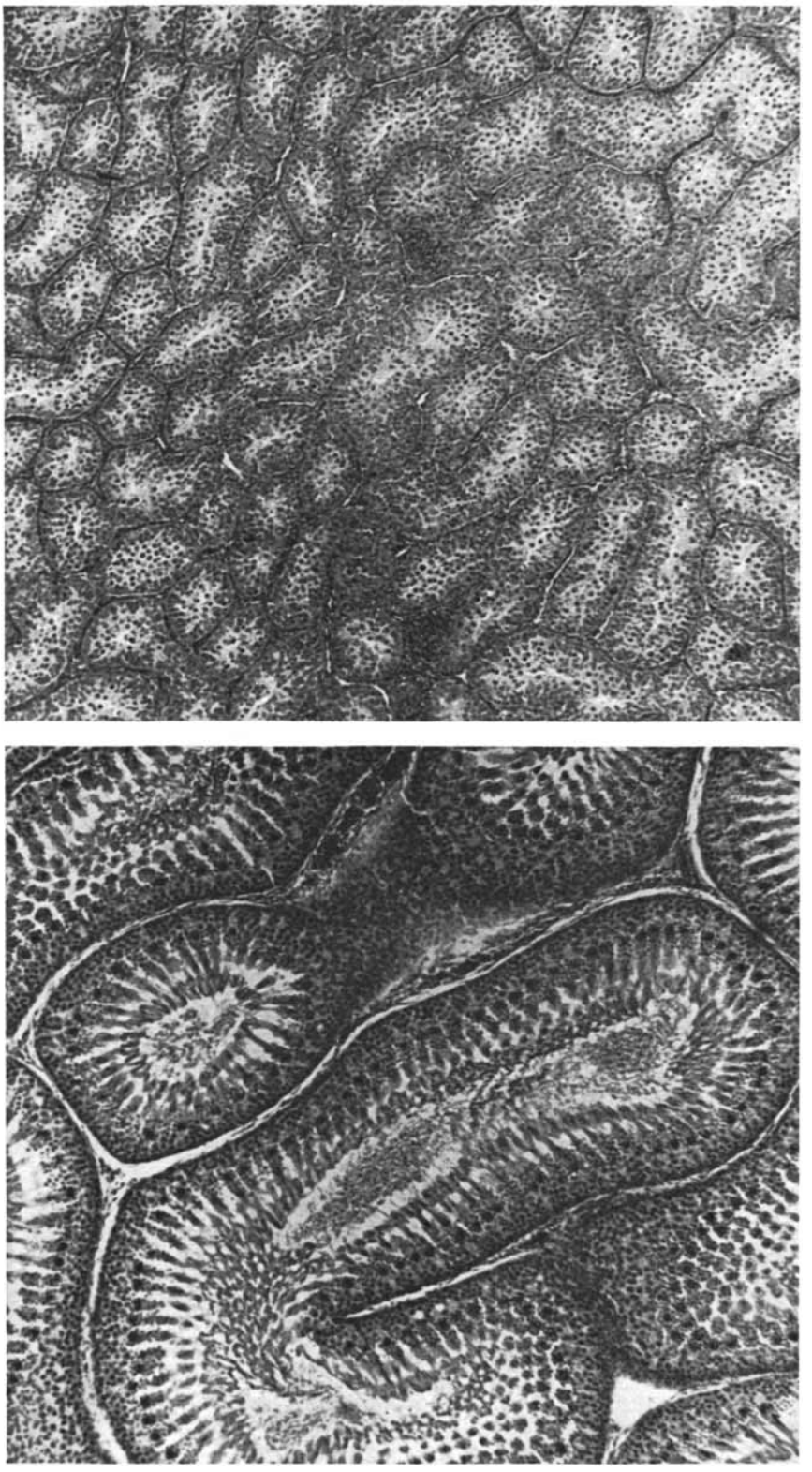
We used a reliable, automatic and rather inexpensive experimental chamber that had worked well with perching birds in other laboratories. Each bird is



RECORDING CAGE housed each of the sparrows used as subjects in the investigation. Movement of the bird from perch to perch or to its supply of food or water closed switch contacts and sent a signal to a recording pen that marked all such activity over a 24-hour period on a revolving drum. When a series of 24-hour records was assembled (*see illustration below*), the bird's near-to-daily, or "circadian," rhythmic pattern of activity and rest was seen to persist over long periods of time. It was also found that the bird's circadian rhythm, although readily shifted, would persist in continuous darkness or when blinded birds were the subjects.



DIFFERING PROPERTIES of circadian rhythms allowed an assessment of the nonretinal response to light by blinded sparrows. If a normal bird is kept in continuous darkness (*left*), its circadian rhythm becomes "free-running"; its daily periods of activity are unchanged in duration but start a little later each day. Even one brief exposure to light (*a*), however, will "reset" a free-running rhythm. The crucial role of light in governing circadian rhythm is further demonstrated (*right*) with respect to "entrainment." When a cage is illuminated for a set time each day (*b*), the occupant's activity develops an equally regular, although not necessarily synchronous, relation with the light cycle: it is entrained. If a blinded, regularly free-running sparrow should react to the cyclic illumination of its cage by exhibiting an entrained pattern of activity, it would be evidence that it is capable of nonretinal light perception.



QUIESCENT AND ENLARGED TESTES of sparrows, shown in section in these micrographs, are proof that a seasonal readiness to mate, although stimulated by longer hours of daylight, does not depend on the birds' eyesight; the donors of these specimens were blinded before the experiment began. Each bird was exposed for two months to a "day" of predetermined length in the laboratory. A bird illuminated for only six out of every 24 hours produced the testis tissue at top; the many small tubules are no more developed than those in the testes of blinded birds kept in continuous darkness for the same period. The tissue at bottom, seen at the same magnification, contains very large tubules. The donor's "days" were 16 hours long, and the testis development equals that of normal wild sparrows when they react to the longer hours of daylight that prepare them to mate.

individually housed in a lightproof box equipped with overhead lights operated by a timer. When the bird is active, it jumps on and off the perches in its cage a good deal of the time. The perches are connected to microswitches that automatically relay the bird's movements to a strip-chart recorder [*see top illustration on preceding page*].

Circadian rhythms are so widespread among animals that we were not surprised to find them in the house sparrow. Its locomotor rhythm is readily entrained by artificial light cycles. If the phase of the entraining cycle is shifted, there are three to six transient, or intermediate, cycles before the bird reorients its activity to the new regime. The phase angle, or temporal relation, between the entraining light cycle and the locomotor rhythm depends on several factors, chiefly the period at which the light cycle is presented and on the ratio of light to darkness.

If a bird that has been entrained to a particular light cycle is suddenly placed in constant darkness, the rhythm persists indefinitely with a period that is no longer exactly 24 hours long and that reflects the free-running rate of its internal clock. If a bird that is free-running in constant darkness is then exposed to light for a single interval of a few hours' duration, the bird responds with several transient cycles that lead to a shift in the phase of its free-running locomotor rhythm [*see bottom illustration on preceding page*]. If a free-running bird is placed in dim constant light, the length of the free-running period is found to depend on the light intensity. In moderately bright constant light free-running birds become continuously active and therefore arhythmic.

Other investigators have reported similar responses to artificial light cycles in the circadian rhythms of many other animals and plants. The observations raise fundamental questions about the nature of the responding biological clock or clocks. Here it is sufficient to note two facts that emerged early in our own work. The first is that house sparrows live perfectly well in constant darkness, untended except for a monthly change of food and water. The second fact is that their free-running period in constant darkness is seldom if ever exactly 24 hours. It is therefore possible to use entrainment to test for the existence of photoreception. The difference between an entrained bird and a free-running one is immediately obvious from inspection of the strip-chart record of loco-

motor activity. A bird will clearly not be entrained to a light cycle it cannot perceive.

The observation that the house sparrow lives well in constant darkness suggested to us that it would also tolerate being blinded. In fact, maintaining blinded sparrows turned out to be no more difficult than maintaining blinded rats or blinded mice. With characteristic adaptability our blinded sparrows learned the location not only of their food and water jars but also of their perches. They stayed healthy indefinitely and continued to produce activity records.

Knowing of Benoit's work with ducks, we placed a sparrow that had been blinded by bilateral optic enucleation in one of our experimental chambers and set the light timers, as we had done so often with birds that could see. Almost immediately the blind sparrow became entrained to the artificial light cycle. Further experiments showed that blinded birds followed the imposed light cycles no matter how the cycles were phased with respect to the real day [*see illustration on next page*].

It was always possible, of course, that the blinded birds were responding not to the light cycle but to secondary cues. We noted, for example, that the fluorescent lamps used in the experimental chambers produced a slight hum and raised the temperature in the chamber one or two degrees Celsius. In an early control experiment we found that if we wrapped the lamps with black tape, the birds no longer became entrained to the light-dark cycle. When we unwrapped the lamps, entrainment promptly followed. The wrapping eliminates both visible and ultraviolet radiation but does not affect the sound or the temperature rise produced by the lamp. Although the wrapping distorts infrared emission, it does not eliminate it. The results of the experiment were consistent with the hypothesis that the sparrows, although blind, were able to perceive visible light.

We had at this point in our study no idea about the sensitivity of whatever sensory system was involved. We therefore decided to try a much dimmer light source that in addition had a less complex emission spectrum. For other reasons entirely we had already built into our light-controlled boxes exactly the kind of light source we needed: small electroluminescent panels made by Sylvania, widely sold as "Nite-Lites." These panels produce a rather narrow band of radiation that peaks in the green part of the spectrum; they produce no sound, do not heat the chamber and emit negligible amounts of ultraviolet or infrared radiation. Furthermore, they produce only a thousandth as much visible light as the fluorescent lamps.

The amount of electroluminescent light received by our birds was roughly equivalent in intensity, although not in color, to bright moonlight. When we exposed blind sparrows to 12 hours of this faint light alternating with 12 hours of darkness, we found that about half of the birds became entrained and that the other half ignored the cycle and remained free-running as if they were in constant darkness. With this series of experiments we not only had strengthened the case for our hypothesis that sparrows possess a nonvisual light detector, which by this time we had begun to call the extraretinal receptor for entrainment, but also we had accidentally determined a crude operational threshold for the level at which the receptor begins to function.

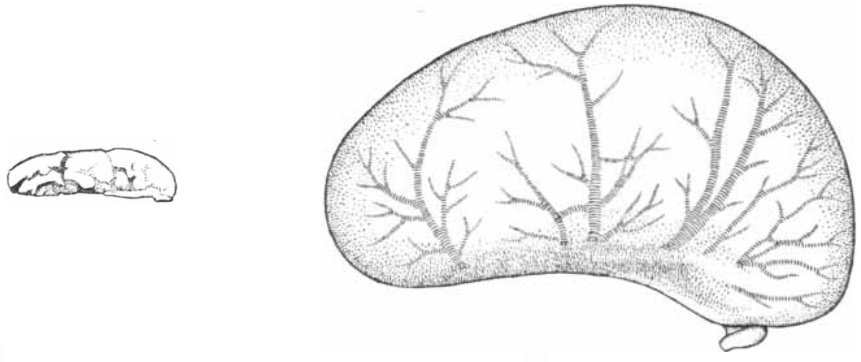
We were now compelled to take seriously a suggestion by one of our more skeptical colleagues that the entire phenomenon was being caused by the sparrow's numerous ectoparasites, which were somehow transmitting information about the light cycle to their host. It was certainly conceivable that the ectoparasites might be roused even by dim light and that their activity would in turn rouse the sparrows. No matter what measures we took, we could never be certain that our sparrows were completely free of parasites. On the other hand, removing as many of them as we could did nothing to alter the entrainment behavior in the slightest. We were finally able to convince ourselves we were dealing with a genuine extraretinal receptor on the basis of another series of experiments.

The new experiments were undertaken by Henry Keatts and me with the primary purpose of anatomically localizing the assumed photoreceptor. Again using electroluminescent light panels to provide the light cycle, we reduced the light intensity by adding pieces of black tape to the panel until, by becoming free-running, each blind bird demonstrated that it was not perceiving the light. Since we suspected that the light receptor was in the bird's brain, we looked for ways to manipulate the amount of light penetrating the skull.

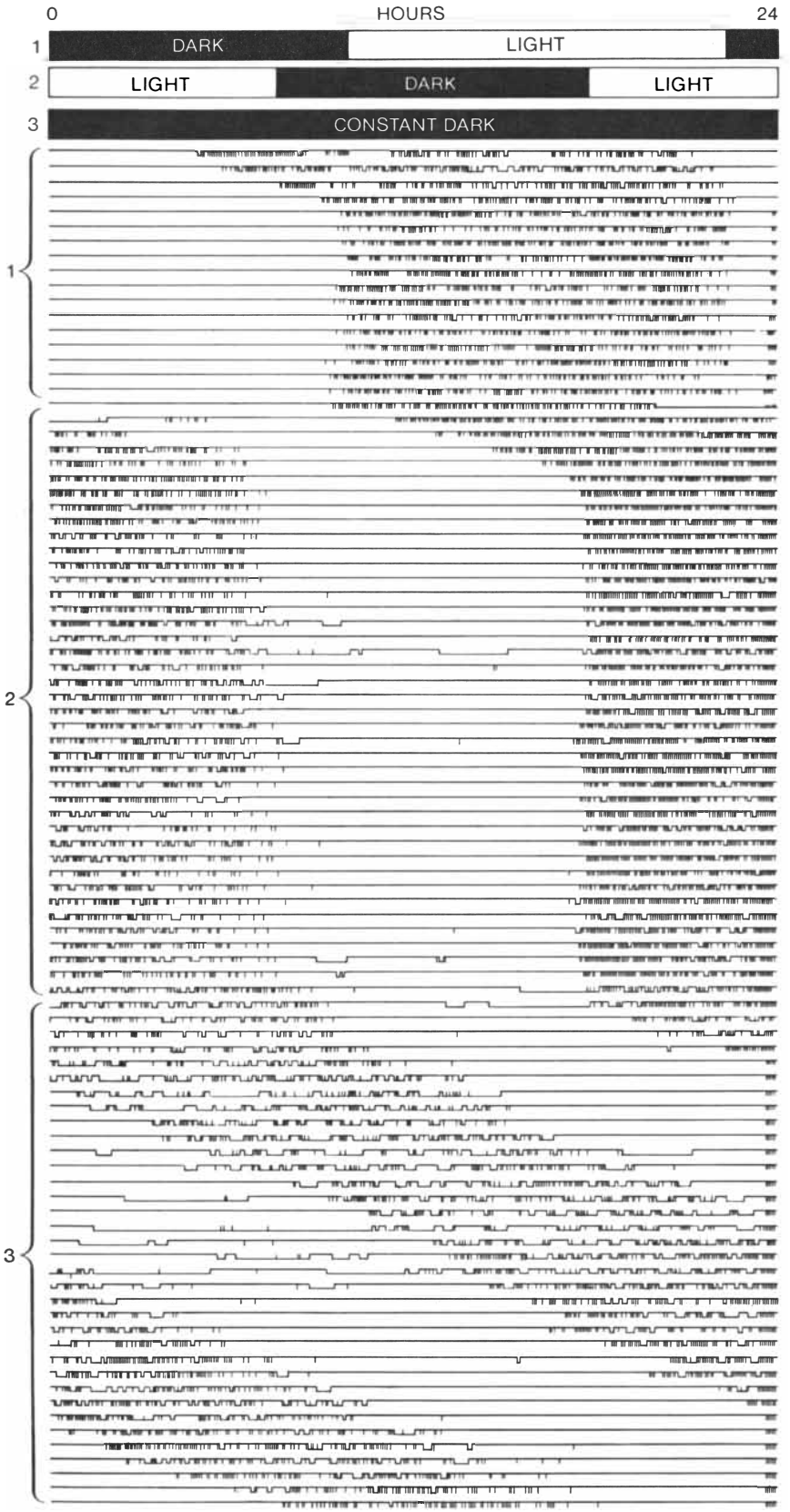
When we plucked feathers from the top of the birds' heads, birds that had previously been free-running became entrained to the light cycle. Feathers plucked from other parts of the body had no effect. Later we measured the amount of light shielding normally provided by the head feathers and were surprised to find that between 100 and 1,000 times more light reached the top surface of the brain when the birds' heads were plucked.

The next step was to restore the opacity of the birds' heads and see if we could reverse the effect of plucking. We did this by injecting small amounts of India ink under the skin of the bald birds and massaging the skin to produce a fairly uniform layer of ink between the skin and the skull. Birds so treated lost their sensitivity to light and again became free-running. Finally we removed a flap of skin and the ink under it; the birds again became entrained to the cycle of very dim light that had been present throughout the experiments.

We were now fairly confident that



EFFECT OF LIGHT in stimulating testis growth through some kind of nonvisual channel was demonstrated in the 1930's by the French biologist Jacques Benoit, who used ducks as experimental animals. The two testes reproduced here are seen one-third larger than actual size; the smaller is from a normal duck that had its head and eyes shielded from light during 28 nights of artificial illumination. The larger is from a blinded duck that was left bareheaded during 27 nights of illumination. Its testes became as big as a breeding duck's.



TWO ENTRAINMENTS appear in this consecutive record of a blinded sparrow's activity over a three-month period. At the start the daily period of cage illumination lasted from late morning to late evening (*Bar 1*) and the bird's circadian rhythm of activity soon came to match this period. After 18 days the pattern was shifted so that nights were illuminated and days were dark (*Bar 2*); prompt entrainment to the new pattern followed. When, 39 days later, the cage was darkened permanently (*Bar 3*), the rhythm became free-running.

light sensitivity was mediated by a photoreceptor in the brain similar to the one reported by Benoit in Pekin ducks but coupled to a different physiological system. In an attempt to further localize the receptor in sparrows we removed the pineal organ, a structure about the size of a mustard seed, from some of our blind sparrows to see if their light perception would be abolished. Entrainment persisted, ruling out the pineal as the sole site of photoreception in the brain. The interpretation of the result of these experiments, however, is complicated by the fact that removal of the pineal organ interferes with the circadian rhythm of locomotor activity itself. We cannot yet rule out the possibility that the pineal organ of sparrows, like the pineal of lower vertebrates, is photoreceptive and may constitute a part of a complex extraretinal photoreceptor for entrainment.

The eyes as well as the extraretinal photoreceptor appear to be involved in the light perception that affects the locomotor rhythm. Although the threshold for entrainment of blind birds is surprisingly low, the threshold of sighted birds is lower still: all sighted birds will entrain to the light from a single electroluminescent panel (about .1 lux, or .01 foot-candle), whereas only about half of blind birds will do so. Moreover, whereas sighted birds become continuously active in constant light of 50 to 500 lux, the activity of blind birds is still rhythmic at 2,000 lux or more. The relative contribution of the eyes and the extraretinal receptor, as well as the location of the extraretinal receptor in the brain and its physiology, remain subjects for continuing research.

Early in the course of our work with sparrows we naturally asked ourselves if brain photoreception was involved in the reproductive response to day length, as it is in ducks. Like most birds of the Temperate Zones, the house sparrow shows a dramatic seasonal cycle in the size and functional state of its gonads. In nature the timing of this cycle is largely regulated by the photoperiod; it can be readily manipulated in the laboratory merely by changing the length of the artificial days presented to the birds. Photoperiodism is least complex in male birds, and the most convenient indicator of the changes produced by light is the size or the weight of the testes.

Although many other organisms have annual reproductive cycles, birds in their seasonal changes in gonad weight sur-

pass by orders of magnitude comparable changes found in other vertebrates. This is easy to understand: natural selection operates with particular stringency in flying animals to reduce total body weight and expenditure of energy. The gonads are an obvious target for such selection because they are active only during a short annual period. For eight months of the year the testes of wild sparrows represent a negligible fraction of the birds' total body weight of about 24 grams. During the breeding season the testes grow rapidly until they average close to a gram.

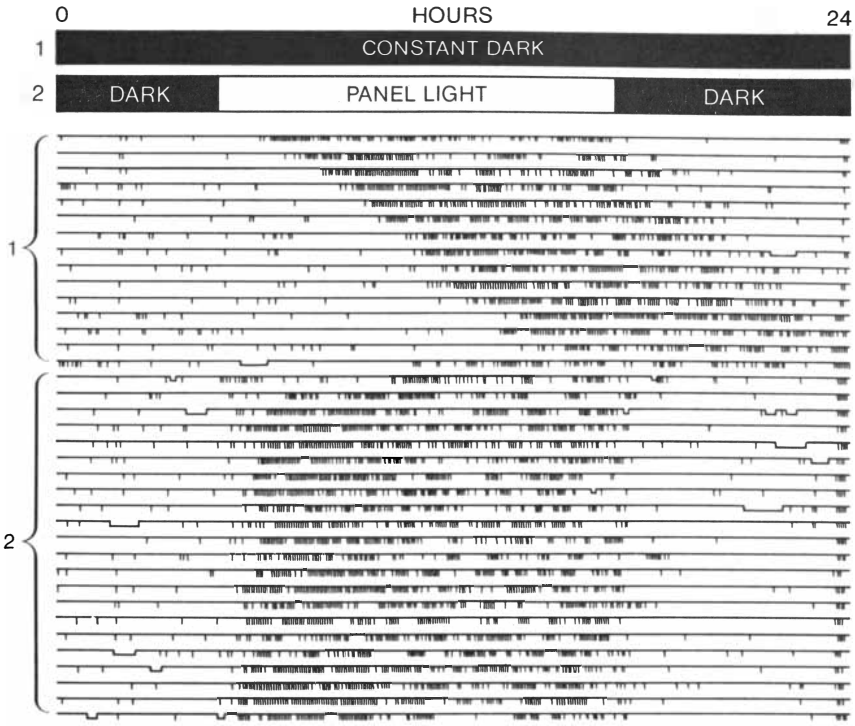
The annual regression of the gonads may enable the sparrow to survive with between 1 and 2 percent less food than it would otherwise need. If this does not seem like a very large saving, think of the commercial advantage that would accrue to a manufacturer who could reduce his budget for raw materials 2 percent below the budget of a competitor. Moreover, much of the effective reduction in energy requirement comes in winter, when food is scarce.

In order to learn whether or not the sparrow required a functioning retina to show testis growth in response to long days, we compared the response of a blind group and a sighted group to two months of artificial 16-hour days, beginning in the month of January. The testes of birds in both groups increased some 40 to 50 times in size. The testes of birds in two control groups, one normal and one blind, exposed to six-hour days for the same length of time showed no significant growth. In these initial experiments the number of birds used was too small to allow a quantitative comparison of the testis growth in blind and normal birds.

Herbert Underwood and I then conducted a series of experiments in which more than 500 birds were exposed to various lengths of day and intensities of light. The results demonstrated that the testes of blind and normal birds grew by the same amount and at the same rate when the birds were exposed equally to long days in winter. We concluded that sparrows, like ducks, have an extraretinal photoreceptor, presumably in the brain, that responds to seasonal changes in the number of daylight hours. We have called this receptor the extraretinal receptor for photoperiodism to emphasize that we reserve judgment about its relation to the extraretinal receptor for entrainment.

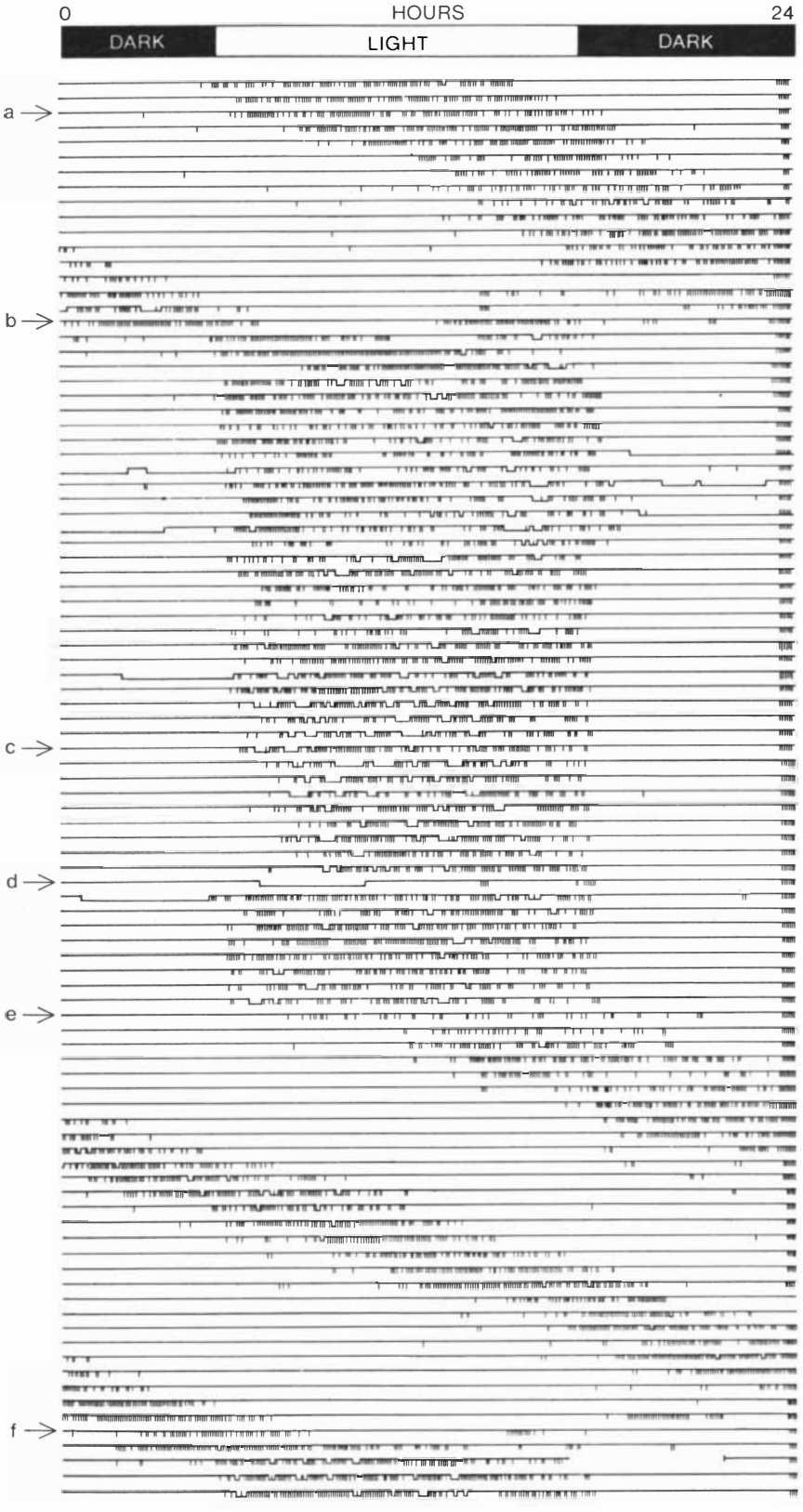
Our results led us to question whether the sparrow's eyes have anything at all to do with testis growth. In order to assess the role of the eyes one would like to remove the extraretinal receptor for photoperiodism. This we could not do, of course, because we did not know the receptor's precise location. Assuming that the receptor was somewhere in the brain, however, we could alter the amount of light reaching it without tampering with the amount of light reaching the eyes. Two experimental tricks were already open to us: we could pluck a bird's head feathers and thereby increase the amount of light passing through its skull, and by injecting India ink under the skin we could reduce the amount of light. We also knew from the work of George A. Bartholomew of the University of California at Los Angeles that in normal sparrows a light intensity of at least 10 lux is needed to stimulate slow but measurable growth of the testes.

Accordingly Richard Roberts, Jeffrey Elliot, Underwood and I set up an experiment in which two large groups of birds with normal sight and with gonads in the regressed condition were exposed to 16-hour days at a light intensity of 10 lux. We plucked the feathers from the heads of the birds in one group, effectively raising the light intensity above 10 lux, and injected ink under the skin of the fully feathered heads of the birds in the other group, effectively lowering the light intensity below the 10-lux threshold. Because the eyes of both groups were exposed to the 10-lux intensity, we reasoned that if birds' eyes were involved in photoperiodic photoreception, we should observe some testis growth in both groups. If, on the other hand, the eyes were not involved, we should see testis growth in the plucked birds and none in the India-ink group. This was exactly the result we obtained [*see bottom illustration on page* 29]. Furthermore, it indicates that the eyes are not involved in the photoperiodic response to light, although they are involved in entrainment.



MUCH FAINTER LIGHT, which also caused neither noise nor a temperature change in the cages, was used to determine the minimum illumination threshold for response. The source was an electroluminescent panel, emitting light with about the intensity of bright moonlight. Of the 20 subjects of the experiment, roughly half responded as shown by this typical record: when a blinded sparrow, free-running for two weeks in the dark (*Bar 1*), was exposed to the faint panel light (*Bar 2*), its circadian rhythm was promptly entrained.

At about the time we finished this study we learned of some ingenious experiments that had just been carried out at the University of Tokyo by Kazutaka Homma. He had made tiny beads incorporating a radioluminescent paint of the kind used to paint numerals on watch dials; he then implanted the beads in various regions of the brain and also directly in the eye of Japanese quail. Homma subjected his birds to short days,



RECEPTOR FOR ENTRAINMENT was proved by this experiment and similar ones to lie somewhere within the sparrow's skull. Light far weaker than the previously determined threshold value illuminated the cages throughout the experiment. When feathers were plucked from the birds' backs (*a*), their free-running rhythm stayed unchanged, but when the head feathers were plucked (*b*), entrainment followed. After 30 days (*c*) regrowth of the feathers started to affect entrainment, but a second plucking (*d*) restored it. When injections of ink made the birds' skulls opaque to light (*e*), the sparrows' rhythm became free-running again. When the ink was removed (*f*), however, entrainment was reestablished.

which would not normally cause their gonads to grow. He could thereby determine if the continuous light emanating from the luminous beads had any growth-stimulating effect. The experiments disclosed two general regions of the quail brain whose stimulation by light induced growth of the gonads. Homma also found that placing the luminous beads in the birds' eyes had no effect.

Our results with sparrows and Homma's with quail both argue very strongly that photoperiodic photoreception is mediated exclusively by brain receptors without any participation by the retina. Benoit, however, has maintained for some time that in ducks both the eyes and a brain photoreceptor are involved. He has come to this conclusion on the basis of experiments in which ducks with severed optic nerves showed slower testis growth in response to light than intact animals did. Do these conflicting results merely reflect differences among the species of bird used in the experiments?

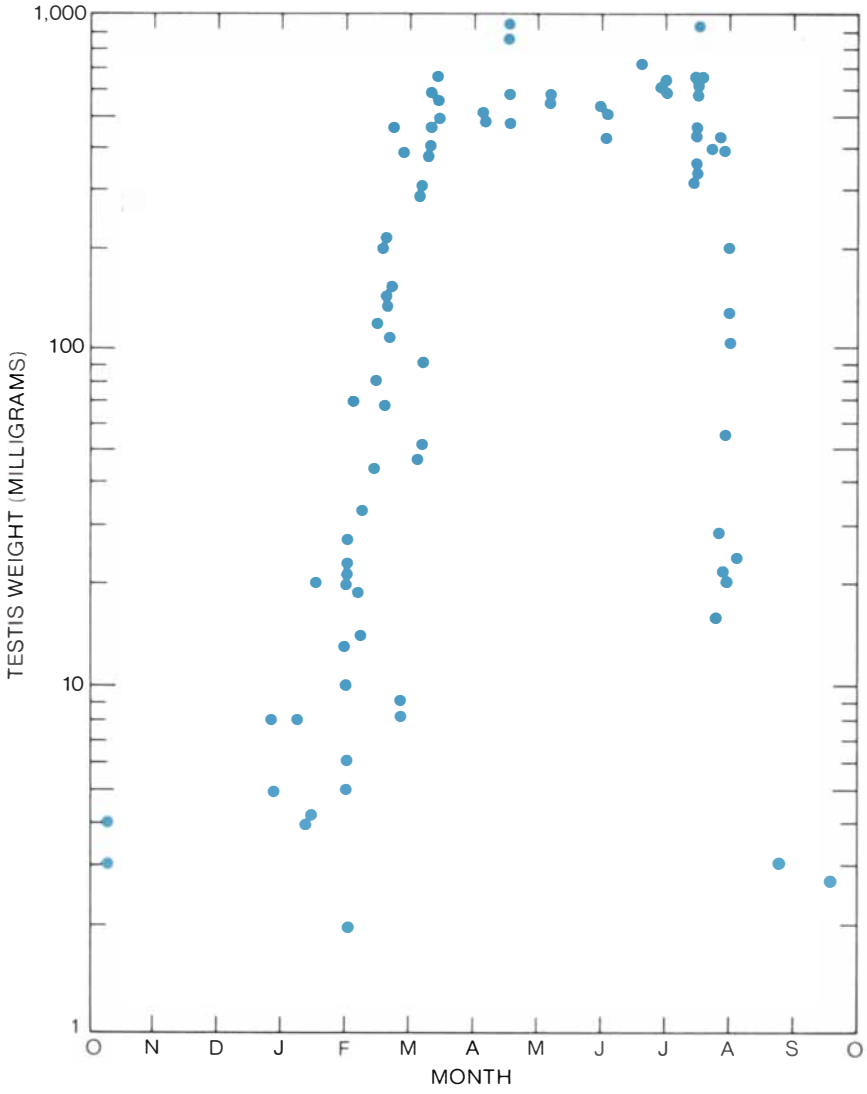
Perhaps, but it also seems possible that the experiments can be reconciled. In a recent extension of their earlier work Benoit, Ivan Assenmacher and their colleagues have come to the conclusion that the retinal response in ducks is much less sensitive to light than the response of the brain photoreceptor and that, unlike the brain receptor, the retina is sensitive only in the orange-red region of the spectrum. If this is true, then neither our results nor Homma's have adequately ruled out retinal participation. Homma's radioluminous beads emit only a very dim light. Our own experiments were based on threshold values that Bartholomew obtained with "daylight" fluorescent sources, whereas our experiments were performed with "cool white" fluorescent lights. Because the "daylight" bulbs emit more strongly in the red than "cool white" sources, it seems possible that there was not enough red light present in our experimental conditions to stimulate the red-sensitive retinal photoreceptors, but that enough light of other wavelengths was present to stimulate the brain receptor, with its greater sensitivity and range of response. If Benoit's proposed red-sensitive retinal receptors do exist generally among birds, their adaptive significance, when it is discovered, ought to be of great interest. Considering the degree of skepticism with which Benoit's findings were first greeted by biologists in 1935, it seems ironic that today it is the role of the

eyes and not the existence of brain photoreception that still remains unresolved.

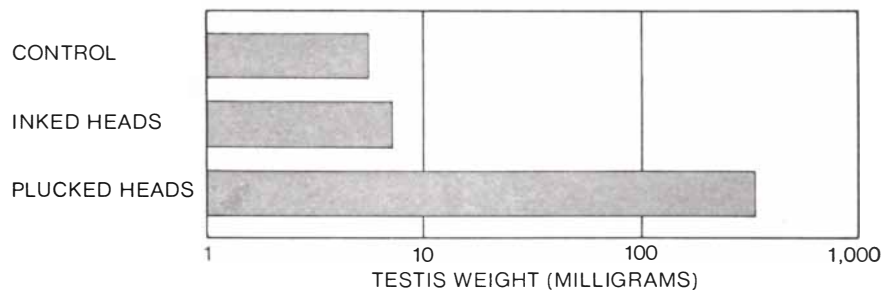
Extraretinal photoreception is not limited to birds. It is common in the invertebrates and may be a very general phenomenon among vertebrates. Underwood and I have shown in experiments with three species of lizard that locomotor rhythms can be entrained by artificial light cycles even after one has removed the lizards' two eyes and also the parietal "eye" on top of the head and the associated pineal organ. Other workers have reported various effects of light on the behavior of blinded fish and amphibians.

Working with sensitive photocells, William F. Ganong of the University of California School of Medicine in San Francisco has measured the penetration of visible light into the brain of the sheep, but in spite of a few tantalizing studies no one has yet demonstrated a clear-cut physiological response to extraretinally perceived light in any mammal. There are, in fact, several studies indicating that the circadian rhythms of blind mammals do not entrain to light cycles even when the light level is very high. Although it is too early to be certain, it seems possible that mammals are exceptional in not utilizing brain photoreceptors to mediate entrainment.

Many people, even when they are faced with the kind of evidence presented here, have difficulty accepting the fact that visible light does penetrate through structures they have always thought of as being opaque. The simplest way to remove any lingering doubts one may have on this score is to perform an experiment that most of us have done as children but have forgotten because the results appear to conflict with our everyday experience. In a completely dark room place an ordinary flashlight against your palm, switch it on and look at the back of your hand. This should convince you that light does penetrate living tissue, and furthermore that it is the long wavelengths that penetrate best. In effect the human hand, although it is almost an inch thick, is a moderately transparent red filter. The experiments described here, in addition to raising many specific questions concerning the ways in which environmental light cycles control rhythmic and reproductive events, underline the fact, of more general concern to biologists, that energy in the visible portion of the spectrum may well have effects on the activities of cells once thought to be completely shielded from it.



SEASONAL TESTIS GROWTH in a population of wild house sparrows in Texas begins in January and February and peaks in March and April. Regression, a return to nonmating dimensions, starts in July and early August. Each dot on the graph records the combined weight of the testes of one bird captured at the time indicated; the scale is logarithmic.



FINAL DEMONSTRATION proved that, like the receptor governing entrainment, the one governing testis growth must be located somewhere within the sparrow's skull. The subjects whose testis weight scarcely exceeded five milligrams (*top bar*) were a control group; these birds were sacrificed before the start of the experiment. The subjects whose testis weight remained below 10 milligrams (*center bar*) retained normal vision but, like the blinded birds in the preceding entrainment experiment, they were given ink injections to render their skulls opaque. The third group also had normal vision (*bottom bar*), but like the blinded birds in the preceding experiment their head feathers were plucked. Their testes enlarged greatly. Because both the inked and the plucked birds' vision was normal the experiment proves that eyesight is not involved in the birds' photoperiodic response.

Geosynclines, Mountains and Continent-building

A geosyncline is a huge deposit of sedimentary rock that forms at the edge of a continent. When it is compressed, it buckles up into a mountain range. It also enables a continent to grow by accretion

by Robert S. Dietz

A geosyncline is a long prism of sedimentary rock laid down on a subsiding region of the earth's crust. It has long been recognized that geosynclines are fundamental geologic units. Furthermore, it has been a dictum of geology that they eventually evolve into mountains consisting of folded sedimentary strata. The laying down of such sediments and their subsequent folding constitute a basic geologic cycle that requires a few hundred million years. Until recently the original nature of geosynclines has been inferred only by studying folded mountains. It was commonly believed that there are no nascent (unfolded) geosynclines in the world today, but this would defy another geologic dictum: that the present is the key to the past.

In recent years the study of marine geology has been revolutionized by the concept of plate tectonics, which holds that the earth's crust is divided into a mosaic of about eight rigid but shifting plates in which the continents are embedded and drift along as passive passengers. With this concept the evolution of ocean basins has been rather clearly resolved. The question arises: Must plate tectonics stay at sea, or is it also the prime mover of the geosyncline mountain-building cycle? In other words, can it account for the collapse of geosynclines and the growth of continents? I am among those who believe it can. Some notable advocates of this new concept of continental evolution are John Dewey and John M. Bird of the State University of New York at Albany, Andrew Mitchell and Harold Reading of the University of Oxford and William R. Dickinson of Stanford University.

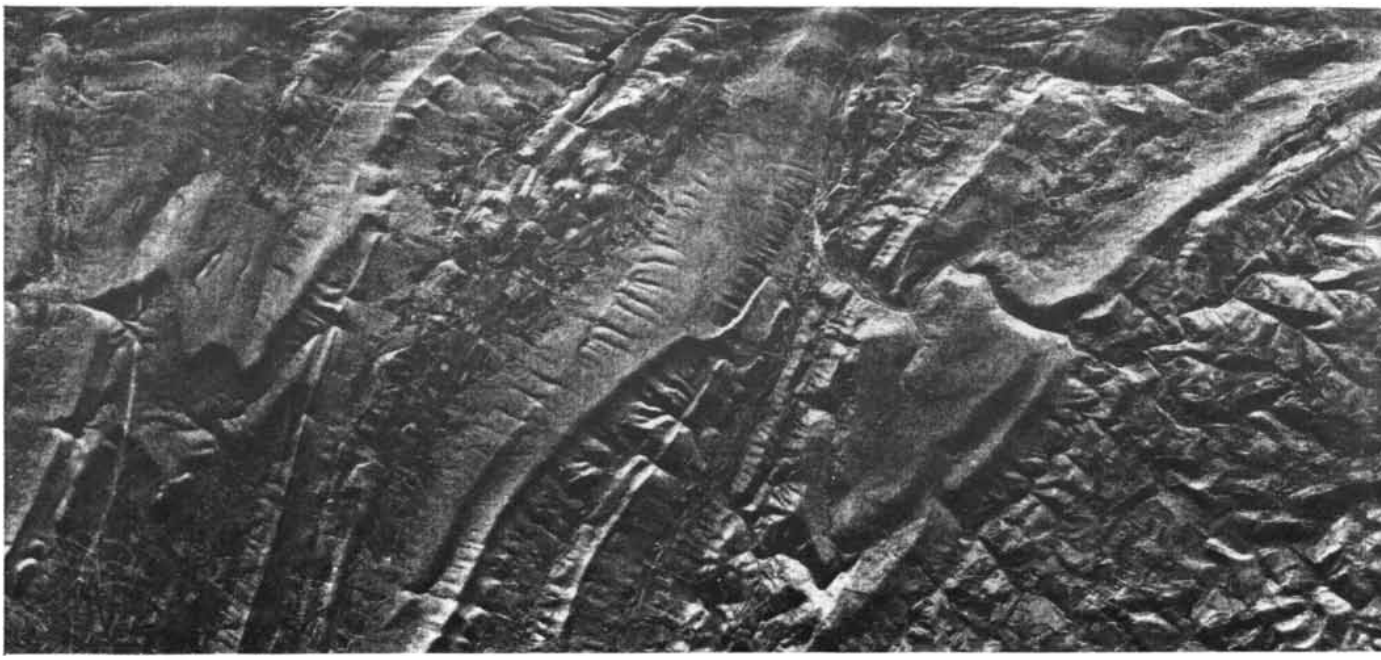
When one examines the structure of ancient folded mountains, one finds that the classic geosyncline is divided into a couplet: two adjacent and parallel structures consisting of a eugeosyncline (true geosyncline) and a miogeosyncline (lesser geosyncline), often shortened to eugeocline and miogeocline. Now that the ocean floor is becoming better known, one need not look far to find an example of the geosynclinal couplet in process of formation. A probable example of a "living" eugeocline is the continental rise that lies seaward of the continental slope off the eastern U.S. Landward of the rise and capping the continental shelf is a wedge of sediments that becomes progressively thicker as it extends toward the shelf edge. This wedge seems to be a living miogeocline.

In dimensions and in the overall character of its rocks and stratigraphy the modern continental-rise prism closely matches typical ancient eugeoclines. It parallels the Atlantic seaboard for 2,000 kilometers, forming an apron 250 kilometers wide from the continental slope to the abyssal plain [*see illustrations on pages 34 and 35*]. Seismic studies reveal that the rise is the top of a huge planoconvex lens of sediments whose maximum depth is about 10 kilometers. The sediments are turbidites, deposited by the muddy suspensions known as turbidity currents. Such suspensions periodically cascade down submarine canyons and pour across the continental rise, depositing sedimentary fans that eventually coalesce into an apron. Turbidites consist of thin graded beds of poorly sorted particles of silt and sand in which coarse material is at the base and finer material is at the top. The gradation in particle size reflects the differential rate of settling from a single injection of muddy sand. Interlayered with the graded beds are fine clays (pelagites) that slowly settle from the overlying water as a "gentle rain" between major influxes of turbidity currents.

Collapsed eugeoclines in ancient folded mountains are similarly composed of thick and repetitive sequences of turbidites; these strata are usually termed flysch or graywacke. Mixed with the graywackes are thin limestones, ironstones and cherts formed from the skeletons of radiolarians, indicating that the sediments were deposited in deep water. True fossils are sparse, but many eugeoclinal sequences of the lower Paleozoic era contain graptolites: extinct plantlike animals that settled down from the surface.

Close examination of the graded beds also reveals what are called sole mark-

COLLISION OF CONTINENTS is depicted in this view of the Zagros Mountains in Iran along the Persian Gulf taken from the spacecraft *Gemini 12* in November, 1966. The mountains are uplifted folds of sedimentary strata, originally deposited as a geosyncline, whose cores have been exposed by erosion. The foldbelt has apparently been thrown up by the collision of the Arabian block, rotating counterclockwise, with the Eurasian block, rotating clockwise. Since the Arabian block is part of the African block, the folding represents the collision between Africa and Eurasia. The Zagros Mountains and the shallow Persian Gulf are both part of the Arabian block that extends to the Red Sea. The suture between the Arabian block and the Eurasian block is marked by a major thrust fault that passes through the upper right corner of the photograph just beyond the mountain chains.

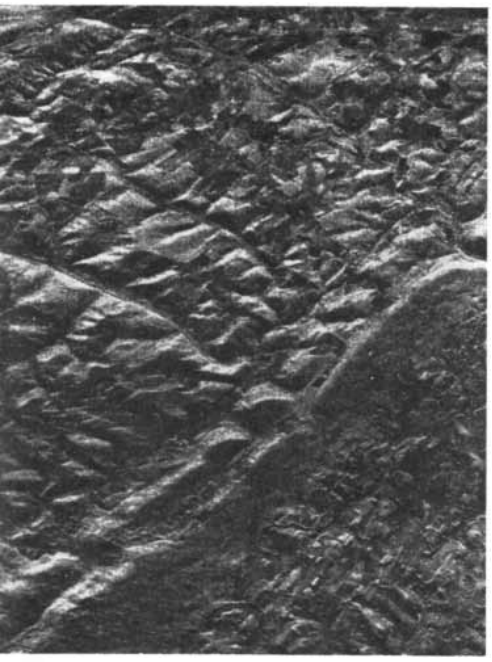


FOLDED APPALACHIAN MOUNTAINS in western Pennsylvania are depicted in this image produced by side-looking radar. The picture covers a region 25 miles long parallel to the Maryland border, centered approximately at 78 degrees 45 minutes west longitude. The picture is printed with north at the bottom so that the landscape appears to be lighted from the top. (The illuminant, of course, is the radar beam transmitted from an airplane.) If the picture is inverted, features that are actually elevated appear depressed and vice versa. This image and the sequence of three views at the bottom of these two pages were made by the National Aeronautics

ings or "flysch figures," for example ripple marks of a kind that could have been produced by turbidity currents. There can be little doubt that most of these sequences are the uplifted and eroded remnants of former continental-rise prisms. The crystalline Appalachians, which are that part of the Appalachians lying seaward of the Blue Ridge Mountains and equivalent ranges to the north and south, bear the clear imprint of being a collapsed continental-rise prism laid down in the early Paleozoic some 450 to 600 million years ago. The original prism has been much altered by intrusions and metamorphism.

The sedimentary wedge that underlies the coastal plain and continental shelf along the Atlantic seaboard appears to be an actively growing miogeocline. The wedge thickens as it progresses seaward, attaining a total thickness of between three and five kilometers along the shelf edge. Laid down on a basement of Paleozoic rocks, the wedge is composed of well-sorted shallow-water sediments deposited during the past 150 million years under conditions much like those of today. The stratified beds exhibit characteristics indicating they were deposited across the continental shelf in alluvial plains, in lagoons, along shorelines and offshore. Taking into account expected changes in the pattern of sedimentation over geologic time, the present Atlantic marine

APPALACHIAN FOLDBELT north of Harrisburg, Pa., is an extension of the foldbelt shown at the top of the page. In these side-looking radar views north is at the right. The three pictures cover a distance of 75 miles from just south of Mechanicsburg to the vicinity of a town called Jersey Shore on the West Branch of Susquehanna River. The Susquehanna River itself appears in the first frame at the left. The folded Appalachians were probably created in a late compressional stage of the collision between Africa and North America more than 450 million years ago, which caused "rugfolds" in the strata of sedimentary rock that formed part of a



and Space Administration in collaboration with the Remote Sensing Laboratory of the University of Kansas. The *K*-band radar system that produced the images was built by the Westinghouse Electric Corporation.

deposits closely resemble the ancient miogeoclinal foldbelts of the Paleozoic era and earlier. For example, the modern sedimentary wedge is much like the one found in the folded Appalachians of Pennsylvania. Both wedges are characterized by "thickening out," signifying that they grow steadily thicker toward the east before they abruptly terminate.

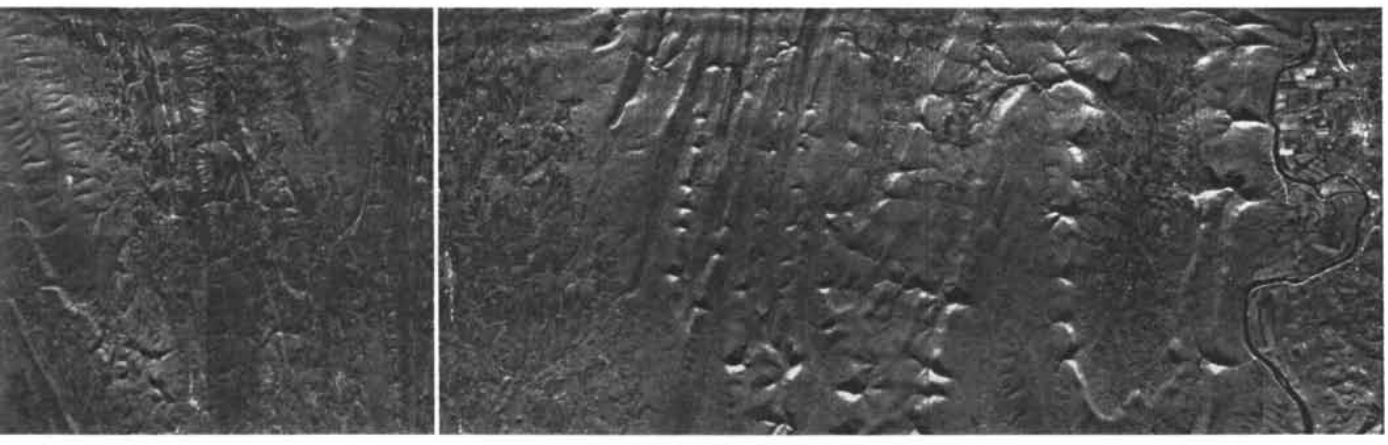
If the foregoing analysis is correct, one must conclude that geosynclines are actively forming along many continental margins today: eugeoclines at the base of the continental slope and miogeoclines capping the continental shelves. It remains to be shown, however, that the crustal shifting associated with plate tectonics can convert these sedimentary prisms into the mountainous foldbelts that make up the fabric of the continents, mostly as ancient eroded mountain roots rather than as modern mountain belts. In order to examine this possibility we must first summarize some of the basic concepts of plate tectonics.

The approximately eight rigid but shifting plates into which the earth is currently divided are thought to be about 100 kilometers thick. Most of the plates support at least one massive continental plateau, often referred to as a craton. We can visualize the ideal plate as being rectangular, although only the plate supporting the Indian craton approaches this simple shape. Along one edge of a crustal plate there is a subduction zone, usually marked by a trench, where the plate dives steeply into the earth's mantle, attaining a depth as great as 700 kilometers before being fully absorbed into the mantle. On the opposite side of the plate from the subduction zone is a mid-ocean rift, or pull-apart zone. As the rift opens, the gap is quickly healed from below by the inflow of liquid basalt and quasi-solid mantle rock. The other two opposed sides of the plate, connecting the rifts to the trenches, are shears called transform faults.

Hence three types of plate boundary are possible: divergent junctures (the mid-ocean rifts where new ocean crust is created), shear junctures (the transform faults where the plates slip laterally past one another, so that crust is conserved) and convergent junctures (trenches where two plates collide, with one being subducted and consumed). Only the last of the three, the convergent juncture, can help to explain how the sedimentary prism of a submarine geosyncline might be collapsed into a folded range of mountains. As the plate carrying a prism collides with a plate carrying a continental craton one would expect the prism to be compressed into folds. Thrusting and crustal thickening would follow, assisted by isostatic forces that act to keep adjacent crustal masses in balance. Such forces would cause the collapsed prism to be uplifted. The entire process would be accompanied by the generation and intrusion of magma, together with extensive metamorphism of the crustal rocks.

A grand theme of plate tectonics is that ocean basins are not fixed in size or shape; they are either opening or closing. Today the Atlantic Ocean is opening and the Pacific Ocean is closing. The drifting of the continents is another theme; every continent must have a leading edge and a trailing edge. For the past 200 million years the Pacific coast of North America has been the leading edge and the Atlantic coast the trailing edge. The trailing margin is tectonically stable, and since the continental divide is near the mountainous Pacific rim, most of the sediments are ultimately dumped into the Atlantic Ocean, including the Gulf of Mexico. Therefore it is primarily along a trailing edge that the great geosynclinal prisms are deposited.

Consider, however, what would hap-



Paleozoic geosyncline (*see illustration on page 36*). After folding the region was eroded to a level plain and then uplifted. The modern mountains were subsequently etched out according to the hardness of the various strata. Thus the ridges are composed mainly of dense sandstone and can be either synclines (troughs) or anticlines (arches). The *V*-shaped chevrons in the first frame are synclines that plunge to the northeast. The Susquehanna River established its course when the entire region was reduced to a level surface, so that its course has been superimposed on the folded structure, thereby cutting directly across the folds and creating water gaps.

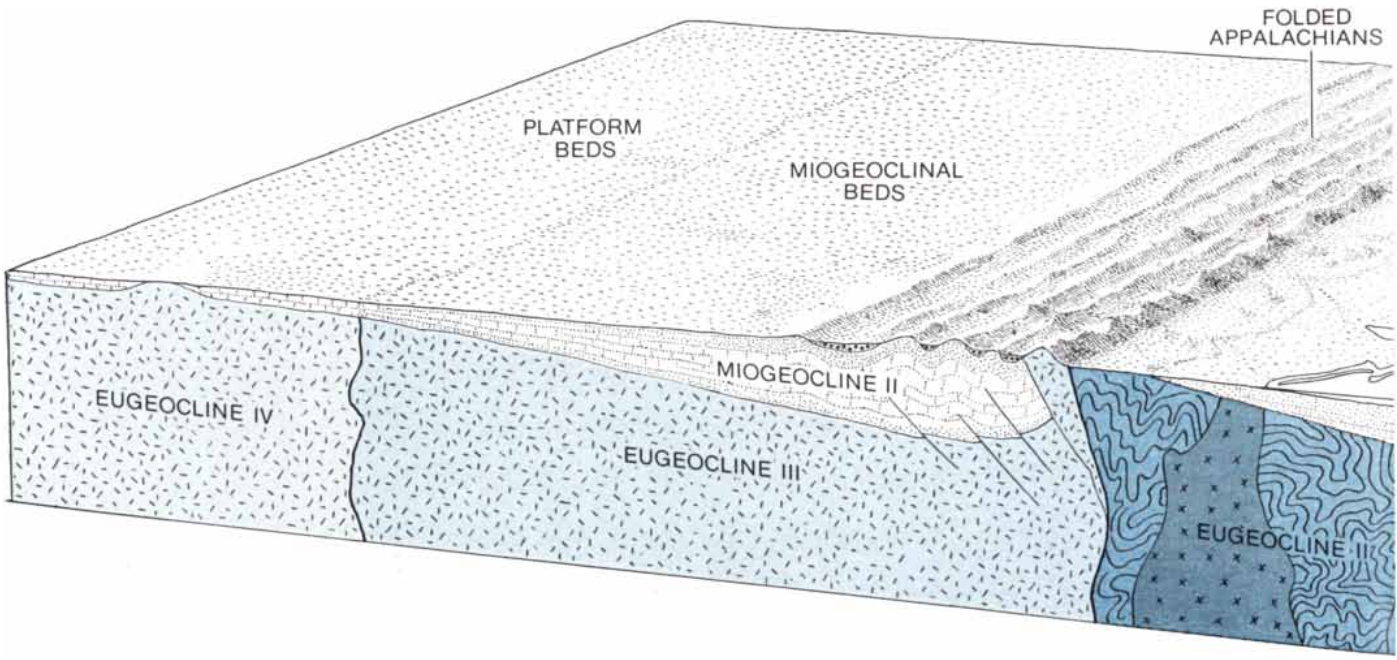
pen if, with the changing patterns of plate motions, a subduction zone (a trench) were created along a former trailing edge, forming a new plate boundary. The Atlantic would be transformed into a closing ocean with its geosynclinal prisms riding toward the trench. The continental margin and the trench would eventually collide, collapsing the eugeocline into a contorted mountainous foldbelt and also folding the miogeocline to a much lesser extent. Before that happened the continental margin would encounter and incorporate an island arc, similar to the island arcs found along the perimeter of the western Pacific. These arcs are created by tectonic and magmatic activity triggered by the plunging crustal plate. It is also quite possible that the Atlantic Ocean would close entirely, causing North Africa to collide with eastern North America. The collision of India with the underbelly of Asia, throwing up the Himalaya rampart, would be a present-day analogy. One can imagine many possible scenarios, depending on the geometry of plate boundaries and other variables.

The creation of a eugeoclinal foldbelt is of course considerably more than simply the accordion-like collapse of a continental-rise prism. The foldbelt is sheared into thrust faults and the landward edge of the eugeocline is commonly thrust onto the adjacent miogeocline. The descending crustal plate is not entirely consumed within the earth's hot mantle, with the result that low-density magmas buoyantly rise and invade the eugeocline. This leads to intrusions of granodiorite (a granite-quartz rock) and the growth of volcanic mountains consisting of andesite (the rock characteristic of the Andes). This lava is highly explosive because it is charged with water sweated out of the descending plate. Magma is not generated from the plunging lithosphere until it has reached a considerable depth. As a result the eugeocline can be subdivided into two parallel geologic belts. Toward the sea one finds sedimentary rock transformed at high pressure and low temperature; farther inland the sedimentary rock has been altered predominantly at low pressure and high temperature by the numerous intrusions of magma. From the new marginal mountain range, delta and river deposits sweep back across the continent, covering the miogeocline with a suite of continental shales and conglomerates collectively called molasse.

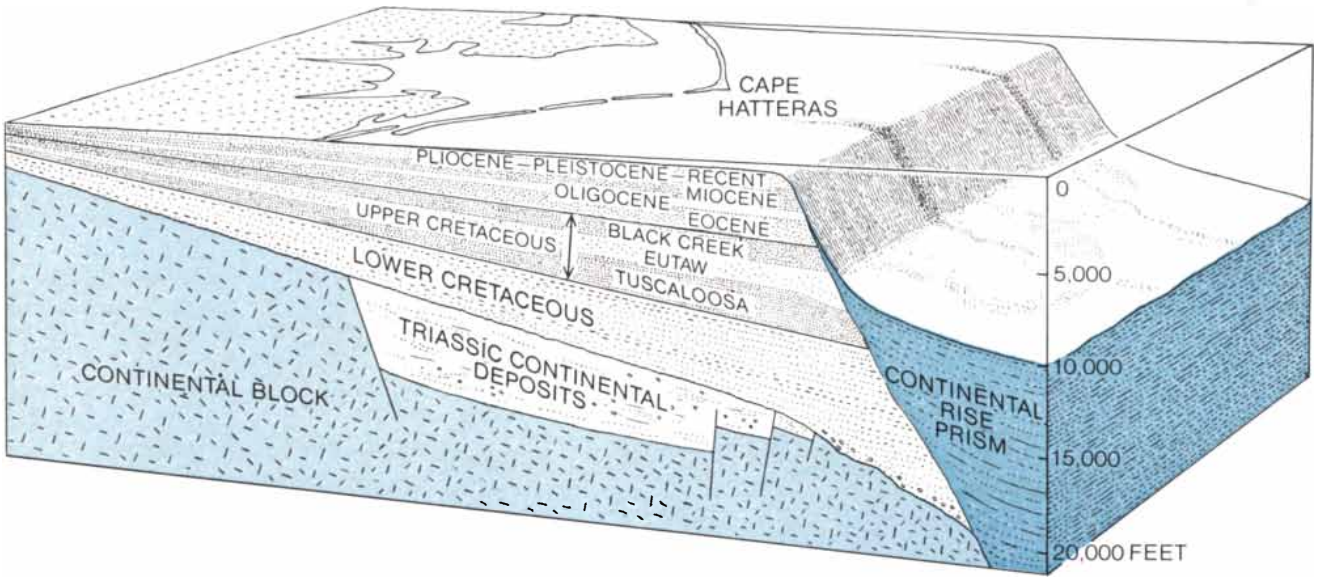
The concept that the geosynclinal cycle is controlled by plate tectonics provides some new answers to old questions about geosynclines. For example, is mountain-building periodic and worldwide or is it random in space and time? The answer must be both yes and no. On the one hand, the crustal plates are highly intermeshed; the drift of any one plate has global repercussions, giving rise to synchronous mountain-building. Any brief interval of rapid plate motion would also be one of widespread mountain-building. On the other hand, the rate of plate convergence is highly dependent on the latitudinal distance from the relative pole of rotation of that plate and on the particular geometry of the plate boundary.

A law of plate tectonics states that seafloor spreading (injection of new ocean crust) proceeds at right angles to a rift; the crustal plate, however, may be subducted into a trench at any angle. The rate of subduction and the attendant distorting of the crust therefore vary greatly from place to place, as can be observed on the perimeter of the Pacific today. Thus it would seem that although mountain-building over the span of geologic time may reach crescendos, it must also be continuous and random.

The plate-tectonic version of the geosynclinal cycle predicts that miogeoclines are ensialic, or laid down on continental crust (sial), whereas eugeoclines are ensimatic, or deposited on oceanic crust (sima). This differs from the earlier view that all geosynclines are ensialic,



SUCCESSION OF EUGEOCLINES underlies nearly all North America below the relatively undisturbed cover beds. These contorted and intruded prisms constitute the fundamental fabric of continents, known as the basement complex. A new geosynclinal couplet is being deposited today. It consists of a miogeocline (lesser geosyncline) of shallow water beds that caps the coastal plain and continental shelf paralleled by a eugeocline (true geosyncline) that is formed at the base of the continental slope by detritus



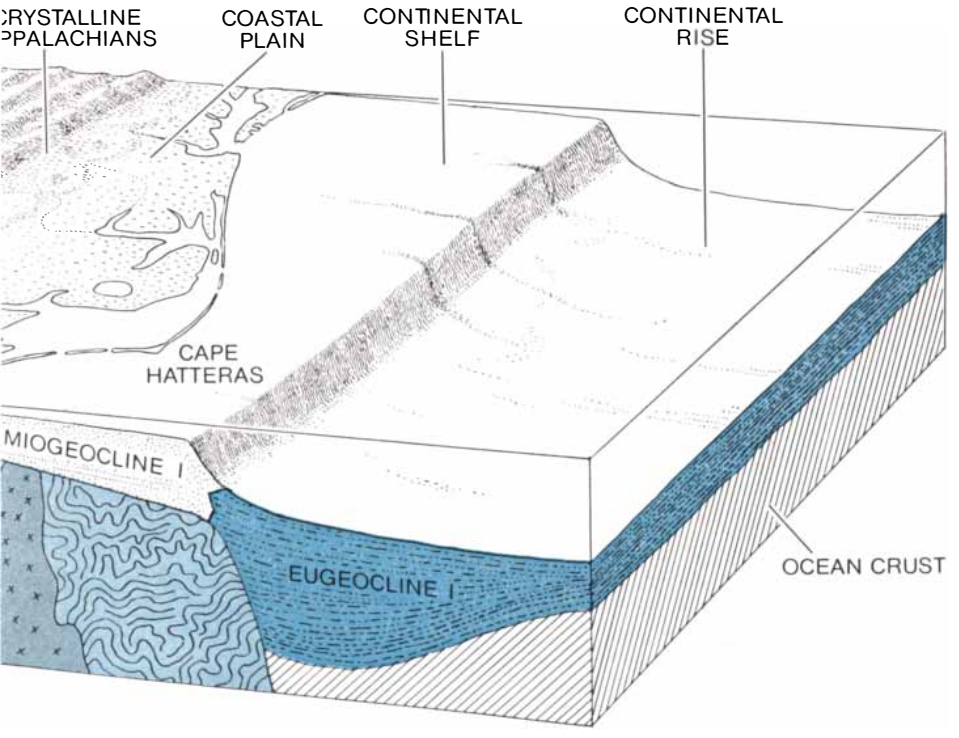
LIVING GEOSYNCLINAL COUPLET off the Atlantic coast of the U.S. consists of a miogeocline, strata laid down on the shallow continental shelf during the past 150 million years, and a eugeocline prism (*dark color*), consisting of thin beds of sand and mud deposited by turbidity currents flowing down the continental slope. The material in the Triassic basin represents continental deposits laid down before the foundering of the continental margin under tension 190 million years ago, prior to opening of Atlantic Ocean.

which is certainly incorrect. Early investigators observed that a granitic basement is invariably present under miogeoclines and evidently reasoned that a similar sialic basement, although it was unseen, must also be present under eugeoclines. A collapsed eugeocline is as thick as the continental plate, about 35 kilometers, so that its basement is beyond the depth of even the deepest boreholes. We can infer that the ultimate basement is simatic, however, by observing detached fragments that are caught up in the contorted mélange of the eugeoclinal pile. These fragments include samples of oceanic crust (for example radiolarian cherts and sodium-altered lavas) and upper-mantle rocks (for example serpentinites and peridotites).

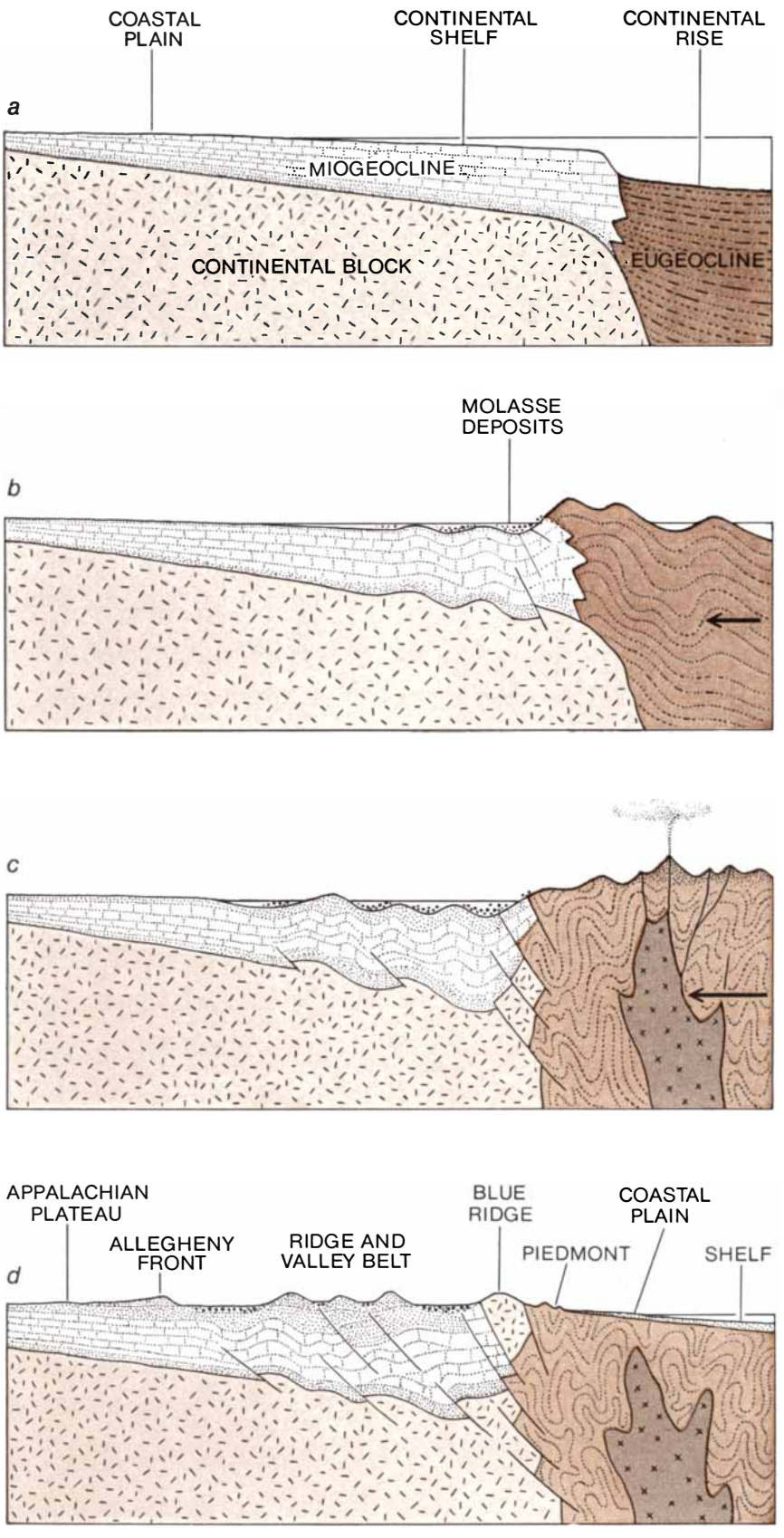
The ensimatic location of eugeoclines can also account for their tectonic style. They are tightly folded, faulted, tumbled and dynamically metamorphosed into an almost unmappable mélange. This contorted state is understandable, since the ocean floor is shearing under the eugeocline and thrusting the sedimentary pile against the continental slope. Extensive tectonic thickening and interleaving must occur before the pile will rise to mountainous altitudes. On the other hand, the miogeocline beds are protected by the stable continental slab, so that they are simply thrown into a series of loose, open, ruglike folds.

It now seems amusing to recall that 19th-century geologists, using a wrinkled apple as an analogy, interpreted folded mountains to mean that the earth was shrinking. Today it seems clear that eugeoclines are deposited at the edge of a continent on oceanic crust seaward of the continental slope, so that folded mountains really show that the continents are growing larger through marginal accretion. Mountain-building is therefore evidence of an even more fundamental geologic process: the growth of continents. The continents grow not as a layer cake but as a craton that is divided vertically into zones with an old nucleus and young margins.

An important aspect of geosynclines requiring explanation is that they are laid down on foundations that are continuously subsiding. This aspect is par-



washed over the shelf edge. If at some future time the sea floor were to thrust against the continent, the modern eugeocline (*I*) would collapse into a new foldbelt like the earlier ones. The hypothetical mechanism that creates foldbelts is shown on page 37. This diagram and others are based on drawings by the author's colleague John C. Holden.



CRUMPLING OF EARLIER GEOSYNCLINAL COUPLET, apparently laid down in late Precambrian and early Paleozoic time more than 450 million years ago, produced the Appalachian foldbelt. The four-part sequence shows how the miogeocline, or western part of the geosynclinal couplet, was folded into the series of ridges between the Blue Ridge line and the Allegheny front. The eugeocline, altered by heat, pressure and volcanism, formed a lofty range of mountains, now almost completely eroded, east of the Blue Ridge line.

ticularly evident in miogeoclines, which can attain a total thickness at their seaward edge of five kilometers even though they are entirely composed of beds deposited in shallow water. This phenomenon is nicely accounted for by plate tectonics: the margins of rift oceans inherently have, as one geologist has expressed it, a "certain sinking feeling."

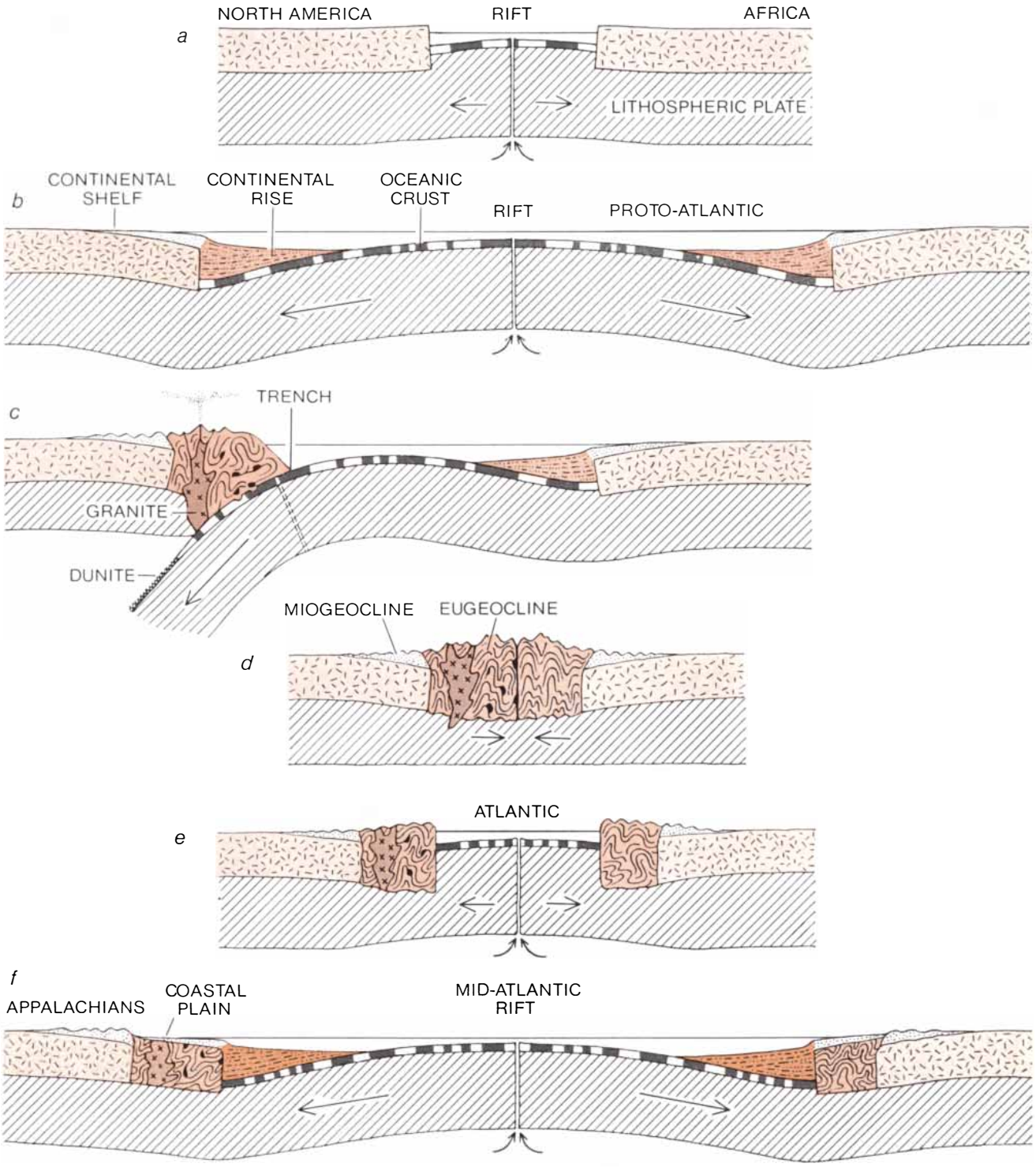
Let us take as an example the Atlantic Ocean between the U.S. and the bulge of Africa. This new ocean basin was created about 180 million years ago by the insertion of a spreading rift that split North America away from Africa [*see illustration on opposite page*]. Attendant swelling of the mantle arched the continents upward along the rift line by about two kilometers. Erosion then beveled the raised edges, thinning the margins of the two continental plates.

A modern example of crustal arching associated with incipient rifting of the crust can be observed in the high dorsal of Africa from Ethiopia southward. The Red Sea provides a more advanced stage of a newly opening ocean basin. Along the flanks of this linear trough crustal arching has stripped away young rocks, exposing "windows" of Precambrian basement.

With the insertion of new oceanic crust by sea-floor spreading, the ocean grew ever wider. In the process the continental edges subsided, as is demonstrated by the sloping flanks of the mid-Atlantic ridge today. When the ocean was smaller, the continental edges had to ride down a similar slope. Eventually the inflated mantle under the ridge reverts to normal mantle, but this takes 100 million years or more. Therefore as a geosyncline is laid down on the trailing edges of a drifting continent it slowly subsides for reasons external to the sedimentary deposit itself.

Additional subsidence, however, is caused by the steadily growing mass of the sedimentary apron, which must be isostatically compensated because the earth's crust is not sufficiently strong to sustain the load. For every three meters of sediment deposited the crust sinks about two meters. This crustal failure, however, is spread over a large geographic area, so that the growth and subsidence of a huge continental-rise prism causes a sympathetic downward flexing of the adjacent continental margin. As the continental shelf slowly tilts, wedges of shallow-water sediments are deposited.

Over the millenniums, as the shoreline transgresses and regresses repeatedly across the shelf, a large composite



MECHANISM OF CRUMPLING that produced the Appalachian foldbelt is depicted on the hypothesis that the Atlantic Ocean has opened, closed and reopened. In the late Precambrian (*a*), North America and Africa are split apart by a spreading rift, which inserts a new ocean basin. By the process of sea-floor spreading (*b*) the ancestral Atlantic Ocean opens. New oceanic crust is created as the plates on each side move apart. As the crust cools, its direction of magnetization takes the sign of the earth's magnetic field; the field periodically reverses, and the reversals are represented by the striped pattern. On the margin of each continent sediments produce the geosynclinal couplet: miogeocline on the continental shelf, eugeocline on the ocean floor itself. The ancestral Atlantic now begins to close (*c*). The lithosphere breaks, forming a new plate boundary, and a trench is produced as the lithosphere descends into the earth's mantle and is resorbed. The consequent underthrusting collapses the eugeocline, creating the ancient Appalachians. The eugeocline is intruded with ascending magmas that create plutons of granite and volcanic mountains of andesite. The proto-Atlantic is now fully closed (*d*). The opposing continental masses, each carrying a geosyncline couplet, are sutured together, leaving only a transform fault (*vertical black line*). The shear contains squeezed-up pods of ultramafic mantle rock. Sediments eroded from the mountain foldbelt create deltas and fluvial deposits collectively called molasse. North America and Africa were apparently joined in this way between 350 million and 225 million years ago. About 180 million years ago (*e*) the present Atlantic reopened near the old suture line. Today (*f*) the central North Atlantic is opening at the rate of three centimeters per year, creating new geosynclines.

megawedge of shallow-water sediments caps the shelf. The abundant supply of sediment usually ensures that the top of the prism is maintained close to sea level. Excess detritus bypasses the shelf, is temporarily dumped on the continental slope and is then carried onto the continental rise by turbidity currents. The shelf edge and the continental-rise prism comprise a couplet within which there is constant interplay.

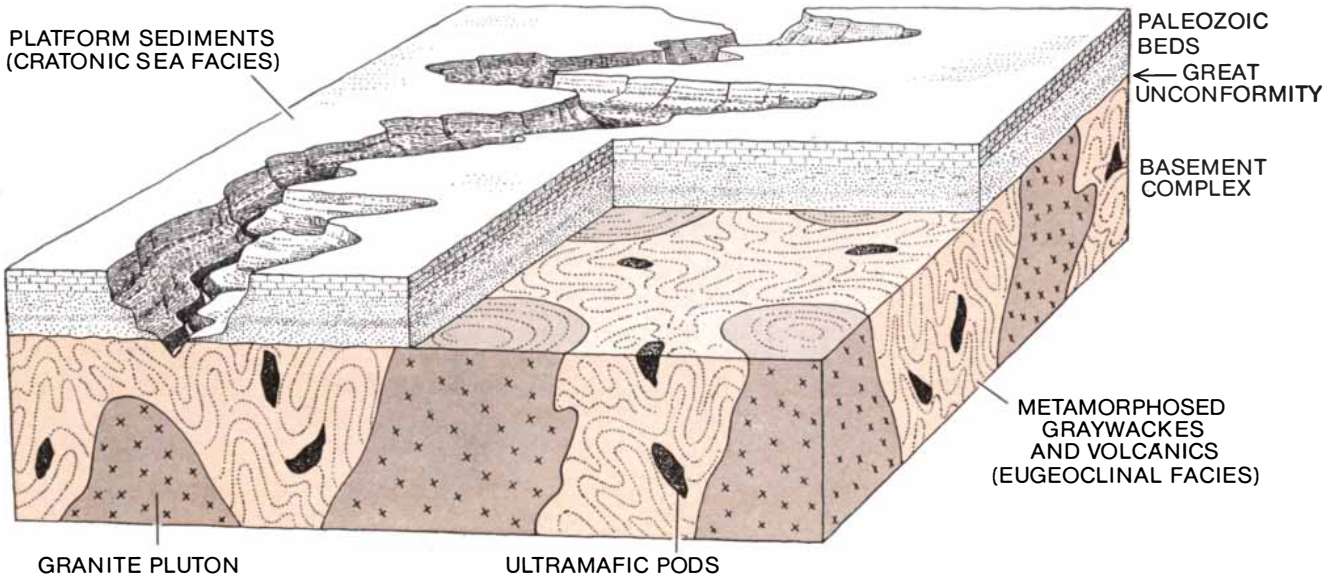
Like the sedimentary wedge under the Atlantic coastal plain today, the early Paleozoic Appalachian miogeocline thickens in the seaward direction. The abrupt termination of this miogeocline was long a mystery to early geologists who mapped it. They suggested that a missing seaward limb had been thrust upward and completely eroded away or that it had foundered into an ancient oceanic basin. This hypothetical land mass of Appalachia was the geological equivalent of the legendary Atlantis. The wedgelike structure of the existing continental-plain prism provides a satisfactory solution to the puzzle: the hypothesized seaward limb never existed. We now see that the thickening out of sedimentary deposits at the shelf edge is a normal mode of sedimentation. One way in which this may happen is that reefs of coral and algae build up along the margin of the continental shelf, creating a carbonate dam behind which other shelf sediments accumulate.

The mechanism of building continents by the peripheral accretion of collapsed continental rises seems also to ensure that the sedimentary deposits become dry land. (We take for granted that continents are above sea level, but it should be remembered that the mid-ocean ridge system, which approaches the continents in importance as a topographic feature, almost never rises above the sea surface.) The sedimentary apron gradually thickens until it approaches the height of the continental slope (about five kilometers), but upward growth ceases once the slope is completed and sea level is attained.

As we have seen, however, isostasy is at work, causing the oceanic crust to subside under the sedimentary load. The result is that a fully developed sedimentary prism attains an overall thickness of about 15 kilometers. When the prism is subsequently collapsed into a eugeosynclinal foldbelt, it becomes thicker still. The attendant metamorphism and granitic intrusion (which increases the total mass of rock) give rise to a monolithic structure that is more than 35 kilometers thick, thicker than a continental plate. Thus new foldbelts not only rise above sea level but also throw up rugged mountain ranges.

The hypothesis that geosynclines are deposited along a continental margin and then crushed against the continent as a result of plate tectonics seems to explain satisfactorily how geosynclines are transformed into folded mountains. The close relation between eugeoclines and foldbelts is not one of cause and effect but a simple consequence of location: geosynclines are laid down along continental margins and such margins are the locus of interaction between continents and subduction zones.

In spite of the vast span of geologic time and the rigors of erosion, the continents remain in a good state of health. We can predict that they always will be: detritus lost to the oceans is eventually carried back to the continents and collapsed into accretionary belts that also incorporate new igneous rock. Although the earthquakes that punctuate mountain-building are sometimes disastrous to man's culture, they are acts of continental construction. The great flood—the complete inundation of the erosionally leveled continents—will always threaten but will never come to pass.

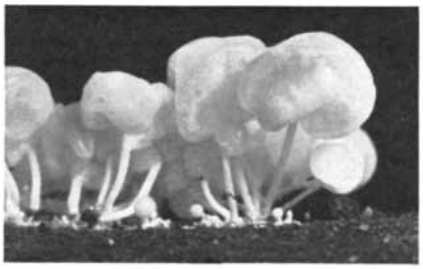


DEEP FABRIC OF CONTINENTS, the basement complex, is the fundamental rock unit of the continental plateaus, or cratons. This complex is usually obscured from view by miogeoclinal beds or by the coating of shallow-sea deposits that have invaded the continents from time to time. (Hudson Bay is a modern example of such an invading shallow sea.) Long a puzzle to geologists, the basement is composed of eugeoclinal foldbelts that have undergone intensive folding, metamorphism and intrusion. Geologists once thought that these "roots of mountains" indicated that the earth had contracted while cooling. The folds were likened to the skin of a dried apple. The present interpretation is that the eugeoclinal facies was laid down on the ocean floor and subsequently was crumpled against the continental margin, building up an onion-like vertically zoned craton. On this view the continents have grown larger rather than smaller with time. Moreover, the basement complex need not be Archean (composed of the oldest rocks) as formerly supposed, because the high degree of metamorphism does not necessarily indicate great antiquity and repeated mountain-building events. Instead it reflects the intensity of the collapsing process; once accreted, the foldbelt is not usually mobilized again. It has long been known that the granites of the basement complex are always younger than the metamorphosed sediments they intrude. On the other hand, the included pods of ultramafic rocks are always older, since they are detached fragments of the oceanic foundation on which the eugeocline was deposited. If geologists ever find "the original crust of the earth," it will be one of these pods within the oldest foldbelts.

We want to be useful...and even interesting

Propaganda

Paxina hispida, by R. H. Runde



Physalacria inflata, by Mrs. R. H. Runde

Marasmius nigripes, by Mrs. R. H. Runde

The Rundes reside in Peoria, Ill. These pictures were made in that vicinity. Dr. Ray Runde is a medical administrator. Mrs. Runde wins prizes in toadstool-picking contests.

See also the good grey shell-collectors patrolling the beaches in Bermuda shorts.

Rich malacologists there may be, but they probably didn't get rich on malacology. Ditto mycology, algology, bryology, and all the other deeply fulfilling, totally absorbing slots where "natural history" academically survives. How dry and dull, how low on the scale of priorities—until the sea shells, the fungal fruit bodies, the algae and mosses grow more distinct in the mind.

A camera finder helps greatly in focusing attention on the fascinating details. There are just so many ways for Pa to photograph Ma in front of the sign marking the easternmost point in the United States. Each year some lucky people discover that their investment in camera and film can pay off spiritually in an awareness of the beauty of the world.

Naturally, Kodak preaches that doctrine. The academic specialist has even more to gain from preaching it. Keep it in mind for the coming season. It is a good doctrine.

Wired optics

The financial seers who counsel on where to put smart money must be boning up furiously on liquid crystals. Cards are being played close to the chest. To get into the game seriously calls for optical talent, logic circuitry talent, and chemical talent.

The chemists deliver a vial of stuff. The opticists spread the stuff out into a thin layer and make an optical device out of it, an optical device that doesn't just sit there passively refracting, reflecting, dispersing, diffracting, scattering, absorbing, transmitting, or fluorescing. This optical device stays alert to the commands of circuitry.

The art rests on molecules which exhibit a mesomorphic phase intermediate between liquid randomness and three-dimensional crystallinity. One out of every 200 or so organic compounds is said to do so. This would open the choice pretty wide were it not for additional constraints on dipole moment and other properties.

So, while the chemists may try for something very special in the nematic or cholesteric line, or some multicomponent mixture of the two, or pleochroic guests among host nematic molecules, their teammates pass the time practicing and polishing their own contributions with commercially available liquid-crystal preparations. There Kodak comes in and has been coming in ever since revival of interest in liquid-crystal phenomena. That occurred in the sixties after several decades of somnolence on the subject.

We hereby announce the commercial availability of a multicomponent nematic that might suffice to call off some chemists. Name: "Nematic Mixture, Dynamic Scattering." Outstanding virtue: a nematic range all the way from 15 to 97 C. When nematic, a 12μm thickness of it, subjected in a NESA-coated cell to more than 4 volts dc (at 25 C), turns turbid from μm-scale turbulent centers of continuously varying refractive index, caused by migrating ionic entities locally perturbing the parallel alignment of molecules. In ac, required voltage rises with frequency to an asymptote at 300 Hz.

At 60 Hz, 200 v peak-to-peak, scatter is such that only 0.1%* of the flux in a parallel beam remains within the forward $f/34$ cone of solid angle. (0.02% at 300 v.) When and where there is no field, it's 78%. Presto, graphic or digital displays with ambient light, no more internal light sources needed than in a grandfather's clock. Contrast ratio at 24 v dc, 26:1†; 39:1 at 38 v (rms), 60 Hz. Resistivity of 6 x 10^9 ohm-cm (rather invariant over the usable frequency range!) keeps power requirements modest.

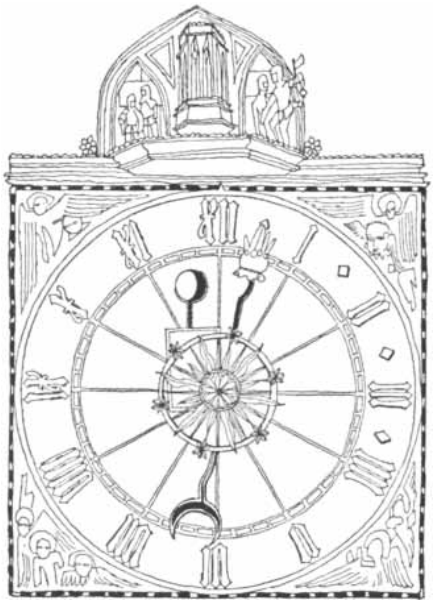
Order as EASTMAN *Organic Chemical No. 11643 from leading lab suppliers, in 5- or 25-gram septum bottles under N_2. Data quoted in this advertisement are to be regarded as typical but not stipulated.*

*90% of this degree of scatter is attained within 10 milliseconds. At break, the drop to 10% takes 400 msec.

†Look Ma, no hands on the watch!

Kodak

Science for pleasure or profit



SCIENCE AND THE CITIZEN

Amazing Mars

Mars continues to surprise the investigators associated with *Mariner 9*, which has been in orbit around the planet since November 13. The first 22 pictures of Mars, transmitted to the earth in 1965 by *Mariner 4*, depicted a heavily cratered surface hardly different from the surface of the moon. Four years later *Mariner 6* and *Mariner 7* revealed that Mars has distinctive surface features quite unlike anything seen on either the moon or the earth. The impression conveyed, however, was still of an essentially dead planet, one whose surface features had been fixed long ago and were subject only to superficial modification.

A total revision of this view began with the very first pictures taken by *Mariner 9*, which revealed that the entire planet was cloaked by vast clouds of dust. The dust storm did not begin to lift until the end of December. Because the first month's pictures showed so little, the spacecraft's orbital low point was raised from 862 miles to 1,025 miles to provide wider coverage per picture in the time remaining. A new mission plan was also devised to provide four 20-day photographic cycles, capable of mapping the entire planet from 65 degrees south latitude to 50 degrees north. By the end of January *Mariner 9* had circled Mars 158 times and had transmitted 5,174 pictures to the earth.

"Perhaps the most amazing thing," says one investigator, Bruce C. Murray of the California Institute of Technology, "is that Mars is so different from place to place. What we had seen before could be described somewhat loosely as subdued lunar uplands, essentially a fossilized surface. Now it appears that Mars is just starting to evolve as a planet; internal processes are just beginning to break through the surface. We see fracturing, collapse, huge volcanic cones and other features of internal origin."

One picture shows a huge chasm several hundred miles long and some 40 miles wide, with branching tributaries that look as if they had been carved by water (*see illustration at top left on opposite page*). Since it still seems unlikely that Mars ever supported large volumes of surface water, the assumption is that the pattern was produced by subsidence along lines of weakness in the planet's crust.

Another view shows a series of parallel rilles, or fissures, that extend almost linearly for more than 1,100 miles (*see illustration at top right on opposite page*). They would have delighted those early observers who believed they had identified dozens of canals, often in parallel pairs, on the Martian surface. The rilles revealed by *Mariner 9*, however, are much too narrow and too low in contrast for them to have been observed from the earth.

Perhaps the most spectacular feature is a volcanic cone at least 300 miles in diameter at the base (*see illustration at bottom left on opposite page*), making it larger than any comparable feature on the earth. Close-ups of the sides of the volcano show a lineated texture almost certainly produced by the flow of lava. The volcano coincides with a circular feature identified on maps since 1879 as Nix Olympica. In pictures returned by *Mariner 6* and *Mariner 7* it appeared to be a giant crater about 300 miles in diameter; now it is seen to be a cone at least four miles high.

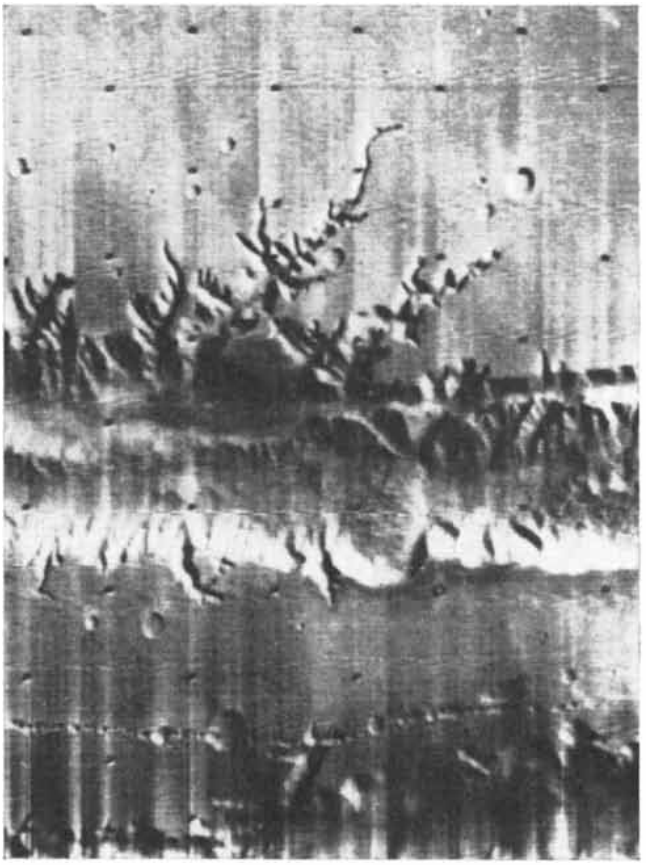
Amorphous Semiconductors Revisited

A potentially important but frequently controversial area of solid-state physics has been examined by a committee of the National Academy of Sciences in a report called *Fundamentals of Amorphous Semiconductors*. Much of the impetus behind the study of electron behavior in amorphous, or glassy, materials has been provided by Stanford R. Ovshinsky, founder of Energy Conversion Devices, Inc. Ovshinsky has pioneered tiny solid-state devices that act as threshold switches and others that act as memory devices, which can be activated either electrically or optically. One of the features of such devices is that, being already disordered (noncrystalline), they are highly resistant to radiation.

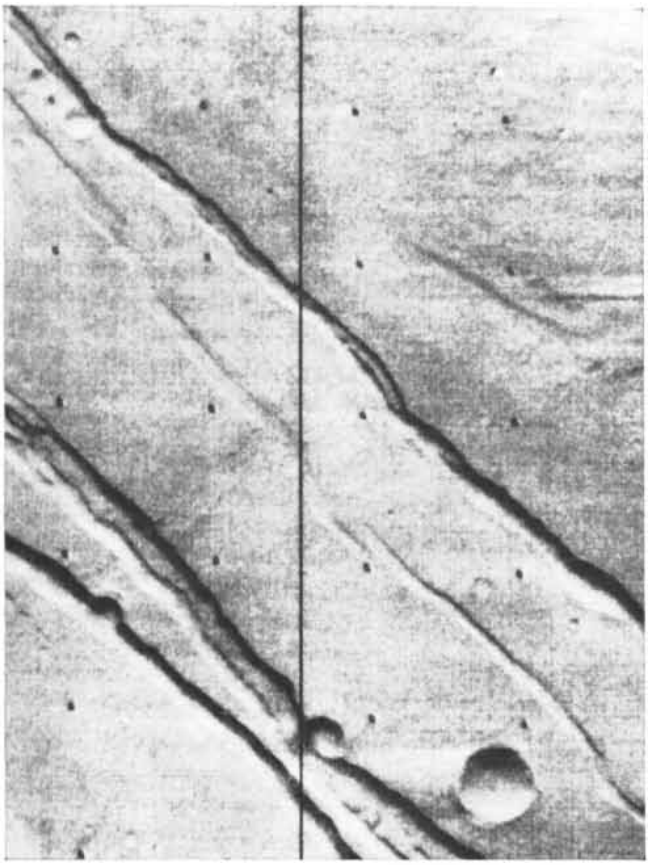
To prepare the report the National Academy appointed an *ad hoc* committee under Henry Ehrenreich of Harvard University. The report reviews the basic structure and electrical properties of amorphous materials and discusses a variety of existing and potential devices. The committee's major criticism of the field is that "a substantial fraction" of the work on glassy semiconductors has been "carried on without a clear understanding of the true nature of" the particular sample whose properties are being reported. As a consequence it is not always easy to distinguish between experimental characteristics that are genuine properties of the material under study and artifacts resulting, perhaps, from poor material preparation. Nonetheless, "amorphous semiconductors lend themselves to a wide array of possible uses."

By far the biggest application of a glassy semiconductor is in xerography, which uses compositions of amorphous selenium and arsenic-selenium mixtures as the photoconductive element on the drum that produces the copy (see "Electrostatics," by A. D. Moore, page 46). The most widely discussed use of the threshold switch devised by Ovshinsky is in display panels, conceivably to produce a television screen of "mural" size. The report reviews the pros and cons of producing such large-area displays by various means but reaches no conclusion on the most promising approach.

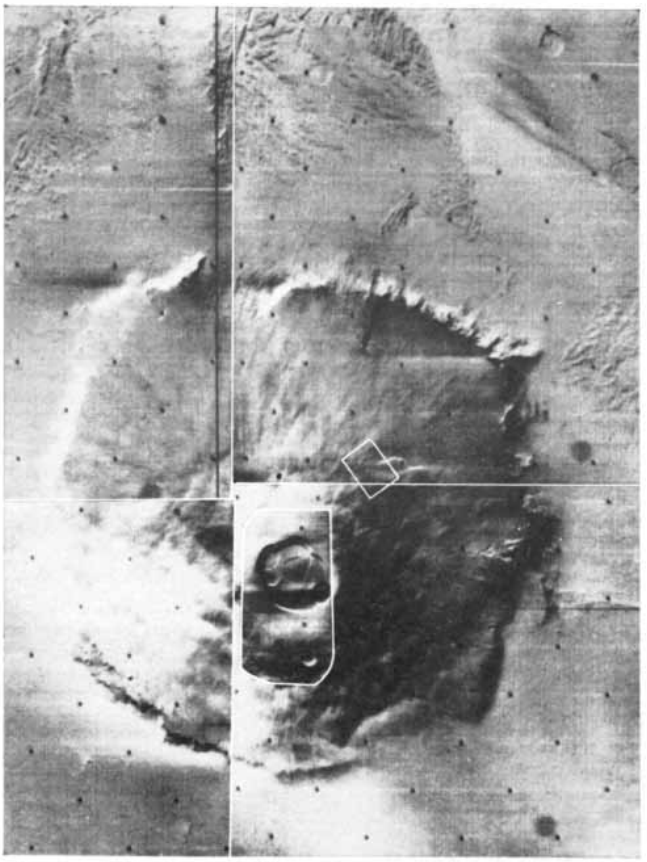
The report suggests that amorphous memory devices have "two possible advantages over most of the comparison candidates." Radiation immunity is one. The other is that the glassy devices do not require power to retain stored information and can be interrogated nondestructively. The report assumes that the amorphous memories will compete not with disk memories and tape memories, which also provide "nonvolatile" storage and nondestructive readout, but



Huge chasm in Tithonius Lacus



"Canals" up to 1,100 miles long



Mosaic of the volcano Nix Olympica

Close-up of area in rectangle at left

with integrated circuit memories. The report observes, however, that "computer memory proposals have, in recent years... gone to *more* volatile schemes for the sake of gains in cost and density. There are [nevertheless] applications, particularly in military systems, where a premium is put on nonvolatility and nondestructive readout."

In conclusion the National Academy committee "urges that a sizable effort of high-quality research regarding the physical, chemical and engineering properties of amorphous materials be encouraged. The present level of financial support should be maintained and probably increased somewhat."

Messenger's Reception

In recent years it has become apparent that many hormones, notably the steroid hormones, do their work by activating certain of the genes in the cells of the particular tissue that the hormone affects. Hormones therefore provide a means of external control enabling the cells composing the tissue to act in concert. A prominent question in biology has been: Exactly how does the hormone activate the genes? Pioneering work in elucidating the mechanism for hormones that affect the uterus has been done by Elwood V. Jensen of the University of Chicago, Jack Gorski of the University of Illinois and Terrell H. Hamilton of the University of Texas at Austin.

A group at the Vanderbilt University School of Medicine has now added to this work by finding evidence that in the cells of the chick's oviduct the hormone progesterone, which activates genes that give rise to the production of the protein avidin, is bound to a receptor molecule of protein in the cytoplasm of the target cell. The hormone-receptor complex then moves into the nucleus of the cell, where it becomes associated with chromatin: the material that incorporates the genes. Thereafter there are quantitative and qualitative changes in the synthesis of nuclear ribonucleic acid leading ultimately to the synthesis of avidin.

The Vanderbilt workers, who describe their experiments in *Nature*, are Bert W. O'Malley, Thomas C. Spelsberg, William T. Schrader, Frank Chytil and Alan W. Steggles. They report that "the oviduct receptor protein is essential for nuclear binding of progesterone" and that "oviduct nuclei but not spleen, lung, intestine or liver nuclei can bind the progesterone-receptor complex." Since the hormone-receptor complex binds much better to target-cell chromatin than to other chromatins, the Vanderbilt workers suggest that "perhaps the target-cell genome contains specific 'acceptor sites' which receive the inducer complex as it enters the nucleus."

The experimenters undertook to determine what part of chromatin was responsible for binding the hormone-receptor complex to chromosomes. They explored first the possibility that the histone proteins of chromatin were involved; they found, however, that the binding was unaffected when the histones of the target cells were replaced by histones from nontarget cells. They then found that if the acidic proteins of chromatin were removed, the chromatin lost most of its capacity to bind the progesterone-receptor complex. Fractionating the acidic proteins into four parts, they found that the principal loss of binding activity occurred after removal of the third part. They therefore concluded that acidic protein No. 3 "contains the great majority of acceptor molecules for the receptor complex."

Software into Hardware

Even though present-day electronic computers contain more circuit elements in a few cubic inches than the first computers contained in a good-sized room, the advances in hardware are generally conceded to have lagged behind the advances in software. The term "software" encompasses the computer languages and detailed instructions that enable the user to communicate his problem to the machine and get an answer. In a major effort to build a computer that is easier to work with, the Fairchild Camera and Instrument Corporation has designed a new computer system in which a large fraction of the programming tasks normally assigned to software are handled by hardware, that is, by logic incorporated directly into the computer. Fairchild calls the new system SYMBOL, which signifies direct hardware symbolic addressing. The name embraces both the design of the computer and a new programming language.

Fairchild engineers, who had helped to pioneer large-scale integrated circuits in which hundreds or even thousands of transistors can be put on a chip only a tenth of an inch square, concluded more than eight years ago that circuit hardware was becoming so cheap it should take over many of the intellectual tasks traditionally performed by programmers. Computer instructions are so complicated that programmers are often baffled when they look at programs they

have written but have not seen for several months, and a third party usually finds them inscrutable.

Perhaps the most important function traditionally assigned to software involves the management of memory. The program must specify, for example, the length of characters, words or groups of words called fields; it must allocate space for these fields in the computer's memory, describe how they are to be entered and recovered, and how the assigned space can be expanded or contracted as needed. In short, the program must prescribe exactly how data space is to be managed. Computer experts debated for years whether or not these and related software functions could be handled by hardware.

Fairchild engineers believe that the debate has now been settled in the affirmative by the SYMBOL system and that the system will handle many critical areas of memory management from 10 to 100 times more efficiently than it is now handled by software. "The SYMBOL philosophy," say Gilman D. Chesley and William R. Smith of Fairchild in a recent paper, "is that user data space management will go the way of manual transmissions on automobiles: fine for special purposes but the average user will be willing to give up some performance and flexibility for ease of use."

The first prototype of the new Fairchild system, SYMBOL IIR, is undergoing tests at Iowa State University. An earlier study made at Iowa State had shown that a typical advanced computer costs about $110,000 per month to operate. The total cost is divided about equally between "problem expense" (primarily the cost of producing a debugged software program), "operation expense" (the cost of operating personnel, space, power and so on) and "rental expense." Fairchild estimates that the SYMBOL system should cut the cost in each category by as much as 50 percent.

Embryo and Tumor

The fact that certain cancer cells possess antigens that also appear in fetal cells has recently attracted much interest among cancer investigators. At the least the similarity of the antigens gives support to the hypothesis of a relation between some types of neoplastic, or cancerous, growth and the growth of an embryo. Moreover, measuring fetal antigens in the tissue fluids of a human patient could prove valuable in diagnosing certain forms of cancer and in evaluating the success of treatment. A longer-range hope is that such investigations

may lead to means of treating certain malignancies by the immunological process in which an antigen gives rise to an antibody that then destroys the antigen, which in this case would mean destroying the malignancy.

An antigen in this context is a molecule that the body treats as foreign, responding to it immunologically. The antigen can be found both inside a cell and exposed at the surface. Presumably such molecules modulate the interactions of the cells of a tissue. In an embryo the cells have a regular, ordered program of exposing molecules that recognize each other and thereby direct the process of cell division and growth whereby a tissue is made. When the tissue is complete, the antigens disappear.

Expression of such an antigen in later life could be a factor in causing such strange properties of cancer cells as their tendency to swarm over one another and to divide uncontrollably. One could explain the exposure of the antigens in cancer cells by three hypotheses: that the cell has a new genetic program, perhaps introduced by a viral gene; that the cancer cell expresses a genetic program that is already in the cell but is normally "silent," and that the cell may, because of an alteration of its genes, produce a new and aberrant protein.

Summarizing the recent developments in *Nature*, Peter Alexander of the Chester Beatty Research Institute in Britain notes that the term "fetal antigens" is used loosely "to describe two quite different situations in relation to cancer." In the first one "immunological methods are used as a method of analysis to establish that a macromolecule found in an embryo is also present in tumor tissue." In the second one "a substance present in both tumor and embryo evokes an immunological reaction in the tumor-bearing host," meaning that for at least a time antibodies are directed against the tumor.

Alexander cites suggestions that various tumor-specific antigens "all have a counterpart in embryonic life and that they are the products of a whole range of repressed genes which are derepressed more or less randomly during carcinogenesis by chemicals." More recent data, he says, "are difficult to reconcile with this hypothesis, and it seems more likely that a number of quite distinct changes in the surface of the cell accompany the malignant transformation." He adds: "The presence in tumors of fetal substances ... may be only one aspect of a much more general phenomenon associated with malignancy, namely the synthesis of aberrant proteins of a wide variety."

The Sweet Smell of Failure

Spectacular failures are not unknown in the history of engineering, but none is more spectacular, in a peculiar way, than one that happened in the North End of Boston, Mass., on January 15, 1919. An undergraduate at the Massachusetts Institute of Technology, John Lange, has recently written an account of the long-forgotten episode for *Tech Engineering News.*

The U.S. Industrial Alcohol Company, a distiller and supplier of raw alcohol to the munitions industry, was the owner of a large molasses storage tank, built in 1914 just off Commercial Street on the waterfront in the North End. Made of half-inch steel plates supported by a heavy concrete foundation, the tank was 58 feet high, 282 feet in diameter and had a capacity of 2.5 million gallons.

Shortly after noon on January 15, while the sole company employee at the site was out to lunch, the tank ruptured with explosive suddenness, its rivets popping out like a fusillade of gunfire. A moving wall of molasses was released; it traveled through the area at 35 miles per hour, sweeping away everything in its path. One two-and-a-half-ton section of tank wall, 400 square feet in area, was deposited 182 feet away from the tank base. A nearby firehouse was pushed off its foundations by the viscid flood and almost swept into the harbor. Another structure was carried down the street until it collapsed; in still another house the molasses mounted to the third floor, rousing a sleeping man. Where the advancing molasses struck the girders supporting an elevated railway it collapsed them; an approaching train was halted just in time. By nightfall 11 dead had been found and 60 injured had been taken to hospitals.

Because of the owner's association with the munitions industry it was quickly conjectured that sabotage had caused the disaster. A series of investigations, as well as legal maneuvers concerning liability, dragged on for years before it was finally discovered that the steel of the tank's wall plates had been considerably thinner than the specifications required. When this fact emerged, Lange recounts, the company settled all claims out of court. The incident was closed except it is said that, when weather conditions are appropriate, one can still detect the scent of molasses in the North End.

1972 TORONADO. THERE'S NOTHING COMMON ABOUT IT.

It has six unique features. Five of them have nothing to do with luxury.

1 Front drive. Instead of the rear wheels pushing the Toronado, the front wheels pull it. You can't appreciate this advantage in traction until you try to dig out of snow. Or mud.

2 Flat floor up front. Front drive eliminates the hump; even the middle-seat passenger rides comfortably.

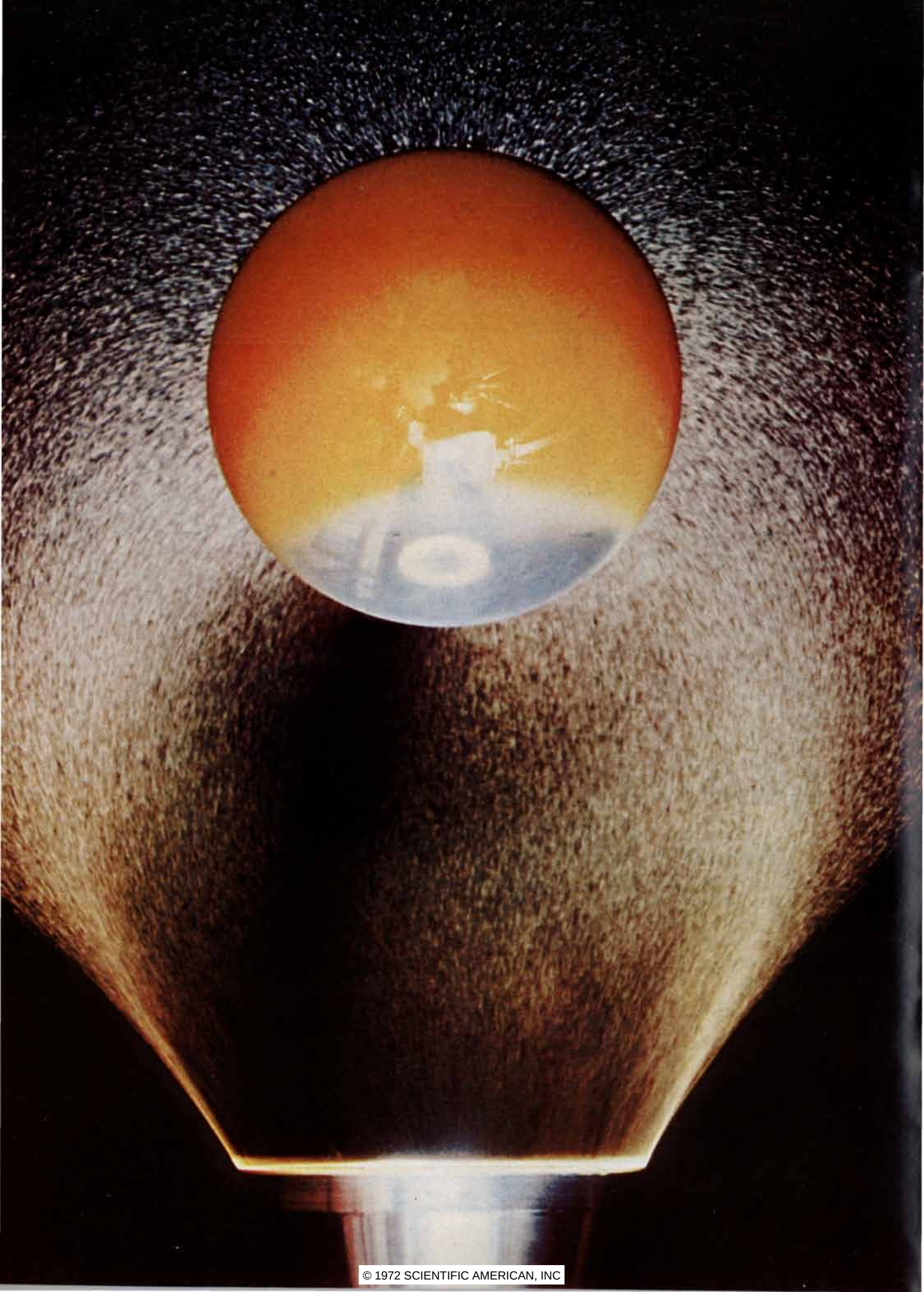
3 High-level signal lights, neatly sculpted into the rear deck of the car. Click on your turn signal or apply your brakes, and the driver in back of you gets a double warning.

4 It's built, patiently and carefully on its own production line, by men who work on no other cars.

5 It breathes differently. Dual air intakes, coupled with an air scoop, direct the air to the radiator. Cooler air is ducted directly to the carburetor.

6 Luxury. You don't just see it, you experience it. When you test drive the Toronado, feel its weight; stretch out in its roominess; listen to its purr; enjoy the coddling ride. The solid handling of the Toronado — combined with all of the GM safety features — helps contribute to a genuine feeling of security. Any way you look at it, we think you'll find this a truly unique automobile.

OLDSMOBILE
ALWAYS A STEP AHEAD



ELECTROSTATICS

The field around a charged body is used to manipulate particles of matter. Among the modern applications of electrostatics are fly-ash precipitators, paint sprayers and Xerox copying machines

by A. D. Moore

Anyone who has ever walked across a rug in a dry room and received a shock from a metal doorknob has had a firsthand experience of an electrostatic effect. On a grander scale, thunder and lightning are examples of electrostatic effects that are familiar to all. Yet it is a curious fact that many manifestations of natural electrostatic processes remain almost completely unknown.

For instance, although most children know from playing with toy compasses that they live within the earth's magnetic field, very few adults realize that we spend much of our lives in a natural electric field as well. The atmosphere of the earth is somehow supplied with a positive charge that sets up a downward electric field amounting to between 100 and 500 volts per meter on a clear day. In an open area at the higher value there would be 1,000 volts between the ground and a point two meters up. If a two-meter man were standing there, would he have 1,000 volts acting on him from head to foot? If so, he would soon be dead. Actually there is no voltage on him. He is a grounded conductor, his skin is an equipotential surface that warps the electric field of the atmosphere and he is totally unaware of the field that surrounds him.

Of course if our two-meter man is a golfer finishing a game under a low thundercloud, the situation can change drastically. The base of such a cloud is usually negatively charged, and the electric field, now reversed, can rise to 10,000 volts per meter. This is still too low to warn the man by making his hair stand up. When he raises his steel-shafted club for the next stroke, however, a different kind of stroke unleashes an electric current measured in the tens of thousands of amperes.

The branch of human curiosity that deals with such matters is called electrostatics. In recent years this enterprise has rapidly expanded, and the scope of modern electrostatics goes far beyond the limited domain of nonmoving electric charges implied by the literal definition. Man has learned to apply his newfound understanding of electrostatic effects in diverse ways. The four main areas of application are electrostatic precipitation of particulate industrial wastes, such as fly ash, dust and fumes; electrostatic separation of mixed granular solids, principally minerals; electrostatic coating, both wet and dry, and electrostatic imaging, the best-known example of which is the xerographic process. Before taking up these and other achievements of modern electrostatic technology, I shall explain some of the general attributes and effects of electric charges and electric fields, and in particular introduce the central concept of corona discharge.

ELECTROSTATIC PAINT SPRAYER shown in operation on the opposite page was photographed in the test laboratory of the Ransburg Electro-Coating Corporation. The atomizer (*bottom*) consists of a bell-shaped member charged to 90,000 volts and spinning at a rate of 900 revolutions per minute; paint is fed into the four-inch-diameter bell at its center, flows out centrifugally to the edge as a film and is there subjected to the high electric field. The paint film at the edge is electrically shaped into a series of spaced liquid streamers (*see illustration on page 53*). Highly charged paint droplets are repelled from the tips of the streamers and follow the lines of force of the electric field to the grounded target (*top*), in this case a steel sphere. Paint that would otherwise miss the target and be wasted is thus made to curl around and coat the sides and rear of the object. The paint droplets were illuminated for this photograph by means of two high-speed stroboscopic lamps, one mounted next to the camera and the other mounted behind the steel sphere. The bright spot near the bottom of the sphere is the reflection of the mouth of the atomizer. Other images appearing on the target are the reflections of various pieces of photographic equipment.

Charges and Discharges

Consider a capacitor consisting of two parallel plates, between which is inserted a long electrically conducting rod [*see upper illustration on page 49*]. The device sets up a uniform electric field except near the center, where the lines of force of the field are gathered inward and end on the rod. Any small element of electric flux ending on the rod exerts an outward pull at right angles to the rod's surface. The pulls on opposite sides of the rod are equal, giving a net force of zero. Positive or negative charges can be induced on the rod because electrons are free to move both within the rod and along its surface.

If the conducting rod is now removed and a dielectric, or insulating, rod is substituted for it, this freedom no longer prevails. Now the field is gathered in toward the rod and passes right through it. Even so, the forces are much the same as before, and again there is no net force acting on the rod.

Next consider a pair of concentric electrodes forming what might be regarded as a kind of coaxial capacitor [*see illustration at lower left on page 49*]. When a voltage is applied to this device, it sets up a radial electric field. When a long conducting rod is inserted in such a nonuniform field, the rod again warps the field, and as a result forces

act on opposite sides of the rod. A new effect comes in here. If two small elements of electric flux ending on the rod have the same flux but different areas, the smaller element will exert the greater pull. Thus the forces exerted by, say, larger outer flux elements ending on positive charges will not be equal to the forces exerted by smaller inner elements ending on negative charges. As a consequence there will be a net radial force acting inward on the conducting rod. Much the same effect holds for a dielectric rod. In other words, both types of object would tend to move inward toward the more intense field.

It might appear that the device just described amounts to an electrostatic precipitator: particles moving through the electric field could simply be forced to move to the inner electrode, where they could be collected. Actually the forces involved in such a simple arrangement are far too weak to be useful for industrial precipitation. The achievement of Frederick G. Cottrell in inventing the first effective electrostatic precipitator in 1905 was finding a way to exert forces large enough to make particles in such a nonuniform field configuration move outward, to be collected on the outer electrode. To do this Cottrell had to exploit the then little-known phenomenon of corona discharge.

Suppose the central electrode is now a thin wire in a duct 20 centimeters in diameter [*see illustration at lower right on opposite page*]. If one applies a negative potential of 100,000 volts to such a wire, the voltage acts over the radial distance of 10 centimeters, yielding an average field strength of 10,000 volts per centimeter. Since the threshold at which air breaks down is about 30,000 volts per centimeter, it might seem that no electrical breakdown of the air would occur. The calculated value of 10,000 volts per centimeter, however, is only an average value, and such averages can be misleading. In this particular nonuniform configuration the field strength is much less than 30,000 volts per centimeter out near the duct wall, whereas it is much more than 30,000 volts per centimeter in the intense-field region near the wire.

The result is a corona discharge, a zone of ionization that manifests its presence in air by a greenish glow around the wire. The placid, inactive field of the earlier examples is replaced by an intensely active field. Billions of molecular ions are produced within the zone of the discharge and move outward. The visible glow comes from photons emitted by excited atoms. The wire is now called the active electrode, whereas the duct becomes the passive electrode.

The physical intricacies of the corona cannot be covered fully here. Suffice it to say that when the active electrode is negative, electrons come whizzing off its surface. Most of the electrons promptly attach themselves to oxygen molecules in the air, making the molecules negative ions. At this point the Coulomb effect takes over: the effect by which the force exerted by the field on a small charged object is proportional to its charge times the strength of the field. As a result the ions move outward toward the walls of the duct, constituting a veritable ionic current.

The operation of an actual industrial precipitator can now be easily understood. The duct carries the flue gas, or other gas, loaded with waste particles. The ions charge the particles. If the velocity of the gas is of the order of 10 feet per second or less, the particles will move across the gas stream by the billions and will collect on the walls of the duct. If the waste particles are solid (as they are in the case of fly ash), the duct is rapped periodically to shake the residue down into a hopper. If the particles are liquid (as they are in most fumes), the residue runs down the duct walls. Industrial precipitators are designed for a negative corona, because higher voltages can then be used without too much sparking. Electrostatic home air cleaners, on the other hand, operate on a positive corona.

The contribution of Cottrell's invention to antipollution efforts would be hard to overstate. Long before most towns and cities stopped dumping raw sewage into streams Cottrell precipitators made life livable near cement mills and ore-processing plants. The first electrostatic precipitator to remove fly ash from coal-burning power plants was installed by the Detroit Edison Company in 1923. It is estimated that at present the fly ash trapped in this country alone —at better than 99 percent efficiency— amounts to some 20 million tons per year.

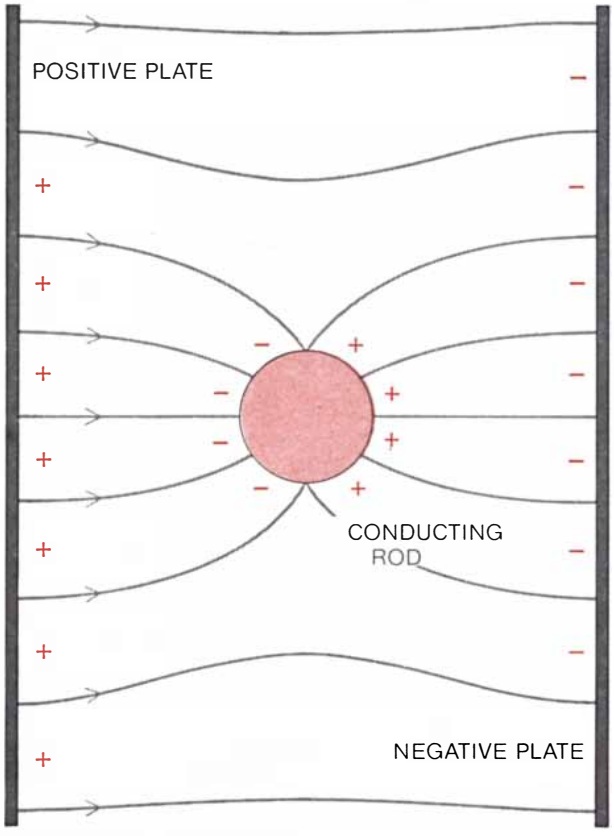
Further Uses of Corona

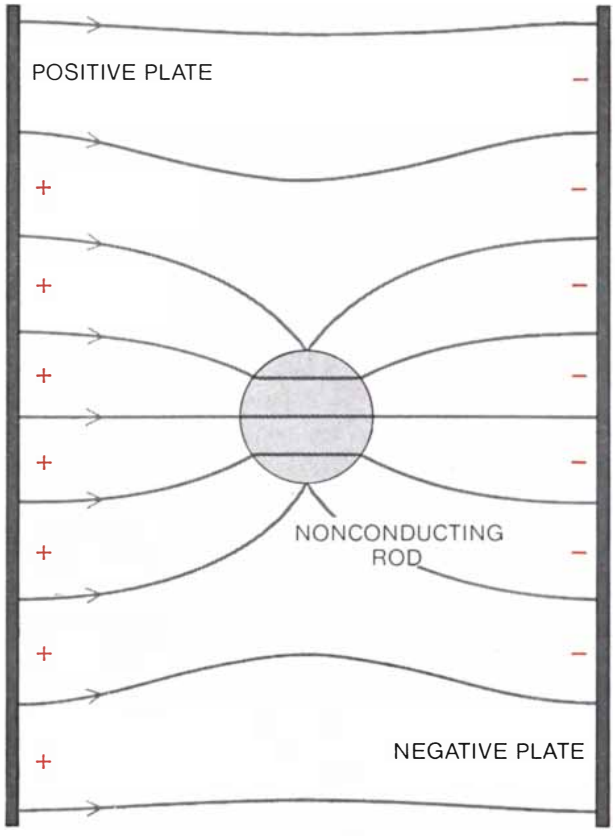
The principle of corona discharge is the key to many other important applications of electrostatics. For example, in the separation of granular mixtures by one electrostatic approach the two kinds of particle to be separated differ from each other in their conductivity, one kind being a conductor and the other an insulator. In a typical separator [*see top illustration on page 56*] the mixture comes down from a hopper to form a thin layer on top of a rotating steel drum. The drum, which is grounded, passes under a wire that is generating a corona discharge. As a result ions flood to the drum. The charges of those ions that hit conducting particles are passed directly to the drum and accordingly the particles fall right off the drum. The charges of ions that hit insulating particles, in contrast, coat the outer surfaces of the particles and "pin" them to the drum, from which they later fall or are scraped. In addition to this common type of electrostatic separator, which is used mainly in the mineral industry, a number of quite different designs exist.

The world's largest electrostatic-separation plant for the beneficiation of iron ores, installed in 1965 at the Wabush Mines in Canada, handles six million tons per year. It has been estimated that some 35 other mineral mixtures are currently being separated electrostatically. Moreover, the electrostatic approach has been adopted for a number of other separation tasks, including the removal of rodent excreta from rice, the extraction of garlic seeds from wheat and the separation of nut meats from shells.

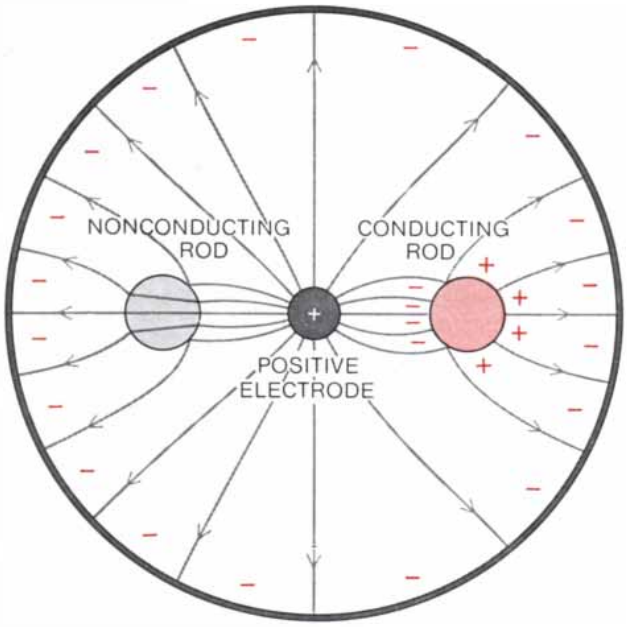
The electrostatic stratagem for pinning insulating particles to a drum in the preceding example can also serve to pin an insulating web to a roller. (In several industries a continuous moving sheet of paper or film is termed a web.) Normally tension is applied to such a web by pairs of rollers at each end. If one side of the web is coated with some sticky substance such as ink or adhesive, however, the two-roller system clearly will not work. Instead the web can be pinned to a single roller electrostatically. A corona bar is installed to supply ions to charge the outer surface of the web. A noteworthy application of this technique is the manufacture of thin films such as Mylar, which in the final operation are stretched to give them high strength. Various other kinds of web are also handled by electrostatic pinning.

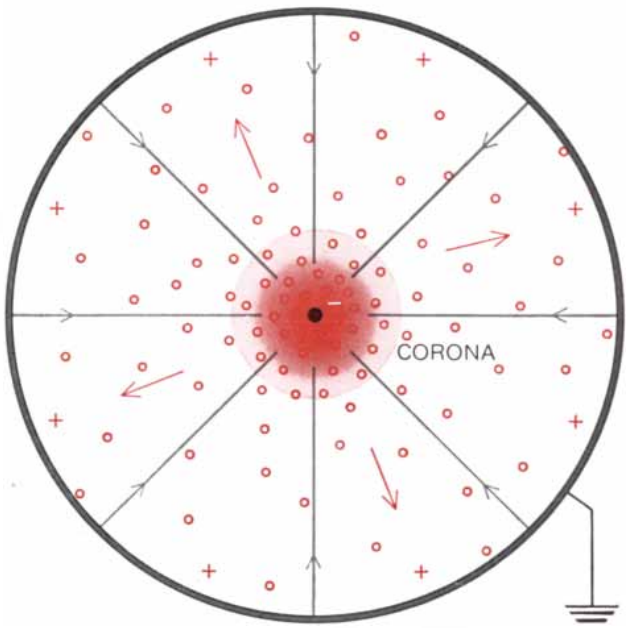
A little-known but important industrial process in which electrostatics, and specifically corona discharge, plays a central role is electrocoating, a process that is currently being harnessed on a large scale to apply various coatings, including wet paint, grit particles, dry powders and even short fibers. In a typical arrangement a spray gun with a corona point emits a fine mist of paint particles; the particles gather the field lines to themselves and attract the ions from the corona, thereby acquiring a charge. The force tending to pull the particles back to the more intense field





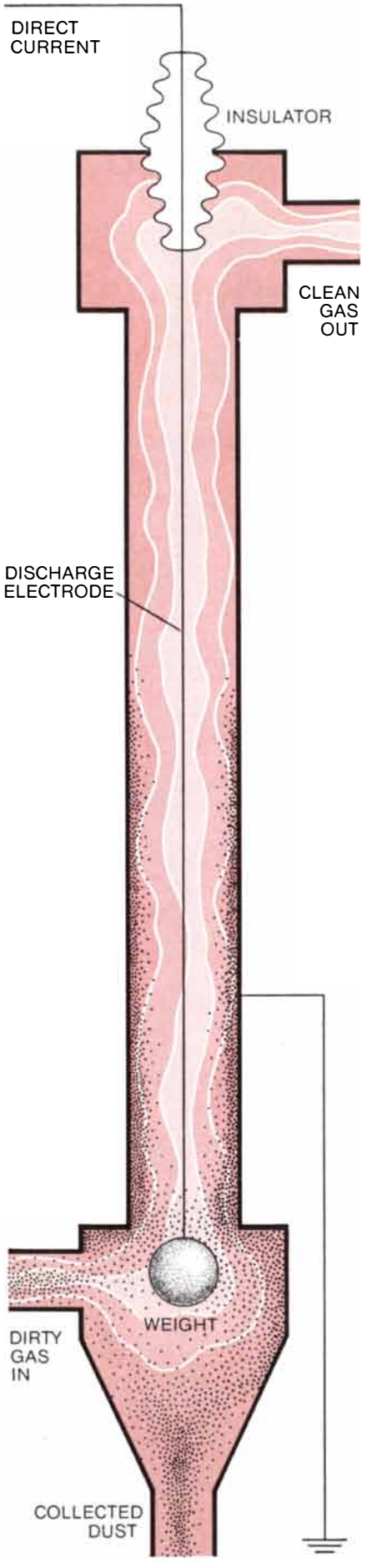
PARALLEL-PLATE CAPACITOR normally sets up a parallel electric field, which becomes warped near the center when a long rod of either a conducting material (*left*) or a nonconducting material (*right*) is inserted between the plates. In the case of the conducting rod the lines of force of the field (*gray arrows*) are gathered inward and end on the rod; positive and negative charges (*colored signs*) can be induced on the rod because electrons are free to move both within the rod and along its surface. In the case of the nonconducting rod the field is gathered inward and passes right through the rod; no charges are induced on the rod. In both cases the outward pulls exerted by the electric flux on opposite sides of the rod are equal, yielding a net force of zero on the rod.





RADIAL ELECTRIC FIELD is set up when a voltage is applied to two concentric electrodes. When a conducting rod and a nonconducting rod are inserted in such a device, they warp the field as shown. If two small elements of electric flux ending on either rod have the same flux but different areas, the smaller element will exert the larger pull; hence there will be a net radial force on each rod tending to move it inward toward the more intense field.

CORONA DISCHARGE (*colored area*) develops when the inner electrode of the preceding example becomes a thin wire and enough voltage is applied to ionize the air molecules in its vicinity. With the inner, or active, electrode negative electrons from the surface of the wire attach to oxygen molecules in the air, making the molecules negative ions (*open circles*). The ions are pushed outward by the Coulomb force to the outer, or passive, electrode.



ELECTROSTATIC PRECIPITATOR is based on the principle of corona discharge. In this simplified diagram a single grounded duct (the passive electrode) carries the flue gas, loaded with solid waste particles. Ions produced around the weighted central wire (the active electrode) charge the waste particles, which move across the gas stream and collect on the walls of the duct. The duct is rapped periodically to shake the residue into a hopper. An actual industrial precipitator consists of a large assembly of units.

is still there, but it is overcome by the far larger Coulomb force driving the particles out more or less along the lines of force of the field toward the grounded target. Paint that would otherwise miss a target and be wasted can thus be made to curl around and coat the sides and rear of the object [*see illustrations on page 52*].

For many purposes a spinning disk or bell serves as the sprayer; the paint is fed in at the center, streams out centrifugally to the edge and is thrown off. On leaving the sprayer the paint forms streamers with a remarkably uniform spacing [*see illustration on page 53*]. The streamers break up into droplets that become charged and hence seek the target. In an automatic operation designed for long runs the targets can be carried past the spray gun by means of a conveyor system; for large objects several spray guns can be used in tandem, and they can be mounted to move up and down.

It has been estimated that the savings in paint alone attributable to the electrocoating approach amounts to some $50 million per year. In addition there is a growing interest in electrocoating as a technique for spraying insecticides on plants in such a way that the undersides of the leaves are also coated. Environmental pollution is thereby reduced, and smaller amounts of insecticide are made to do a better job.

A variation of the electrocoating approach called flocking can be used to give a surface such as a wall a velvet finish. If the wall is nonconducting, it is first covered with aluminum paint, to which an adhesive is applied. The operator holds a hopper filled with short fibers. As the fibers are shaken out of the hopper they are charged by ions from a set of corona points mounted on the hopper. At this stage three important effects take over. First, the fibers are impelled along the field lines by the Coulomb force. Second, the mutual repulsion of like charges keeps them apart. Third, they align themselves with the field, arriving end on to stick to the adhesive. In this way more than 200,000 fibers per square inch can be applied. The process can also be used to put pile on carpeting, to make artificial suede and to cover the interior of instrument cases. By printing the adhesive in patterns on cloth one can create decorative designs.

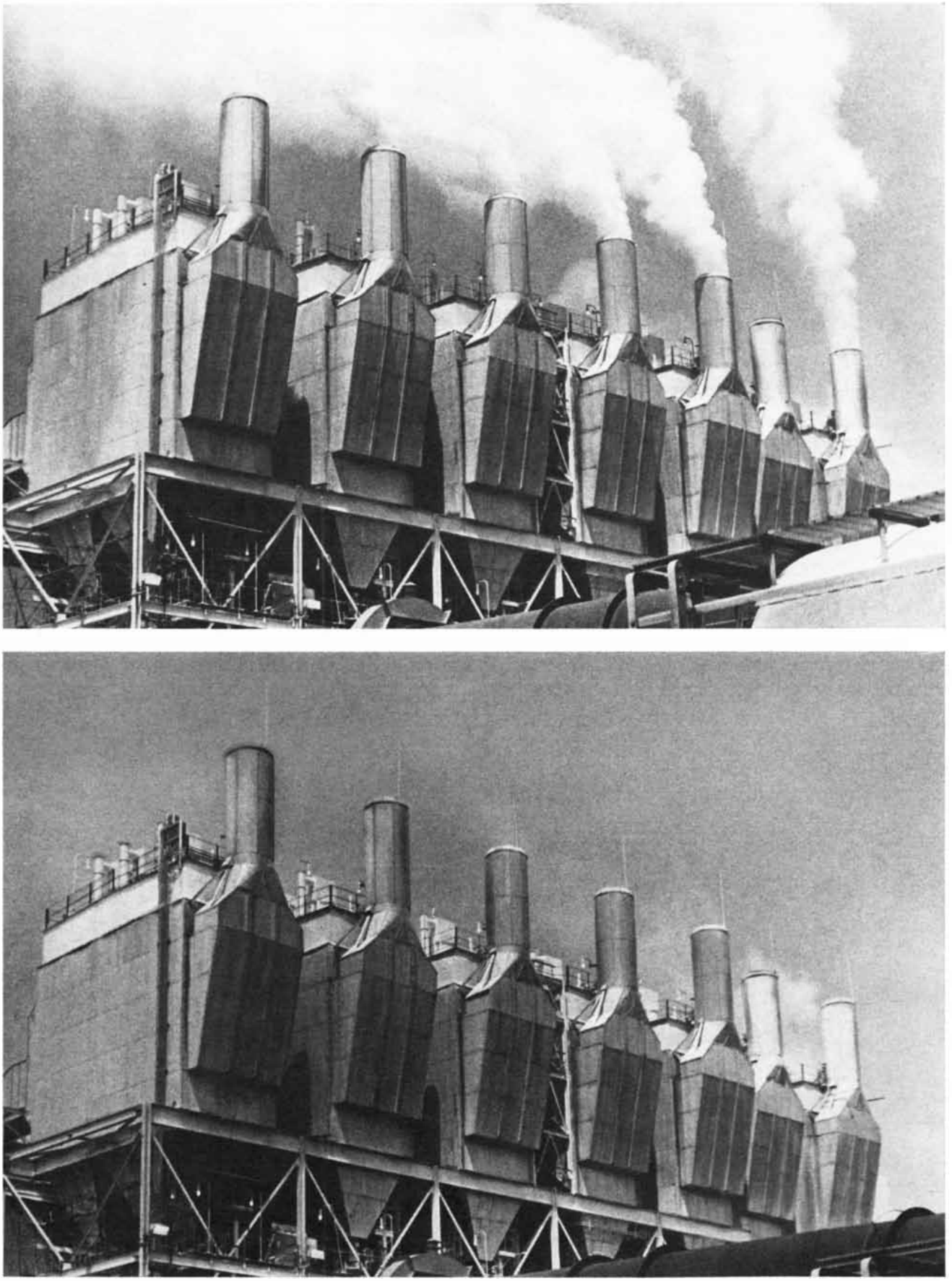
Somewhat the same process is used to make coated abrasives, a business amounting to some $200 million per year. Superior sandpapers and coated cloth abrasives are turned out electrostatically on huge machines. For a "filled" abrasive the operation continues until all the spaces are filled. For "open" coatings the rate of feed is lowered, and the electrostatic-spacing effect makes for a uniform rather than a patchy coating.

Xerography

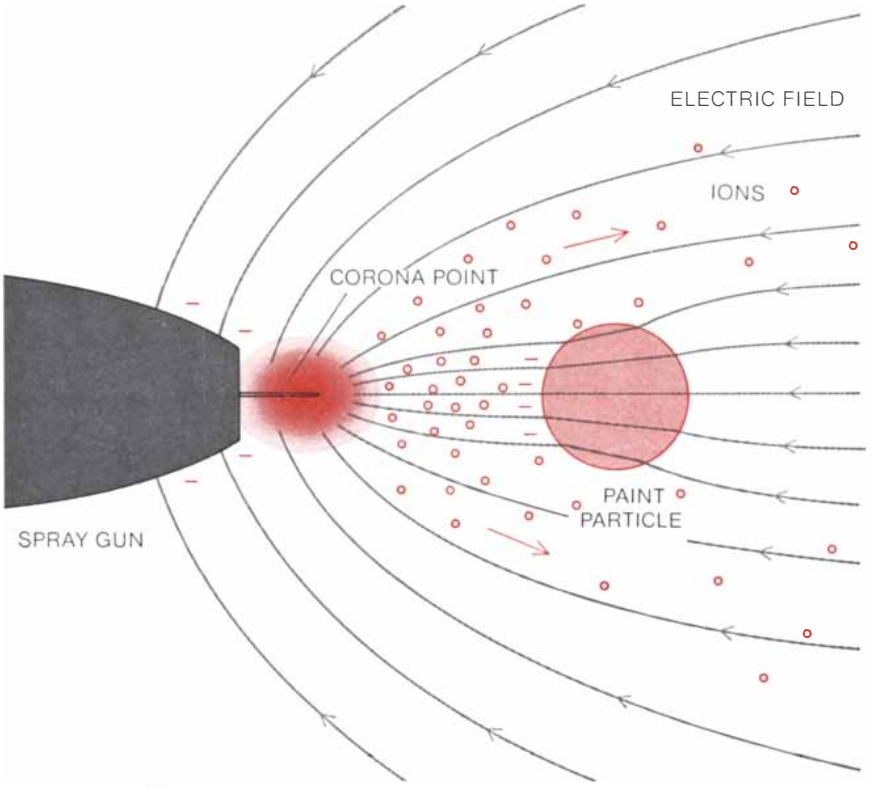
Perhaps the best-known application of modern electrostatics is the dry-copy imaging process known as xerography. This is not to say that many people know how a Xerox machine works, even though they are expert at operating it. The xerographic process, invented by Chester Carlson in the mid-1930's, depends on the photoelectric property of certain materials such as zinc oxide or selenium. In a typical Xerox dry-copying machine [*see illustration on page 55*] a rotating selenium-coated drum is charged in the dark by a corona bar. Next an optical system focuses an image of the page to be copied onto the drum. The light removes all the charge except where the images of the black areas appear. These images are then made to attract a black dust called the toner. Farther along the paper, which is precharged, is fed in, makes contact with the drum and attracts the toner from the drum to itself. The paper then moves through a rapid-heating stage that fuses the toner to itself to make a permanent copy. Meanwhile the drum passes a brush that removes the remaining dust and is exposed to light once again to remove the remaining charge. The drum is then recharged by the corona bar and is ready to make the next copy.

In early models of Xerox copiers a problem was encountered in trying to apply the toner evenly. This was solved by mixing the toner with a "carrier" consisting of tiny glass spheres. The spheres and the dust stick together, being oppositely charged by contact electrification. When the drum comes to the toner-application stage, the toner is attracted and the carrier is repelled.

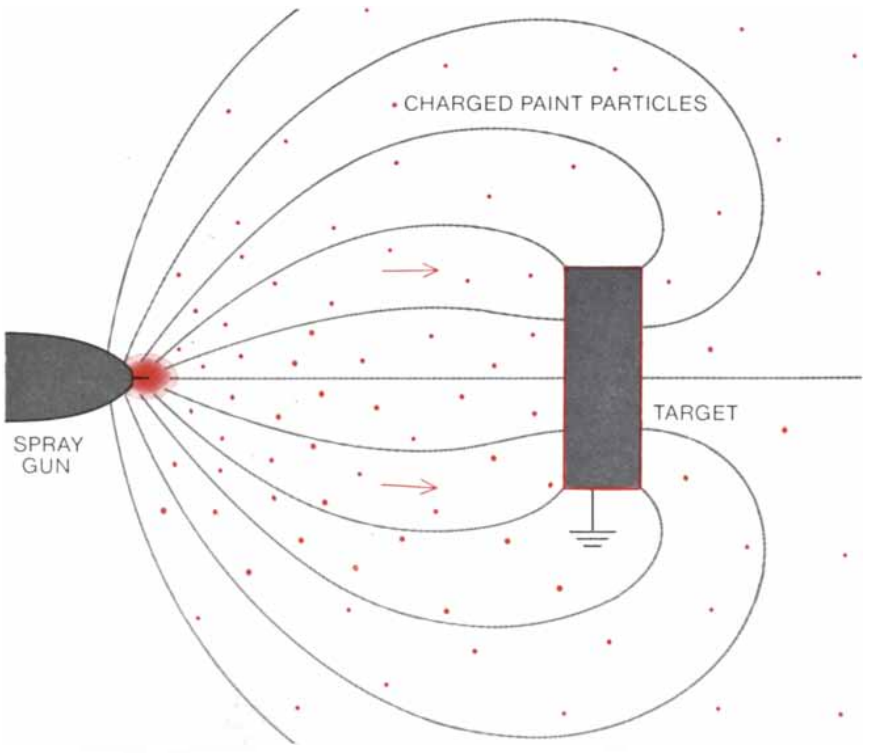
If all of this sounds simple, take a look inside a large Xerox machine and see the



EFFECTIVENESS OF ELECTROSTATIC PRECIPITATOR in preventing air pollution is dramatically demonstrated by this pair of photographs, which show a battery of electrostatic precipitators that were momentarily turned off (*top*) and then on again (*bottom*) during the normal operation of large alumina-processing plant. The system was designed and built by Research-Cottrell, Inc.



PAINT PARTICLE leaving an electrostatic spray gun is shown greatly enlarged in this schematic diagram. The negative corona point on the spray gun emits electrons (*not shown*), which attach themselves to the air molecules in the vicinity of the corona, producing a flood of negative ions (*open colored circles*) that stream outward more or less along the lines of force of the electric field. The paint particle gathers the lines of force to itself, thereby attracting the ions and hence acquiring a negative charge of its own.



CHARGED PAINT PARTICLES (*solid colored dots*) are shown more to scale in this diagram. The paint particles, impelled toward the grounded target by the Coulomb force, tend to follow the electric-field lines, curling around to coat the sides and back of the target.

complex arrangement of automatic-control devices required to achieve the final copy. Various aspects of the xerographic process are even now being intensively developed. For example, it seems only a matter of time before we have full-color electrostatic reproduction of high quality. The social and economic impact of the success of the electrostatic dry-copying process can be roughly gauged from the fact that the worldwide business of the Xerox Corporation reached a new high last year of almost $2 billion.

In all four of the major applications involving corona discharge taken up so far, no mention has been made of space charges. As a matter of fact space charges, in the form of ions and other charged particles, can radically alter the electric field. As a result any attempt to make an accurate mathematical analysis of an electrostatic process is forbiddingly complicated. Fortunately persistent inventors, experimenters and other assorted tinkerers have rushed in where theoreticians have feared to tread, and as a result many marvelously workable devices have been achieved.

Some Noncoronal Applications

Up to this point I have described only those electrostatic techniques employed for moving particles in gases. Particles can also be moved electrostatically in liquids without a corona. To cite the simplest example first, in electrophoresis charged particles are moved through a liquid to an electrode of opposite charge.

Dielectrophoresis is a more complex process whereby uncharged particles in a liquid are moved to the more intense part of a nonuniform electric field [*see bottom illustration on page 56*]. The field shifts the electron cloud surrounding a neutral molecule slightly to one side, in effect shifting the positive nuclei the other way. Such a shift of charge takes place in any field, even a uniform one. In a nonuniform field, however, the force on the inner charged side is slightly stronger than the force on the outer charged side. As a result there is a very weak net force in the direction of the more intense field.

The phenomenon of dielectrophoresis has long been known, but it has been rather neglected because the effect was so small. Nonetheless, according to Herbert A. Pohl of Oklahoma State University, one of the pioneers in opening up this area to further investigation, "nonuniform electrical fields produce unique and frequently mystifying effects on matter—even neutral matter. With non-

uniform fields one can pump liquids or powders, clean up suspensions, classify organisms and even separate live from dead cells. And this just starts the list. Its applications in biophysics and cell physiology to studies of normal and abnormal cells are at an early but exciting stage. In colloid science the new technique is helping to resolve surface properties. At the molecular level, nonuniform effects are seeing renewed use in determinations of molecular polarizabilities, in maser operations and in laser control."

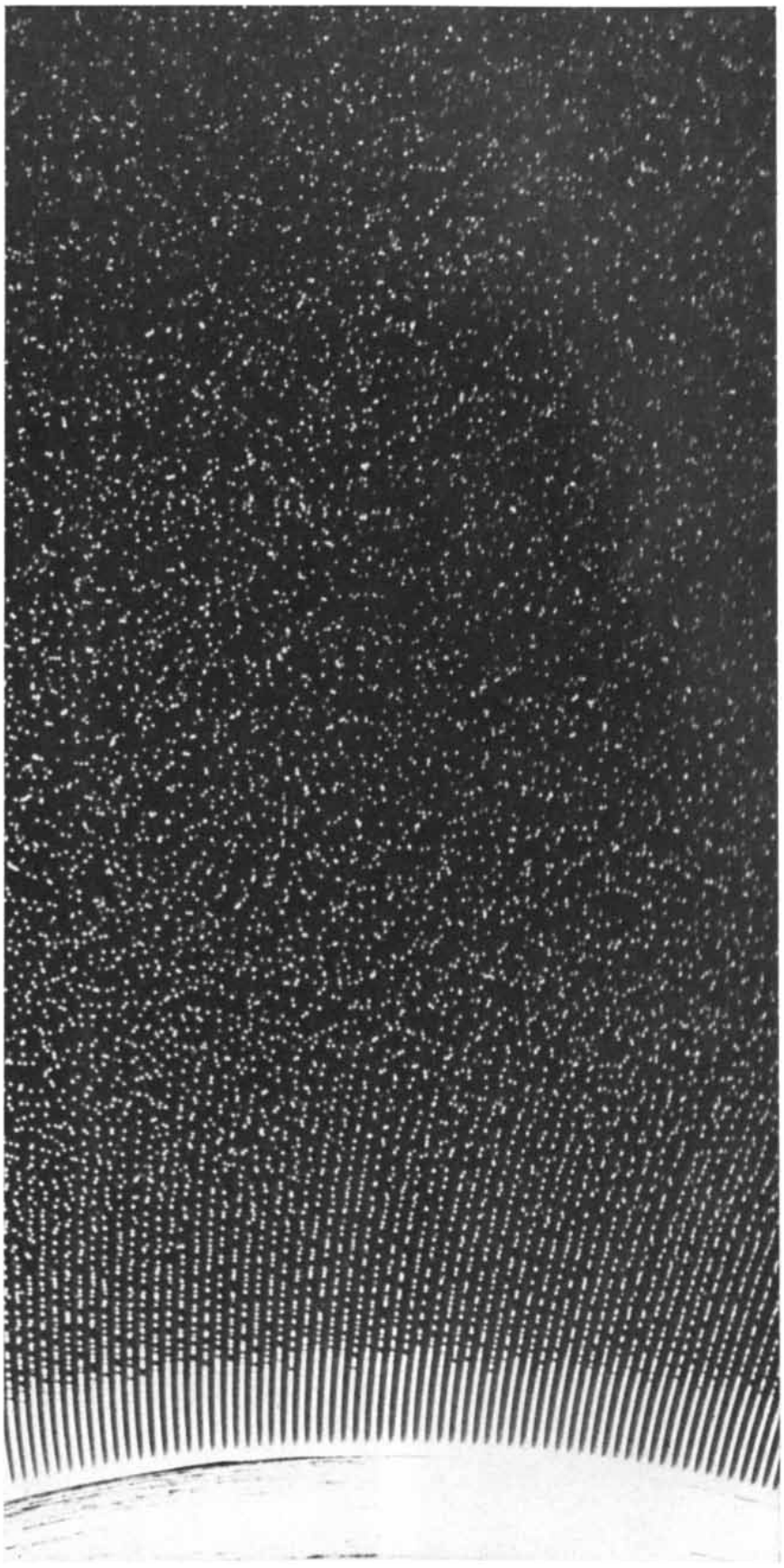
In passing it should be mentioned that whereas the electrostatic manipulation of charged particles in air requires a constant, or at least a unidirectional, field, dielectrophoresis can operate with either a direct-current or an alternating-current field. With alternating current the molecular polarization simply reverses with the field, up to high frequencies.

Contact Electrification

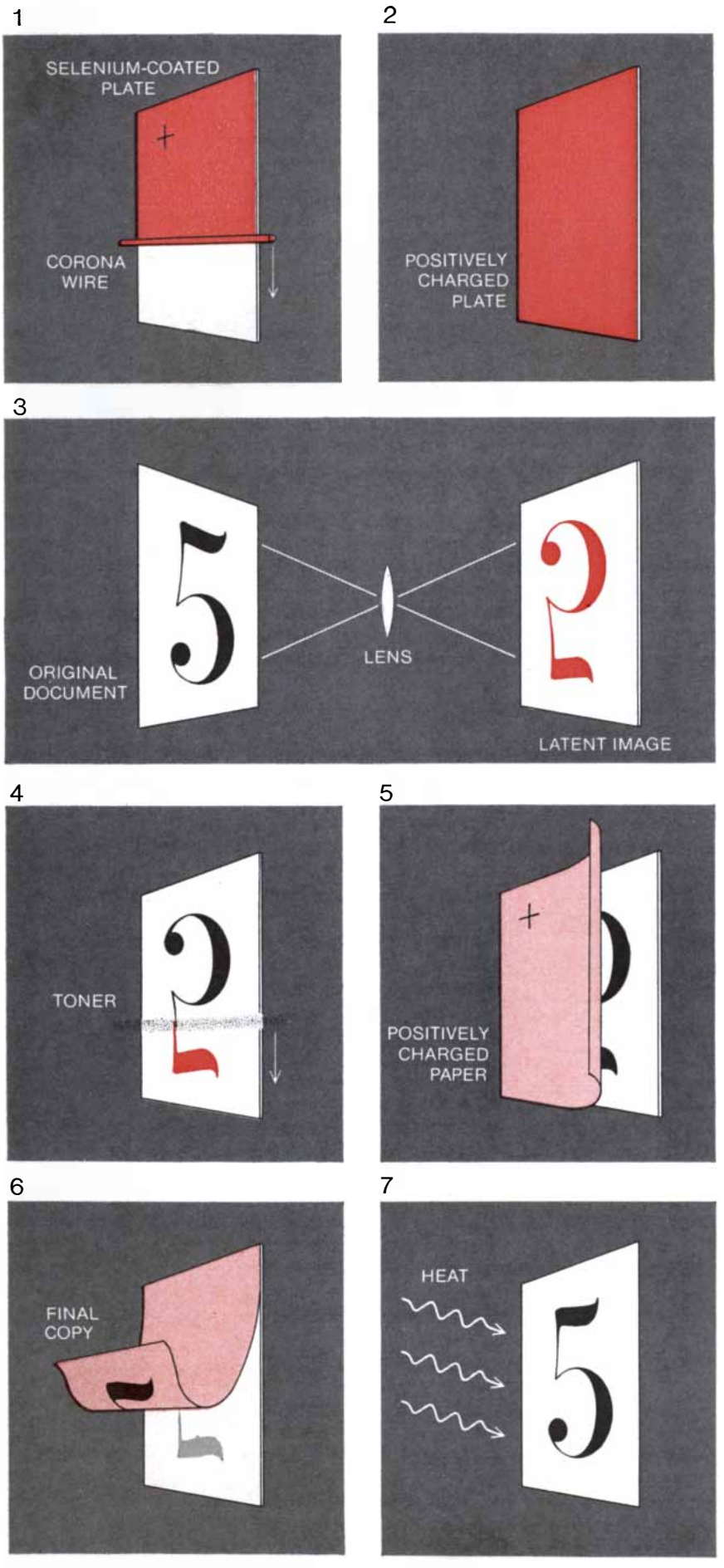
If this article had traced a historical development, it would have started with frictional electricity. Such electricity was supposedly known to the ancient Greeks, particularly Thales of Miletus, who observed about 600 B.C. that when amber was rubbed, it would attract small bits of matter. The term "frictional electricity" gave way to "triboelectricity," although since "tribo" means "to rub," the newer term does little to change the concept. Then it was discovered that charge transfer can be achieved without rubbing—that mere contact between unlike surfaces would do the job. The prevailing view today is that rubbing merely enhances the effect by bringing more tiny areas together.

When two metals are placed in contact, electrons pass from one to the other because of the difference in the metals' quantum-mechanical work functions. This process continues until an equilibrium is reached. With the metals still in contact there can be a potential difference between them of anywhere from a tiny fraction of a volt to several volts. When the metals are separated, they have opposite but equal charges. If one member of the pair is an insulator, with its comparatively fixed electrons, the situation is much less clear, and it is even less so when both members are insulators. In both cases continuing investigation is gradually yielding clarification.

Contact electrification is a subject of considerable importance, since it accounts for many of the nuisances and hazards of electrostatics. Industry is



PAINT STREAMERS leaving the edge of a bell-type electrostatic atomizer exhibit a remarkably uniform spacing. Each streamer becomes the source of a string of droplets, which are highly charged and hence seek the target. The spark flash used for this shot lasted only a third of a microsecond. Photograph was supplied by the Ransburg Electro-Coating Corp.



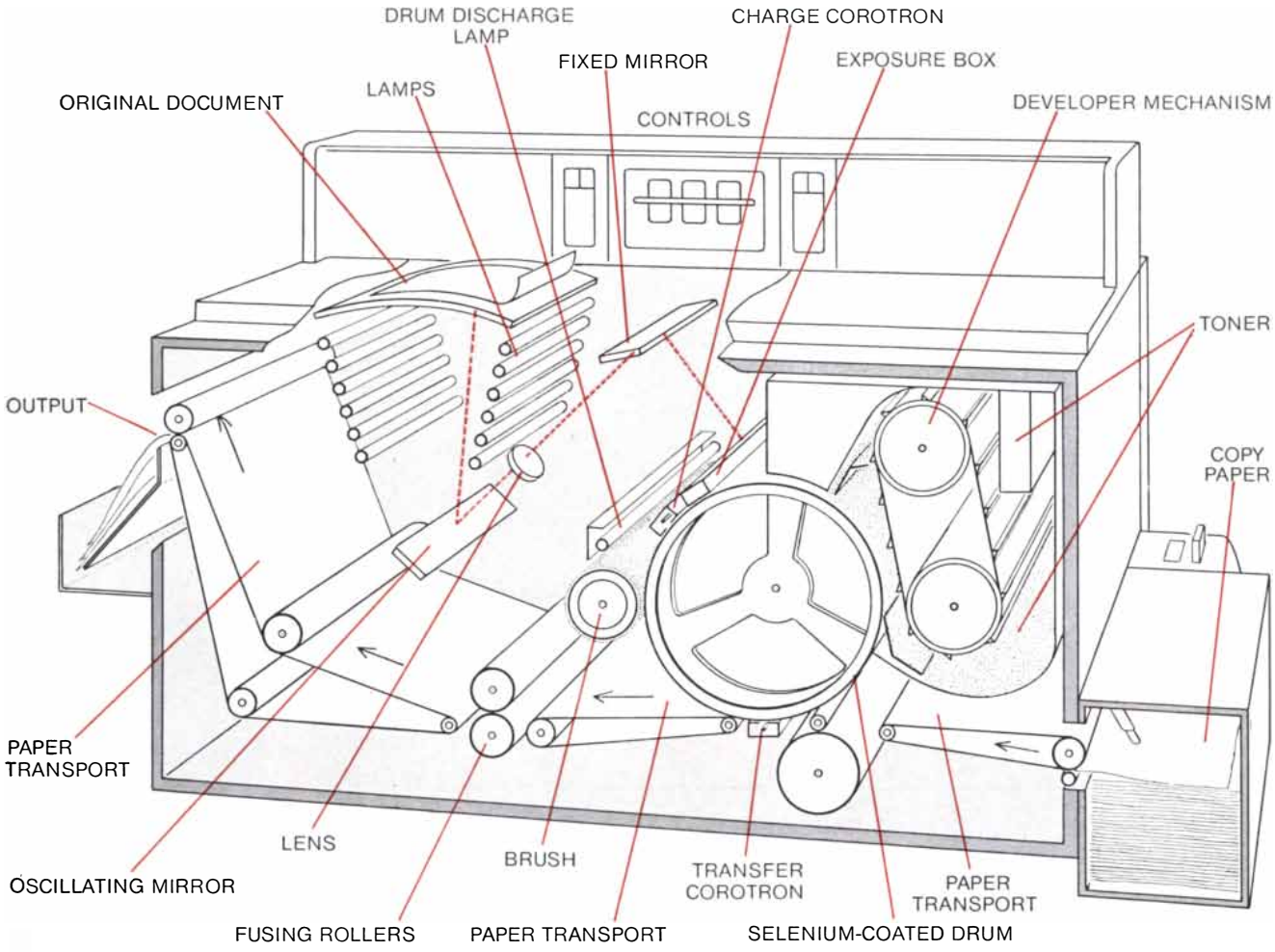
forever building electrostatic generators without intending to do so. It is impossible, for example, to move sugar, flour or any similar dry powder through ducts without charging the powder. Belts running over pulleys develop high voltages, giving rise to sparks. Yarns, plastic films, paper and kindred materials moving at high speeds inevitably lead to triboelectric troubles. If a resulting spark has enough energy, inflammables are ignited and explosions are set off.

The infrequent disasters, however, are far outnumbered by the nuisance effects. Charge may accumulate to the point where an industrial process will have to shut down temporarily. Sometimes simply raising the humidity increases conductivities enough to "bleed" the excess charge away. Another means of getting rid of the charge is to place grounded "tinsel" in contact with the charged objects. Finally, the local air can be made more conductive by ionizing it with a corona discharge or a radioactive isotope.

Airplanes and helicopters can also become highly charged, either from flying through dust, sleet or snow or from encountering strong electric fields in clouds. Their communications equipment can be knocked out or their control system can be damaged by subsequent discharges. Moreover, they become a prime target for a lightning stroke. The protection of aircraft and their electrical equipment from the hazards of contact electrification is an exceedingly complex problem, of great interest in both industrial and government circles. Electrostatics is also strongly suspected in several recent disastrous explosions aboard oil tankers, and an intensive research effort is under way in this area as well.

Indoor and outdoor dusts are almost invariably charged. Dust blown in contact with almost anything will acquire a charge. Further, an uncharged particle in the air will pick up ions. Cosmic rays

XEROGRAPHIC PROCESS is illustrated in this sequence of highly simplified diagrams. First a selenium-coated plate is charged in the dark by a corona wire (*1, 2*). Next an optical system focuses an image of the page to be copied onto the plate (*3*); the light removes all the charge except where the images of the black areas appear. These latent images are made to attract a black dust called the toner (*4*). Precharged paper is then brought in contact with the plate (*5*) and attracts the toner from the plate to itself (*6*). The paper is finally exposed to a rapid-heating stage that fuses the toner to itself to make a permanent copy (*7*).



TYPICAL XEROX DRY-COPYING MACHINE (designated Model 2400) employs a rotating selenium-coated drum instead of the flat plate shown in the preceding illustration. After transferring the toner to the paper the drum passes a brush that removes the remaining toner and is exposed to light once again to remove the remaining charge. The drum is then recharged by a corona bar (called the charge Corotron) and is ready to make the next copy. This diagram shows only a few of the major components of the Model 2400 machine; actually a much more complex arrangement of automatic-control devices is required to achieve the final copy.

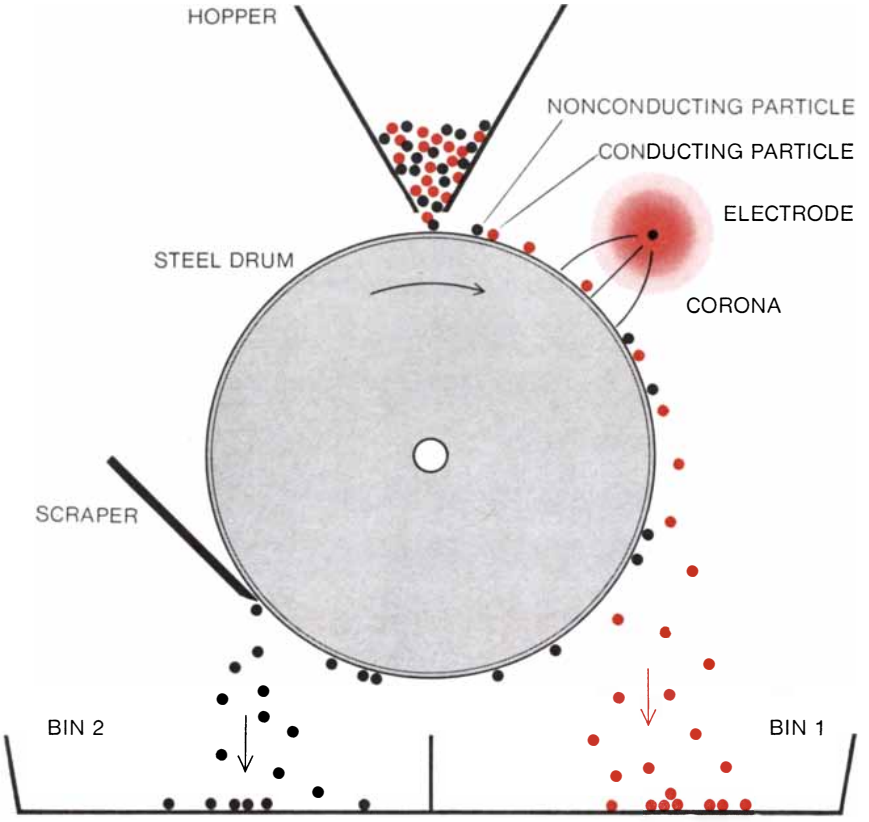
and terrestrial radioactivity constantly break up air molecules, forming an average of between 10 and 20 new pairs of ions per cubic centimeter per second. The ion pairs constantly recombine, but nonetheless clear air may contain anywhere from 100 to 500 pairs per cubic centimeter. Thus an uncharged particle in air is very likely to pick up an excess charge of one sign or the other. The same probability applies to water droplets in a cloud.

Charge transfer is not confined to contacts between solids. Many insulating liquids flowing through pipes or hoses can also become charged, leading to fires and explosions. Charges can even be separated when a liquid separates from itself! A fascinating and important instance of this phenomenon has been studied by Duncan C. Blanchard of the State University of New York at Albany. He observed that when sea waves break, large numbers of bubbles are buried. The rising bubbles, on reaching the surface, burst. At a certain stage a jet from the center of the bursting bubble goes up at a high velocity, and from the tip of the jet four or five droplets continue upward [*see illustration on page 58*]. It turns out that most of these droplets are positively charged. The droplets evaporate, leaving behind positively charged salt nuclei, which are carried farther upward by the wind and eventually become the nuclei for the formation of rain droplets. All of this amounts to a partial answer to a question raised at the beginning of this article: How does the earth's atmosphere maintain its positive charge?

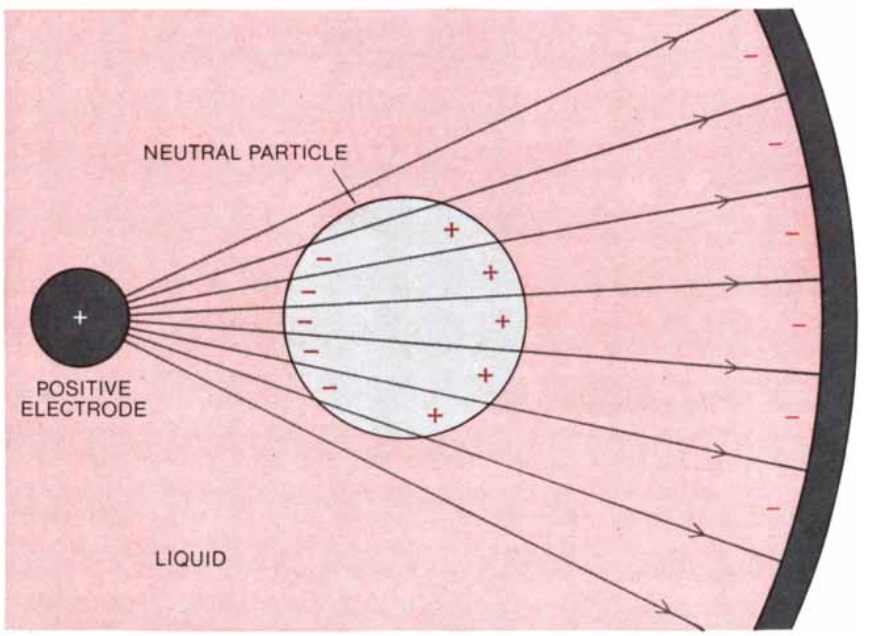
Meteorological Effects

The huge space charges represented by thunderclouds are partly drained by lightning within the clouds. As for lightning that strikes the ground, there has been a certain amount of confusion about it over the years. Even though a cloud's electric potential may be as high as 150 million volts, the field just under the cloud is usually less than 10,000 volts per meter, or 100 volts per centimeter. Since the breakdown strength of air is 30,000 volts per centimeter, how could a lightning stroke ever happen? Studies of this problem finally led to the discovery that a "leader" stroke develops inside the cloud in a stepwise fashion and comes to the ground first. Its ionized path is then followed by the main upstroke or by repeated strokes.

The negative base of such a thundercloud induces positive charges in objects on the ground. A field of 100 volts per centimeter is enough to give rise to tiny point discharges on the tips of leaves and grass. The positive ions produced in such discharges can be carried by updrafts to add positive charge to some other part of the cloud, thereby helping to main-



ELECTROSTATIC MINERAL SEPARATOR also employs a corona discharge. The mineral mixture, consisting of conducting particles (*color*) and nonconducting particles (*black*), comes down from a hopper to form a thin layer on top of a rotating, grounded steel drum. The drum passes under a corona-discharge wire that sends a flood of ions to the drum. The charges of ions that hit conducting particles are passed directly to the drum and hence these particles fall right off the drum into Bin 1. The charges of ions that hit nonconducting particles "pin" these particles to the drum, from which they later are scraped into Bin 2.



DIELECTROPHORESIS is a noncoronal application of electrostatics whereby uncharged particles in a liquid are moved in a nonuniform electric field. The field shifts the negative electron cloud surrounding the neutral molecules slightly to one side, in effect shifting the positive nuclei the other way. Since the force on the inner side is now stronger than it is on the outer side, there is a weak net force in the direction of the more intense field.

tain the net positive charge of the atmosphere.

Recently Roger V. Cheng of the State University of New York at Albany studied a mechanism whereby charges can be separated inside a cloud. When a supercooled water droplet freezes, its outside freezes first; this freezing is followed by the development of spicules, which emit large numbers of far smaller droplets. These "daughter" droplets are positively charged, whereas the "mother" droplet remains negative. If such a process really does operate on a large enough scale, along with gravity separation, very large charge separations would result.

An electric field in a cloud can have precisely opposite effects on a water drop, depending on its size. Large enough drops in a strong enough field can be pulled apart, giving birth to oppositely charged twins, whereas small enough drops lined up with the field can be drawn together to form a larger drop [*see illustration on opposite page*]. The latter process, called electrostatic coalescence, is precisely the same as the "desalting" process used to remove water and salts from crude oil before it goes to the refinery.

While on the subject of meteorological effects involving electrostatics, the last and perhaps the strangest effect of this type I should like to mention is earthquake lightning. It has long been observed, particularly in Japan, that earthquakes are sometimes accompanied by flashes of light, in the atmosphere, even when the sky is clear. During a Japanese earthquake in 1930, for example, more than 1,500 such flashes were recorded. The area in which the event took place is characterized by quartz-rich lavas. It has been suggested that given the right kind of crystalline order and the right kind of seismic waves, rock formations of this type may be capable of generating millions of volts of electrostatic energy by the piezoelectric effect. In view of this strong possibility it seems reasonable to propose that if there are quartz-rich layers anywhere in the vicinity of the San Andreas fault, stations for continuous monitoring of the electric field of the atmosphere could be set up; if field disturbances are found to coincide with ground tremors, an earthquake early-warning system might come of it.

The Diversity of Electrostatics

The world's electrical energy is supplied by electromagnetic generators

made of iron and copper, and much of it goes to motors of the same type. Both the generators and the motors have their electrostatic counterparts. The best-known electrostatic generator is the giant belt-driven Van de Graaff machine, used for high-voltage testing and the acceleration of subatomic particles. A tabletop version of the Van de Graaff generator and the classic rotary Wimshurst generator are widely seen in classroom demonstrations. A newcomer to the tabletop family is my own Dirod electrostatic generator, a rugged and reliable new design of the old induction generator.

Electrostatic motors, on the other hand, are little known. Yet the world's first electric motor, invented by Benjamin Franklin, was an electrostatic device! Much imagination was displayed in the early days of electrical research in turning out a surprising variety of electrostatic motors. This neglected art has been revived in recent years, chiefly by Oleg Jefimenko of West Virginia University, who has been busy making models of both old and new designs. One of his new electrostatic machines, a corona motor about five inches long, has developed a tenth of a horsepower. Recently Jefimenko accomplished the remarkable feat of putting up a wire by balloon and running a motor on energy from the atmosphere's electric field. Because of their very light weight and high speed, such motors may find special applications, perhaps in space research.

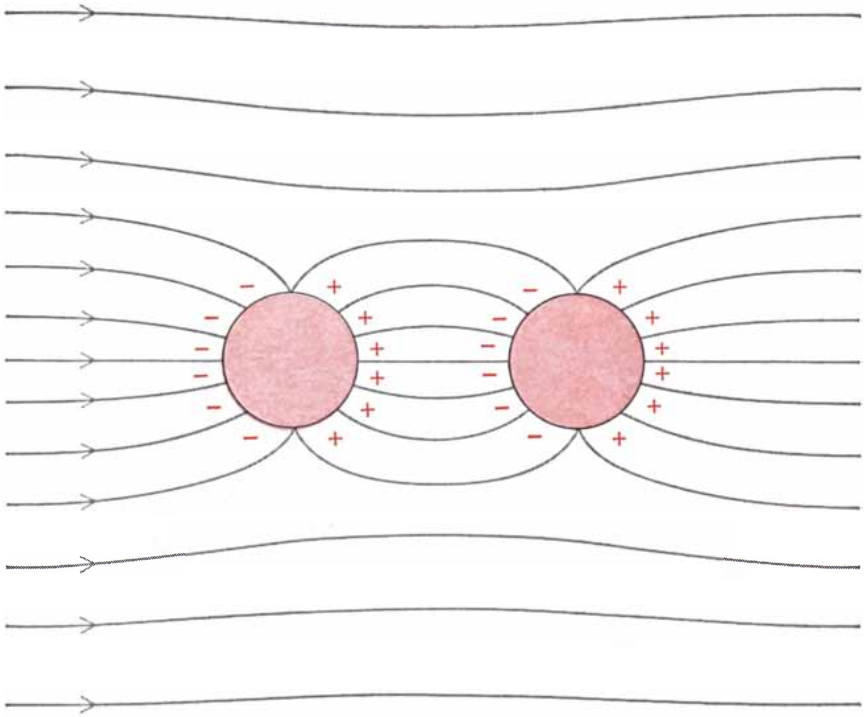
The diversity of modern electrostatics is by no means exhausted by the foregoing examples. None of them, for example, hints at nonimpact printing, an application that, like the horseless carriage, is named for what it is not. This article was born in the noisy impact printing of a typewriter; computer print-out machines also employ noisy and still comparatively slow impact-printing methods. A recent form of nonimpact printing produces a jet of ink with a tiny tube; the tube is vibrated at rates as high as 100,000 times per second to break the jet up into uniform, evenly spaced droplets. The droplets are then charged by induction. As they pass through two pairs of charged plates they are successively deflected back and forth and up and down in order to write characters silently at a high speed. A somewhat different electrostatic process, developed recently by Electroprint, Inc., can write 20,000 lines per minute on wide computer paper. Undoubtedly far higher nonimpact-printing speeds are possible. As computer speeds rise they strain the limits of impact printing, and electrostatic nonimpact printing stands ready to take over.

Other industrial applications that illustrate the versatility of electrostatics include corona devices that hold large sheets of paper to flat surfaces; electrostatic "chucks" that hold nonmagnetic objects in place for machining; electrostatic fluidized beds that spread powder evenly on a surface, preparatory to the fusing of the powder by heat to form a coating; electrostatic spinning machines that make cotton yarn, and electrostatic transducers, such as microphones, that are based on the electret, the electrostatic equivalent of the permanent magnet.

It is worth mentioning a number of other fascinating effects under the heading of electrokinetics. For example, one effect of moving gases and liquids electrostatically is to increase convective heat transfer. When a corona point drives ions away from itself, the ions drag the rest of the air along with them. One would expect such an electric wind from a corona point to enhance heat transfer. Actually the enhancement turns out to be much greater than the enhancement such a wind could accomplish alone. It may be that ions arriving at a heated surface break up the stagnant film of air that characteristically forms around such surfaces. Inter-Probe, Inc., has applied this cooling effect associated with corona probes to plasma-arc welding with several beneficial results: better control of the weld, the reduction of gaseous inclusions and greater strength due to the finer grain structures. This success may perhaps be explained in part in terms of corona chemistry, another old art much neglected until recently, in which a corona discharge creates numerous free radicals that go to work as chemical catalysts [see "Corona Chemistry," by John A. Coffman and William R. Browne; SCIENTIFIC AMERICAN, June, 1965].

A curious new device in this general category, developed by Senichi Masuda of the University of Tokyo, is called the electric curtain. In one form there is a series of rings set one below the other in order of decreasing size to form a funnel-like skeleton. Three-phase 60-cycle voltage is applied to create an electric field that sweeps downward, without a corona. If a cloud of charged particles is poured into the top of the funnel, the particles do not scatter. Instead they flow down through the rings and come out the bottom in a thin stream. The device as it stands constitutes a promising new particle-transport system.

The virtuosity of modern electrostatics research is further illustrated by an application totally different from any cited

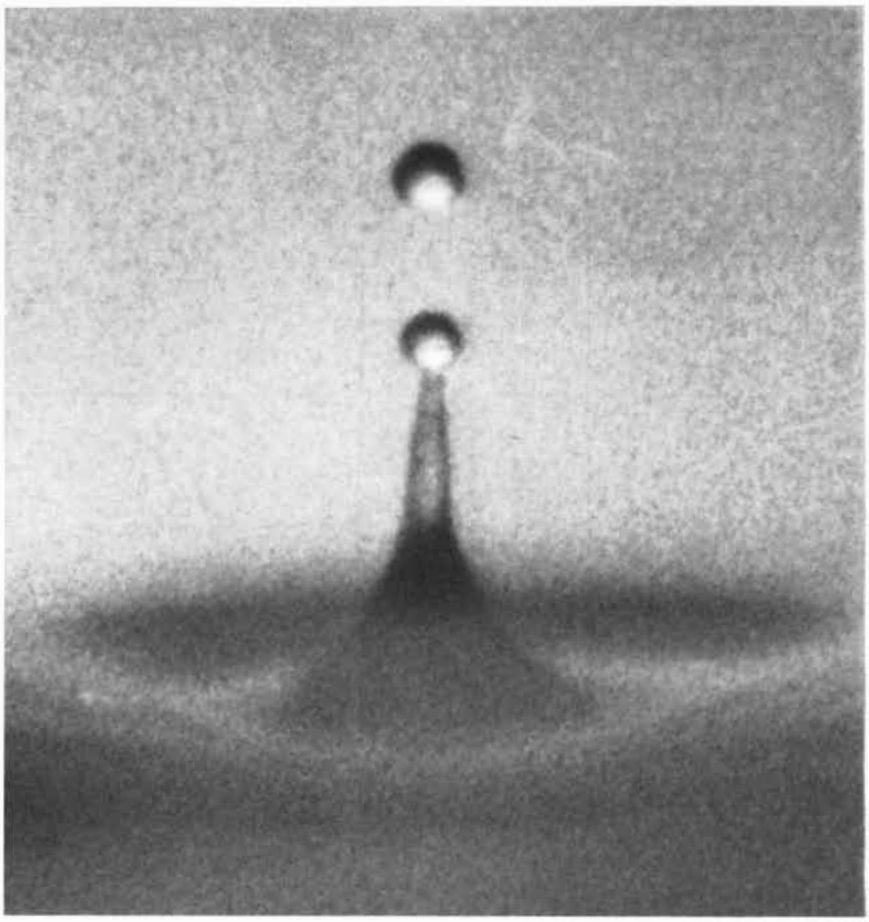


ELECTROSTATIC COALESCENCE of rain droplets under the influence of an electric field inside a cloud, illustrated schematically here, is precisely the same process as the "desalting" process used to remove water and salts from crude oil before it goes to the refinery.

so far. Electrostatically operated signs, developed by the Display Technology Corporation, are just now entering the market. In these devices small blackened vanes are placed edge to edge to make a square array of any size. When all the vanes are flipped to be seen edge on, a white background shows through; when they are flipped back to be seen face on, they appear black. One can electrostatically flip selected vanes to make a mosaic, creating letters, cartoons or other designs. Commands can be given locally by various means or from a distance by teletype. The many advantages of this electrostatic innovation suggest that a real revolution in signs lies ahead.

Last but hardly least, man's internal electrostatic processes rival in ingenuity any that man has been able to devise. Every cell of the body is surrounded by a membrane capable of maintaining a potential difference of about 70 millivolts between the contents of the cell and the intercellular environment. For all cells but nerve cells this potential difference remains constant. In the case of nerve cells, when a neuron is fired by an incoming stimulus, an electrostatic discharge races along the long tubular membrane of the cell at speeds as high as 100 meters per second to deliver the message to, say, a muscle fiber. Then the membrane restores itself in due course to carry succeeding impulses many times per second.

The diverse effects and applications of electrostatics cited in this article, together with many others I have not mentioned, indicate that the field of electrostatics research is in a stage of rapid development. With so many electrostatic effects available, there is obviously much more to come in the way of ingenious applications. I have no doubt that such contributions will continue to flow, as in the past, not only from highly trained workers but also from persistent amateur experimenters, among whom I have always counted myself.



CHARGE-SEPARATION MECHANISM thought to be responsible in part for maintaining the positive charge of the earth's atmosphere is represented by this high-speed photograph of a bubble bursting after rising to the surface of a body of water. A jet from the center of such a bursting bubble goes up at a high velocity, and from the tip of the jet several positively charged water droplets continue upward. When this happens at sea, the droplets, on evaporating, leave behind positively charged salt nuclei, which can be carried farther upward by the wind and eventually become the nuclei for the formation of rain droplets.

THE CAVE BEAR

This large-headed species lived from the Pyrenees to the Caspian Sea during the Ice Age. One cave alone has yielded the remains of 30,000 such bears. What caused the species to become extinct?

by Björn Kurtén

Many species of bears spend part of their lives in caves, but only one has ever been known as the cave bear. That species is *Ursus spelaeus*, which is now extinct. It was given this name by its 18th-century discoverers because they based their description on a skull found in a cave. The name became widely established during the next century as European fossil collectors turned up thousands of the animals' bones in caves all the way from the Spanish Pyrenees in the west nearly to the coast of the Caspian Sea in the east. In most of the caves where the bones of cave bears were found they comprised 90 or even 99.9 percent of all the fossils present. The largest single accumulation was discovered in the Austrian province of Steiermark, in a cave near Mixnitz known as the Dragon's Lair. The deposit contained the fossil remains of no fewer than 30,000 cave bears!

By paleontological standards *Ursus spelaeus* is a species with a very brief history. The first cave bears probably evolved at a time near the end of the second great ice advance of the Pleistocene epoch. That was the Mindel glaciation, which began some 700,000 years ago. The bones of an immediately ancestral bear species have been found in older fossil strata. Deposits laid down in the subsequent interglacial period, the Mindel-Riss, contain a few fossil remains of a true cave bear. A bear skull preserved in Mindel-Riss sediments at Swanscombe in England shows the domed forehead that is characteristic of the species, and other cave bear fossils of similar age have been found in a cave in Württemberg in Germany.

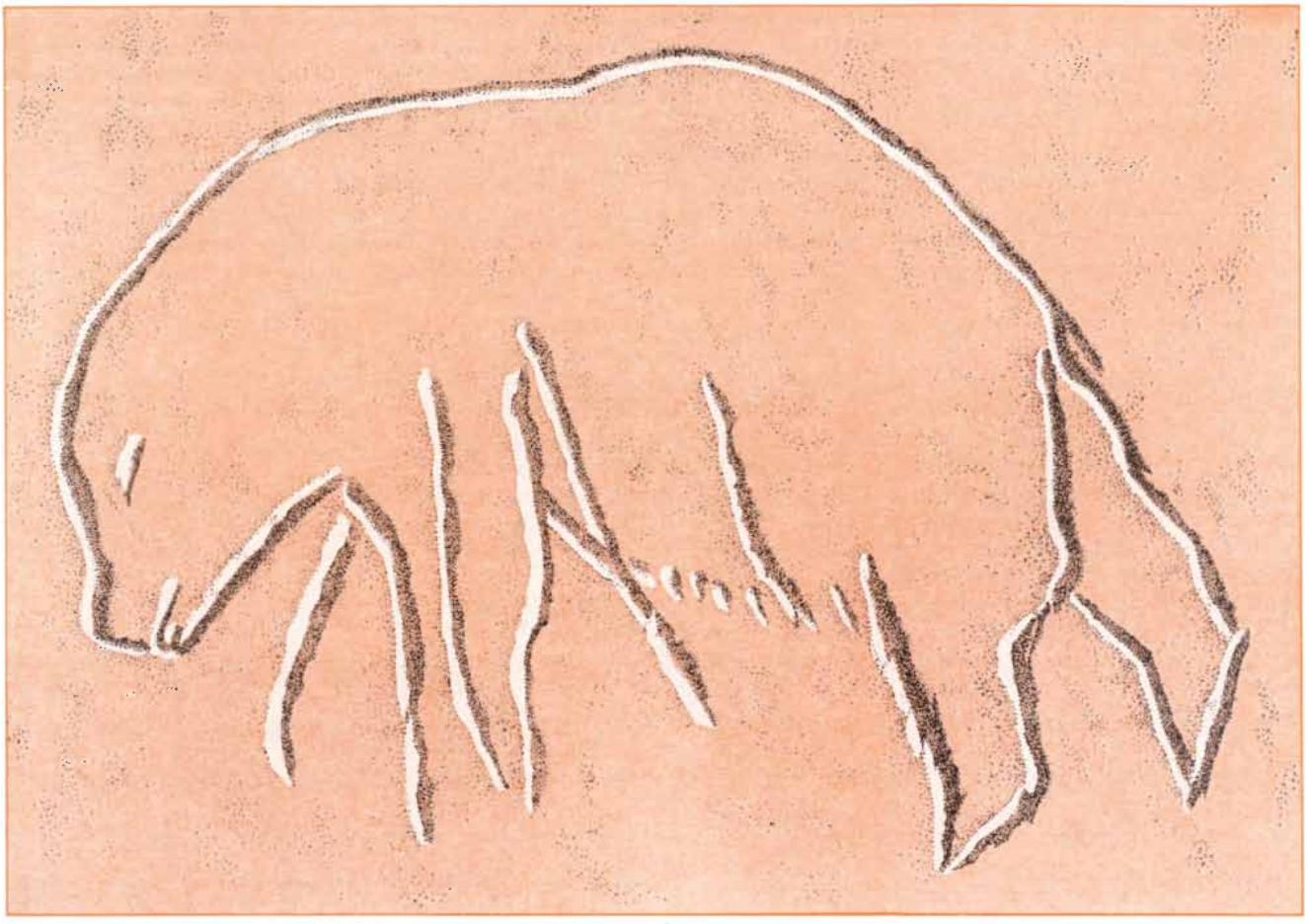
It is from sites of later Pleistocene times, however, that the cave bear fossils have come in the greatest abundance. For example, the very numerous remains found in the Dragon's Lair were evidently deposited there during the final Pleistocene ice advance, the 60,000-year Würm glaciation that ended some 12,000 years ago. This fossil accumulation and others of equivalent antiquity make it plain that the species flourished during the late Pleistocene. Yet by the end of the Würm glaciation, or at most a few hundreds of years later, *Ursus spelaeus* was entirely extinct. What can the fossil record tell us about the life of the cave bear? What accounts for its ultimate disappearance?

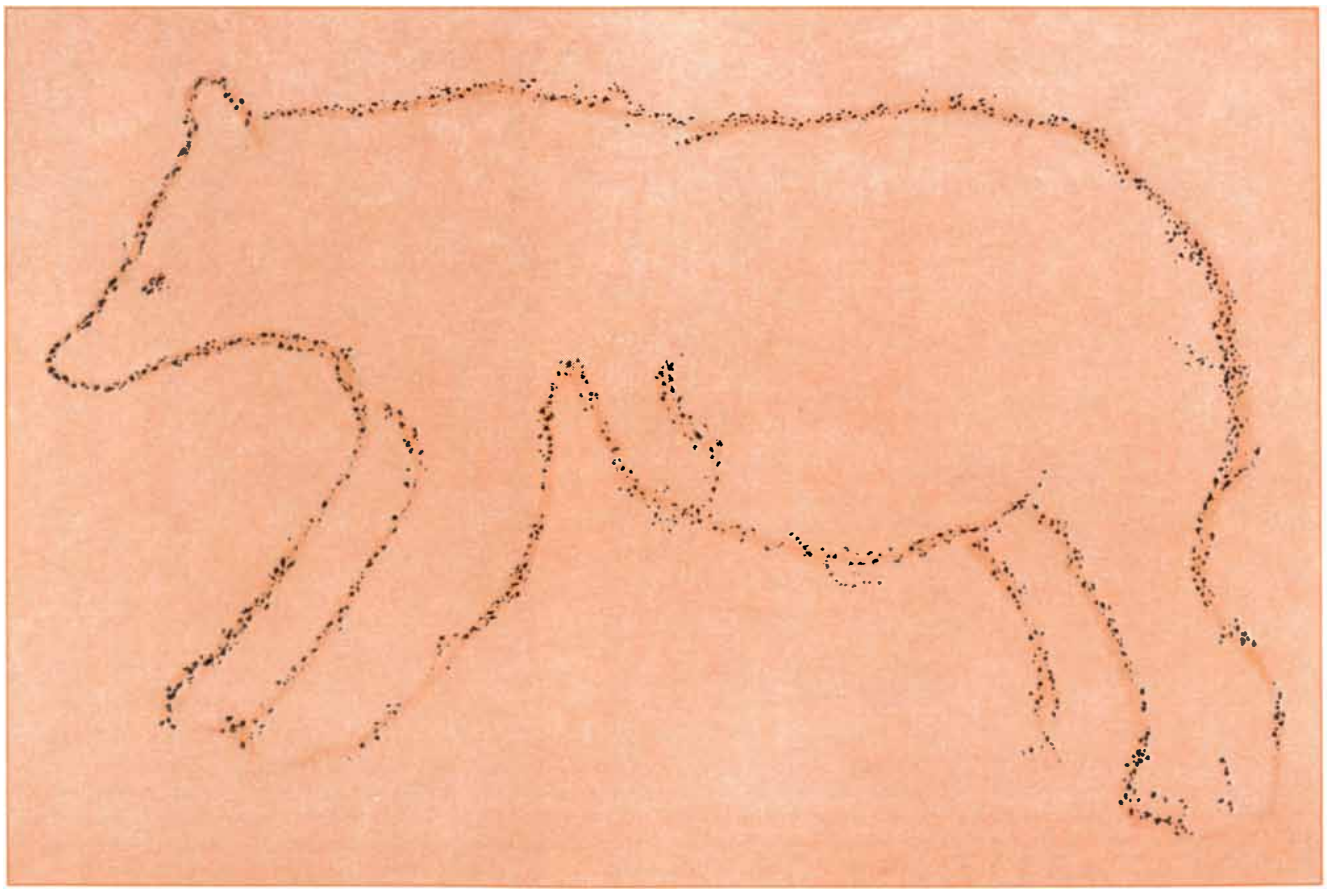
Kurt Ehrenberg of the Vienna Natural History Museum has made a detailed study of the thousands of fossils from the Dragon's Lair. His findings present a vivid picture of cave bear life. The most numerous specimens from the site are bear teeth. Ehrenberg found that the unworn milk teeth of very young bears—newborn animals and possibly even some fetal ones—were relatively abundant. The presence of this earliest age class at the site indicates that the cave bears came to the Dragon's Lair to hibernate during the winter months. It is during this winter interval that living bears of the Temperate Zone drop their cubs; there is no reason to believe the same was not true of the cave bears. The thousands of milk teeth at the site are the remains of bear cubs whose lives ended before they had ever seen the world outside the cave. Because the bones of newborn animals are extremely fragile almost no trace of the young cubs except teeth has been preserved in the Dragon's Lair. Ehrenberg was successful, however, in recovering an almost complete skeleton of a seven-month-old cave bear cub the size of a St. Bernard puppy—scarcely two feet long and a foot high—from another cave deposit in Steiermark [*see illustration on page 62*].

One-year-old bears, animals that had summered for one season outside the cave and then returned for their first hibernation, made up the next age class at the Dragon's Lair. A few of them still had some of their milk teeth but all had begun to cut their permanent teeth. The bones of the one-year-old bears are as rare as those of the newborn ones, so that almost no evidence of pathology is available to explain why bears of this age failed to survive their second winter.

Ehrenberg has identified two additional age classes—two-year-olds and three-year-olds—but thereafter he has difficulty distinguishing between sexually mature cave bears of different ages. No detailed census was made at the Dragon's Lair. Studies at other caves, however, indicate that roughly 70 percent of the bear fossils were from cubs that had died before reaching sexual maturity at age four or five. Nearly 30 percent of the remaining fossils were from bears that had died when they were quite old. Only a few of the fossils represented bears in the prime of life that had either been maimed in some fashion or were suffering from disease. Evidently once a cub reached maturity it might

PALEOLITHIC PORTRAITS of bears, by artists of the period who left their works in the caves of western Europe, include the engraved image of a cave bear (*rendered in the top illustration on the opposite page*), a species characterized by its domed forehead. The 20-inch-long likeness is one of the many Ice Age animal images at La Combarelle, near Les Eyzies in France. A brown bear (*bottom illustration*), with its receding forehead, is an Ice Age animal that still flourishes today. Its 12-inch-long image, outlined in black pigment, appears on the stalagmite-coated wall of a cave at Santimamiñe, a site near Santander in Spain.



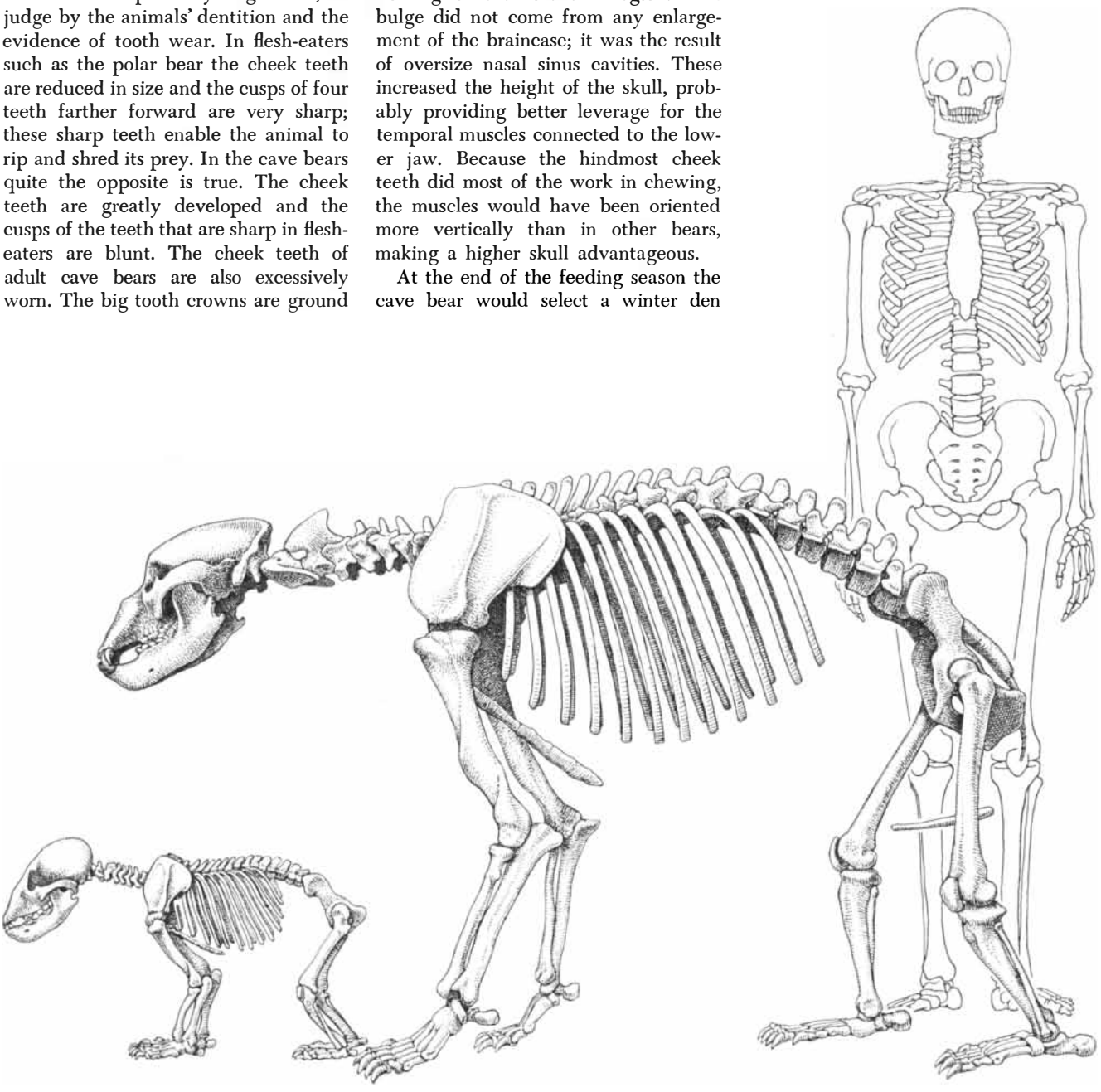


reasonably be expected to survive for another 15 years or so. Taking the cave bear population as a whole, the gross annual mortality appears to have been about 20 percent.

Cave bears were free-ranging during most of the year; they took shelter in caves only during their winter hibernation. Each spring they dispersed, the males moving off alone and the females traveling with their surviving cubs, and began a season of feeding and fat accumulation. Their diet seems to have been primarily vegetarian, to judge by the animals' dentition and the evidence of tooth wear. In flesh-eaters such as the polar bear the cheek teeth are reduced in size and the cusps of four teeth farther forward are very sharp; these sharp teeth enable the animal to rip and shred its prey. In the cave bears quite the opposite is true. The cheek teeth are greatly developed and the cusps of the teeth that are sharp in flesh-eaters are blunt. The cheek teeth of adult cave bears are also excessively worn. The big tooth crowns are ground down; in some of the older animals they are completely worn away and even the roots of the teeth show signs of wear.

Although female cave bears were consistently smaller than the males, both were formidable animals. From nose to tail they were about the same size as a grizzly bear, but their body was much heavier, with a deep, barrel-like chest. Their paws were short but broad, and the toed-in stance was more pronounced than it is among other bears. The cave bear's most striking characteristic, seen in all but a few fossil skulls, was the doming of the forehead region. This bulge did not come from any enlargement of the braincase; it was the result of oversize nasal sinus cavities. These increased the height of the skull, probably providing better leverage for the temporal muscles connected to the lower jaw. Because the hindmost cheek teeth did most of the work in chewing, the muscles would have been oriented more vertically than in other bears, making a higher skull advantageous.

At the end of the feeding season the cave bear would select a winter den and hibernate. The caves show no evidence of strict segregation by sex. In most natural populations of mammals males and females are born in approximately equal numbers; the cave bear apparently was no exception. As a result the fossil remains in many caves show a 50–50 ratio of male to female. This, however, was not true everywhere. In upper strata of the Dragon's Lair males



CAVE BEAR SKELETONS are those of an adult male, some five feet long from nose to tail and four feet high, in the collection of the Smithsonian Institution, and of a seven-month-old cub, only two feet long and one foot high, in the collection of the Vienna Natural History Museum. For purposes of comparison a human skeleton five feet 10 inches in height is included. Between the hind legs of the adult bear is the animal's penis bone. Many of these fragile bones had reknit after fracturing during the bear's lifetime.

If it costs 46% more to repair a car today than in 1961, what about a telephone truck?

With 100 million phones to care for, our fleet now numbers 110,000 trucks.

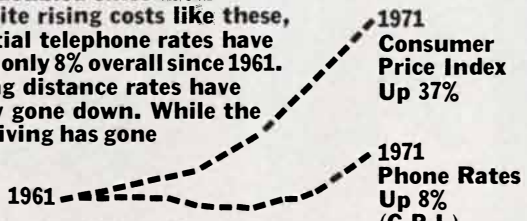
And with auto repair costs rising 46% since 1961, you can imagine what that's done to our budget.

We've been doing some of our own repair work, but our costs went up just as much—from $230 to $340 per truck.

Our total running expenses from $612 to $823 a truck.

And to get the money to buy new trucks, we're having to borrow at interest rates that have almost doubled since 1961.

Despite rising costs like these, residential telephone rates have gone up only 8% overall since 1961. And long distance rates have actually gone down. While the cost of living has gone up 37%.

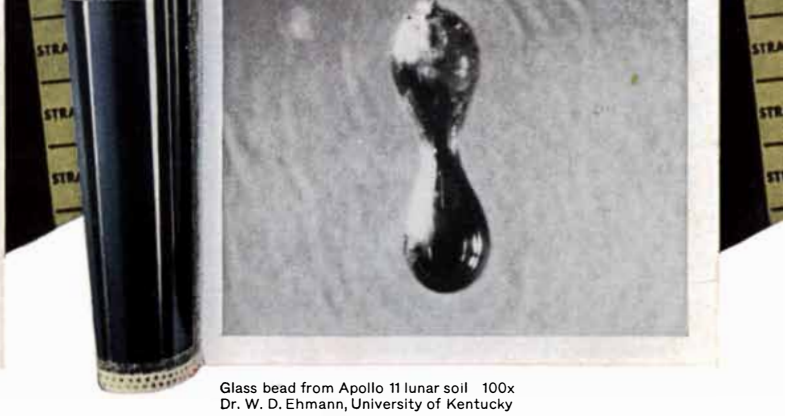


Obviously, this can't go on.

Because the cost of providing you good telephone service is going up, telephone rates are going up, too—but based on the last ten years, far less than most things you buy.

AT&T and your local Bell Company.

Mold spore from soil sample 430x
David McCurdy, Middletown, N.J.



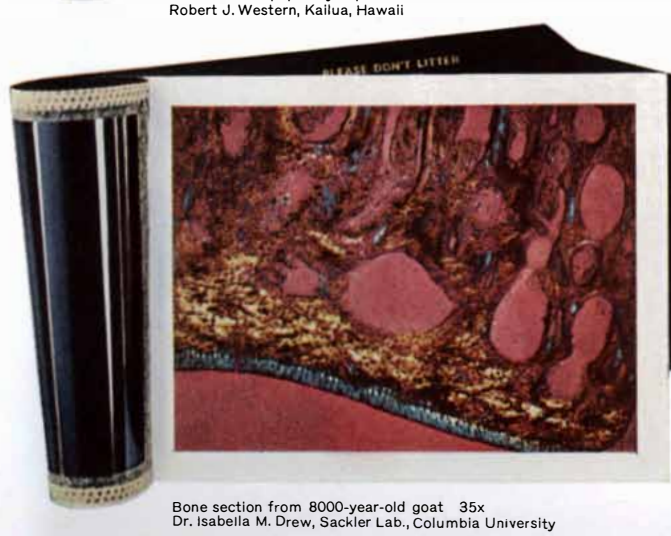
Glass bead from Apollo 11 lunar soil 100x
Dr. W. D. Ehmann, University of Kentucky

Spicule growth on radiolarian 1000x
Selwyn R. Mather, Elmhurst, Ill.

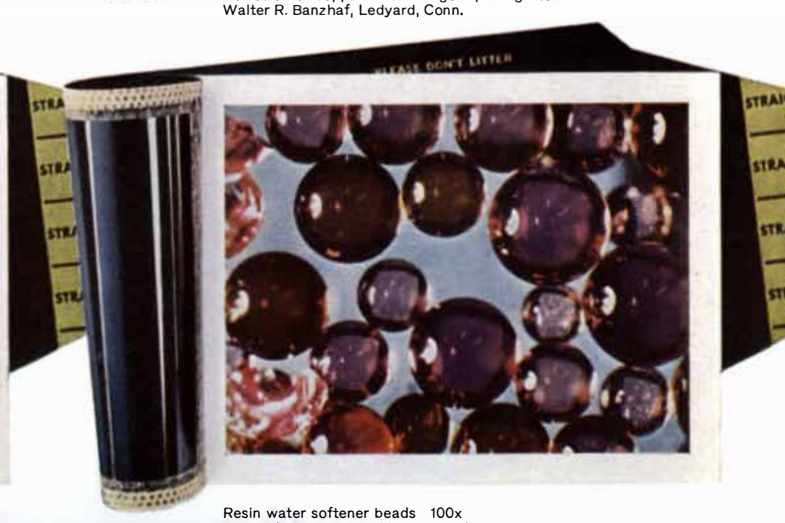
Larva of monarch butterfly 10x
J. Roger Matkin, Santa Ana, Calif.

Freshwater copepod cyclops 80x
Robert J. Western, Kailua, Hawaii

Corrosion of copper wire with gold plating 13x
Walter R. Banzhaf, Ledyard, Conn.



Bone section from 8000-year-old goat 35x
Dr. Isabella M. Drew, Sackler Lab., Columbia University



Resin water softener beads 100x
Benjamin B. Bonadio, Madison, Ind.

All it takes is a microscope, a minute, and less than $70.

You're looking at some winners in a nationwide photomicroscopy contest conducted by Polaroid Corporation.

All entries were taken with the Polaroid ED-10 Land Instrument Camera. Despite its low cost, you can see that it produces pictures as high in quality as those taken with the most expensive microscope cameras. Because the quality depends on the microscope's own optics.

The ED-10 fits virtually any microscope. Whether monocular, binocular or trinocular. And can be attached in moments. A universal adapter attaches to the eyepiece, and you simply slide the camera over the adapter.

You don't even have to know photography to use it. All you do is frame and focus the specimen in the microscope, slip on the focusing tube to refocus in the camera's film plane, replace the tube with the camera, and snap the picture.

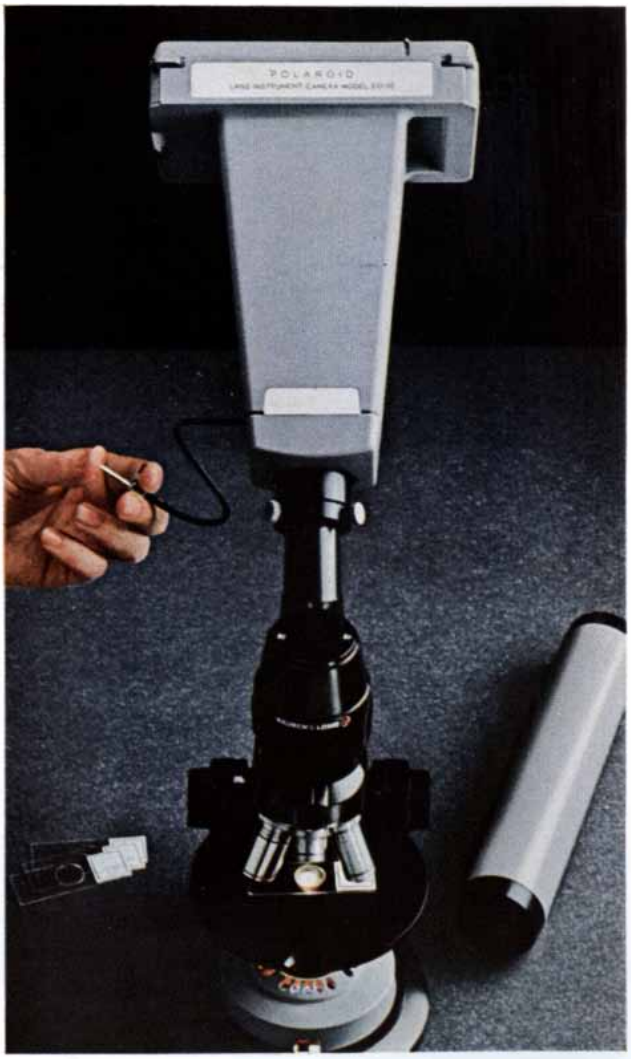
And you don't wait hours or days for results. All it takes is a minute for a full-color, finished photomicrograph and only 15 seconds for black-and-white.

You use Polaroid Land pack film, either Polacolor Type 108, or black-and-white Type 107 (3000 speed). Both are available almost everywhere film is sold.

The ED-10's body is molded of high-impact ABS material. So it's rugged, lightweight and remarkably easy to handle.

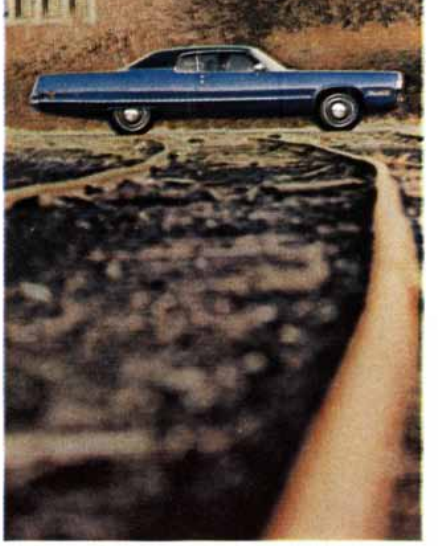
You can buy it at scientific supply houses, microscope dealers and many camera stores. But if you'd like more information first, write to Polaroid Corporation, Dept. 57-213, 549 Technology Square, Cambridge, Mass. 02139. In Canada, 350 Carlingview Drive, Rexdale, Ont.

It's well worth looking into. "Polaroid" and "Polacolor"®



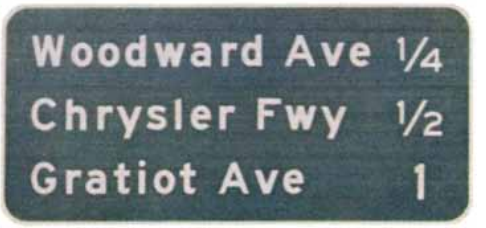
The Polaroid Land Instrument Camera.

It doesn't matter who you are. You can't get a Chrysler Imperial that hasn't had a shakedown in Detroit traffic.



ARTHUR GODFREY

Nobody, but nobody gets an unused Imperial. Every Imperial that Chrysler builds gets somewhere between twelve and fifteen miles put on it before it ever gets delivered. If there are any problems, Chrysler wants to find them before you do.



Each car is checked on city streets and freeways, in all kinds of traffic conditions. The test drivers listen for wind noise and check for rattles. They even inspect *under* the car: Things like the fuel line and brake line. There's a check sheet with more than 200 items that gets filled out on each car. Believe me, if it's not right, they don't want you to have it.

CHRYSLER
Plymouth

Coming through with the kind of car America wants.

That's their slogan this year, and I think they're living up to it.

I've seen the way they build the 1972 Chryslers. And believe me, they want these cars to last longer and perform better than any car they've ever built before.

Your Chrysler-Plymouth dealer can let you take an Imperial out for a road test of your own. You'll see what I mean.

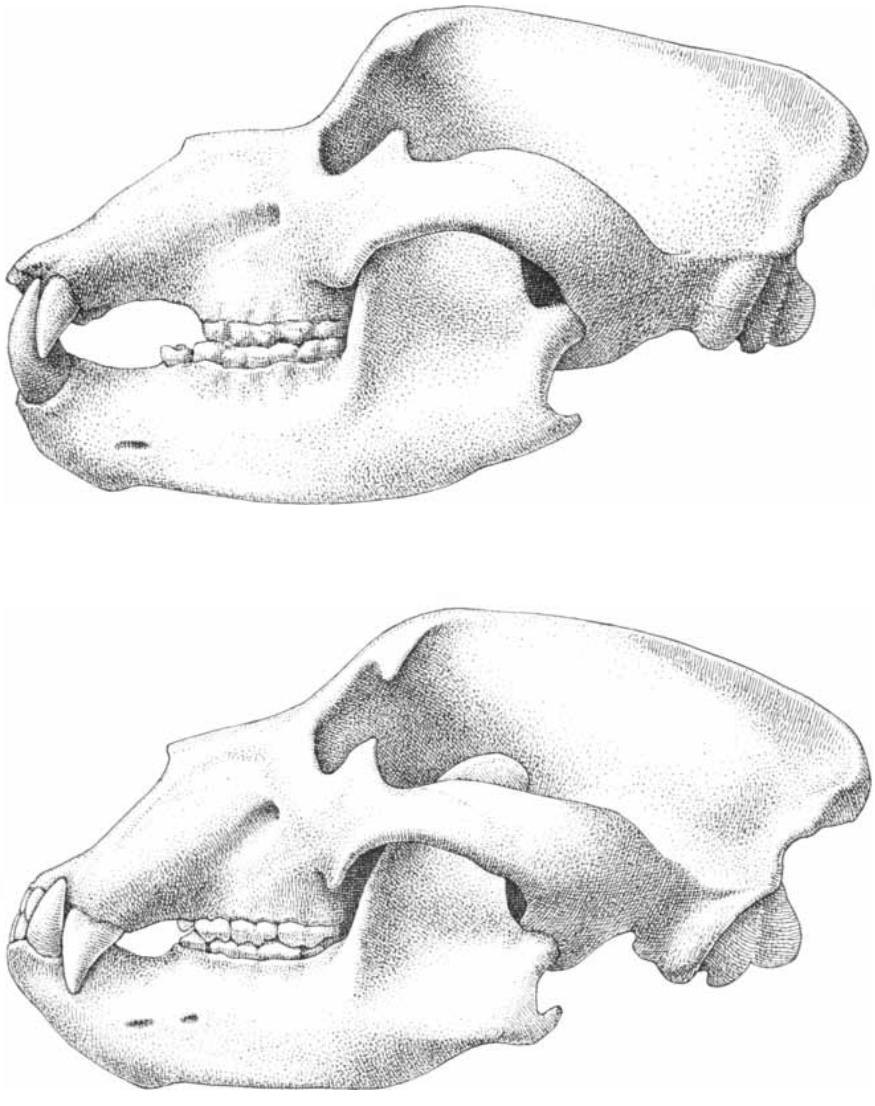
outnumbered females three to one; for the fossil collection as a whole the ratio is 60 percent male to 40 percent female. At Cueva del Toll in Spain 52 percent of the fossils are male. At Cotencher in Switzerland the male percentage is 48; at Gondenans, Montolivot and Saint-Bras in France the percentages are respectively 44, 33 and 28. When the dimensions of these caves are compared, it appears that sexual preference may explain the different ratios. Females predominated in the small caves and males in the large ones. Perhaps females with cubs to guard tended to select smaller, more easily defended caves where danger was not likely to lurk around the corner.

In some instances the sex ratios are quite bizarre, reflecting a human bias rather than any natural phenomenon. For example, a cave bear site at Hohlestein in Germany was excavated in the 19th century by a party from the Stuttgart Natural History Museum in cooperation with some amateur fossil-hunters. The fossils were then divided among the museum and the private collectors, with the museum having first pick. As a result the Hohlestein fossils at the museum are almost all large, showy specimens, which is to say males. Most of the private collections have since been lost, so that the true ratio of male to female cave bears at Hohlestein will never be known. Another joint venture undertaken by the Stuttgart museum, at a cave bear site named the Sibyls' Cave, gave first pick of specimens to the amateur diggers; the leftover fossils that the museum received were almost exclusively those of female cave bears. One unwary paleontologist, analyzing the collection without taking this fact into account, came to the conclusion that the bears of the Sibyls' Cave were a dwarf race.

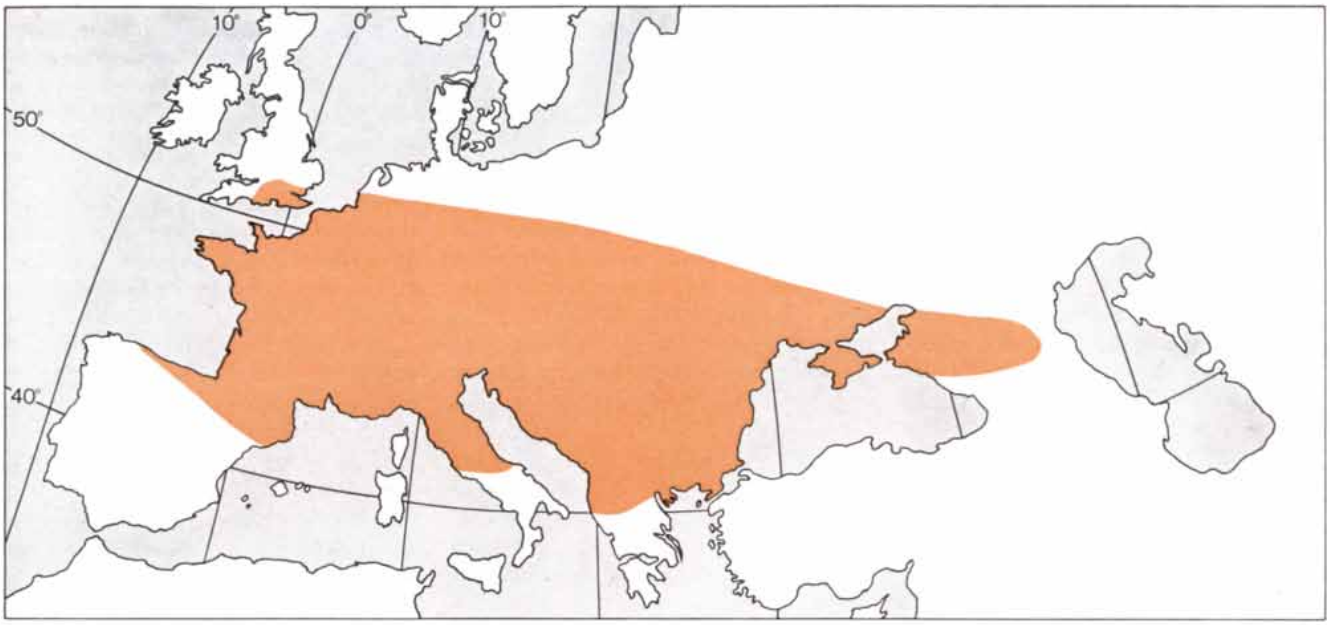
The cave bears' way of life, as revealed by the fossil record, thus seems to have been a simple enough annual cycle of wandering and feeding from spring to fall and hibernating and whelping in shelter during the winter months. Once past the perils of immaturity, little lay ahead for the adult except an uneventful repetition of the cycle for a decade or two. If the same bear made its winter den in the same cave two years in succession, this would perhaps have been no more than coincidence. Nonetheless, the large number of remains in many of the caves suggests that few of them were unoccupied for long.

Why is it that the very young and the very old bears were the ones most likely to die in their winter den? It must be that the bears of these age classes were the ones most likely to have experienced an unsuccessful summer season, and hence to have failed to build a store of fat sufficient to last them through the long and severe Ice Age winter. It is easy to see how this fate could befall the very old; among bears of advancing years tooth wear was increasingly severe. The completely worn-down teeth characteristic of old bears must have greatly hampered their summer feeding and left them ill-prepared for winter's rigors.

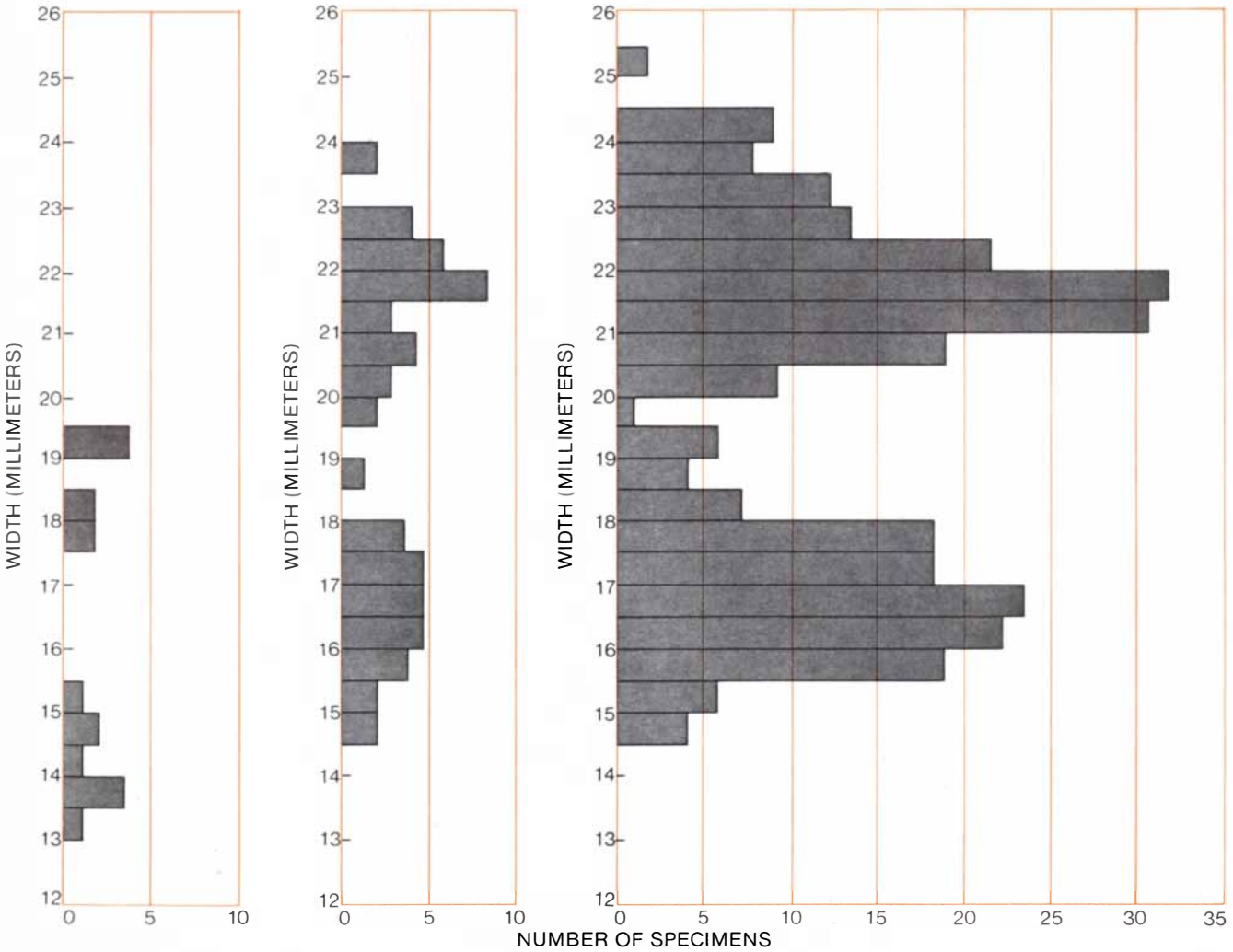
As for the high mortality rate among immature bears, some or perhaps most of the deaths can be attributed to simple inexperience and inadequate summer feeding. Yet accident, conflict and disease—which are apparent among the relatively rare remains of mature bears—may also have played a part. The roster of known cave bear diseases owes much to the careful fossil analyses of the Austrian pathologist Richard Breuer. Among his findings are instances of gout and related changes (which sometimes led to the fusion of limb bones and vertebrae), of rickets and of tooth decay and damage accompanied by inflammation of the jaw. For example, one well-preserved skull from a Swiss cave, which I have before me as I write, belonged to a powerful young male probably no more than five or six years old. One of its canine teeth is broken and there are traces of serious inflammation around the stump. If a sore jaw meant a bad summer season, the broken canine may have been the cause of the young male's death. If death from any such disability should



MALE BEAR SKULLS are shown here at the same scale for comparison; the cave bear skull (*top*) is actually some 20 inches long and the brown bear skull (*bottom*) is only 18 inches long. The more rounded forehead of the cave bear is due to an enlargement of the nasal sinus cavities. Its cheek teeth are larger and fewer in number than the brown bear's.



RESTRICTED TERRITORY of the cave bear (*color*) contrasts unfavorably with the far wider distribution of relatives such as the cosmopolitan brown bear. Even within its limited European territory the species was divided into small localized populations.



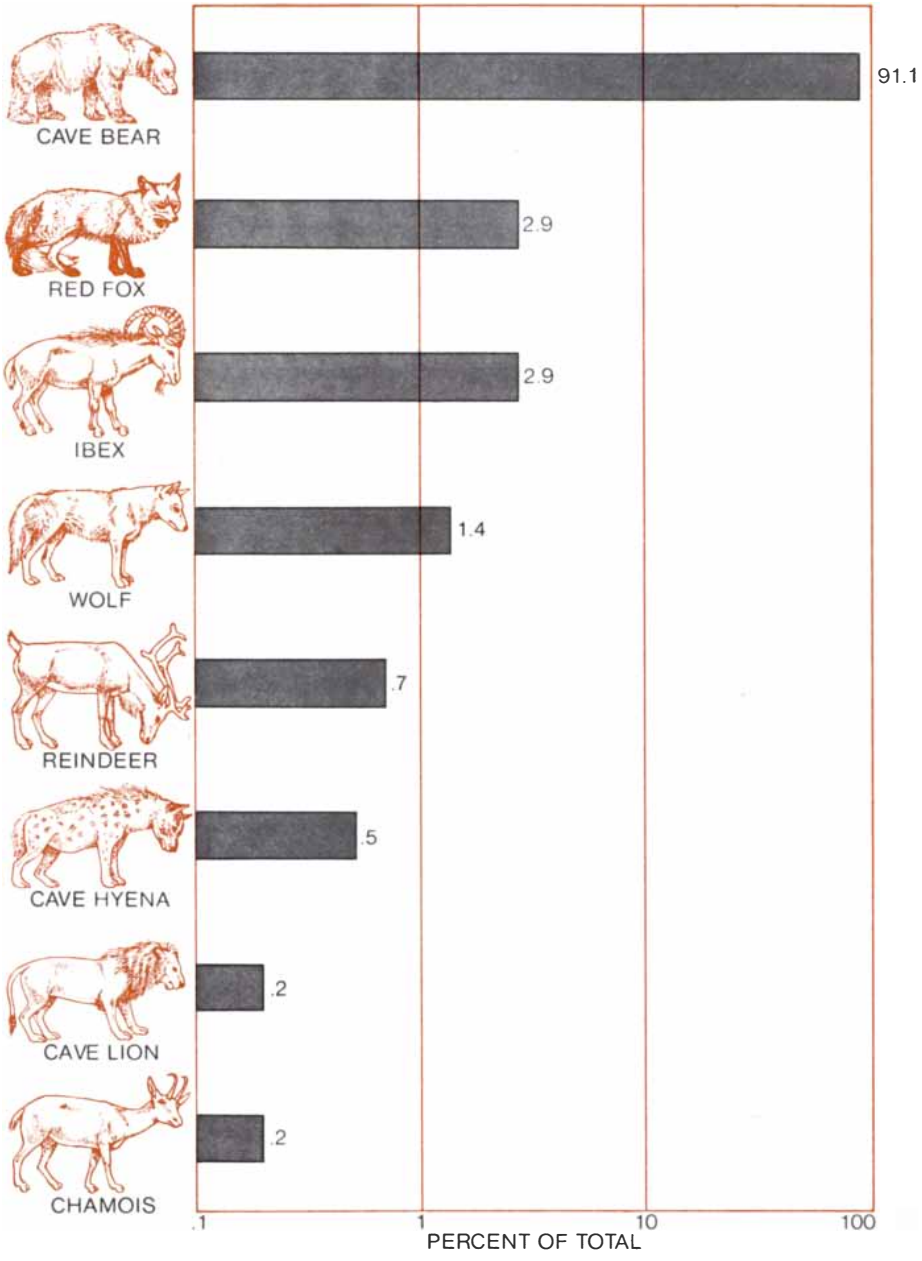
MALE CAVE BEARS were consistently larger than the females in the same population. The graph shows the average width of cave bear lower canine teeth in a Ukrainian fossil collection (*center*), in the collection from the "Dragon's Lair" near Mixnitz in Austria (*right*) and in another Austrian collection from Dachstein (*left*). Each collection shows two clusters of measurements. The larger sizes comprise the male teeth, the smaller the female. The dimorphism is apparent even among the dwarf bears from Dachstein.

have happened to strike a nursing female, her cubs too would have been doomed.

Accidental deaths must have been far less common, but there is evidence of fatalities caused by rocks falling from the ceiling of caves. A famous cave at Gailenreuth, near Würzburg in Germany, has yielded evidence of another kind of mishap: many cave bears apparently fell down an open shaft there from time to time. Either the animals were killed by the fall or, unable to escape, they died of starvation.

Cave bears would rarely have been killed by other animals; a mature bear in good health was much too powerful an animal to have fallen prey to most other Ice Age carnivores. In terms of size the only species in late Pleistocene times that might have been antagonists of the cave bear were the cave lion (*Felis leo spelaea*) and the leopard (*Felis pardus*). Conflicts between bears and lions or leopards are virtually unknown among living species, and there is no evidence to suggest that circumstances were different during the Ice Age.

There is evidence of fights between rival male cave bears during the mating season. Like many living carnivorous mammals, the cave bear had a penis bone. When these rather fragile bones are found in a fossil collection, some of them show signs of having been fractured and then having reknit, apparently as a result of the mating-fight behavior of the animals. If the other injuries in these encounters were severe enough to interfere with the animals' summer fattening, death might have followed during the winter. It does not seem improbable that a certain number of cubs would have died each year because of chance contact with a hostile adult; such incidents are known among living bears.



PREPONDERANCE OF CAVE BEARS among the fossil animals unearthed in the caves of Europe is typified by this inventory from Tischofer Cave, near Kurstein in Austria. Eight species of mammals, represented by some 420 individuals, were collected from one level there. Bears outnumbered even the next most abundant species more than 30 to one.

The great number of cave bear remains might easily suggest that in late Pleistocene times there was a cave bear "population explosion." Any such conclusion would be wrong. As Wolfgang Soergel of the University of Freiburg has demonstrated, no such hypothesis is needed. For example, consider the largest accumulation of all, the fossils from the Dragon's Lair. The death of no more than one cave bear at that site every other winter during the 60,000 years of the Würm glaciation would have been enough to produce the fossils found there.

Calculating on the basis of an estimated cave bear mortality rate of 20 percent per year, how large a standing population was needed in order to supply the necessary carcasses? The year-to-year census at the Dragon's Lair, it turns out, might never have exceeded two or three individuals—say a mother with one or two cubs one year and some solitary males the next. At other sites, where there are fewer cave bear fossils, the accumulations were evidently produced by even smaller numbers of bears taking up residence in a far more haphazard manner. The correct conclusion to draw, therefore, is that, far from growing explosively, the cave bear population of Ice Age Europe was probably always rather small.

At this point it is worth emphasizing a common item of paleontological knowledge. This is that the bones likeliest to be preserved as fossils are the bones of animals whose carcasses are protected against destructive processes soonest after death. Such protection can be the result of, among other causes, accidental drowning in a swamp, burial in flood sediment or tar or volcanic ash or, as is pertinent here, concealment in a cave. Now, it seems quite clear that most cave bear fossils are the remains of bears that died during winter hibernation. It appears probable, moreover, that rela-

tively few mature bears died before they reached old age. Admittedly the bones of any cave bears that died outside their winter shelters had only a small chance of preservation, so that our ability to estimate the number of such deaths is extremely limited. Nonetheless, the rugged physique of the mature cave bear suggests that the mortality rate was low during this part of the animals' life cycle. Finally, the fossil record of the caves includes a significant number of old bears, senile animals that met their end nearly toothless and almost certainly hungry. When one combines the large number of juvenile cave bears that died in the winter before they reached sexual maturity with the smaller but still substantial number that died of old age, it seems likely that the final figure represents a large percentage of all the cave bears that ever lived. If this interpretation is correct, and the caves of Europe have indeed sheltered the bones of most Ice Age cave bears, that would go a long way toward explaining the mass occurrences of these fossils. It would also be an instance of selective species preservation unique in the fossil record.

What are the major factors that brought about the extinction of the cave bear? One is that as a species the cave bear inhabited a surprisingly restricted geographic range compared with the ranges of other bear species. For example, in the British Isles the remains of cave bears are found only in a narrow strip of southern England. Their penetration into Spain was similarly limited. They are unknown to the south of Monte Cassino in Italy or south of Macedonia in Greece or much north of 51 degrees north latitude in the rest of Europe. The species' most easterly extension was a narrow corridor running from east of the Sea of Azov nearly to the Caspian Sea. The geographic range of the cave bear's evolutionary predecessors, *Ursus minimus* of the late Pliocene and *Ursus etruscus* of the early Pleistocene, was also essentially European. The other bear species that rose from the same stock, the Asiatic and American black bears and the cosmopolitan brown bear, have a much wider range.

Not only was the cave bear restricted in range as a species but also individual cave bears, unlike most large carnivores, had a rather small home territory. The evidence is the development of several local races of cave bears, most of them distinguished by subnormal size. Unlike the Stuttgart museum's hand-picked assortment of females from the Sibyls' Cave, these animals were genuine dwarfs. The males are about the size of normal female cave bears and the dwarf females are smaller still. The skull from the Mindel-Riss interglacial deposits at Swanscombe is that of a dwarf male. The female of the same dwarf English race may be represented by a skull from the lowest level in Kent's Cavern at Torquay; it is the smallest adult cave bear skull I have ever seen.

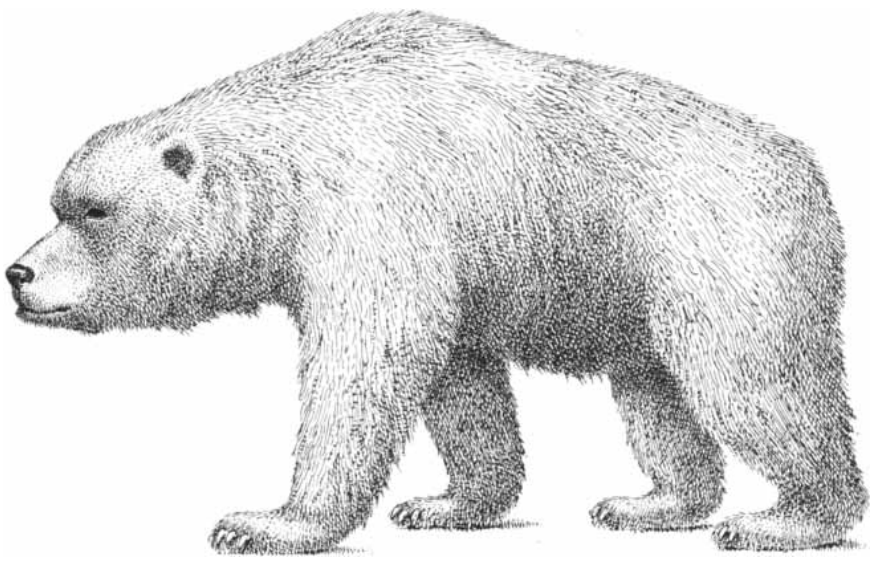
Austrian cave deposits of late Pleistocene age are notable both for their dwarf races of cave bears and for a puzzle the dwarfs present. If one selects some suitable index of the animal's size, such as the length of the crown of the last upper molar, and plots this measurement for each of the Austrian dwarf races according to the altitude at which it is found, an unmistakable negative correlation appears. The measurements range from a mean length of 45.5 millimeters at 1,000 meters above sea level to a mean length of 40 millimeters at 2,200 meters. Evidently the higher the site, the smaller the bear.

How is one to interpret this finding? It is scarcely credible not only that the bears sorted themselves out according to size and elevation but also that no valley bear ever went uphill, and no dwarf bear went downhill, to breed. A more logical assumption is that the larger and smaller races were not contemporaries; indeed, it is likely that the dwarf forms flourished only during warm intervals when the higher alpine sites where their remains are found were more inhabitable.

The result of isolation was not invariably dwarfism. The upper strata at Kent's Cavern and at Wookey Hole, another late Pleistocene cave deposit in England, contain the fossils of cave bears that may fairly be described as a giant race. Whether such races are large or small, however, their development bespeaks small home territories, stationary habits, breeding isolation and a minimum of contacts among the many widely scattered and not particularly abundant stocks of cave bears.

A species with a limited range that is further subdivided into a series of isolated races is ill-prepared to cope with the shock of drastic changes in climate and environment. The final millenniums of the Ice Age, which saw a European landscape that had consisted mainly of tundra, subarctic taiga and steppe transformed into all-enveloping Temperate Zone forest, brought with them just such a shock. Simultaneously the cave bears began to disappear from many regions. Some isolated groups may have vanished even before the changes became severe; when the number of animals in a local population falls below a critical minimum, the remaining animals are liable to accidental extinction.

There is some suggestive evidence to support the proposal that the cave bears' extinction was caused by environmental change. Rudolf Musil of the Moravian Museum in Brno has made an interesting discovery concerning the fossil remains in a bear cave in Czechoslovakia: the uppermost strata show an increase in the mortality rate for juvenile cave bears. In view of the fact that the normal juvenile mortality rate was already about 70 percent of births, an increase in the rate, presumably reflecting new environmental pressure, may in itself



CAVE BEAR IN THE FLESH probably looked very much as it is shown here. This is a sketch of the life-size reconstruction of a male at the Natural History Museum of Basel.

HOMING SALMON, THE, by Arthur D. Hasler and James A. Larsen, 1955 Aug. 72. [411]
HOMO ERECTUS, by William W. Howells, 1966 Nov. 46. [630]
HOMO MONSTROSUS, by Annemarie de Waal Malefijt, 1968 Oct. 112. [639]
HOMUNCULUS OR MONKEY? OREOPITHECUS, by Loren C. Eiseley, 1956 June 91.
HONEYBEE, THE, by Ronald Ribbands, 1955 Aug. 52.
HONEYBEES, SOUND COMMUNICATION IN, by Adrian M. Wenner, 1964 Apr. 116. [181]
HOOKE, ROBERT, by E. N. da C. Andrade, 1954 Dec. 94.
HOPEWELL CULT, THE, by Olaf H. Prufer, 1964 Dec. 90.
HOPI AND THE TEWA, THE, by Edward P. Dozier, 1957 June 126.
HORMONE, THE JUVENILE, by Carroll M. Williams, 1958 Feb. 67.
HORMONE, THE PARATHYROID, by Howard Rasmussen, 1961 Apr. 56. [86]
HORMONE, THE THYMUS, by Raphael H. Levey, 1964 July 66. [188]
HORMONES, by Sir Solly Zuckerman, 1957 Mar. 76. [1122]
HORMONES AND GENES, by Eric H. Davidson, 1965 June 36. [1013]
HORMONES AND SKIN COLOR, by Aaron B. Lerner, 1961 July 98.
HORMONES IN SOCIAL AMOEBAE AND MAMMALS, by John Tyler Bonner, 1969 June 78. [1145]
HORMONES, PLANT, by Victor Schocken, 1949 May 40.
HORN OF THE UNICORN, THE, by John Tyler Bonner, 1951 Mar. 42.
HORNS AND ANTLERS, by Walter Modell, 19[illegible] Apr. 114. [1139]
HOT ATOM CHEMISTRY, by Willard F. Li[illegible] 1950 Mar. 44.
HOT SPOTS IN THE ATMOSPHERE OF THE SUN, by Harold Zirin, 1958 Aug. 34.
HOUSE MICE, POPULATIONS OF, by Robert L. Strecker, 1955 Dec. 92.
HOUSE MOUSE, THE PHYSIOLOGY OF THE, by Daniel S. Fertig and Vaughan W. Edmonds, 1969 Oct. 103. [1159]
HOW A RATTLESNAKE STRIKES, by Walker Van Riper, 1953 Oct. 100.
HOW A TADPOLE BECOMES A FROG, by William Etkin, 1966 May 76. [1042]
HOW AN EGGSHELL IS MADE, by T. G. Taylor, 1970 Mar. 88. [1171]
HOW AN INSTINCT IS LEARNED, by Jack P. Hailman, 1969 Dec. 98. [1165]
HOW ANIMALS CHANGE COLOR, by Lorus J. and Margery J. Milne, 1952 Mar. 64.
HOW ANIMALS RUN, by Milton Hildebrand, 1960 May 149 [illegible]
Sept. 100. [93]
HOW CELLS MAKE ANT[illegible] Nossal, 1964 Dec. 10[illegible]
HOW CELLS MAKE MO[illegible] Allfrey and Alfred [illegible] [92]
HOW CELLS MOVE, [illegible] 184. [97]
HOW CELLS RECE[illegible] Miller, Floyd [illegible] 1961 Sept. 222
HOW CELLS SPE[illegible] and Antonie [illegible]
HOW CELLS [illegible] Lehninger, [illegible]
HOW CHILD[illegible] CEPTS, b[illegible]
HOW DO [illegible] Waddi[illegible]
HOW DO [illegible] 1958
HOW [illegible] Wi[illegible]
HOW [illegible] A[illegible]
HO[illegible]
H[illegible]
HOW LIVING [illegible] man K. Wessells, 19[illegible]
HOW MAN CAME TO NORTH [illegible] Ralph Solecki, 1951 Jan. 11.
HOW "NEWER ALCHEMY" WAS RECEIVED, by Lawrence Badash, 1966 Aug. 88.
"HOW NICE TO BE A PHYSICIST," songs by Arthur Roberts, 1948 Sept. 50.
HOW OPIATES CHANGE BEHAVIOR, by John R. Nichols, 1965 Feb. 80.
HOW PEOPLE INTERACT IN CONFERENCES, by Robert F. Bales, 1955 Mar. 31. [451]
HOW PROTEINS START, by Brian F. C. Clark and Kjeld A. Marcker, 1968 Jan. 36. [1092]
HOW REPTILES REGULATE THEIR BODY TEMPERATURE, by Charles M. Bogert, 1959 Apr. 105. [1119]
HOW SAP MOVES IN TREES, by Martin H. Zimmermann, 1963 Mar. 132. [154]
HOW SLIME MOLDS COMMUNICATE, by John Tyler Bonner, 1963 Aug. 84. [164]
HOW SNAKES MOVE, by Carl Gans, 1970 June 82. [1180]
HOW THEMISTOCLES PLANNED THE BATTLE OF SALAMIS, by Michael H. Jameson, 1961 Mar. 111.
HOW THINGS GET INTO CELLS, by Heinz Holter, 1961 Sept. 167. [96]
HUMAN [illegible] Dec. 104.
HUMAN THERMOSTAT, THE, [illegible] er, 1961 Jan. 134. [129]
HUMAN WALKING, THE ANTIQUITY OF, by John Napier, 1967 Apr. 56. [1070]
HUMMINGBIRDS, THE METABOLISM OF, by Oliver P. Pearson, 1953 Jan. 69.
HUMOR, RESPONSES TO, by Jacob Levine, 1956 Feb. 31. [435]
HUNTERS' CAMP, A STONE AGE, by Grahame Clark, 1952 May 20.
HUNTERS' VILLAGE IN NEOLITHIC TURKEY, A, by Dexter Perkins, Jr., and Patricia Daly, 1968 Nov. 96.
HUNTING IN NORTH AMERICA, ELEPHANT-, by C. Vance Haynes, Jr., 1966 June 104.
HUNTING SOCIETY, THE FLOW OF ENERGY IN A, by William B. Kemp, 1971 Sept. 104. [665]
HURRICANE MODIFICATION, EXPERIMENTS IN, by R. H. Simpson and Joanne S. Malkus, 1964 Dec. 27.
HURRICANES, by R. H. Simpson, 1954 June 32.

account for the cave bears' extinction.

Was Paleolithic man instrumental in the disappearance of the cave bear? Most probably his major influence was indirect. Certain evidence found in Upper Paleolithic paintings, engravings and sculptures indicates that the early Europeans who produced these works were familiar with the appearance of both the cave bear and the brown bear [*see illustrations on page 61*]. Paleolithic hunters may have attacked cave bears; some bear skulls show lesions that could have been produced by stone-tipped projectiles. Such evidence of hunting is rare, however, and the possibility that Paleolithic hunters exterminated the cave bears is remote. There is one site in Germany (Taubach, near Weimar) where an accumulation of bear bones seems to reflect the hunting activities of Paleolithic man, but almost all the bones at the site are those of brown bears, not cave bears.

Much has been said and written about a supposed "bear cult" among the peoples of Upper Paleolithic Europe. Most of the evidence that is presented in support of the cult hypothesis, however, can just as easily be explained on other grounds. For example, cave bear skulls have been found in natural crannies and niches in the walls of caves, as if men had placed them there. Taking into consideration the year-to-year sequence of natural events in a bear cave, the Swiss speleologist F. E. Koby has concluded that it is unnecessary to invoke a human agency to account for such findings. When the onset of cool weather caused the bears to start looking for a winter den, they must often have entered caves containing bears that had died the preceding winter. By the end of summer, thanks to the scavenging of smaller animals, any such carcass would have been picked clean; the skeleton might already have been partly disarticulated, with some of its bones crushed by hyenas and wolves. Now the skeleton would be further trampled, broken up and swept aside as the new settlers prepared their den. Koby calls this disarticulation process "dry transport." He believes it can account for not only the discovery of cave bear skulls in cave crannies but also a second kind of evidence cited by bear-cult advocates: the occasional finding of a bear skull with limb bones pushed into the eye sockets, the nasal opening or the opening for the spinal cord.

Another piece of evidence that is advanced by supporters of the bear-cult hypothesis is a life-size clay sculpture of a headless bear found in the cave at Montespan in the French Pyrenees. Whether or not this example of Upper Paleolithic sculpture, which is certainly ambitious in scale, should be given the status of a cult object seems open to question. Other animal species are even more frequently represented in the art of the period.

There remains a discovery made in the 1920's by the Swiss prehistorian Emil Bächler in the Drachenloch cave near Vättis in the Tamina Valley of Switzerland. What Bächler found in the cave were several coffin-like enclosures made with slabs of stone. A diagram that he published in 1923 represents one of the enclosures as containing the skulls of two cave bears, visible in profile. Here too Koby has challenged the bear cultists. His criticism is based on a later diagram of the same stone enclosure published by Bächler in 1940. In the enclosure there are now six cave bear skulls rather than two, and they are seen full face rather than in profile. One must add in fairness that this criticism does not in itself dismiss Bächler's original observation.

What if, in spite of such doubts, one accepts the hypothesis that an active and widespread bear cult flourished at the end of the Ice Age? Do the various bear skulls and limb bones that served as cult objects represent cave bears that were first hunted down and slain by the cultists and then preserved as trophies? In view of the number of bones already at the cultists' disposal in the bear caves, such a pattern of ritual trophy-hunting seems scarcely probable. Thus the bear cult, even if it did exist, could have flourished without producing significant pressure on the cave bear population.

Where Upper Paleolithic man may have helped to tilt the scale in favor of the cave bears' extinction was in competition for natural shelter. Unlike other bear species, which can and generally do hibernate successfully in an open-air den, the cave bears seem to have been entirely dependent on caves for winter refuge. Of course, not every such natural shelter was occupied in the winter by the people of the Upper Paleolithic. Nonetheless, human interference that affected only a percentage of the available bear caves at irregular intervals could have helped to reduce the numbers in local populations of cave bears below the minimum survival level. Ultimately this indirect human influence may have acted, in combination with the pressures of climatic and environmental change that marked the end of the Ice Age, to bring about the cave bears' extinction.

New... AUERBACH EDP Users' Notebook

The desk top consultant that costs you only pennies per day.

AUERBACH has prepared, especially for EDP users, a unique, ready reference service that provides the information you need to operate your EDP facility at maximum efficiency and at minimum cost.

The *AUERBACH EDP Users' Notebook,* in addition to saving you time and money, buys you protection against future information slippage. As the recognized watchdog of the EDP industry, we'll see that nothing important slips by without being brought to your attention.

We have been the leader in providing EDP information for nearly 15 years, but we have never offered a service for Data Processing Managers that is as useful, comprehensive, timely, complete and authoritative as the *AUERBACH EDP Users' Notebook.*

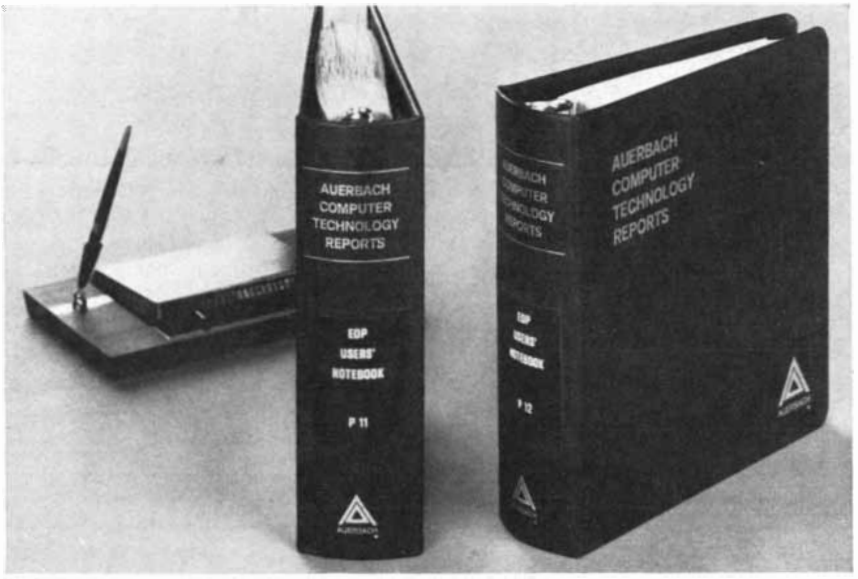
What's in it for you

The *AUERBACH EDP Users' Notebook* initially contains highly detailed reports on 25 important general-purpose computers, small business computers and minicomputers; 33 commercial system support software packages; 74 major and independent peripherals, which include devices, magnetic tape, disc storage, data handling, OCR and mark sense peripherals; and 17 data communications terminals. These reports will be updated and new reports added monthly.

In addition more than a dozen search charts pinpoint the characteristics and capabilities of 1,000 different devices, including 286 computer systems, 540 peripherals and 225 data communications terminals *providing you with the single most comprehensive source of information available today.*

Prepared by a research staff of 30 with a combined experience of more than 200 years in the EDP field, the *Notebook* covers every significant product offering in hardware, software, peripherals and data communications.

You could subscribe to 26 industry publications, pour over hundreds of EDP reference books, spend thousands of costly manhours . . . and still not get the breadth or caliber of EDP user information found in the new *AUERBACH EDP Users' Notebook.*



The two-volume set comes in simple, looseleaf form. It's organized by product categories for easy use. And it's updated monthly — our guarantee against your development of an information gap.

A one-year subscription is $245. A two-year subscription is $425, a $65 saving. With your subscription you'll receive our monthly newsletter, *AUERBACH Reporter,* Computer Technology Edition—absolutely free.

To order your *AUERBACH EDP Users' Notebook* or request additional information, send the coupon today.

To: AUERBACH Publishers Inc.
121 North Broad Street
Philadelphia, Pennsylvania 19107

Gentlemen : I wish a ☐ 1-year $245 ☐ 2-year $425 subscription to the *AUERBACH EDP USERS' NOTEBOOK,* and the free monthly newsletter, *AUERBACH Reporter.*

☐ Payment enclosed ☐ Bill my company ☐ I wish more information.

Ship to:

NAME _____

TITLE _____

COMPANY _____

ADDRESS _____

CITY _____

STATE _____ ZIP _____

AUTHORIZING SIGNATURE _____

AUERBACH ®

Do Infants Think?

They supposedly know the world only in terms of sensory impressions and motor activities. New experiments suggest that their cognitive, or hypothesis-forming, development begins at the age of nine months

by Jerome Kagan

The sturdiest knowledge we have about the development of infants is the exquisitely invariant order in which they master complex motor coordinations. Most infants can accurately reach for an object by the age of six months, can sit erect by eight months, can crawl by 10 months, can stand by 12 months and can walk steadily by 15 months. Although there are slight differences in the ages at which these abilities are attained, the sequence of events is rarely altered. This fact has tempted some psychologists to propose that deviations from the normal attainment of motor coordinations may be indicative of more general dimensions of intelligence. Except for the 7 percent or so of the infants who are retarded in their passing of these milestones because of damage to the central nervous system, there is no strong relation between the age at which these super-reflexes emerge and the child's later language capacity, richness of memory and problem-solving abilities. In spite of the obvious developmental control of the universal motor patterns, most psychologists have not been interested in asking if early cognitive functioning is similarly yoked to the maturation of the central nervous system.

My colleagues and I at Harvard University have been working in the area of infant cognition for some 10 years. Initially we believed, as many other psychologists do, that mental phenomena are much more plastic than motor skills and are influenced primarily by experience. The data we have collected, however, have forced us to adopt the view that certain cognitive processes in infants are to a large extent controlled by maturation. When we say that aspects of cognitive development are controlled by maturation, we mean that biological factors limit the earliest appearance of certain functions, that the order of appearance, given the appropriate inducing environment, is controlled by the growth of the central nervous system. This does not mean that experience is irrelevant.

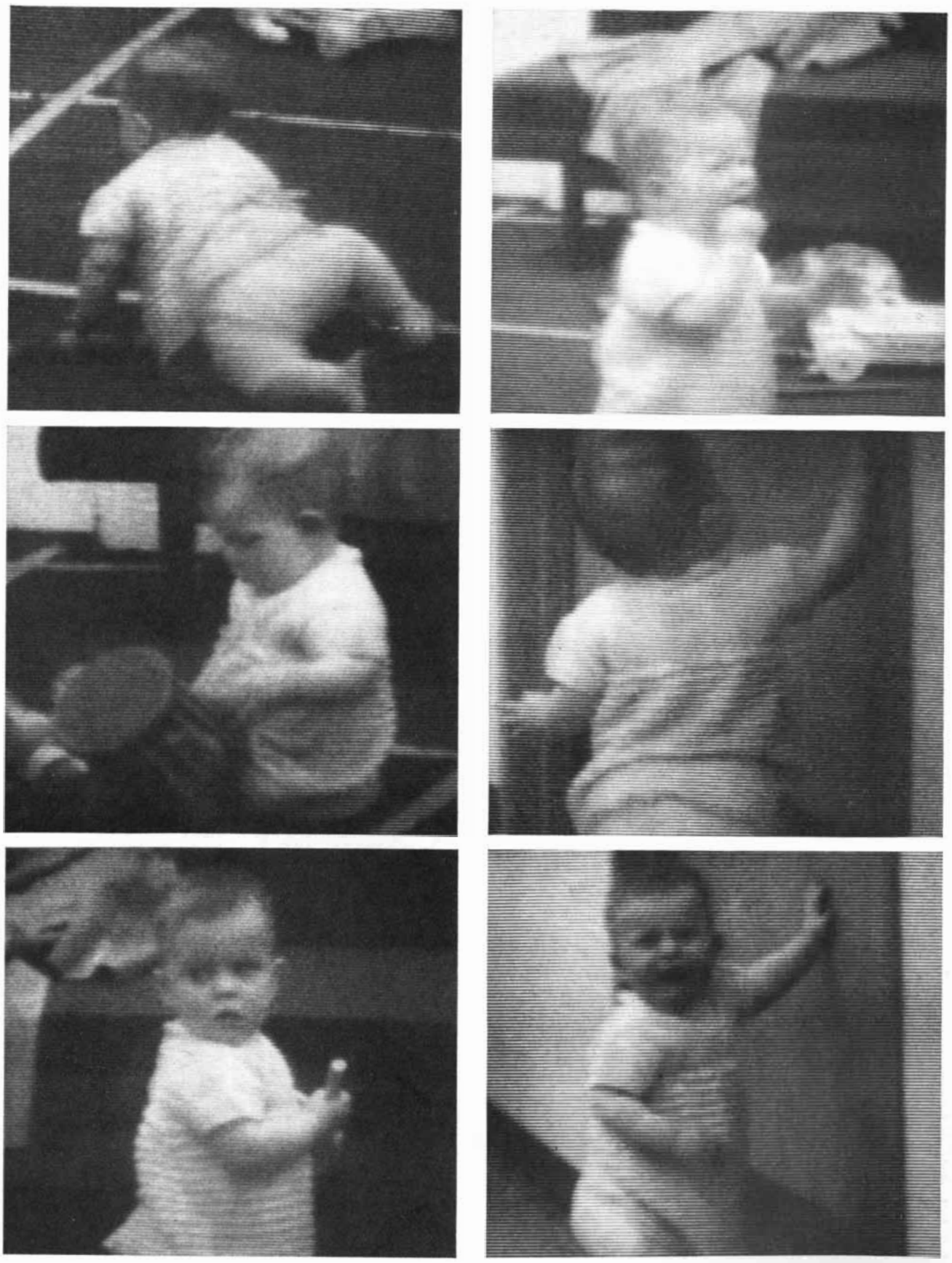
In our studies the primary measures of the infant's cognitive processes are selected kinds of behavior that accompany attention to an interesting event. By "attention" we do not mean the brief two-second orienting reaction of an infant to any sudden change in stimulus energy but rather the duration of sustained orientation that follows the initially orienting response. We believe that during the period of sustained orientation an infant more than 30 days old is trying to build a representation of the event. The duration of sustained attention is a rough index of how easy or how difficult it is for him to understand a new experience.

The attention of newborn infants is attracted by objects that move or have sharp contours and light-dark contrast. These perceptual preferences seem to be inherent in the structure of the visual system. A two-day-old infant is more attentive to a moving or intermittently flashing light than to a steady light; he looks longer at a solid black figure on a white background than at a low-contrast gray figure. The rate of stimulus change is also important during this first era of growth. If a stimulus is introduced too rapidly, the infant may become fearful. Similarly, as Richard Kearsley of Harvard has found, if an unexpected sound, say 70 decibels of "white" noise, reaches maximum intensity within a few milliseconds, a newborn infant closes his eyes, starts and shows an increase in heart rate, all of which are signs of a defensive response. If the same sound reaches its maximum intensity in two seconds, the infant then opens his eyes, looks around and is likely to show a decrease in heart rate, all of which are signs of interest.

The attention-recruiting power of contrast, which is maximal at birth or soon after, loses its force to a second property as early as the second month. As a result of the infant's encounters with the environment he acquires mental representations of events. We call these representations schemata. Toward the end of the second month the infant begins paying more attention to stimuli that differ moderately from those he usually encounters. The functional relation between duration of attention and the nature of the external event is summarized by the discrepancy principle: events that are moderately different from an infant's schema elicit longer spans of attention than either totally familiar events or totally novel events. Moreover, we suggest that the time of emergence of the infant's special reaction to discrepancy is controlled by maturational processes.

There is some controversy over the nature of the relation between the degree of discrepancy and the duration of attention. We believe it is curvilinear; others believe that increasing discrepancy is correlated with increasing attention. We do not have a way to measure degree of discrepancy, and we must be satisfied—only temporarily, we hope—with the qualitative statement that a discrepant stimulus is one that shares some aspects with its referent. A novel stimulus shares very few attributes with the standard or none at all. This definition always assumes the perspective of the viewer: an event that is novel to a six-month-old may be only mildly discrepant to a 26-month-old.

Data from several investigations directly support the curvilinear nature of the discrepancy principle or at least reveal that the infant reacts differently to a discrepant event compared with a novel event. In one experiment the infant



INFANT RESPONSES in experimental situations often are recorded on videotape by investigators for later analysis. By replaying the tape, the frequency and duration of many different kinds of behavior can be accurately measured. In these pictures, taken from videotape recording of an experiment by Milton Kotelchuck of Harvard University, a 12-month-old girl plays happily (*top left*) until she notices her mother leave (*bottom left*). Later the infant begins to cry and runs to the door through which mother had gone.

was shown a two-inch orange cube on six separate occasions. Then the infant was shown either a smaller orange cube (a discrepant event) or a yellow rippled cylinder (a novel event). Infants between seven and 12 months, particularly females, became excited by the discrepant small cube and showed their excitement by vocalization; the appearance of the novel rippled cylinder did not induce any vocalization.

A more direct demonstration of the discrepancy principle in auditory perception has been achieved recently by Dennis K. Kinney of Harvard. He had groups of seven-and-a-half-month-old infants listen to repetitions of four-syllable phrases spoken in a distinctive rhythm. There were four groups of infants, and each group heard repetitions of a different phrase. After eight repetitions of the phrase all the infants then heard exactly the same four-syllable phrase. For one group the final phrase was identical with the one they had been hearing. For two groups the final phrase was either slightly or moderately discrepant from the repeated phrase. For the fourth group the final phrase was markedly different from the original and could be classified as novel.

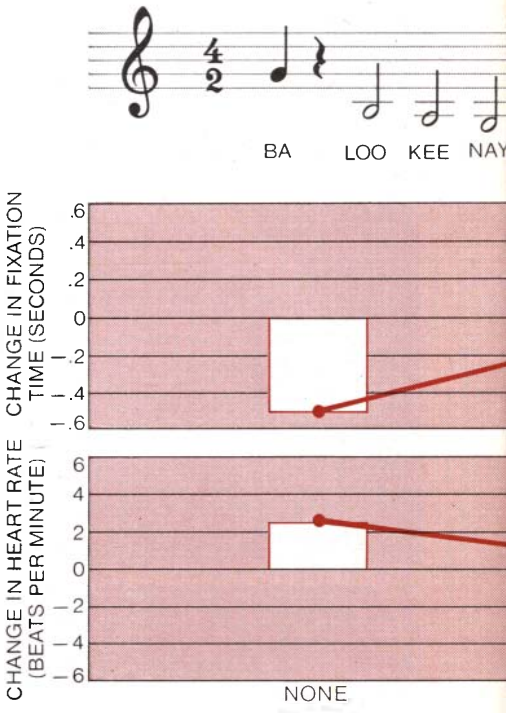
Two measures of attention were used: orientation of the head and eyes toward the sound source and a decrease in heart rate. The infants who heard the slightly or moderately discrepant spoken phrase displayed a longer orientation to the sound source and a lower heart rate than the infants who heard either the identical phrase or the novel one [*see top illustration at right*]. Both measures are indicators of the degree to which the infants were attempting to understand the new event or, to put it differently, how much attention was invested in the changed phrase.

An equally convincing example in visual perception of the curvilinear relation between discrepancy and attention is provided by a study of four-month-old infants. Initially they were shown an arrangement of three colored geometrical objects in a "mobile." Each infant was exposed to 12 half-minute viewings of the mobile, and the amount of attention, as measured by fixation time, was recorded. Then some mothers were given a mobile to hang above the child's crib for 30 minutes a day for three weeks. Other mothers did not receive a mobile. All the mothers were asked to bring the infant back in three weeks for further testing. At home some infants saw a mobile with the same arrangement of shapes they had viewed initially in the laboratory. Other infants saw mobiles at home that differed from the original either slightly, moderately or greatly.

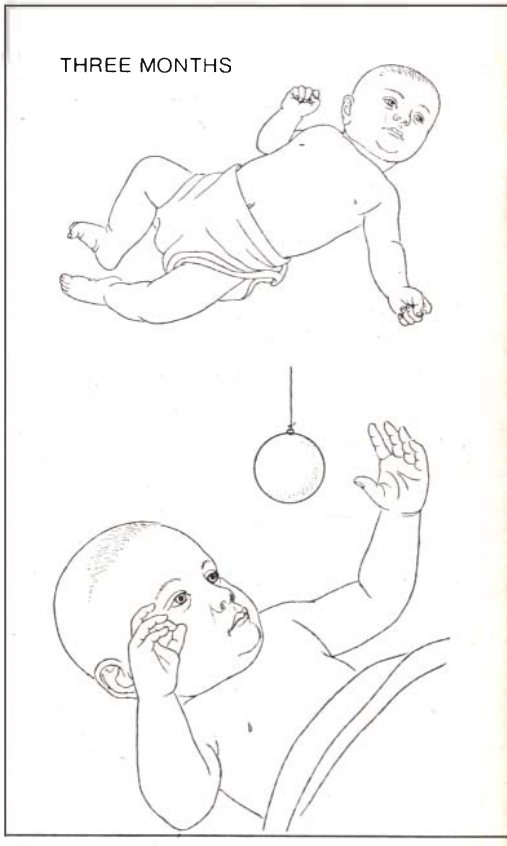
Each infant was then returned to the laboratory and was shown the same arrangement of objects he had seen three weeks earlier. The duration of sustained attention again was measured, and changes in attention between the first and the second viewing were assessed. Infants who had not been shown a mobile at home exhibited virtually no change in their attention to the objects. Infants who had seen moderately discrepant mobiles at home paid more attention to the laboratory stimulus than infants who had seen novel mobiles or mobiles with only minor variations. In short, the moderately discrepant mobile attracted more attention than one with great discrepancy or slight discrepancy [*see illustration on page 78*].

The discrepancy principle may help to explain the regular occurrence of fear responses in very young animals. A novel event rarely engenders fear, whereas a discrepant event often does engender it. Eric A. Salzen of the University of Liverpool has shown that alteration in the immediate environment of chicks gave rise to fear only in animals raised in a normal environment. Chicks reared in isolation and deprived of seeing any moving objects did not respond fearfully to a moving object such as a cardboard cylinder. Chicks that had lived under normal conditions did respond fearfully to the moving cylinder. We believe chicks raised in isolation perceived the moving object as a novel event and were not alerted by it. Chicks from the normal environment had developed a schema for moving objects and had some basis for regarding the moving cardboard cylinder as being discrepant. Because they could not assimilate this discrepant event they reacted with fear. Research on "imprinting" in birds also suggests that a bird does not fear a decoy on hatching but does react with fear a few days after the animal has had the opportunity to become familiar with its environment. It is the relation between an external event and a schema for that event that seems to be the critical determinant in provoking fear, as well as attracting and maintaining attention.

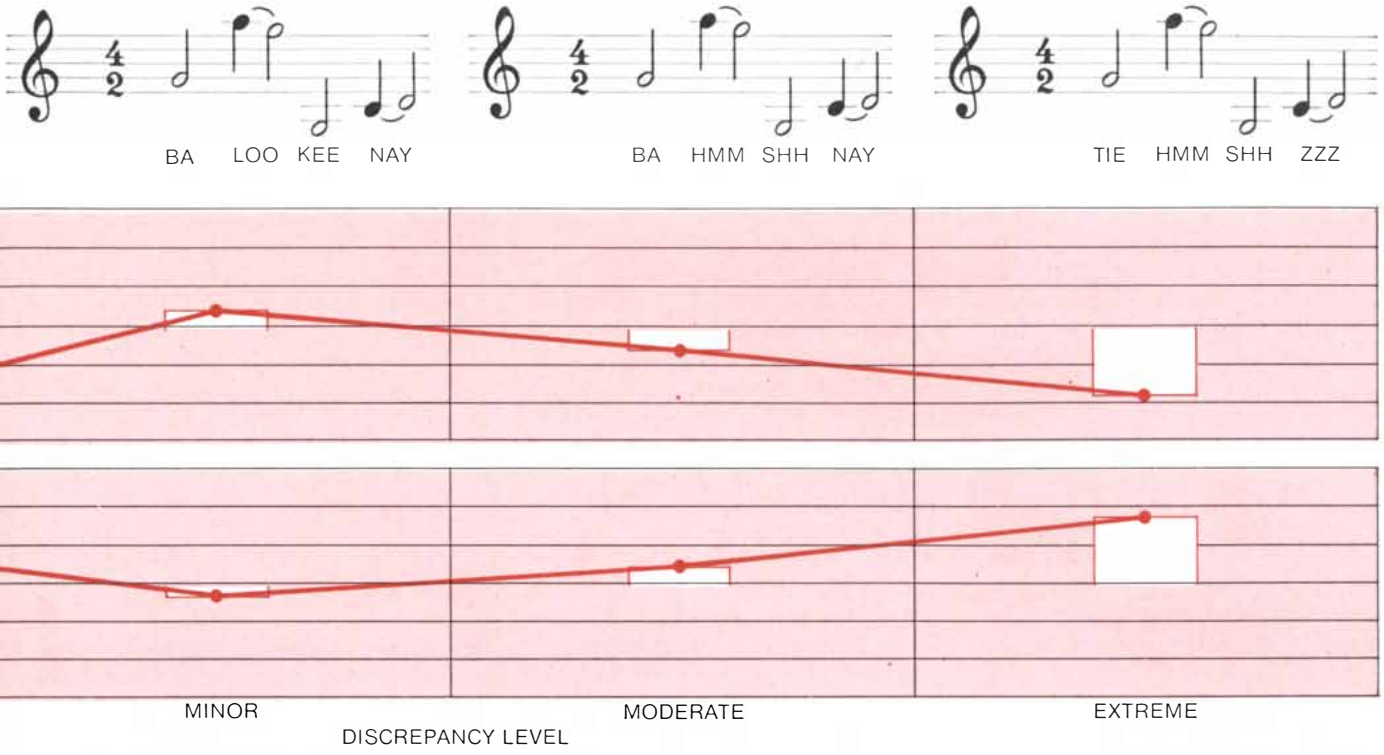
One of the significant problems in this area of research is how to quantify degree of discrepancy. Currently this classification is done after the fact. There are some hints that for visual discrepancy, at least in infants, the shape and arrangement of elements are among the



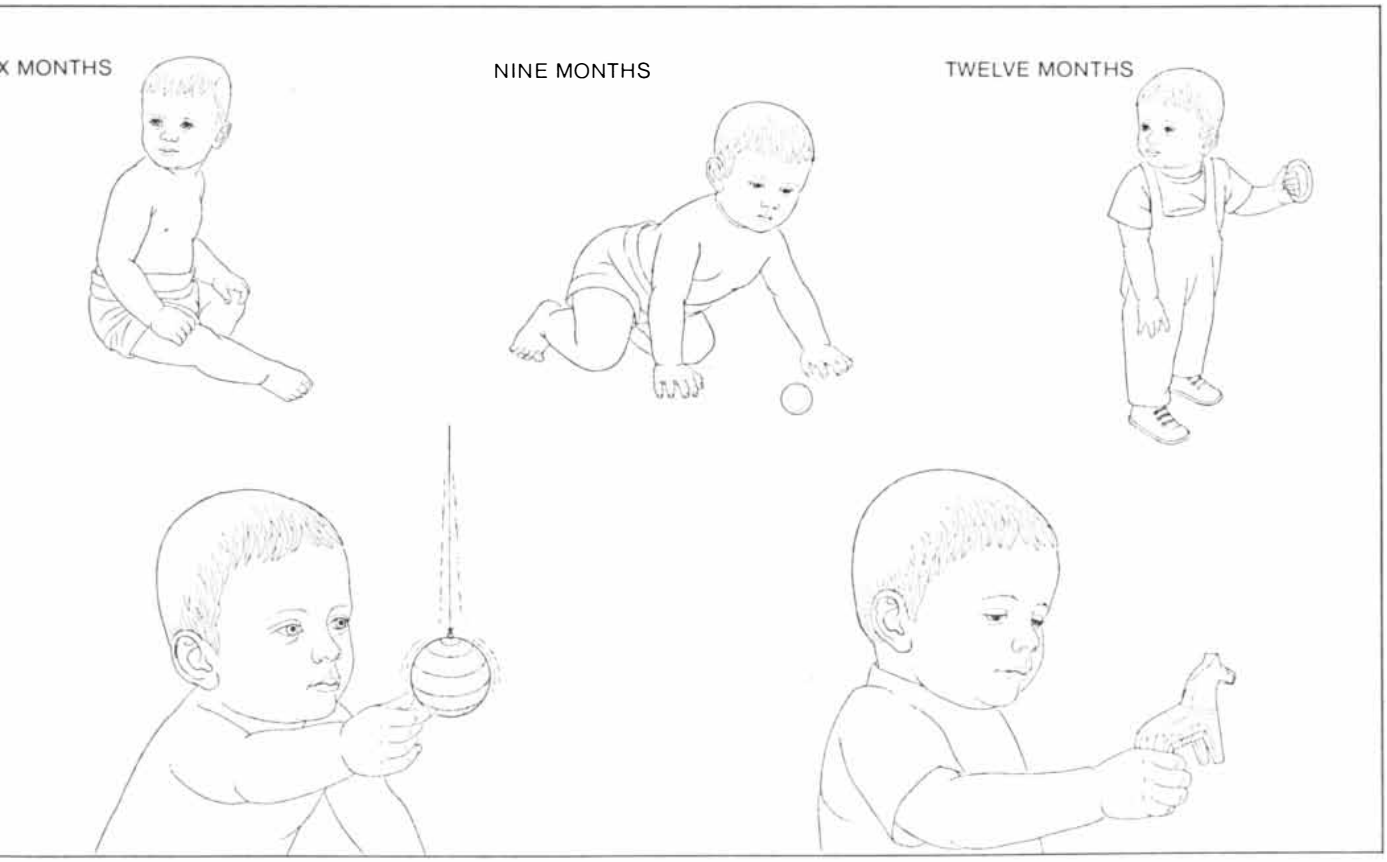
AUDITORY DISCREPANCY test involved four groups of seven-and-a-half-month-old infants. Each group listened to a different



DEVELOPMENT of motor abilities in infants (*top row*) is controlled by maturation of the nervous, muscular and other body systems. Most infants can roll over at about three months of age, sit erect between six



spoken rhythmical phrase (*top*). After eight repetitions all then heard the first phrase. Attention paid to the last stimulus varied according to the discrepancy of the phrase (*abscissa*). The slightly discrepant phrase drew the most attention as evidenced by the increase in orientation to the sound source (*top graph*) and the decrease in heart rate (*bottom graph*), both indicators of attention.



and eight months, crawl by 10 months and stand by 12 months. The order of motor development is rarely altered. Certain cognitive developments (*bottom row*) appear to be linked to maturation of the central nervous system. As early as the second month infants begin forming "schemata," or mental representations, of events and objects. For the next six months they pay attention to events and objects that differ moderately from these schemata. They pay less attention to familiar or totally novel events. At about eight or nine months appear the first signs of active mental work, in which the infant tries to generate hypotheses to explain novel events.

important dimensions to be quantified. Joseph Fagan of Case Western Reserve University has found that if six-month-old infants are repeatedly shown a pair of identical objects and minutes later one of the objects is altered either in shape or in arrangement, the infant looks longer at the changed member of the pair. For auditory perception rhythm seems to be as critical as the nature of the sounds in the recruiting of attention.

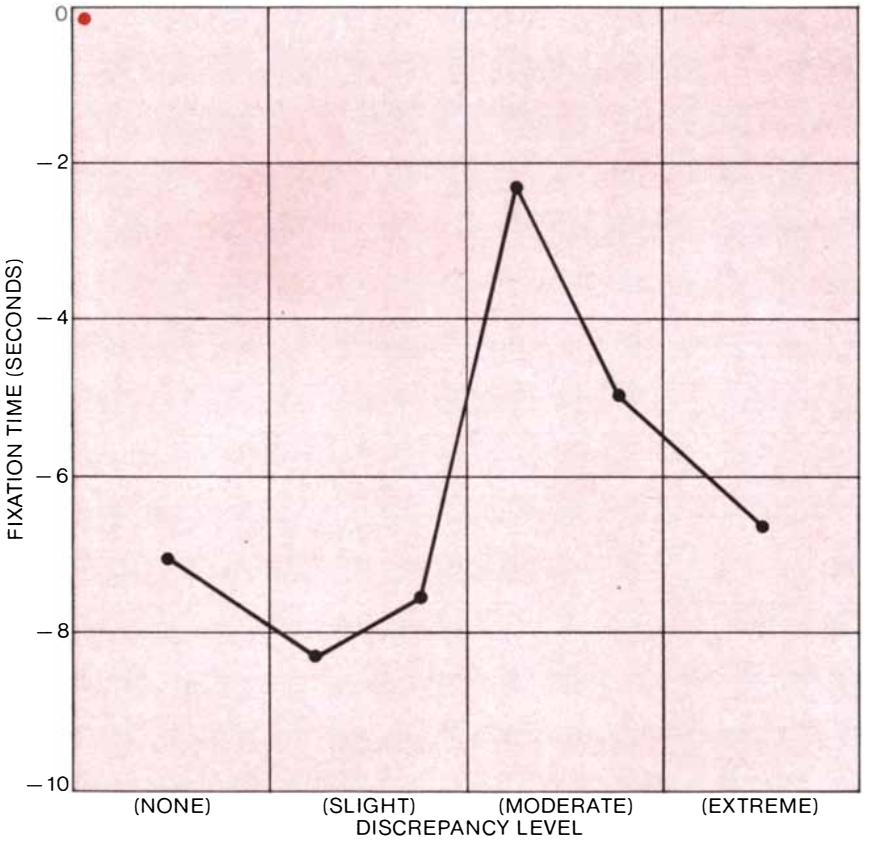
We believe the essential dimensions to be quantified for both visual and auditory events are the distribution of the stimulus elements in space and in time. This approach is intuitively attractive because of the history of the physical sciences. Chemistry had a major breakthrough in understanding when it was found that compounds could be classified by the spatial arrangement of atoms. Chemists assigned related names to compounds that had the same basic architecture but differed in minor ways. Physics used the arrangement of energy bands to classify complex sights and sounds. We are suggesting that psychology requires an analogous set of constructs that describes the spatial and temporal distribution of the events that comprise its basic forms.

Two empirical facts require the postulation of a third process that begins to control attention toward the end of the first year. Duration of attention to masklike representations of the human face decreases from two to nine months of age because the child's schema for a face becomes more mature and the masks become less discrepant. From nine to 36 months, however, duration of attention to the same facial masks increases. The *U*-shaped curve of this relation held true not only for American children but also for children living in extremely isolated rural areas of eastern Guatemala and for Bushman children living in isolated regions of the Kalahari Desert in Africa. If discrepancy from schema exerted the major control over attention, then the amount of attention should have continued to decrease after the first year as the infant's schema of the human face became better formed and the masks became increasingly less discrepant.

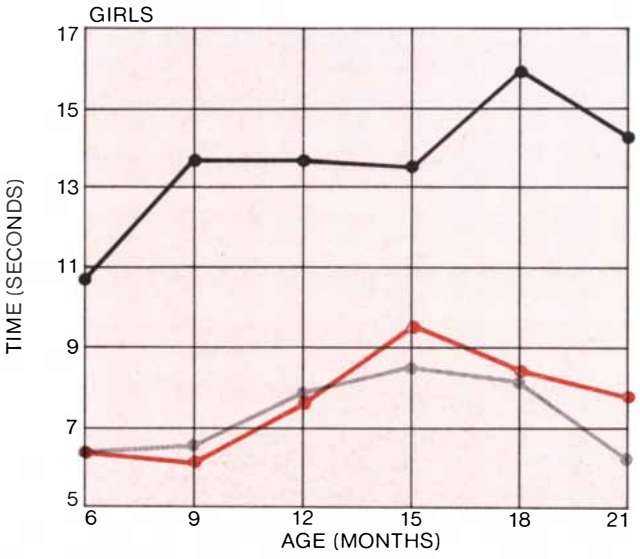
In order to explain the unexpected increase in attention toward the end of the first year, we postulated the emergence of a new cognitive structure or process. We call this presumed structure a hypothesis. The child, we suggest, tries to mentally transform the discrepant event into the form with which he is familiar, the familiar form being the schema. The cognitive structure involved in this mental transformation is the hypothesis. To be able to recognize that a sequence of high-pitched sounds is human speech rather than a birdsong requires schemata for the human voice and for birdsongs. On the other hand, the interpretation of why the speech is unusual requires the generation of hypotheses, which is an attempt to understand why the event is odd and how it might be related to the event that generated the original schema.

A quite different set of facts also supports the idea that "activation of hypotheses" emerges toward the end of the first year. Observations in varied cultural settings suggest that between eight and 15 months children are likely to show fear when they are separated from their primary caretaker.

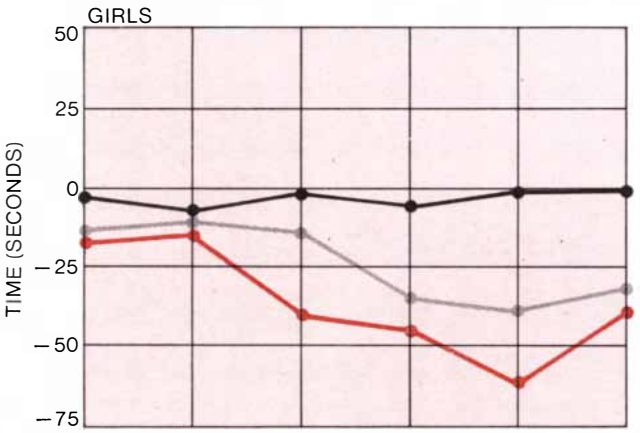
The extensive studies with infant monkeys by Harry F. Harlow and Margaret Kuenne Harlow of the University of Wisconsin have tempted psychologists to assume that fear following separation reflects the intensity of the infant's attachment to his mother [see "Love in Infant Monkeys," by Harry F. Harlow; SCIENTIFIC AMERICAN, June, 1959]. With human infants the intensity of attachment of the infant to his mother often is measured by how long the infant cries when he is separated from her. Our experiments suggest that crying may not necessarily index intensity of attachment but can be the result, at least in part, of the infant's inability to interpret the discrepant event described by watching his mother leave him alone, or with a stranger, in an unfamiliar situation. The arousal of fear under these conditions seems to be caused by the same mechanism that elicits fear in much younger infants, namely an encounter with a discrepant event that cannot be assimi-

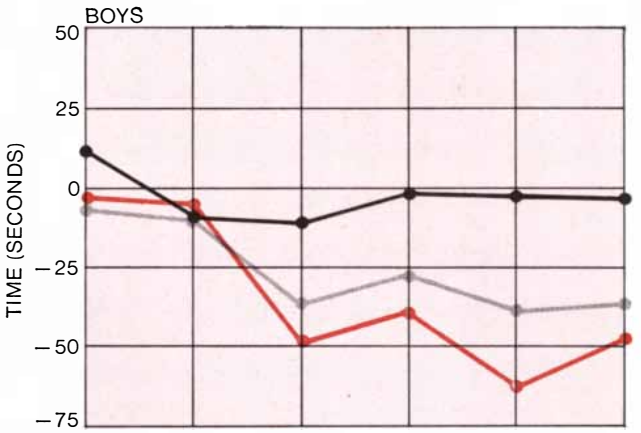


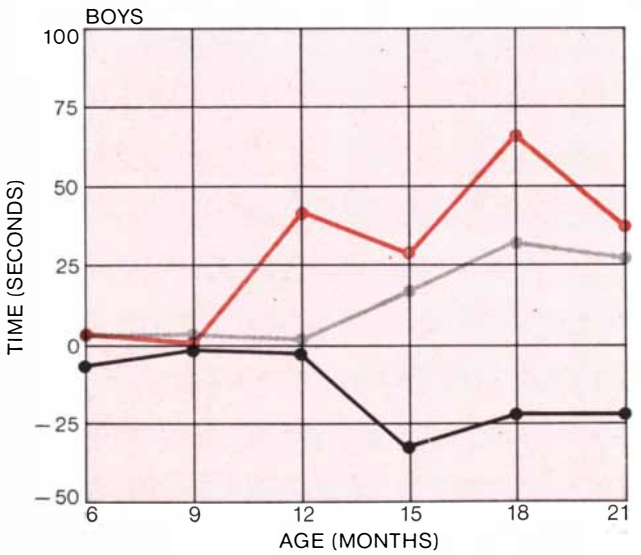
VISUAL DISCREPANCY study with four-month-old infants involved measuring the amount of attention (eye fixation) given to an arrangement of geometrical objects. After baseline measurements were made in the laboratory the infants were returned home with special "mobiles" to be hung above their crib for 30 minutes a day. Some infants had an arrangement identical with what they had viewed in the laboratory. Others had slightly, moderately or extremely discrepant arrangements. A control group had no home mobile. After three weeks changes in attention to the initial arrangement were determined. There was no change in the attention span of infants who did not view a mobile at home (*colored dot*). Of the infants who had viewed a mobile at home, those with the moderately discrepant mobiles showed the smallest drop in attention. Infants who had identical or similar mobiles showed the greatest drop in attention. Extreme discrepancy also drew less attention.



ARRIVAL OF A STRANGER drew more attention (*black curve*) from both male and female infants than arrival of either mother alone (*colored curve*) or father alone (*gray curve*). Infants 18 months old paid the most attention following the stranger's arrival.







EFFECTS OF DEPARTURE, as measured by change in infants' play (*upper graph*) and change in crying (*lower graph*) during the first minute after departure, are shown. In infants younger than nine months the departure of the mother (*colored curve*) or the father (*gray*) or the stranger (*black*) resulted in a relatively small change in playing and crying. In older infants the departure of mother or father resulted in a large decrease in play and a large increase in crying; departure of the stranger resulted in less crying.

lated, interpreted, rejected or destroyed.

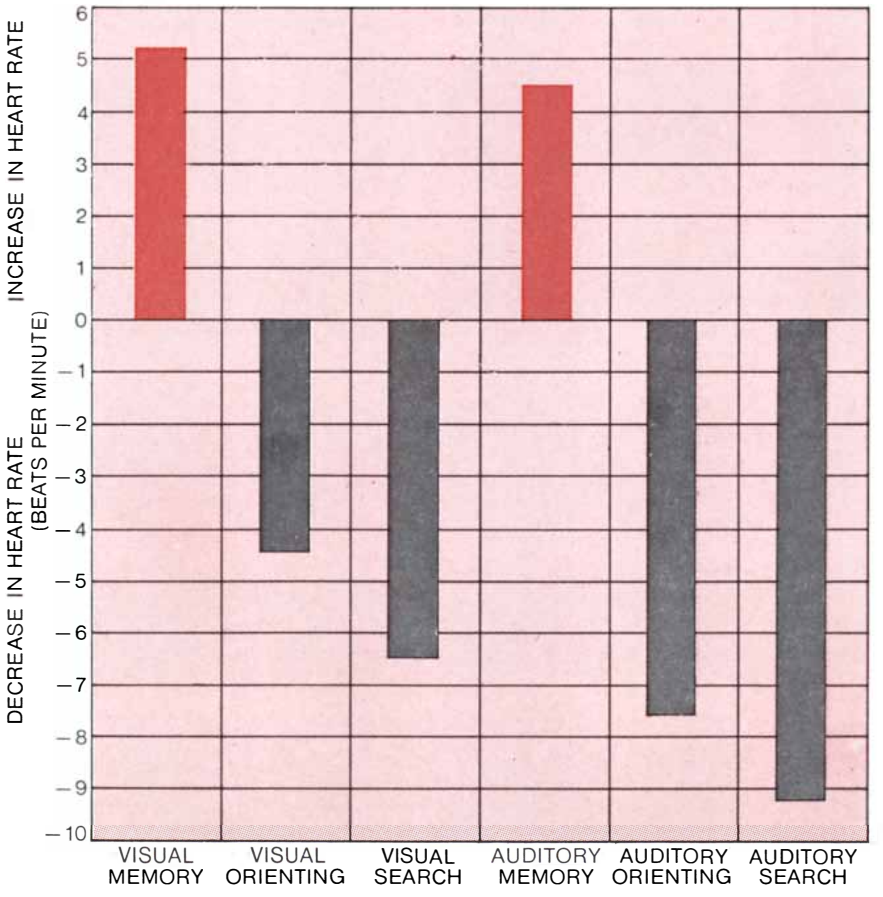
Milton Kotelchuck of Harvard recorded on videotape the responses of children between the ages of six and 21 months to the departure of their parents in the presence of a female stranger. The observations were made in a living-room setting that included toys appropriate to the age of the child. During the first three minutes the child was with both his father and his mother, who sat near the wall while the child played. Then one parent left the room. After three minutes a female stranger entered and three minutes later the second parent departed, leaving the child alone with the stranger.

The episodes were repeated several times. In every case when the stranger entered the room, the infant looked more often at the stranger than at either of his parents, indicating that even the youngest child could discriminate the stranger from the parents [*see illustrations on preceding page*]. The presence of the stranger when neither parent was present produced little crying or other signs of fear in infants either six or nine months old. Among infants either 12 or 15 months old the presence of the stranger alone caused the most crying and the most marked inhibition of play. At 18 and 21 months the infants cried less and showed less inhibition of play in the presence of the stranger.

The most marked inhibition of play, a sensitive sign of fear, occurred when the child was left alone with the stranger. We suggest that these behavioral signs of fear were the result of the child's inability to understand where his parents were or whether or not they would return. The hypotheses he generated to understand this discrepant event were insufficient, and faced with an unresolved discrepancy the child became frightened.

These data are relevant to the earlier assumption that the process of activating hypotheses does not emerge until the end of the first year, since we have to ask why Kotelchuck did not find much separation fear in the six-month-old infants. We know that four- and five-month-old infants will cry and show signs of distress to discrepant events they cannot assimilate. We can partly resolve the paradox if we analyze the nature of the discrepancy. The fear displayed by the one-year-old infant following the departure of his parents may be due to the activation of hypotheses that cannot be resolved involving the immediate future: What will happen to me? Will my parents return? What will the stranger do? If the child has no answer to these troublesome hypotheses, he may become afraid, stop playing and begin to cry. The cause of his distress is uncertainty generated by unanswered questions about the future rather than direct encounter with an unusual, external event.

In cultures where a mother carries her young child for most of the day (as she does in parts of Uganda and Central America) a four-month-old infant will cry when he is removed from his mother. This crying is unlikely to be mediated by the same class of discrepancy that elicits crying in a one-year-old infant when he is left alone in a room. The fear at four months is a result of the discrepant kinesthetic and tactile stimuli that follow separation from the mother's body and not the activation of hypotheses as to her whereabouts or the probability of her returning. Therefore crying following separation from the parent may be mediated at different ages by qualitatively different sources of discrepancy.



HEART RATE AND MENTAL ACTIVITY in children six to 10 years of age are related in this graph. Mental activities such as memorizing or calculating increase the heart rate. Paying attention to an event, either by looking or by listening, decreases the heart rate.

The strongest support for the notion that a one-year-old is capable of generating hypotheses about his experiences comes from studies of heart-rate changes in response to discrepant events. John and Beatrice Lacey of the Fels Research Institute have found that when an adult looks at or listens to an interesting stimulus, his heart rate decreases; when he is actively thinking—either memorizing verbal material or performing arithmetical calculations—his heart rate increases. Recently Kathleen van Hover of the National Institute of Mental Health has extended the validity of this finding for children from six to 10 years old. The children showed an acceleration of heart rate in response to memory problems and a deceleration in heart rate to orienting and search situations for both visual and auditory tasks [*see illustration at left*].

The relation between active mental work and heart-rate acceleration observed in children and adults may hold for infants as well. Since generating hypotheses is a form of mental work, there should be a rise in the heart rate of infants as they become maturationally able to activate these structures. This prediction was affirmed in a study of infants ranging in age from 5½ months to 11½

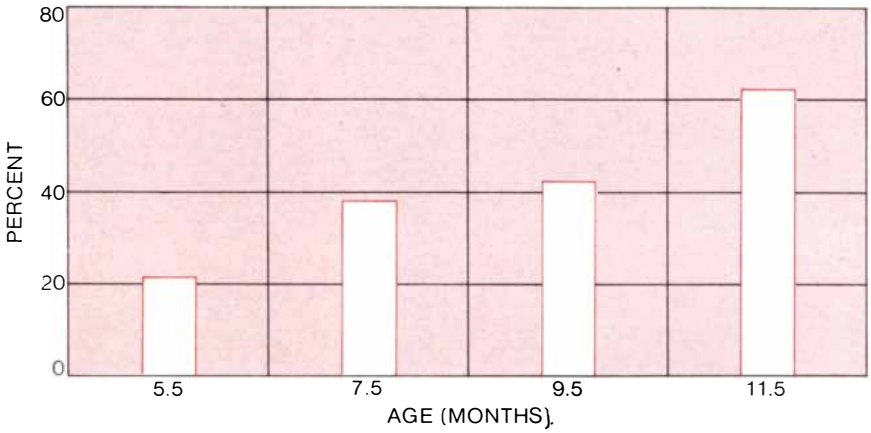
months. The infants heard from eight to 12 repetitions of a meaningful phrase of speech followed by a discrepant speech phrase. The older the infant, the more likely his heart rate was to increase in response to the discrepant stimulus, suggesting active mental work in the form of activation of hypotheses [*see top illustration on this page*]. In another study infants of the same age range heard repetitions of a nonsense phrase followed by a different nonsense phrase. Again the discrepant event produced the greatest rise in heart rate in the older infants. Moreover, there was an increasing tendency with age for infants to vocalize more after the discrepant speech phrase had ended than during its presentation [*see bottom illustration on this page*]. This behavior suggests that the older infants were working, in a cognitive sense, on the speech that had just terminated and their vocalization during this period was a reflection of the excitement that accompanied the work.

When the initial series of phrases is spoken by a male voice and the discrepant series is spoken by a female voice, older infants did not show an acceleration of heart rate. Similarly, infants do not show a rise in heart rate if the series is an ascending scale played on a cello and the discrepant series is a random arrangement of the same notes. This may mean that the infant does not perceive these events as discrepant transformations of the standard and thus does not attempt to generate hypotheses to assimilate the new stimulus.

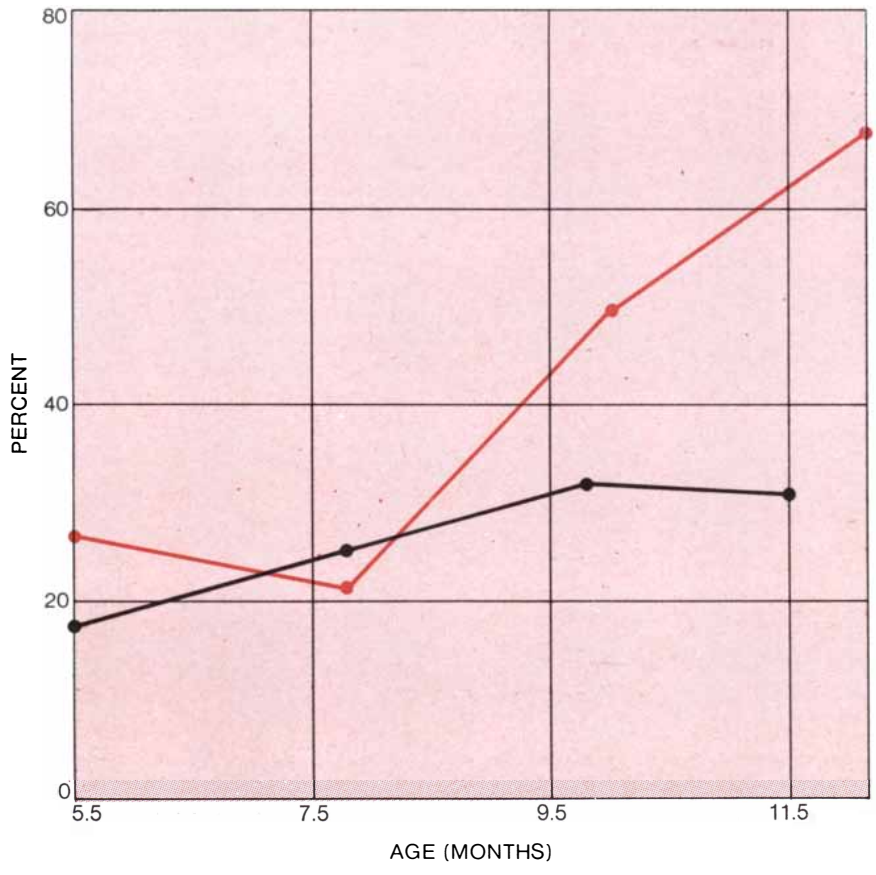
Philip Zelazo of Harvard recorded eye fixation and changes in heart rate of 11½-month-old infants as they watched a toy car roll down an incline and knock over a plastic object. After a few trials some infants began to anticipate the car's motion and looked toward the object. In 73 percent of the anticipatory fixations Zelazo found an accompanying increase in heart rate. There was a decrease in heart rate in only 14 percent of the cases. If the increased heart rate means the same thing in infants as it does in older children, namely an increase in mental activity, it is reasonable to suggest that the child begins thinking about unusual events toward the end of his first year of life.

The suggestion that eight to nine months marks the maturational frontier when infants begin to generate hypotheses is supported by the fact that intra-individual continuity of behavior changes toward the end of the first year. We recently completed a long-term study of 180 firstborn infants whose reactions to a set of visual and auditory stimuli were assessed at four, eight, 13 and 27 months of age. We found virtually no relation between attentiveness or vocalization at four or eight months and attentiveness or speech in the same child at 27 months. We found moderate stability of these reactions, however, when we compared the kinds of behavior displayed at 13 and 27 months. These data suggest that the relation of attentional phenomena to mental structure changes toward the end of the first year.

Both theoretical and practical implications flow from these findings. It ap-



HEART-RATE INCREASE of infants who have just heard a discrepant sound is linked to age. Here the percentage of infants in each age group who exhibited cardiac acceleration is shown. A larger proportion of the older infants showed a rise in heart rate in response to the discrepant event. This could be a sign of greater mental activity in older infants.



INCREASE IN VOCALIZATION following a discrepant event also is linked to the age of the infant. The percentage of infants in each age group who vocalized during the stimulus (*solid curve*) did not vary much. A large proportion of older infants vocalized after the stimulus had ended (*colored curve*). The poststimulus vocalization could be an indication that the older infants were engaged in some mental activity related to the discrepant event.

pears that before an infant is a year old he has become a thinking creature who activates cognitive structures to resolve discrepancies and solve problems. Over the past 20 years many students of child development have held a different view. For example, the eminent Swiss psychologist Jean Piaget has argued that during the first 18 months of life an infant knows the world only in terms of his sensory impressions and motor activities. Cognitive development, says Piaget, begins after the sensorimotor period ends. These results provide a mild challenge to his view. The infant may be more thoughtful than most psychologists have surmised.

A second implication takes the form of a suggestion to developmental physiologists and anatomists to search for important changes in the function and structure of the central nervous system between eight and 12 months of age. The infant becomes consistently reactive to discrepancy between two and three months, and the emergence of this competence is accompanied by a cluster of biological phenomena (including a mature form of the visual evoked potential and consistent occurrence of sleep spindles in the electroencephalogram). Nine months seems to be a frontier that announces the child's ability to generate simple hypotheses, but the existing literature does not mention a comparable set of biological changes in the central nervous system. We suggest that this new competence is dependent on correlated changes in the brain. A new cluster of biological phenomena should appear at that time. It would seem appropriate for neurophysiologists and developmental anatomists to look for changes in the structure and function of the central nervous system of infants between eight months and one year old.

The final suggestion is more practical. Up to six months of age the differences among infants in motor and cognitive development are fairly independent of the child's social class, his ethnic origin and even some aspects of his rearing conditions. By one year of age, however, differences in rearing experience seriously affect cognitive functioning. This means that pediatricians and psychologists might assess the reactions of one-year-old infants to discrepant stimuli to obtain information about their developmental status, particularly the status of infants who may have been exposed to psychic risks or who suffer from metabolic or structural defects.

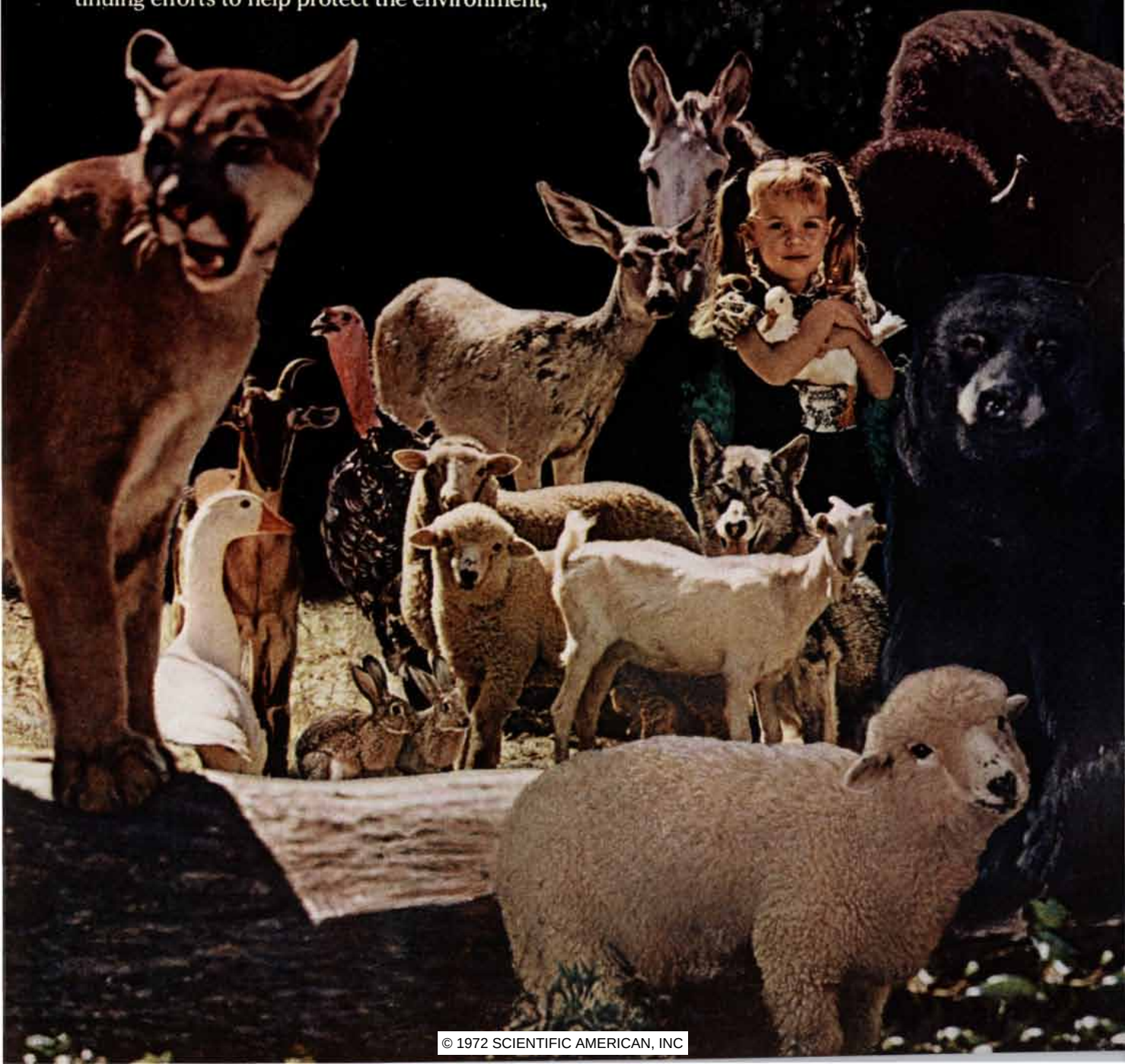
Texaco announces some good news for all living things on this earth.

If man continues to let pollution go unchecked, all of us, even creatures in the most faraway forests, will one day feel its effects. So, for all living things on this earth, Texaco would like to announce some good news concerning our environment.

At Texaco, we've created a new department for environmental protection. It's made up of Texaco people who are dedicated to finding ways to protect the earth, the air, and the water. This department will guide the Company's continuing efforts to help protect the environment, while Texaco provides the petroleum energy people need. We're all aware of the problems of pollution. We hope teaming up all our experience will help us find practical solutions.

TEXACO

We're working to keep your trust.



The Sources of Muscular Energy

The immediate source is adenosine triphosphate (ATP). The ultimate sources are the combustion of food and the breakdown of glycogen. The time relations of these processes offer some practical hints

by Rodolfo Margaria

A muscle can be regarded as an engine, and like any engine it obtains its energy essentially from the "burning" of fuel. Its performance, or capacity for work, depends on the nature and availability of the source of energy. In the case of an artificial engine the energy input is easily identified and measured: it is simply the rate of consumption of the supplied fuel. The muscle engine, however, is much more complex. It uses several different fuels, and it regenerates some of them itself. As a result the evaluation of the factors involved in muscle performance is far from a simple matter. It requires detailed, quantitative analysis of each of the sources of energy and their relative timing and collaboration. Through systematic investigations over the past few years a reasonably comprehensive picture of the operation of the energy sources in muscular exercise has been obtained, and it has become possible to suggest rational regimes for employing muscles most efficiently. The new knowledge should be widely applicable in athletics and in physical work, thus improving the productive capacity, comfort and health of people engaged in such activities.

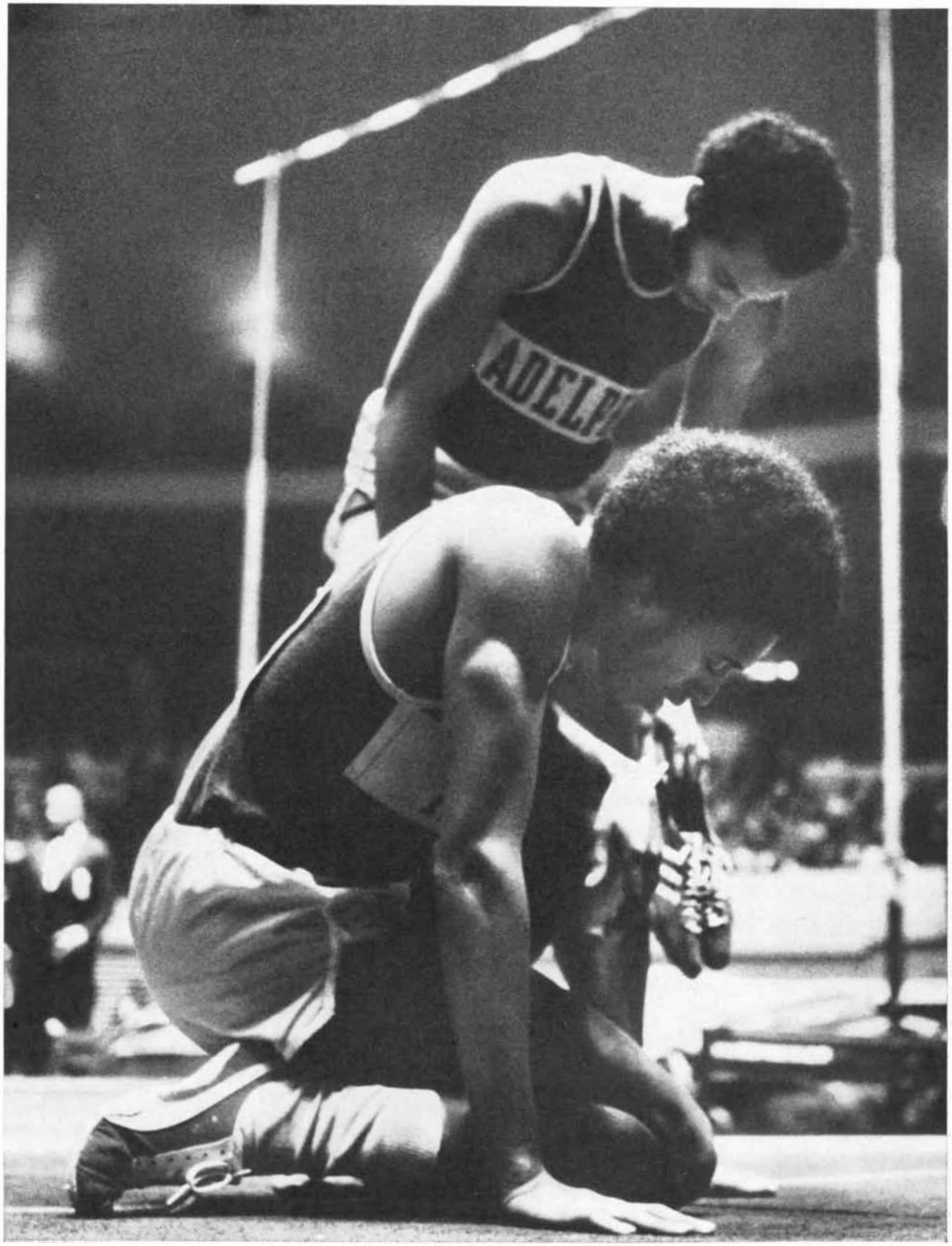
The direct source of energy for a muscle's activity is adenosine triphosphate (ATP); the release of energy from the splitting of ATP into adenosine diphosphate (ADP) and phosphoric acid is what powers muscle contractions. The ATP must be synthesized continuously, as there is no appreciable store of it in the muscle. It is actually resynthesized from its products as soon as it is broken down. The energy needed for the recombination of ADP and phosphoric acid into ATP is supplied by another energy-yielding reaction in the cells: the splitting of creatine phosphate. This auxiliary "phosphagen" is likewise in short supply in the muscle and needs to be resynthesized continuously. There are, in turn, two ultimate sources of energy for the resynthesis of the phosphagens: (1) combustion of food, measured by the consumption of oxygen, and (2) glycolysis, the breakdown of glycogen resulting in the formation of lactic acid. The second of these processes is reversible: with an input of energy from food combustion, lactic acid is reconstituted to glycogen.

The system, then, consists of five reactions, three of which (phosphagen-splitting, food combustion and glycolysis) yield energy and two of which (phosphagen and glycogen resynthesis) absorb energy [*see top illustration on page 86*]. We need to obtain a measure of each of these five quantities in order to compute the net total of energy and power (energy per unit of time) available to the muscle for given levels of performance.

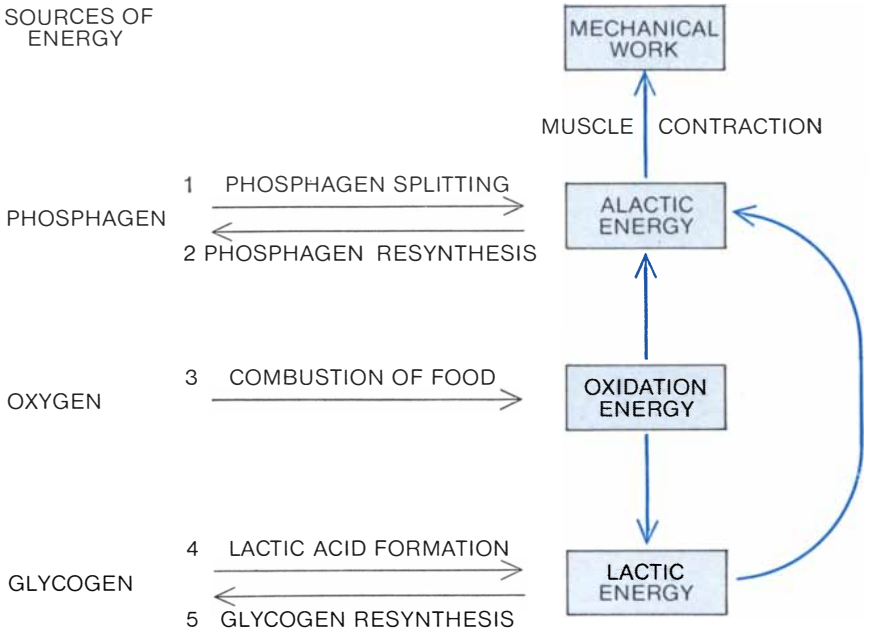
How can we disentangle the factors for separate measurement? Fortunately we can set up experimental conditions in which some of the factors can be disregarded. For example, it is known that during moderate exercise (that is, below a certain level of exertion) the muscles do not produce lactic acid; therefore glycolysis and the reverse do not enter into the equation. Furthermore, if the exercise is maintained at a constant level of energy consumption, the splitting and resynthesis of the phosphagens soon balance out in energy terms, so that we can disregard those quantities also. In exercise under these conditions the net expenditure of energy can be calculated simply from the consumption of oxygen. It has been known for some time that the energy yield from food combustion in the muscles is five calories per milliliter of oxygen consumed. Hence it is easy to calculate that during exercise at a constant, submaximal level the energy employed for the given work load is a number of calories amounting to five times the number of milliliters of oxygen consumed. This formula has been known and employed in physiological studies for many years. In such studies the work load is commonly measured on an ergometer such as a treadmill or a bicycle.

When exercise is raised to a strenuous level at which energy can no longer be provided in sufficient quantity through oxidation (because of the limit on the rate of delivery of oxygenated blood to the tissues), the muscles begin to supplement the energy supply by means of glycolysis. The quantitative details of this process are difficult to get at and have only recently been determined.

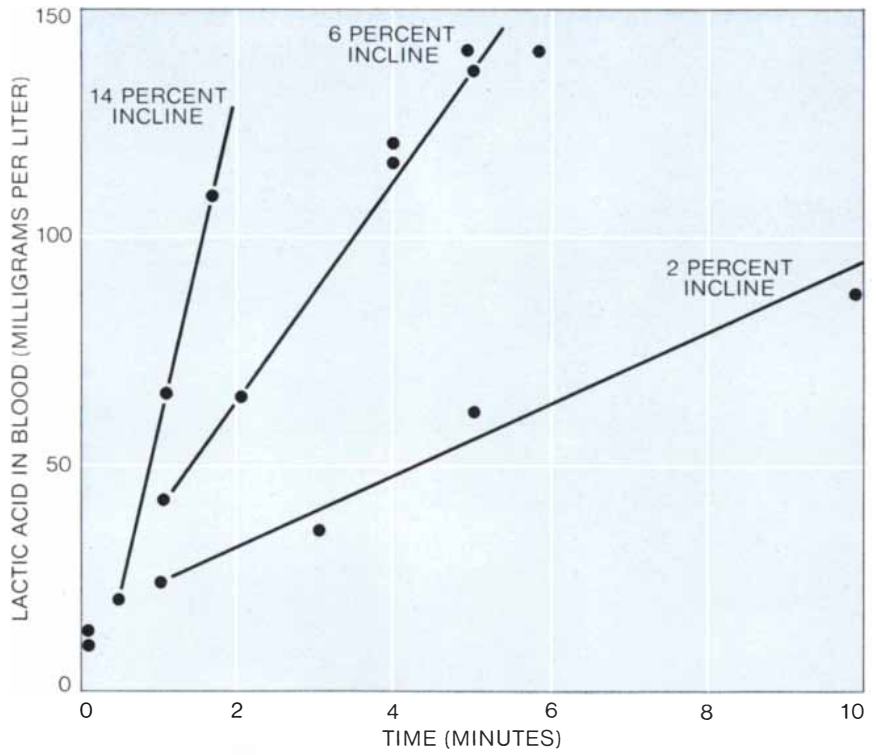
Let us consider a maximal effort such as a runner's sprint of 100 or 400 meters at top speed. Within a few seconds his muscles' energy requirement has established an equilibrium between the splitting and the resynthesis of phosphagen and also has passed the limit of production of energy by oxidation. The only variable, then, is the amount of energy generated by glycolysis. Can this be estimated from measurement of the amount of lactic acid (the breakdown product) in the blood? If glycogen were resynthesized from lactic acid about as rapidly as it is broken down, such a measurement would be meaningless. Actually, however, the resynthesis is very slow; it takes about 15 minutes, by exponential increase, to reach half its maximal rate. Therefore in a test of strenuous exercise lasting only a few seconds or a few minutes the amount of resynthesis is so small it can be disregarded and the concentration of lactic acid in the blood can be



AFTER A HARD RUN an athlete may be near exhaustion because his muscles have depleted their supply of ATP, and oxidation and glycolysis have not yet been able to replenish the muscles with their primary source of energy. Here the kneeling runner, comforted by one of his Adelphi University teammates, has just finished running the fourth half-mile leg of a two-mile relay at the Wanamaker Millrose Games, which were held at Madison Square Garden in New York on January 28. His team finished fourth.



RELATIONS OF SOURCES of muscular energy are outlined. "Phosphagen," the direct source of energy, is a general term for the high-energy phosphates such as ATP and creatine phosphate (CP) that are found in cells. Energy is released by the splitting of phosphagen molecules. The split phosphagen is almost immediately resynthesized. Energy for the recombination is supplied by the two secondary energy sources: the combustion of food and the breakdown of glycogen, a carbohydrate stored in the muscle, into lactic acid. Glycogen in turn is resynthesized from lactic acid with energy from oxidation. The overall cycle consists of five reactions, three of which release energy and two of which absorb it.



HEAVY EXERCISE leads to the formation of lactic acid. For a given work load the concentration of lactic acid in the blood increases linearly with the duration of the exercise. The exercise consisted in running to exhaustion at a constant speed of 12 kilometers per hour on a treadmill tilted at inclines of 2, 6 or 14 percent. At these work loads the rate of increase of lactic acid, indicated by the slope of the curves on the graph, is proportional to the intensity of the exercise. Blood samples were taken a few minutes after the exercise to allow the lactic acid from the muscles to become uniformly distributed throughout the body.

taken as a good indication of the energy contribution of glycolysis beyond what is supplied by oxidation. For a given subject the production of lactic acid per minute should vary directly with the magnitude of the energy requirement or with the length of time a given requirement is imposed on the muscles.

Of course, the subject's blood cannot be sampled while he is running. Furthermore, a sample taken during the exercise would be misleading, because it takes time for the lactic acid to diffuse from the muscles into the body fluids. As a result the drawing of the blood samples can conveniently be postponed until two or three minutes after the run, and the lactic acid concentration in the blood at that time can be considered a reasonably accurate representation of the total quantity of lactic acid formed as a result of the exercise.

Experiments have proved these assumptions to be correct. Subjects were tested in runs on a treadmill, tilted at different inclinations to call for different intensities of exercise, and the exercise levels were strenuous enough to lead to exhaustion in one minute to 10 minutes. Measurement of the lactic acid concentration in the blood after the run indicated that the amount of glycolysis for a given work load did indeed increase linearly with time, as the premises predicted [*see bottom illustration at left*]. On the basis of assumptions about the diffusion of lactic acid to the various tissues and organs of the body it was possible to calculate how much lactic acid the muscles produced in relation to the body's total weight. This calculation likewise showed that the amount of lactic acid produced per minute per kilogram of body weight varied in direct proportion to the energy requirement [*see illustration on opposite page*].

The measurements established several interesting facts. They confirmed that in steady exercise the body meets its energy needs exclusively by means of oxidation up to a certain level of requirement for energy, and they showed that ordinarily the maximum provided by oxidation is about 220 calories per minute per kilogram of body weight. The production of lactic acid usually began when the energy requirement passed that level. Furthermore, the amount of lactic acid produced in relation to the energy need indicated that the energy yield from glycolysis is about 230 calories per gram of lactic acid produced. This determination of the energy yield

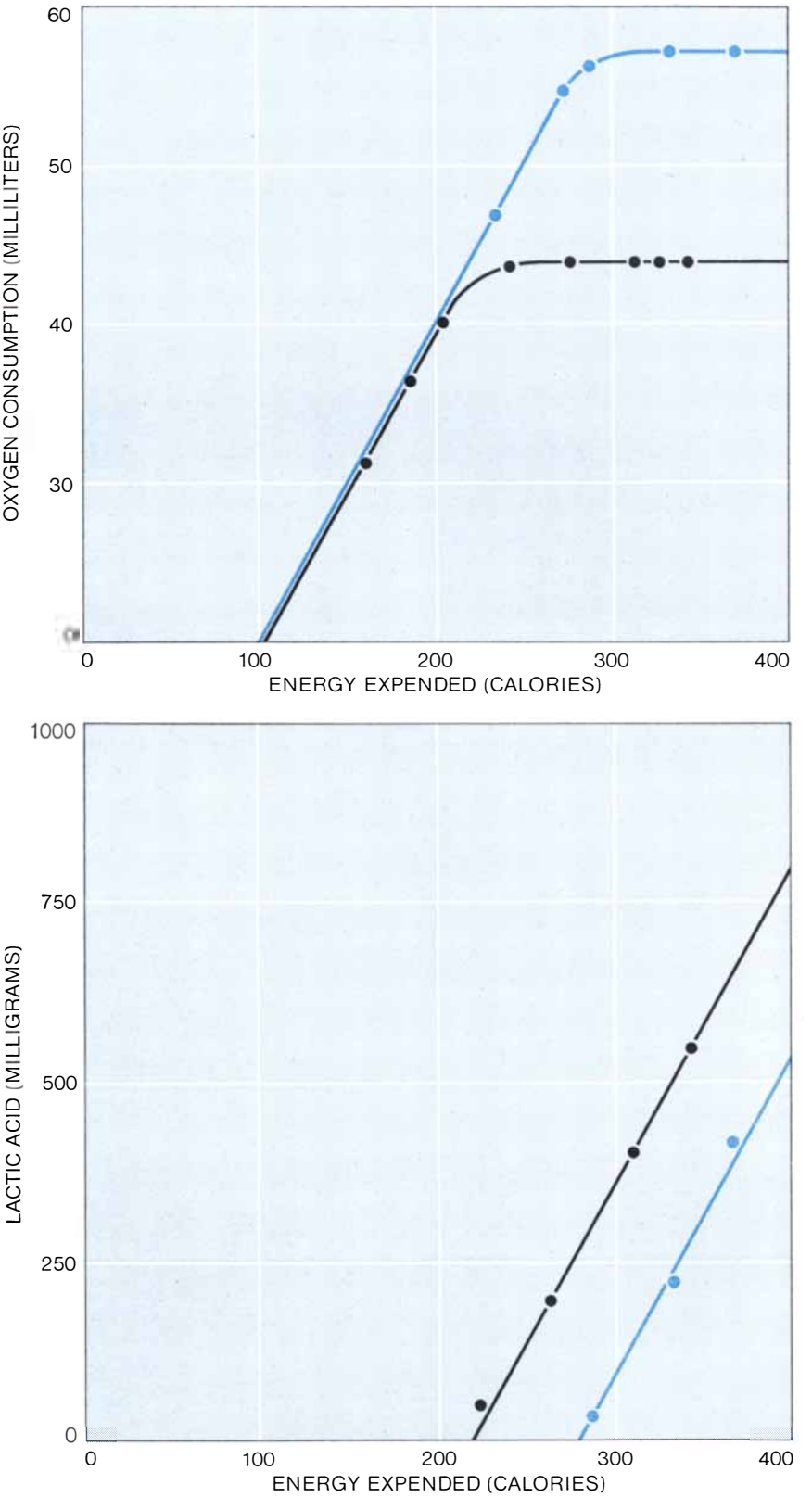
could be considered a close approximation of the true value, since it was obtained in vivo with the subjects showing an entirely normal physiological state in terms of their body temperature and the physical characteristics of their body fluids. Moreover, the energy yield was the same in trained athletes as it was in other persons.

Another significant finding is that trained athletes (middle-distance runners) were able to obtain an unusually large share of their energy need from oxygen consumption, so that lactic acid production began at a somewhat higher level of exertion than it does in nonathletes. The same effect was obtained by nonathletes by their breathing oxygen during exercise. Breathing oxygen before the performance, however, apparently is not helpful, and there is no physiological basis for believing it should be (although it is sometimes resorted to by athletes). Oxygen respiration during a performance is useful because it provides an extra current supply for the continuing need for chemical energy; by the same token, air with a subnormal content of oxygen, such as the athletes had to contend with at the high altitude of Mexico City in the Olympic Games of 1968, understandably must reduce performance.

As the trained athletes' performances in our tests indicated, people differ in their capacity for performing work without resorting to glycolysis. When we tested subjects with different capacities in this respect on the same work load, we found that individuals differ in the rate of lactic acid production. A person with a low capacity for oxygen consumption produces considerably more lactic acid.

What, then, is the total contribution glycolysis can make to the muscle machine? Measurements of the lactic acid concentration in the blood after strenuous exercise have shown that the maximum increase in this content is about 1.5 grams per liter, or about 1.12 grams per kilogram of body weight. Putting the yield per gram at about 230 calories, we can compute that the maximum amount of energy obtainable from the formation of lactic acid is about 260 calories per kilogram of body weight.

Obviously the rate at which lactic acid can be produced must have an upper limit, just as there is a limit on the rate of oxygen consumption by the tissues. In the case of lactic acid the limit is imposed by the rate of the chemical processes involved in its production. We find that increasing the intensity of ex-



ENERGY CONTRIBUTIONS of oxygen (*upper graph*) and a glycolysis (*lower graph*) during exercise are plotted. In moderate exercise the oxidation of food, as measured by increased oxygen consumption (in milliliters per minute per kilogram of body weight), provides virtually all of the energy requirement (shown on the abscissa in calories per minute per kilogram of body weight). When the energy requirement surpasses the upper limit of oxygen consumption, the additional energy is provided by formation of lactic acid (in milligrams per minute per kilogram of body weight) from glycogen. Athletes (*colored curves*) have a higher oxygen-consumption capacity, which delays the formation of lactic acid.

ercise raises the rate of lactic acid production only to a certain level; beyond that the speed of its production cannot be pushed further. The maximal rate is about 1.7 grams per kilogram of body weight per minute, which corresponds to a power output of 390 calories per kilogram per minute. Thus the power available from glycolysis turns out to be about 50 percent higher than the power from oxidation (220 calories per kilogram per minute).

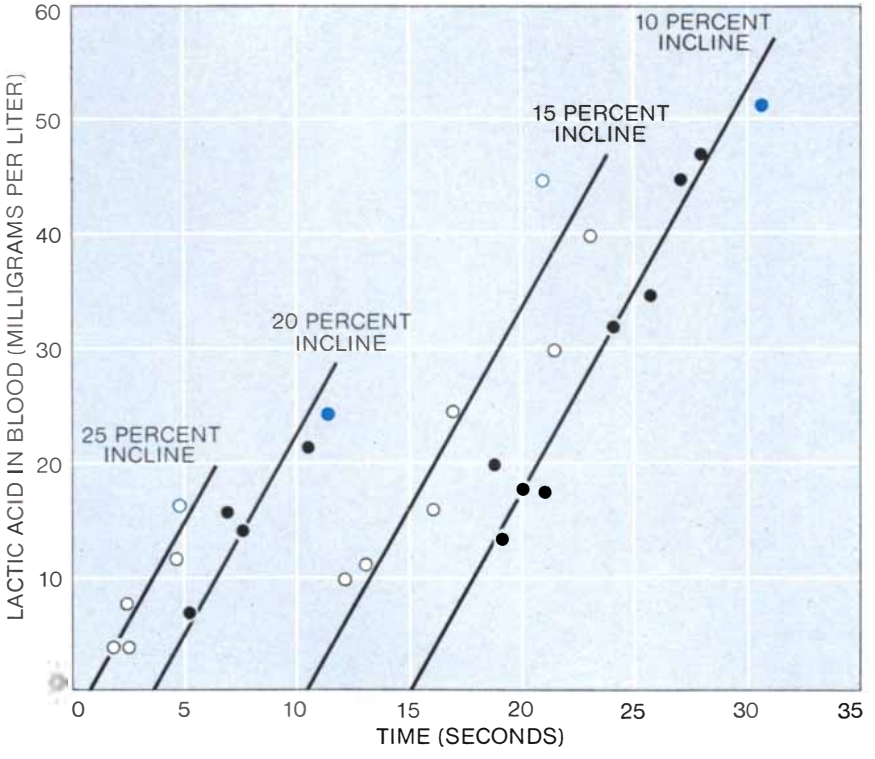
Consider now the physiological debts the body contracts in strenuous exercise that leads quickly to exhaustion, say within 35 seconds, when lactic acid production has reached its maximal rate. We know that in exercise of such intensity there is for the first 15 seconds or so no energy contribution from glycolysis. The oxidative mechanism also is rather sluggish and in the first 15 seconds it contributes only a small portion of the energy, although its contribution rises at an exponential rate. On the basis of these known facts we can calculate the energy contributions made in exhausting exercise by each of the three energy-yielding mechanisms: oxygen consumption, lactic acid production and phosphagen-splitting, the last called the "alactic" mechanism [*see illustration on opposite page*].

The calculation on this basis indicates that theoretically the maximal contribution of phosphagen-splitting itself (the alactic source) would be about 200 calories per kilogram of body weight, which would amount to the splitting of almost all the available phosphagen. We find in experiments that the amount of split phosphagen at the end of a maximal aerobic exercise (exercise that can be supported by energy from oxidations only) is only about half this theoretical quantity; the amount corresponds to an energy yield of about 100 calories per kilogram of body weight instead of 200 calories. At the end of supramaximal exercise that involves energy from anaerobic sources and leads to exhaustion in a short time, the energy yield from phosphagen breakdown may go up to 150 calories per kilogram of body weight; this has been found recently by Pietro di Prampero and others in my laboratory in measurements of the alactic oxygen debt. Under no conditions, however, did they find that the energy yield from phosphagen reaches 200 calories per kilogram of body weight, which would be the case if all the phosphagen in the muscle had been broken down.

In spite of this experimental finding in strenuous supramaximal exercise to exhaustion, I believe practically all the available phosphagen has actually been split. The reason the experiments show otherwise probably lies in the fact that the theoretical estimate of the energy balance overstates the amount of contribution by glycolysis during the exercise. It overlooks the fact that some of the lactic acid is produced afterward during the early period of the muscles' recovery. This production of lactic acid represents an energy debt that is used to pay a corresponding amount of the alactic debt. In other words, a fraction of the alactic debt is shifted into a lactacid debt. That such a process of anaerobic recovery (contributing energy for phosphagen resynthesis by a means other than oxidation) takes place after exercise was observed many years ago in studies of the isolated frog muscle by the author and Gianni Moruzzi and more recently in our laboratory at the University of Milan by Paolo Cerretelli. It is estimated now that during recovery from very strenuous exercise the delayed production of lactic acid may contribute about 50 calories per kilogram of body weight to the resynthesis of phosphagen, which would explain why Prampero did not find complete breakdown of phosphagen at exhaustion.

In short, the "oxygen debt" that the British physiologist A. V. Hill first described in the 1920's has two aspects: oxygen is needed for the dual functions of (1) reconstituting glycogen from lactic acid (as Hill observed) and (2) providing energy directly on its own account for the resynthesis of phosphagen (as H. T. Edwards, D. B. Dill and I found later in studies at the Harvard Fatigue Laboratory). We call these aspects the lactacid oxygen debt and the alactic oxygen debt.

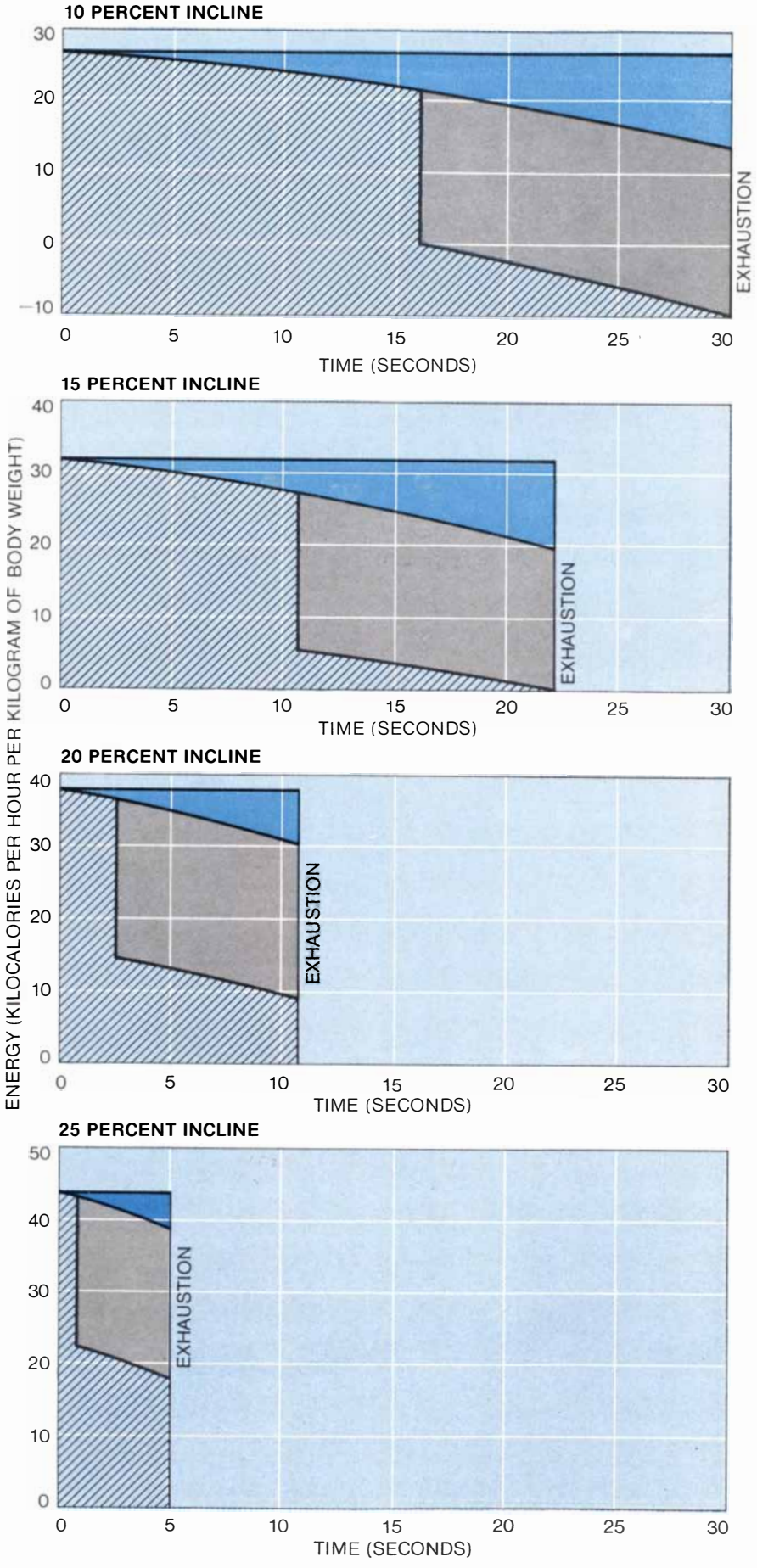
From the standpoint of provision of energy we can distinguish three phases in the operation of the muscle engine in strenuous exercise. During the first phase, lasting only a few seconds, all the energy is provided solely by the splitting of phosphagen, as the oxidative and glycolytic reactions have not yet got under way. How much power can be provided by this alactic mechanism? We can obtain a measure of the maxi-



SUPEREXERTION pushes the rate of lactic acid formation to its upper limit, after which increasing the intensity of the exercise no longer affects the rate, as is indicated by the parallel slope of the curves in the graph. The work load consisted in running to exhaustion at a constant speed of 18 kilometers per hour on a treadmill tilted at inclines of 10, 15, 20 or 25 percent. The point of exhaustion (*colored dots and colored circles*) is reached more rapidly as the exercise becomes more strenuous. Although rate of lactic acid production does not change, the more strenuous the exercise, the more quickly formation of lactic acid begins.

mum output by using an exercise in which the subject climbs stairs two steps at a time at top speed. The maximum speed is reached within three seconds and is maintained for one or two seconds after that. Since the work consists almost entirely in lifting the body, it can be calculated as the kilogram-meters of rise per second per kilogram of body weight. The maximum efficiency in the use of muscle energy by a man running uphill has been found to be about 25 percent. Dividing the work performed in the tests by this figure, we calculate that the maximum of muscular power obtainable from the splitting of phosphagen is about 800 calories per minute per kilogram of body weight. (Of course, this test tells us only about the power in the muscles of the lower limbs; work involving mainly the arms might give a different figure.)

The second phase in the energy economy, chronologically speaking, is the arrival of oxidation as a source of energy for resynthesizing phosphagen. The maximum power available from this contribution too can conveniently be measured experimentally. The rate of the heartbeat gives a sufficiently accurate measure of the rate of oxygen consumption by the body, and it has been found that there is a maximum for the heart rate during exercise, depending on the person's age. A convenient and well-accepted exercise for this test is stepping up on and down from a stool of a given height at a given frequency dictated by a metronome. The subject's heart rate is measured after three or four minutes, when the rate has reached a constant level. The test shows that this rate in an average young man indicates a maximum oxygen consumption of about 40 to 45 milliliters per minute per kilogram of body weight, providing a power contribution of about 220 calories per minute



CONTRIBUTIONS of the three energy sources during exercises calling for superexertion are plotted in the four graphs at the right. The exercises consisted in running to exhaustion at 18 kilometers per hour on a treadmill inclined at 10, 15, 20 or 25 percent. Alactic energy (*hatched areas*), released by the splitting of phosphagen, is both the initial and the major source of energy during superexertion. Energy from the oxidation of food (*colored area*) increases exponentially from the onset of strenuous exercise, but the mechanism is sluggish and provides only a small portion of the required energy during the first few seconds. The remaining energy is provided by breakdown of glycogen into lactic acid (*gray area*).

per kilogram, as we have already noted.

The third energy contribution for exercise, which enters only after oxidation is no longer able to keep up with the muscles' needs, is glycolysis. I have already mentioned that the maximum quantity of lactic acid produced in strenuous exercise is about 1.12 grams per kilogram of body weight, corresponding to an energy production of about 260 calories per kilogram. In recent studies Jan Karlsson and Bengt Saltin of the Gymnastic School in Stockholm found that on the average human muscle contains about 10 grams of glycogen per kilogram of muscle tissue, which suggests that only about a third of the glycogen in the muscles is broken down to lactic acid to furnish energy for work. These general characteristics are not very significant for judging an individual's capacity for exercise, however, because people vary greatly in the glycogen content of their muscles. The glycogen content depends on the state of the individual's nutrition; obviously, then, an athlete or anyone engaging in strenuous exercise should pay careful attention to nutrition.

The measurements I have described give us a balance sheet showing the capacity and power of each of the three contributing mechanisms and some information about the contraction and repayment of the alactic and lactacid oxygen debts. We are interested, however, not just in the overall balance sheet but, more important, in the timing of the respective processes for supplying energy. From the various experiments and analyses we can now sketch a rough, well-confirmed picture of the energy events that take place during exercise and the following recovery period.

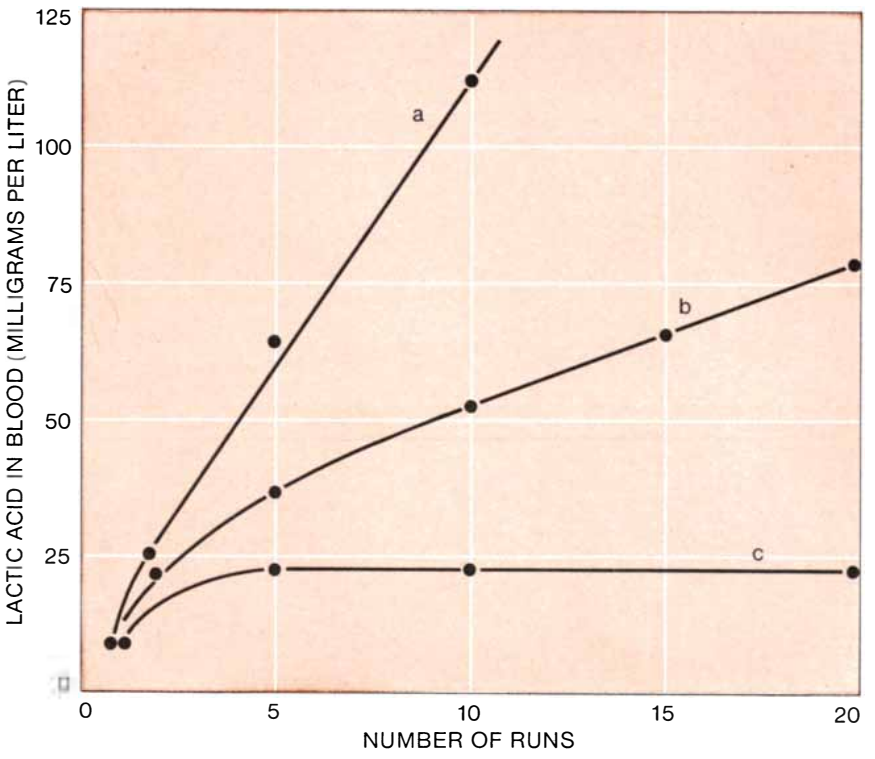
In exercise that calls for maximal consumption of oxygen but no glycolysis, the oxygen consumption rises exponentially to its limit (40 milliliters per minute per kilogram of body weight) and then, at the end of the exercise, falls back exponentially to the rate of a resting muscle. The oxygen debt contracted is entirely alactic; it amounts to about 20 milliliters and is quickly paid off. Only half of the phosphagen content of the muscle is split during the peak period of activity, and all of this is resynthesized by means of energy from oxidation in the brief recovery period [*see top illustration on opposite page*].

The picture in exercise that requires the additional input of energy from glycolysis is considerably more complex and has only recently been traced out. When the energy requirement is twice as great as what can be furnished by oxidation (that is, when the requirement is equivalent to an oxygen consumption of 80 milliliters per minute per kilogram of body weight), oxygen consumption rises to its maximum rate in half the time, and the formation of lactic acid, apparently triggered by the splitting of 50 percent of the muscle phosphagen, begins at that point. If the exercise proceeds to exhaustion, presumably the remaining 50 percent of the muscle phosphagen is split. The alactic oxygen debt at the onset of the recovery period after exhaustion amounts to about 40 milliliters per kilogram of body weight, the lactacid oxygen debt depends on the amount of lactic acid built up during the exercise. In the recovery period of the alactic energy debt 30 milliliters are paid by energy from oxidation and about 10 milliliters by energy from the delayed production of lactic acid through glycolysis [*see bottom illustration on opposite page*].

What does all of this tell us about how to make efficient use of the muscle engine? Obviously it is advisable to avoid, if possible, driving the muscle to the pitch of incurring a lactacid debt. The payment of that debt is a very slow process, taking more than an hour after the exercise has been concluded, and the lactic acid in the body induces a state of acidosis that hampers muscle performance, causes great discomfort and produces other disagreeable symptoms. On the other hand, if the oxygen debt is only alactic, there is no acidosis and the debt is paid off in a few minutes.

Because there is always a certain period of delay in the onset of lactic acid production, even in highly strenuous exercise, one can avoid this production by limiting the activity period to a short enough time. This suggests that, in the case of exercise or work that may push the energy requirements beyond what can be furnished by oxidation, the muscles can be used most effectively by adopting a schedule of intermittent activity and rest, each activity period being short enough to forestall lactic acid production and each rest period long enough to pay off the alactic oxygen debt contracted during the period of strenuous activity.

We have conducted experiments that



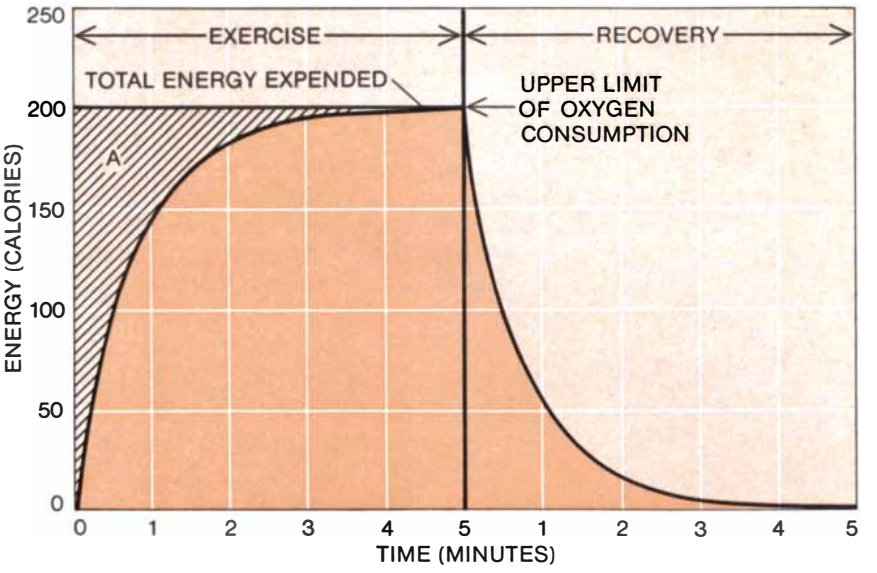
REST INTERVALS between sessions of strenuous running can greatly reduce the production of lactic acid and increase the total distance that can be run. The exercise consisted in running on a treadmill at 18 kilometers per hour for 10 seconds followed by a rest period. When the rest period was 10 seconds (*a*), the runner could complete 10 cycles before exhaustion. When the rest period was 20 seconds (*b*), exhaustion occurred after 20 cycles. With a rest period of 30 seconds between runs (*c*) the lactic acid level remained constant and the exercise could be continued indefinitely. Athletes in training can increase the amount of strenuous exercise they can perform by judicious timing of rest periods.

strikingly demonstrate the value of this tactic. The test consisted in running on a treadmill at a speed and incline that would lead to exhaustion in about 35 seconds if it were continued without rest. We limited the run to 10 seconds at a time and varied the experiment by trying rest periods of different lengths between the runs. The object of the test was to find out how many 10-second runs could be accomplished (in effect, how much distance could be covered in total) before exhaustion.

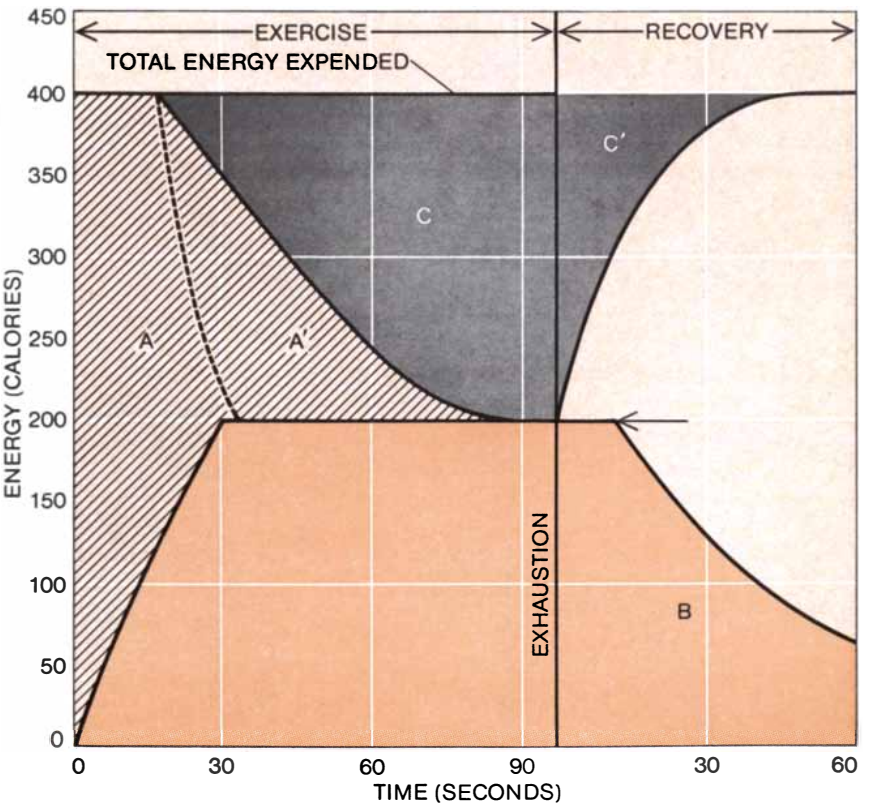
When the rest period was 10 seconds, the subjects could run about 10 cycles (totaling 100 seconds of exercise) and cover a distance corresponding to 500 meters. The lactic acid in the blood at the end of the performance amounted to 1.15 grams per liter. When the rest period was increased to 20 seconds, the subjects accumulated less lactic acid and covered a considerably greater distance. With a rest period of 30 seconds between runs the subjects were able to go on with runs indefinitely and showed a lactic acid content in the blood of only about .2 gram per liter; this content could be attributed to the fact that leg muscles usually generate a little lactic acid at the beginning of exercise because of sluggishness of the oxidative mechanism [*see illustration on opposite page*].

These findings can be projected to predict potential performances in sports events on a track. For example, a trained runner sprinting 400 meters at top speed will finish the race with a state of acidosis that will require a rest of at least an hour and a half for recovery; thus in four hours he could make only three such runs at most, covering a total distance of 1,200 meters. If instead the runner cut the runs to shorter ones, sprinting only 100 meters each time (at the same speed) and resting for 30 seconds, he would not accumulate any lactic acid and in four hours could make 360 runs, covering a total distance of 36,000 meters. In other words, by limiting the individual runs to 100 meters with the short resting intervals he could accomplish 30 times more work than with 400-meter runs. Thus a program of intermittent 100-meter sprints could be more effective as training for a 400-meter runner than 400-meter sprints.

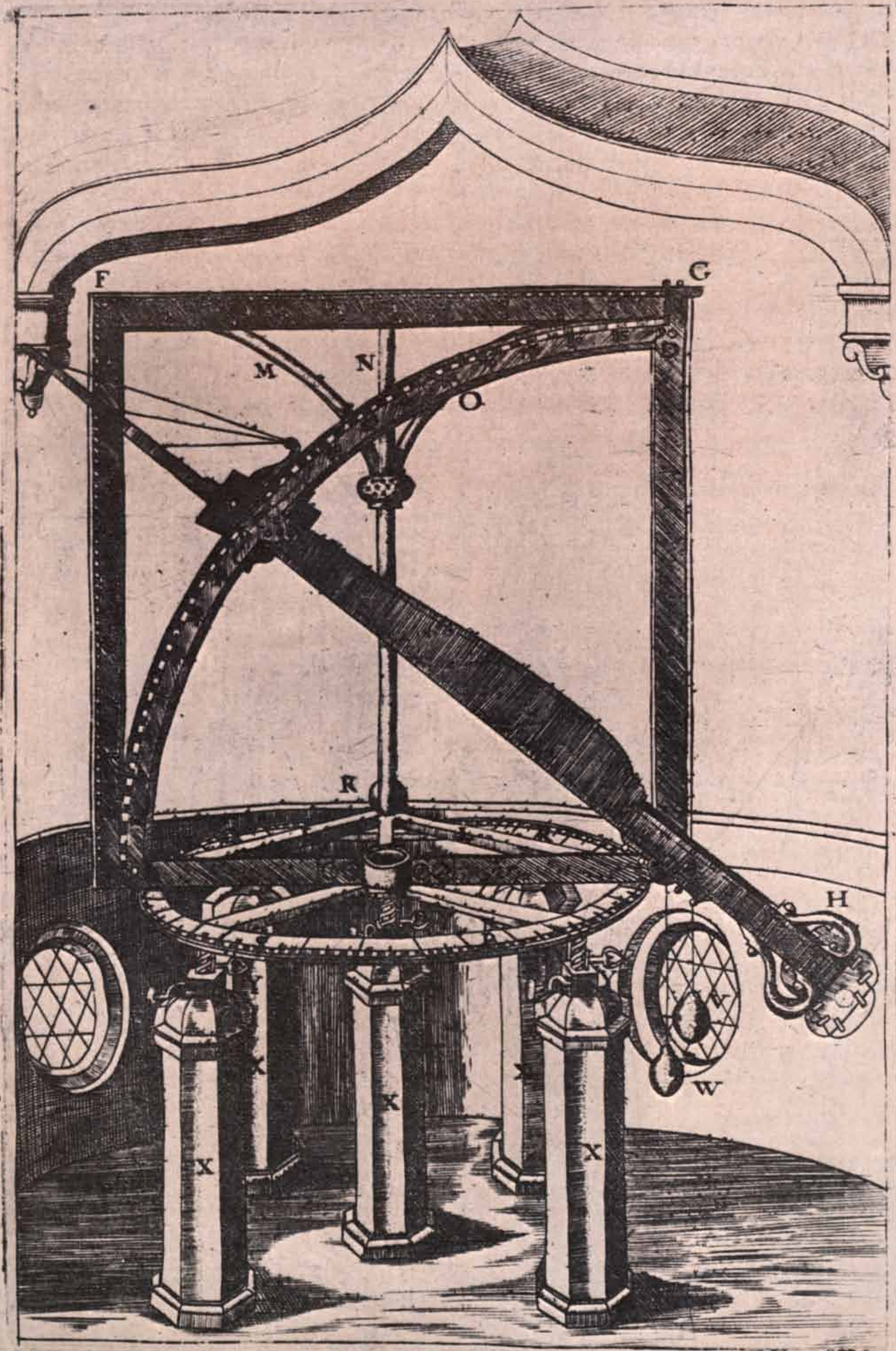
The same principle very probably applies to any kind of muscular work that can be taxing for the worker. By proper pacing of the work and the intervals of rest a person can produce more work than by driving himself relentlessly.



ALACTIC OXYGEN DEBT is created by muscles drawing energy from phosphagen-splitting (*hatched area A*) while the oxidation mechanism is being activated. In moderate exercise that does not lead to the formation of lactic acid, oxygen consumption quickly rises to its upper limit of about 40 milliliters (equivalent to 200 calories) per minute per kilogram of body weight and provides all of the energy required by the muscles. At the end of the exercise additional oxygen is consumed (*colored area B*) to repay the alactic oxygen debt (*A*). The repayment energy is used to resynthesize the phosphagens that were split.



ALACTIC AND LACTACID DEBTS are incurred when the exercise calls for superexertion to exhaustion. The alactic oxygen debt (*hatched areas A and A'*) is formed by the muscles obtaining energy from phosphagen-splitting while energy from oxygen consumption (*colored area*) is rising to its upper limit (*arrow*). The lactacid oxygen debt (*gray area C and C'*) is the result of energy obtained from the formation of lactic acid. During recovery from superexertion oxygen consumption remains at a high level for 15 to 20 seconds and then begins to drop. Energy from the delayed oxygen consumption (*colored area B*) pays the greater part (*area A*) of the alactic oxygen debt. Continued production of lactic acid (*C'*) after the exercise is over provides energy to pay the fraction *A'* of the alactic debt.



How Did Kepler Discover His First Two Laws?

It is generally assumed that he did so by calculating the distances between a planet and the sun and then perceiving that the distances fitted into an ellipse. It is more likely that the ellipse came first

by Curtis Wilson

Every plank of [science's] advance is first laid by Retroduction alone, that is to say, by the spontaneous conjectures of instinctive reason.

—Charles Sanders Peirce

Is a scientific theory formed by the patient gathering of facts until a grand design emerges? Or is it made by the pursuit of an inspired guess based on a few facts? Science probably advances both ways. The classic example of working from inspired guesses is the attempt of the Renaissance astronomer Johannes Kepler to arrive at the laws that specify the paths and motions of the planets. It was this achievement that Peirce, the 19th-century American philosopher of science, called "the greatest piece of Retroductive reasoning ever performed."

Kepler's first law states that a planet orbits the sun in an elliptical path, with the sun at one focus of the ellipse. The second law states that a line extended from the sun to the planet would sweep out equal areas of the ellipse in equal times, as the planet speeds up and slows down in its orbit when approaching and receding from the sun.

One might think that establishing the ellipse as the shape of a planet's orbit would be a straightforward geometric affair. From observations of the planet's position in the sky it should be possible to calculate the distances between the sun and the planet by a process of triangulation (such as is used in mapmaking) and perceive that these distances fit into an ellipse with the sun at one focus. Given the elliptical path, one could go on to verify the equal-areas law. Whatever the mistakes, confusions and speculative vagaries of Kepler's journey of discovery, the foregoing steps would presumably be the essential ones in the establishment of Kepler's first two laws. Indeed, some such interpretation is implicit or explicit in most accounts of Kepler's work.

A study of Kepler's own account in his *Astronomia Nova* of 1609, with special attention to the problem of observational error, has convinced me that this interpretation is mistaken. Trigonometric determinations of distance do not in themselves lead to the concept of an ellipse. They are too imprecise to give the dimensions of the ellipse, and they do not provide what Kepler regarded as an adequate confirmation of the elliptical form. What emerges from Kepler's own account is that he goes on his journey laden with theory, and that he manages to arrive at the two laws only because he approached the problem with a preconception. It is an initial hunch, a physical hypothesis, that guides him throughout; every step is taken deliberately, not only in confrontation with the data but also in pursuance of his hunch. In order to understand the logical status of Kepler's laws it is necessary to retrace the logical map of his journey of discovery.

With the benefit of hindsight we can say that all the planets known to Kepler have elliptical orbits that are nearly circular. The most elliptical of these orbits is the orbit of Mercury, but since Mercury is always close to the sun it is difficult to observe. The next most elliptical orbit is the orbit of Mars, which Kepler studied. The minor axis of the ellipse of Mars's orbit is about one two-hundredth, or .5 percent, shorter than the major axis. It would be difficult, merely by looking at a picture of this orbit drawn to scale, to detect its departure from circularity.

Hence a planetary theory that relies on circular orbits can be surprisingly good. It is not quite right, therefore, to talk about errors arising from the assumption of circular orbits alone. In any planetary theory—and this is a crucial point—an additional assumption has to be made about the motion of the planet. It can nonetheless be said that, if one did have a theory for Mars that erred only in assuming a circular orbit, the discrepancies between prediction and observation could never rise above 10 minutes of arc: about a third the diameter of the moon as it is seen from the earth. The discovery of the elliptical orbit would be possible only if the errors in the observations were notably less than this amount.

The observations that Kepler worked with were made by his contemporary Tycho Brahe from 1575 to 1600. They were made by the unaided eye with large instruments—quadrants, sextants, octants—on which the scales could be finely divided. Since many of Tycho's determinations of the angular distances between stars were checked by two other observers, it was known that his results could generally be trusted to with-

STEEL QUADRANT was among the instruments used by Tycho Brahe to make the observations on which Johannes Kepler later based his laws of planetary motion. These large precision instruments could establish the position of a celestial object to within two minutes of arc. One of the quadrant's arms was made precisely horizontal and the other precisely vertical. The entire instrument was pivoted so that it could be rotated through 360 degrees. When the alidade, or sighting arm, was pointed at a star or a planet, the altitude was read off the 90-degree arc and the azimuth off the 360-degree circle within which the quadrant revolved. This engraving is from Tycho's *Astronomiae Instauratae Mechanica*, published in 1598.

in two or three minutes of arc. Both Ptolemy, the second-century Alexandrian astronomer, and Copernicus speak of a 10-minute error as being tolerable for observational purposes. Without Tycho's reduction of the expected error Kepler's discoveries could not have been made.

When Kepler went to Prague in 1600 to work under Tycho, he was already an ardent Copernican; he had no doubt that the earth revolved around the sun. Copernicus put the sun at the center of the planetary system, and he postulated that the apparent daily revolution of the firmament was attributable to the earth's rotation on its own axis. The sun sat like a lamp in the middle of the solar system, lighting and heating the world; otherwise it was functionless. The centers of Copernicus' planetary circles did not even lie in the sun, and in accounting for what we now think of as variations in orbital speed he still found it necessary to make use of the tradition-sanctioned mechanism of the epicycle: a small circle that travels along a large circle.

Kepler felt that Copernicus had failed to realize the full meaning of his new system, and he believed the sun was more than a lamp. He noted that the actual linear speeds of the planets farther out from the sun are less than those closer in: Venus moves more slowly than Mercury, the earth more slowly than Venus, Mars more slowly than the earth and so on. To Kepler this correlation between distance and speed suggested a causal relation: he hypothesized that the sun was somehow causing the planets to move about, and that the "virtue," or motive force, fell off in strength farther out from the sun. Thus when he started his work on Mars, Kepler already had the germ of the concept of a celestial physics: an account of the planetary movements in which the planets do not have within them the source of their motion (as they had in earlier cosmological theories) but are moved from without by pushes and pulls. What we need to see is how Kepler, faced with a mass of confusing data and theory inherited from his predecessors, was guided by this physical hypothesis to his laws.

So much for preliminaries. We turn now to what Kepler jocularly called his war on Mars. The war can be divided into seven phases.

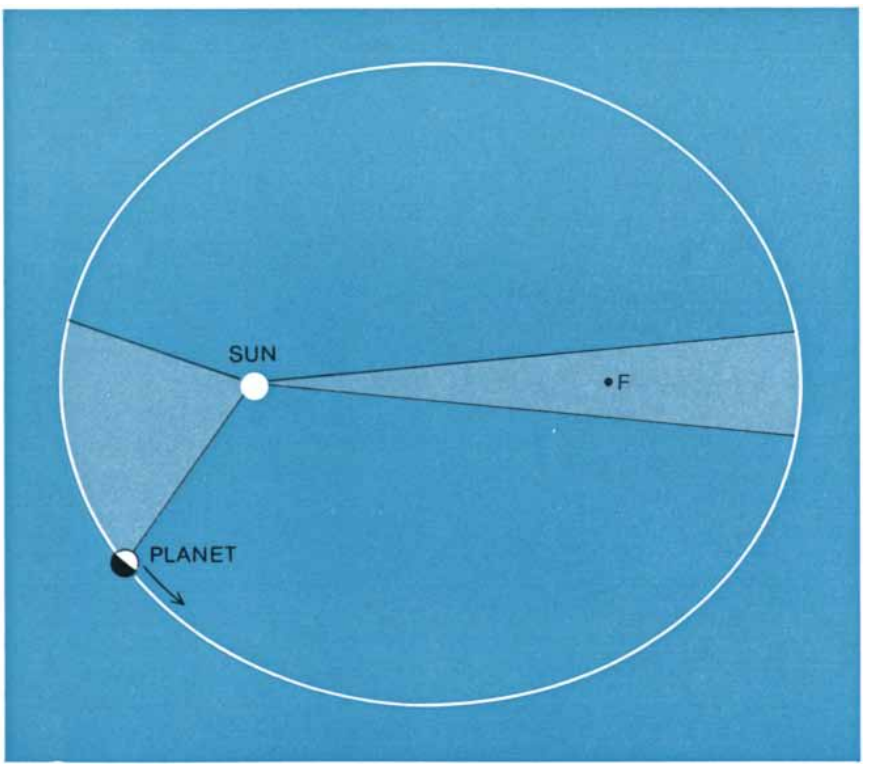
The first phase concerned the latitudes of the planets. The observed position of a planet is specified by latitude and longitude. For the purpose of making planetary observations the earth can be imagined as being suspended in the center of a much larger celestial sphere. The stars are embedded in the surface of the sphere; the Equator of the earth is projected onto it. Inclined to the equator is the circle of the ecliptic: the apparent path the sun makes across the sky in the course of a year, projected against the background of the stars. The sun's position along the ecliptic, measured in degrees eastward from the point called the vernal equinox (one of two points where the ecliptic crosses the line of the equator), is called its longitude.

The planets, like the sun, generally move eastward, approximately along the ecliptic. They depart from the ecliptic by small amounts, however, gradually going above and below it. These departures, measured in degrees from the ecliptic, are the latitudes of the planets. Copernicus' theory of the planetary variations in latitudes had been frightfully complicated, involving oscillations in the planes of his theoretical wheels: the epicycles and the circles on which they traveled. Kepler enormously simplified matters by hypothesizing that the plane of Mars's orbit is inclined at a constant angle (of about one degree 50 minutes) to the plane of the ecliptic and passes through the sun.

Kepler was able to verify this hypothesis for certain "privileged" observations. For one of these observations the earth had to be in the line of nodes: the line on which the plane of Mars's orbit cuts the plane of the earth's orbit (which is the ecliptic). At the same time Mars had to be on a line that was both perpendicular to the line of nodes and passed through the earth. At that time the observed angular distance of Mars from the ecliptic would be equal to the angle between the planet's orbital plane and the ecliptic: one degree 50 minutes.

Note that in all these observations Kepler used the sun as a fixed reference point. Earlier theories had always relied on an imaginary point called the mean sun rather than on the real or visible sun. This substitution was one of the reasons for the complications in the Copernican theory of the latitudes. Kepler's establishment of the constancy of the inclination, and the passage of the plane of the planetary orbits through the real sun, was the first victory for his hunch about the important role of the sun.

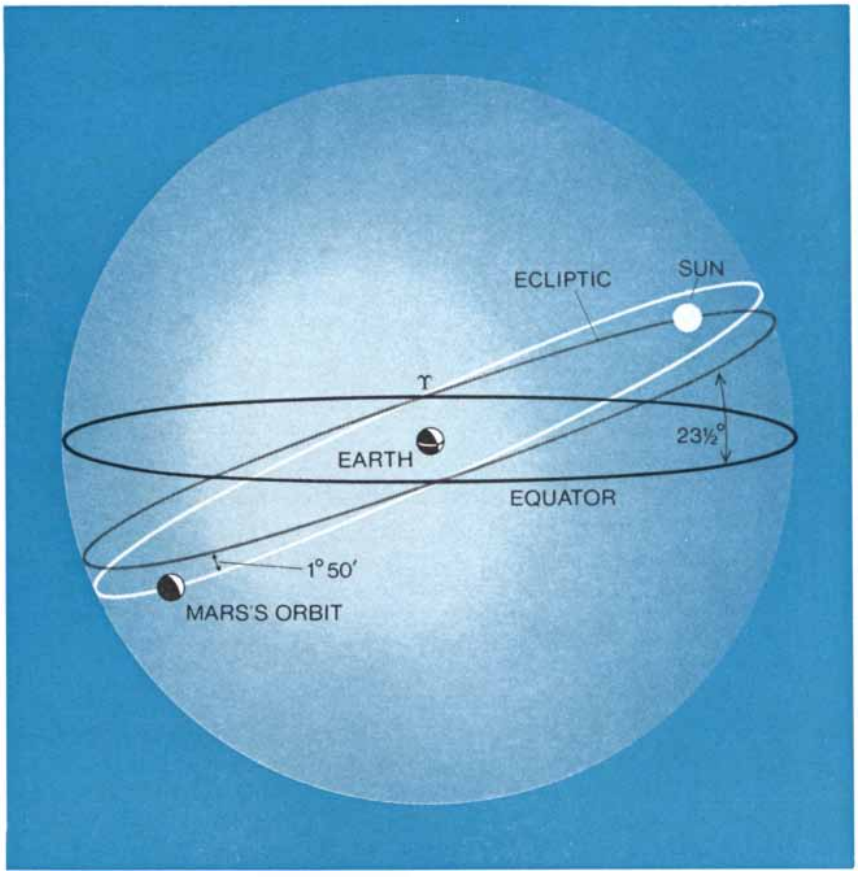
In the following discussion of the other phases of Kepler's war on Mars, the problem of latitudes will be left out. We shall deal only with the longitudes: the positions of Mars as projected onto the ecliptic. In the second phase Kepler set up a theory for the longitudes of Mars that succeeded in one way and failed in



KEPLER'S TWO LAWS state that every planet orbits the sun in an elliptical path with the sun at one focus of the ellipse, and that a line (called a radius vector) that was extended from the sun to the planet would sweep out equal areas of the ellipse in equal times.

another. He had two kinds of observations of Mars at his disposal.

Let us take the Copernican standpoint and think of the sun as being a fixed point, with the earth and the other planets moving in orbits about the sun. Ordinarily when an observation of Mars is made, the earth, the sun and Mars form a triangle; one side is the distance between the sun and the earth, the second side is the distance between the earth and Mars and the third side is the distance between Mars and the sun. The observation gives only the direction of Mars from the earth against the background of the stars. At this point we are far from knowing everything about the triangle. Kepler had a theory of the earth's motion—actually a theory of the sun's motion—that he had inherited from Tycho. Tycho's theory was a compromise between the Ptolemaic and the Copernican: all the planets except the earth orbited the sun and the entire system revolved around the earth. A simple geometric transformation turns this theory into a theory of the earth's motion about the sun. Tycho's theory was known to give the heliocentric longitudes of the earth (that is, the earth's position on the ecliptic as seen from the sun) with considerable accuracy; the principal verification lay in the theory's prediction of the midday altitudes of the sun throughout the year. For a reason soon to emerge, Kepler did not entirely trust Tycho's theory to give him the earth-sun distance, which varies around the circle of the earth's orbit because the sun is off center. And even if he had accepted Tycho's theory, in order to solve the triangle he would still have needed to know the direction of the line between Mars and the sun. How could he determine that direction?

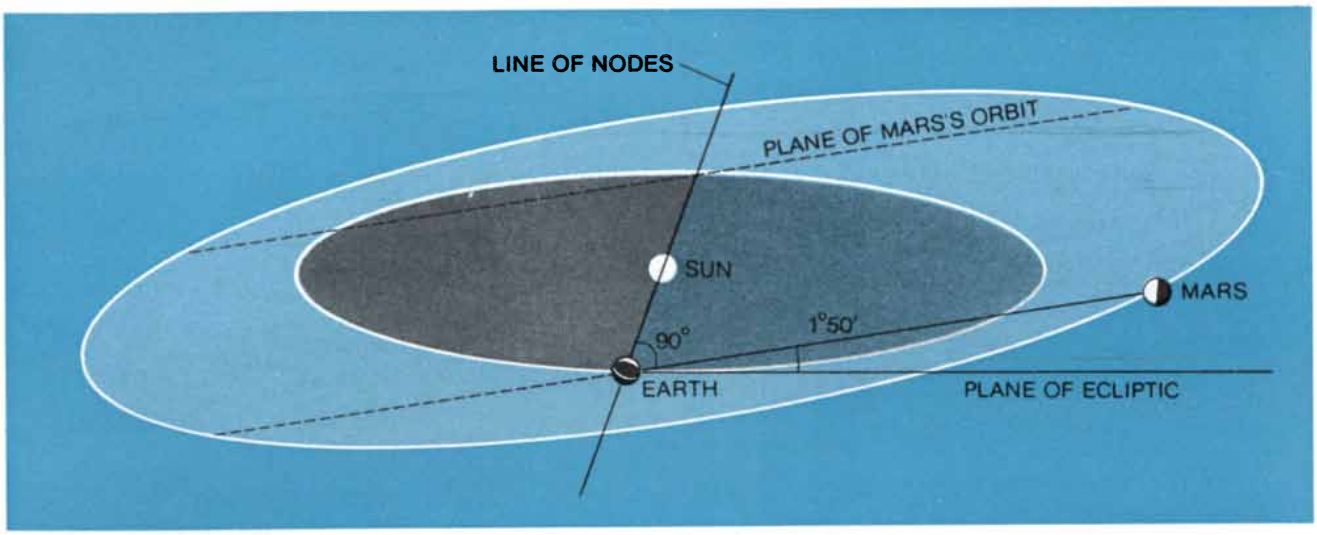


CELESTIAL SPHERE represents the way the heavens look from the earth. The earth is suspended in the center of the sphere; the stars and planets are embedded in the sphere's surface. The earth's Equator is projected onto the sphere (*black*). Inclined to the equator by 23½ degrees is the circle of the ecliptic (*gray*): the apparent path the sun makes across the sky in the course of a year. The longitudes of the sun and planets are measured in degrees eastward along the ecliptic from the vernal equinox (marked by the sign of Aries), one of the two points where the ecliptic crosses the equator. Mars and the other planets depart slightly from the ecliptic, traveling above and below it as well as along it. Kepler hypothesized that the plane of Mars's orbit (*white*) is inclined to the ecliptic plane.

This problem brings us to the second kind of observation of Mars. About every 780 days Mars, the earth and the sun come into opposition and form a line, with the sun on the far side of the earth from Mars. At about that time Mars can be observed approximately on the meridian overhead at midnight. At the exact time of opposition an observer on the earth is seeing Mars against the background of the stars in just the position he would see it if he were standing on the sun. Kepler had 12 such observations to work with; they were made between 1580 and 1604. Actually one seldom has perfect conditions for making an observation at the exact time of opposition; the position of the planet at opposition has to be calculated from a group of observations made at roughly the time of opposition. Kepler was the first to calculate the oppositions to the real sun, rather than the mean sun; this procedure was in keeping with his hunch about the role of the sun. He hoped that this change would lead him to a new and better theory.

But how to set up a theory? At the time of opposition one knows the heliocentric longitude of Mars, that is, its position on the ecliptic as seen from the sun. One does not, however, know its distance from the sun. The only thing to do is to make an assumption. Kepler first proceeded to try to fit a theory of Ptolemaic type to the data. He assumed that Mars moves in a circular path with a certain center point; the sun is off center at some other point, and the motion of Mars is uniform not necessarily about the center of its orbit but about another point called the equant point [*see illustration on pages 102 and 103*]. All these features were present in the original Ptolemaic theory of the planetary system, except that the sun was replaced by the earth. What one needed to find out was the direction of the line formed by these three points (the center of Mars's orbit, the location of the sun and the equant point) and the ratios of the distances between the points to the radius of Mars's orbit. The procedure for finding these things, for both Ptolemy and Kepler, was the horrendous one of trial and error: make a guess, then alter it if the theory fails to fit the facts.

There is one difference between Ptolemy and Kepler that should be mentioned. Ptolemy, having discovered that the center of Mars's orbit and the equant point could not be assumed to coincide, had hypothesized that the center of Mars's orbit was midway between the sun and the equant point. Kepler did not



PLANE OF MARS'S ORBIT (*color*) is inclined to the plane of the earth's orbit or the plane of the ecliptic (*gray*) by a constant angle of one degree 50 minutes. Kepler was able to verify this hypothesis for certain "privileged" observations. For one of these observations the earth had to be in the line of nodes: the line on which the plane of Mars's orbit cuts the plane of the ecliptic. At the same time Mars had to be on a line that was both perpendicular to the line of nodes and passed through the earth. At that time the observed angular distance of Mars from the ecliptic would be equal to the angle between Mars's orbital plane and the ecliptic: one degree 50 minutes. Four such observations can be made; broken lines indicate line of sight at times of other three observations.

want to make this assumption; he wanted to place the orbit's center between the other two points at exactly the right place to yield the best fit between observation and theory.

Kepler used four of the 12 observations at opposition to set up his theory, and after 70 trials he found a theory that fitted these observations. He then checked the theory against the remaining eight observations and found the concordance very good; the average discrepancy between theory and observation was 50 seconds of arc, the maximum discrepancy being two minutes 12 seconds. What the theory did, then, was give the heliocentric longitude of Mars (Mars's positions as seen from the sun) with a precision equal to that of the observations.

Kepler next discovered that this theory was false! He came to call it his "vicarious" theory, because although it was false it served to give him the heliocentric longitudes of Mars for all his later work. What it failed to give correctly were the distances of Mars from the sun.

How could Kepler learn anything about the distances of Mars from the sun? Suppose for a moment we take Tycho's solar theory and transform it into a theory of the earth's motion. Consider two observations of Mars when it is not in opposition. The earth, the sun and Mars in each of the two observations form a triangle. The position of each sun-Mars distance is now known from the vicarious theory; the position of each earth-Mars distance is given by the observation in each case, and the position of the earth-sun distance is determined by Tycho's solar theory. The angles in these two triangles are therefore known to within, say, four minutes of arc. The ratios of the sides of the triangles can then be calculated trigonometrically. If one could trust Tycho's theory to also give the ratio of the two earth-sun distances (they are not equal because the sun is not in the center of the earth's circular orbit), one could determine the ratio of the two sun-Mars distances.

Kepler knew that he could trust Tycho's theory in its prediction of earth-sun distances within certain limits. One basis for this trust was provided by Kepler's measurements of the apparent diameter of the sun's disk, which becomes larger and smaller as the earth in its orbit approaches and recedes from the sun. These measurements, although they were rough, showed that the distance between the earth and the sun does not change very much; in fact, it changes less than Tycho's theory predicted. The orbit must therefore be nearly circular. The possible errors were small enough so that Kepler could apply Tycho's theory to learning something about sun-Mars distances. He was particularly interested in checking the position of the center of Mars's orbit. The vicarious theory had put it about six-tenths of the way from the sun to the equant point. Kepler computed the sun-Mars distances near the line of apsides, that is, near the line joining the perihelion of Mars's orbit (the point of the orbit closest to the sun) and the aphelion (the point of the orbit farthest from the sun). The calculations showed that the center of Mars's orbit must be much closer to halfway between the sun and the equant point. If Kepler altered the vicarious theory accordingly, the altered version would no longer predict the heliocentric longitudes correctly but would give errors as large as eight minutes of arc, whereas Tycho's observations were not in error by more than about two minutes. The eight-minute error was what forced Kepler to go on to a total reformation of astronomy.

Kepler's situation at this point can be summarized as follows. He had two theories: Tycho's solar theory for the earth and the vicarious theory for Mars. Both were Ptolemaic-style theories, involving circular paths and uniform angular motion about a point within the circle. Each theory, Kepler believed, predicted the heliocentric longitude of its planet with errors not exceeding about two minutes of arc, as confirmed rather directly by observation. But if Tycho's theory for the earth were to be taken as being correct for earth-sun distances, then the vicarious theory would be wrong with respect to sun-Mars distances. In fact, the vicarious theory was definitely wrong, because the possible errors in the distances in Tycho's theory were relatively small compared with those deduced for the vicarious theory. Hence of the two assumptions on which the vicarious theory rested, the circular

300 gallon garbage can.

Our Marlex Marketing Team helped the container makers create it. How can our team help you?

Making garbage cans is not our business. What we do is supply high quality resins and the services of our Marlex Marketing Team to help create engineered plastic applications like outsize garbage cans, stadium seats, fuel tanks, gas pipe, cattle feeders. Lots of things.

So when low cost municipal waste programs called for a special giant garbage can, we came into the picture quite early. What the container makers wanted was a sanitary can that would be weatherproof, rustproof, damage resistant. Durable. And large. Very large.

With the help of the Marlex Marketing Team's field sales engineers and technicians, resin technologists, production experts, a 300-gallon can was rotationally molded from a new cross-linkable resin: Marlex® CL-100. It's a high density polyethylene that has what it takes, including superior resistance to impact and stress cracking.

That's what Marlex Marketing Team performance is all about. And that's how it helps make Marlex® resins worth more to you, whatever your product. For help call:

Robert F. Lynch, Fairfield, N. J. 201 227-6600
Bill Shriver, Southfield, Mich. 313 353-6666
Rod Ricketts, Villa Park, Ill. 312 834-6600
Harry House, Memphis, Tenn. 901 682-8317
David Day, Los Angeles, Calif. 213 685-6380

PHILLIPS 66

At Phillips 66 it's performance that counts®

You have known us for our brawn. Now know us for our brains.

You know Manpower.

We're the people who lift crates. And type letters.

Now meet our Technical Services Division: draftsmen, designers, engineers, research experts, scientists, technical writers, architects, even nuclear physicists.

And our Data Processing Division: keypunch and computer operators, programmers, systems analysts.

Think of them as a team of minds, not bodies. Minds that you can tap on a project basis. Minds that are there when you need them, gone when you don't.

Our computer will be glad to match wits with your job description. And, if we can't find the mind you need across the street, we will find him across the sea.

We'll always be glad to help you type your letters and lift your crates.

But the real muscle of the 70's is going to be brains.

And we've got plenty of those, too.

MANPOWER®
Temporary Help Services,
Technical Services Division.
In Canada, Manpower Services Ltd.

path and the uniform motion about an equant point, one or the other or both must be wrong. The vicarious theory remained useful, indeed indispensable, for finding the heliocentric longitudes of Mars, but it was a false theory.

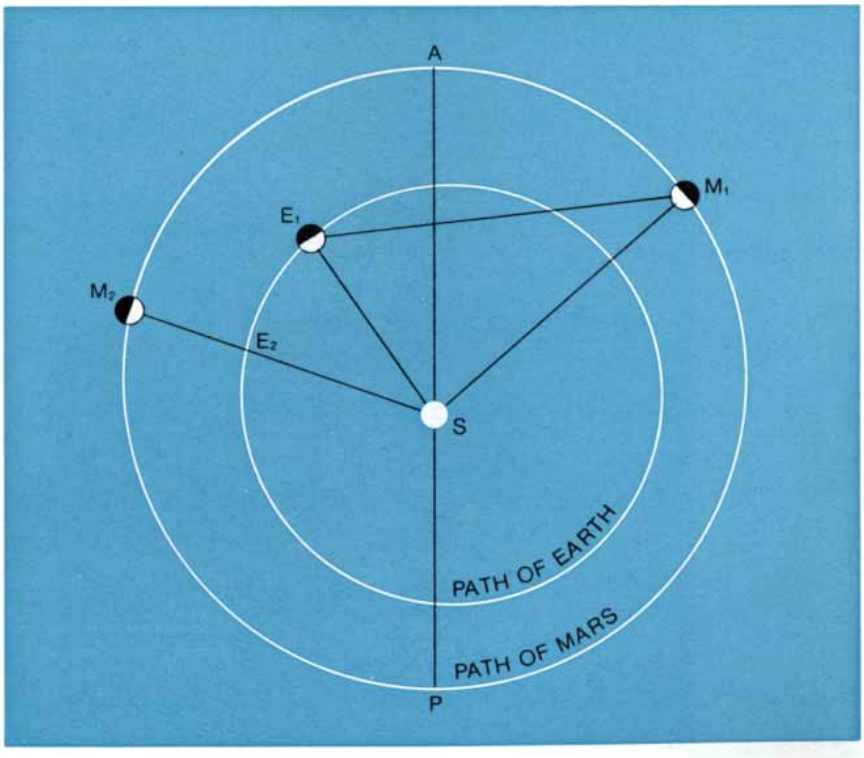
Which of the two Ptolemaic principles, circular path or equant point, should be abandoned? Kepler had no doubts on this score. If one has in mind the possibility of a celestial physics, the equant principle appears artificial; there is no body at the equant point, no believable mechanism of which the equant principle would be the expression. Moreover, the work on the vicarious theory provided Kepler with a clue to a possible substitute for the equant principle, and in the third phase of his war on Mars he devised a replacement.

Kepler found that the midpoint of the center of the line of apsides, corresponding to the center of Mars's orbit, could not be where the vicarious theory put it but must be more nearly midway between the sun and the equant point. Now, it is a peculiar fact, on which Kepler had meditated, that Ptolemy in his theories of Venus, Mars, Jupiter and Saturn had assumed what is called the "bisection of eccentricity." If we translate the earth-centered Ptolemaic theories into heliocentric Copernican form, what "bisection of eccentricity" means is that the center of a planet's orbit was assumed to lie at the midpoint between the sun and the equant point. Kepler had obtained a rough verification of this assumption for Mars. Among the principal planets the one exception to which this assumption had not been applied was the sun in its course around the earth—or, in our Copernican translation, the earth in its orbit about the sun. Tycho's solar theory, like Ptolemy's, was a simple "eccentric" theory: the earth is off center but the sun goes uniformly about the center of its circle, so that the equant point and the center of the circle coincide. When we transform this theory into a theory of the earth's motion, we again get a simple eccentric theory, with equant point and center of the orbit coinciding.

What if the eccentricity were bisected in the case of all the planets? Each planet was assumed to move at a uniform angular rate about its equant point. In a given time the planet would go through a small angle; in a later equal time it would go through an equal angle. The equality of the angles meant that, for small arcs near the line of apsides, the ratio of the lengths of the two arcs would equal the ratio of the distances between the planet and the equant point. In other words, near the line of apsides the length of the arcs traveled by the planet in equal times varied inversely with the distances of the orbit from the sun. This finding meant that the velocity of the planet varied inversely with respect to its distance from the sun.

This was a new hypothesis that Kepler believed might be generally applicable throughout the orbit for every planet. It was not precisely equivalent to the equant principle, except in the line of apsides when the eccentricity was bisected. But this hypothesis was in accord with Kepler's hunch, the physical hypothesis he brought with him to Prague: the notion that the sun, by means of some mysterious immaterial "virtue," pushes the planets about, and that its virtue falls off in strength with greater distances from the sun. Kepler felt that he would be justified in trying out this new hypothesis if he could establish that the eccentricity was bisected for the earth, just as Ptolemy had assumed it would be for the other planets. He took three of Tycho's observations of Mars made 687 days apart, the interval of time it takes Mars to make a single orbit. At all three times Mars would be at the same place. The earth, however, would be at a different part in its orbit for each of the three observations (since 687 days is not a multiple of 365 days). Tycho had thus observed Mars from three different positions. The direction of the earth-Mars line was determined by observation; the direction of the sun-earth line in each case was determined by Tycho's solar theory, and the single sun-Mars line (since Mars was in the same spot in its orbit at each observation) occupied a position determined by the vicarious theory of Mars. The observations therefore yielded three triangles with all the angles known and one side—the sun-Mars line—in common. Kepler could then find the ratios of the lengths of each of the sun-earth lines to the single sun-Mars line, and hence to each other, and thus he could determine the position of the earth in its orbit at each of the three observations. Three points determine a circle; Kepler could determine the position of the center of this circle, and hence the distance between the sun and the center of the circle in relation to the radius of the circle.

It is important to note that these trigonometric calculations are much less satisfactory than the usual accounts of Kepler's work suggest; into the solution of the three triangles went seven pieces

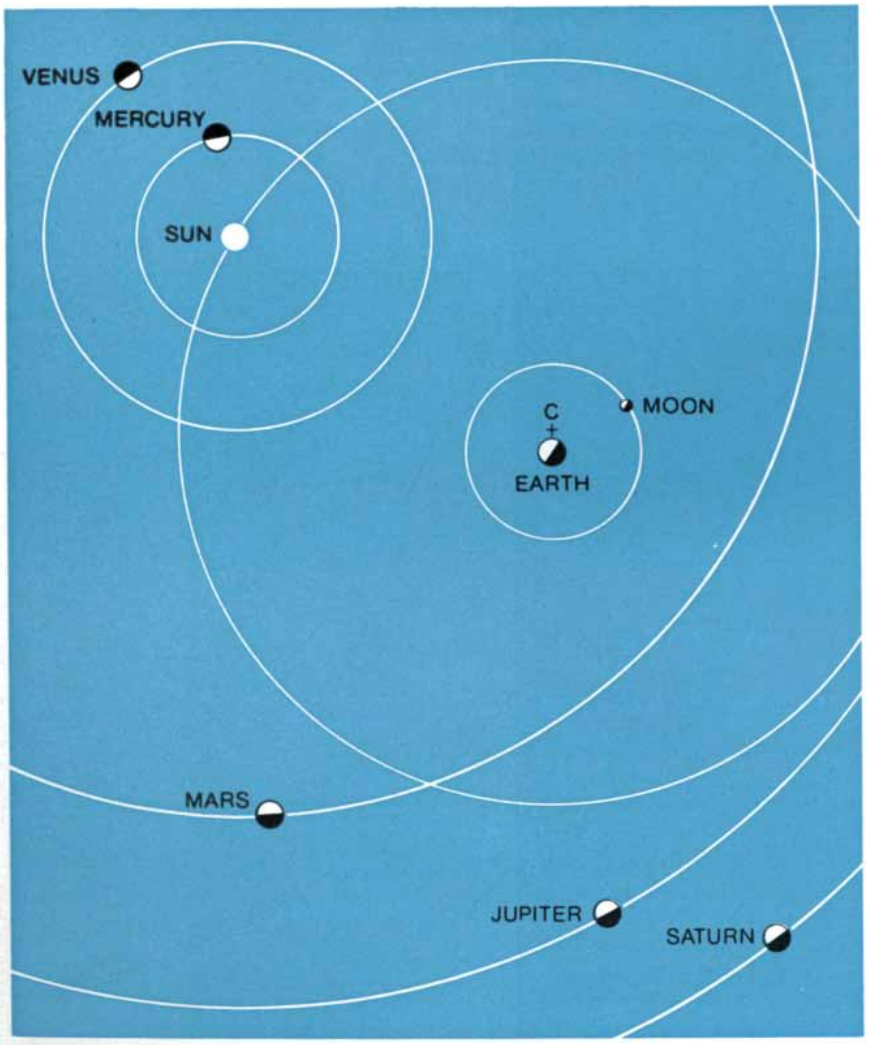


OBSERVATIONS OF MARS are of two types. When Mars is at M_1 and the earth at E_1, Mars, the earth and the sun form the triangle SE_1M_1. When Mars is at M_2 and the earth at E_2, Mars is in opposition to the sun and is observed in the same position against the background of stars where it would appear if the observer were located on the sun.

of data, each of which could be in error. The trigonometric calculation could greatly magnify the initial errors, particularly where small angles were involved. Kepler went through procedures of this kind with several sets of observations, and he got divergent results; the largest result for the distance from the sun to the center of the circular orbit was two-thirds again larger than the smallest result.

All the results, however, showed that the eccentricity of the orbit was less than the value assigned by Tycho. The eccentricity determined by Tycho was essentially the eccentricity of the equant, which (it can be shown) could always be determined with greater precision than the center of the orbit. Kepler's results, although they were markedly divergent among themselves, showed that the center of the earth's orbit did not coincide with the equant point, as Ptolemy, Copernicus and Tycho had assumed. Rather, the center lay somewhere in the middle between the equant point and the sun. Kepler, then, *assumed* that the eccentricity was exactly bisected, so that he could proceed to try out his new hypothesis (that the velocity of the planet in its orbit varied inversely with its distance from the sun).

The new hypothesis had one great disadvantage: it was difficult to apply. It stated how the speed of the planet varies as a function of its distance from the sun, but what we need to know is how far the planet goes in a given time along its path. To solve this problem one would need the procedures of the calculus, which had not yet been invented. Kepler attempted an approximation: The times needed for the planet to traverse small equal arcs were approximately proportional to the distances of those arcs from the sun; the greater the distance, the greater the time. In applying this notion Kepler went through very tedious calculations, dividing the semicircle into 180 arcs of one degree each, computing the distance of each of these arcs from the sun, adding up successive sums of these distances and making the sums proportional to the times. The result, in the case of the earth, was encouraging: predictions based on the new principle differed from predictions based on Tycho's theory by at most nine seconds of arc, which in Kepler's day was below the level of observational detection. The new theory was satisfactory in predicting the heliocentric longitudes of the earth because it agreed with Tycho's solar theory. And Tycho's solar theory, a simple eccentric theory in which the equant point coincided with the center of the orbit, was satisfactory for heliocentric longitudes because the eccentricity of the earth's orbit is small, less than a fifth the eccentricity of Mars's orbit. What mattered, therefore, was only the placement of the equant point, which Tycho had determined with fair precision. Kepler knew this and depended on it.



TYCHO'S SOLAR THEORY, which Kepler inherited when he began his work on Mars, stated that all the planets except the earth orbited the sun on Ptolemaic epicycles, and the entire system revolved uniformly in a circular path about a point off center to the earth.

Then Kepler had another thought about abbreviating the calculations. The distances from the sun to the points in one of these equal arcs are all contained in the area of the sector determined by the arc. The area of this sector suggested itself as a measure of all the distances within the sector. Just as earlier the distances of the equal arcs were proportional to the times, so now the areas might be assumed to be proportional to the times for these equal arcs. Here is the origin of Kepler's second law: the law of areas. It is usually stated today that the areas swept out by the radius vector—the line from the sun to the planet—are proportional to the times. Actually this hypothesis was a new principle, not equivalent to the former one. Again in the case of the earth it rhymed satisfactorily with Tycho's theory: the maximum discrepancy was 34 seconds of arc. In the case of Mars the differences would prove to be much larger.

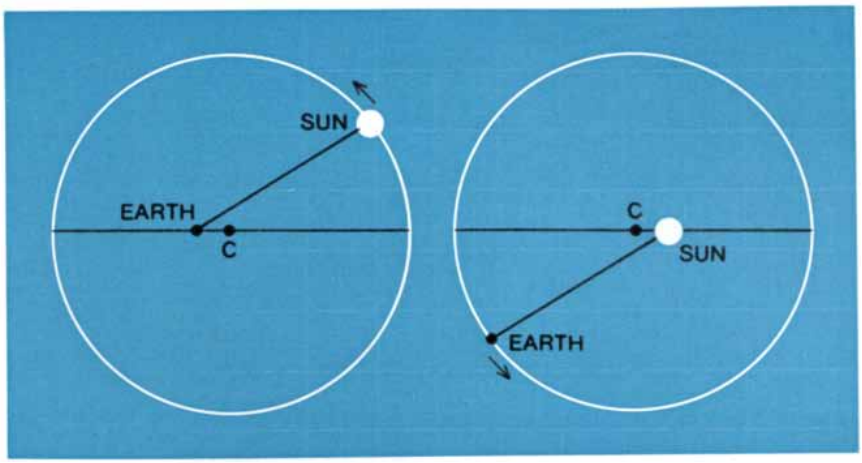
In the fourth phase of his war on Mars, Kepler renewed his attack using one or the other of his new principles that the times for equal arcs to be traversed along the orbit vary as the distances of the arcs from the sun, or that the areas swept out are proportional to the times. (Initially Kepler regarded the first of these principles as being the more valid; throughout his discussion he remained concerned with the differences in the consequences of the two principles. For the sake of

simplicity, here and in the consideration of the next two phases we shall deal only with the application of the area law.) In Kepler's new attack he first assumed a circular orbit for Mars with the sun off center. The sun presumably pushed the planet around by its immaterial "virtue," the strength of the push decreasing with increasing distance from the sun.

At this point Kepler asked himself a question. Why should the distance of the planet from the sun vary? Why does it not simply move in a circle concentric with the sun? The only account Kepler could think of was an adaptation of Ptolemy's epicyclic theory for the sun. In that theory the small circle of the epicycle had a radius equal to the eccentricity (the distance between the center of the circular orbit and the sun). The center of the epicycle moved counterclockwise around a circle with the earth at the center, while the sun moved clockwise at the same angular rate on the epicycle. At every moment the angle between the radius vector and the line of apsides equaled the angle the sun described on the epicycle with respect to the radius vector. On these assumptions the sun simply moved in a circle eccentric to the earth [*see illustration on page 104*]. Now, this same mechanism could be adapted to make Mars move in a circle eccentric to the sun. Besides being moved around the sun, Mars then had a "mover" of its own, a soul with sensitive, motile and possibly intellectual faculties that moved the planet in the tiny epicyclic circle. Kepler worried about how the planetary mover might accomplish this. He knew that, as Tycho's observations on the parallaxes of comets had indicated, comets moved right through the planetary region. The heavens were not solid; there were no crystalline spheres holding up the planets as most followers of Aristotle assumed. Moreover, the planetary mover had no feet or wings. There was another difficulty. The angle the radius vector made with the line of apsides did not increase uniformly with time but rather followed the dictates of the area law. For Mars to remain on a circle, however, this angle must at each instant be equal to the angle that Mars described on the epicycle with respect to the radius vector. Therefore the planet had to move nonuniformly about the epicycle. Did the planetary mover have to study planetary tables, trying to find out where it ought to be? Kepler was skeptical about that, but he proceeded anyway to apply the area law to the eccentric circular orbit.

With the area rule Kepler found out where Mars was in its circle at given times in the entire period of 687 days in which it completes its orbit. He checked the results against the predictions of the vicarious theory. As it turned out, there was agreement of the two theories in the line of apsides and in the quadrants (a line drawn perpendicularly to the line of apsides) but a discrepancy of about eight minutes in the octants (at 45-degree angles to the apsides). The new theory put the planet ahead of where it should be in two of the octants and behind where it should be in the other two. Hence the planet, on the assumption of a circular orbit and the area law, was being made to move too rapidly about the apsides and too slowly about the quadrants. Therefore either the circular orbit must be wrong, or the area rule, or both. If it were only the circle that was wrong, then the orbit had to be brought within the circle around the quadrants so that the area was reduced there, and hence the times for given arcs shortened. The orbit would be oval.

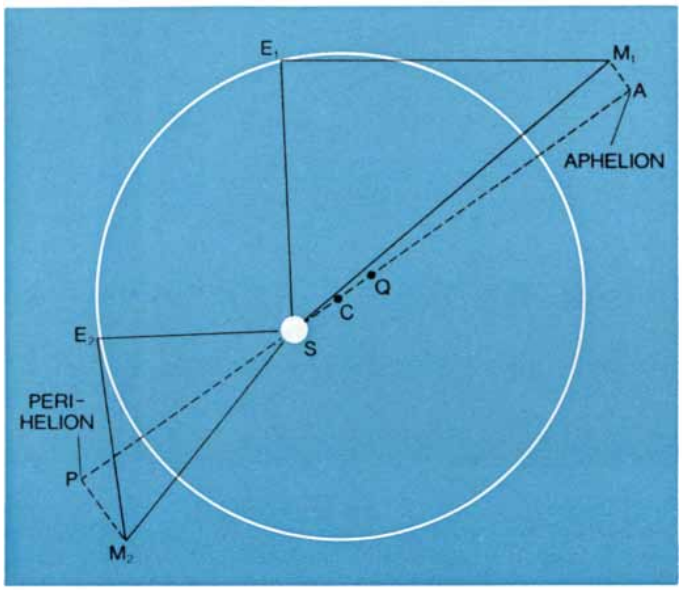
Kepler performed a number of calculations with triangles involving the earth, the sun and Mars in order to determine the distance of Mars from the sun at the quadrants. As we have seen, this type of calculation can involve considerable error. Kepler's modification of Tycho's solar theory with the introduction of the bisection of the eccentricity made that theory a better predictor of earth-sun distances, but the error was still troublesome. Kepler's results did show that Mars came within the circle at the quadrants—in other words, that the orbit *was* some kind of oval. The area law could still be right; it predicted that the orbit was oval, and the distance determinations, although they were rough, confirmed this prediction.

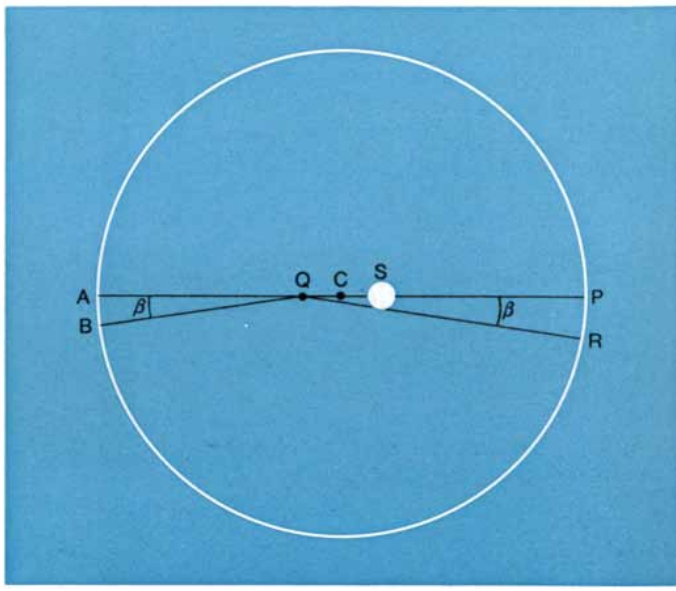


KEPLER TRANSFORMED TYCHO'S THEORY of the sun's motion into an equivalent theory of the earth's motion by interchanging the positions of the earth and the sun. If the diagram of Tycho's theory at left is rotated through 180 degrees, leaving zodiac fixed in position, sun-centered theory thus obtained at right is observationally equivalent to Tycho's earth-centered theory, that is, it predicts the same solar and planetary positions in the sky.

Kepler now entered the fifth phase of his war. At the beginning of Chapter 45 of the *Astronomia Nova* he tells us that, having discovered that the orbit was not circular, he felt he knew the cause of the departure from circularity. We recall that, in the preceding phase of the war, the planetary mover that was moving Mars on its little epicycle was having a rather difficult time of it. Not only did it lack feet and wings but also it was having to move the planet nonuniformly on the epicycle in order to keep to the eccentric circular path. Its job would be easier (although still impossible) if it only had to move the planet uniformly on the epicycle. If it were to do this, however, the resulting path would be oval, as required. The actual orbit turned out to be egg-shaped, with the sharper end at the perihelion, or the point of closest approach to the sun.

Kepler proceeded to calculate the consequences of this new hypothesis, assuming as before either the area rule or the rule according to which the velocities varied inversely with the planet's distance from the sun. The calculations were horrendous. To simplify matters Kepler substituted an ellipse (he called it the "auxiliary ellipse") for the egg-shaped oval; the difference in shape is small. Even then the problem remained difficult, and Kepler tried a number of different routes to its solution. The results in which he finally reposed trust





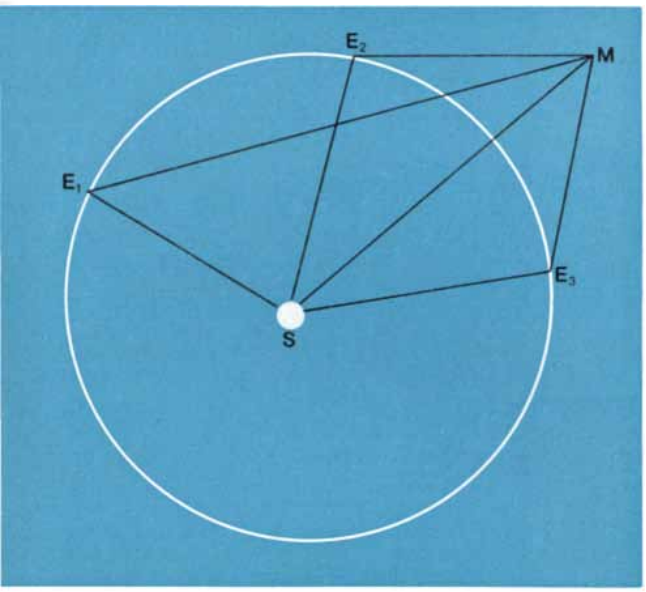
"VICARIOUS" THEORY, Kepler's first attempt to define Mars's orbit, stated that Mars moved in a circular path with center *C*, the sun was off center at some other point *S* and the motion of Mars was uniform about the equant point *Q*. Kepler calculated that the center point *C*, which was also the midpoint of the line of apsides *AP*, lay .61 of the way from the sun *S* to the equant point *Q*. To check his work Kepler assumed Tycho's solar theory and took observations of Mars near the perihelion *P* and the aphelion *A*; he computed Mars's distance from the sun at each of these times and found that *SC* must be more nearly half rather than .61 of the distance *SQ*, that is, that the eccentricity *SQ* of Mars's orbit was bisected. This fact showed that the vicarious theory was wrong. Because the eccentricity was bisected, the sun-center distance *SC* was equal to the center-equant distance *CQ*. Since the planet was assumed to move at a uniform angular rate about its equant point *Q*, during equal time intervals the planet would go through equal small angles. The equality of the angles meant that, for small arcs near the line of apsides *AP*, the ratio of the length of the arcs *AB* to *RP* (*second illustration*) was equal to the ratio of the distances *AQ* to *QP*. Since the eccentricity was bisected, placing *Q* and *S* symmetrically about *C*, it followed that the aphelion-equant distance *AQ* was equal to the sun-perihelion distance *SP*, and that the sun-aphelion distance *SA* was equal to the equant-perihelion distance *QP*. Substituting these new distances in the ratio of the arcs, Kepler discovered that the ratio of the arcs *AB* and *RP* was equal to the ratio of the distances *SP* to *SA*. In other words, the lengths of the arcs traversed in equal times at aphelion and perihelion are inversely proportional to the distances of those arcs from the sun. This finding, which Kepler called his "radius rule," meant that the velocity of a planet varies inversely with respect to its distance from the

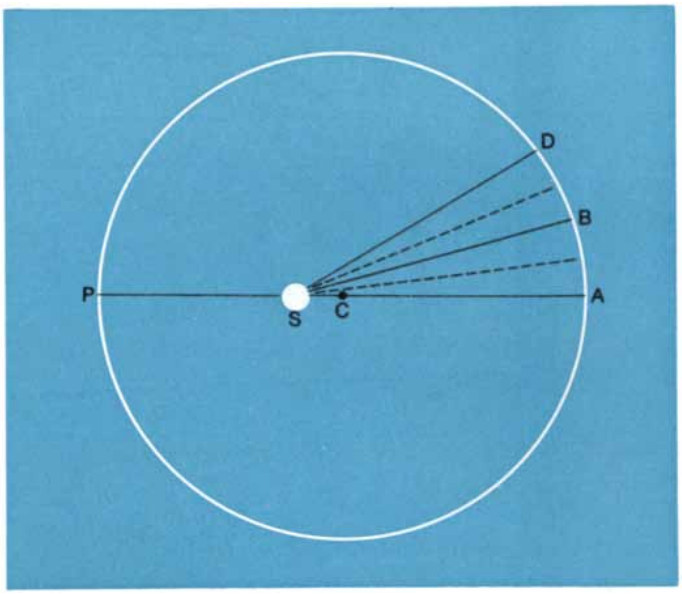
again showed agreement with the vicarious theory in the apsides and quadrants and showed discrepancies in the octants. The errors in the oval hypothesis, on the assumption of the area rule, were almost the exact opposite of the errors found in the circular hypothesis: the planet was going too slowly about the apsides and too rapidly at the quadrants. If the area law were right, then the orbit should be less narrow; areas being proportional to times, there needed to be more area between the sun and the orbit at the quadrants in order to slow the planet down there.

In the sixth phase of his war on Mars, Kepler saw that, on the assumption that the area law was right, he would get a theory that agreed with the vicarious theory in its prediction of heliocentric longitudes if he chose an orbit just midway between the circle of the fourth phase and the auxiliary ellipse of the fifth. The resulting elliptical orbit reduced the symmetrical errors to zero. In effect the area law controls the shape of the orbit. The areas swept out around the sun are assumed to be proportional to the times. Variously shaped orbits distribute the total area of the orbit in different ways; only one shape of orbit will get the planet to the right place at the right time. On the assumption of the area law the right orbit can differ only negligibly from the intermediate ellipse.

Two remarks are in order. First, the correct ellipse is the one that has the sun at one focus. The term "focus"—first introduced into European mathematical literature by Kepler himself in the *Astronomiae Pars Optica* of 1604—does not appear in the *Astronomia Nova.* At the phase of the war we have now described (the sixth) Kepler seems not to have realized that the sun was located at the focus. The focal properties were not involved in the discovery of the correct ellipse.

Second, what was the possible role of sun-Mars distances, determined trigonometrically, in the discovery of the correct ellipse? From three of Kepler's letters written in December, 1604, and January, 1605, it is apparent that the distance determinations were misleading to him. They gave the wrong value for the amount of ingression of the orbit within the circle. Let the mean distance from the sun to the earth be 100,000. Then the mean distance from the sun to Mars comes out to be approximately 152,350. The correct intermediate ellipse comes within the circle by about 660 of these parts. Kepler was getting values of 800 or 900 parts. In a passage of a letter written in May, 1605, when the war on Mars was at last over, Kepler said that the distance determinations generally left him in doubt by about 100 or 200 parts. In contrast, the assumption of the area law together with the intermediate ellipse led to predictions that agreed with the vicarious theory, in Kepler's words, "to the nail." The distance determinations played an essential role, but this role was mainly negative and admonitory. They showed that the vicarious theory was wrong in its predictions of sun-Mars distances, and that Tycho's solar theory was wrong in its predictions of earth-sun distances. In both cases the distance determinations indicated that the eccentricity was more





sun. Kepler thought he would be justified in trying out the radius rule if he could establish that the eccentricity was bisected in the case of the earth. He selected three observations of Mars separated by intervals of 687 days. Since 687 days is the orbital period of Mars, at each of these times the heliocentric longitude of Mars and its distance from the sun would be the same (*third illustration*). The direction of the line *SM* was given by the vicarious theory. The directions of the lines E_1M, E_2M and E_3M were given by the observations of Mars. The directions of the lines SE_1, SE_2 and SE_3 were taken from Tycho's solar theory. In the triangles SE_1M, SE_2M and SE_3M all the angles were known to within about four minutes of arc, and the side *SM* is common. It was thus possible to compute the ratios of SE_1, SE_2 and SE_3 to one another, and so to locate the points E_1, E_2 and E_3. Three points determine a circle, and hence the center of the circle could be located. Kepler found that this center did not coincide with Tycho's equant point but fell more nearly midway between this point and the sun. Kepler thought he had shown that the eccentricity of the earth's orbit was bisected, and he went on to apply the radius rule to Mars's orbit. He divided the circle of the orbit into 360 one-degree arcs (*fourth illustration*), computed the distance of each of these arcs from the sun, added successive sums of these distances starting with aphelion *A* and made these sums proportional to the times. The distances—the radius vectors—from the sun to the points in one of these equal arcs are all contained in the area of the sector *ASB*, *BSD* and so on described by the arc. The area of this sector suggested itself as a measure of all the distances within itself. Kepler concluded that the areas swept out might therefore be assumed to be proportional to the times for these equal arcs to be traversed. This theory is his "area rule" and is the origin of his second law of planetary motion.

nearly bisected. And in the case of Mars they showed that the orbit was not a circle but some kind of oval. After the discovery of the correct elliptical orbit of Mars and after Kepler had, arguing from analogy, altered his theory of the earth to make the shape of its orbit oval, he used the distances in the two orbits to predict positions of Mars as seen from the earth. He checked these predictions against the observations; in 28 comparisons the average discrepancy was less than three minutes of arc and the maximum discrepancy was less than six minutes. Kepler saw these results as confirming his Mars and earth theories jointly, not proving the Martian theory by itself.

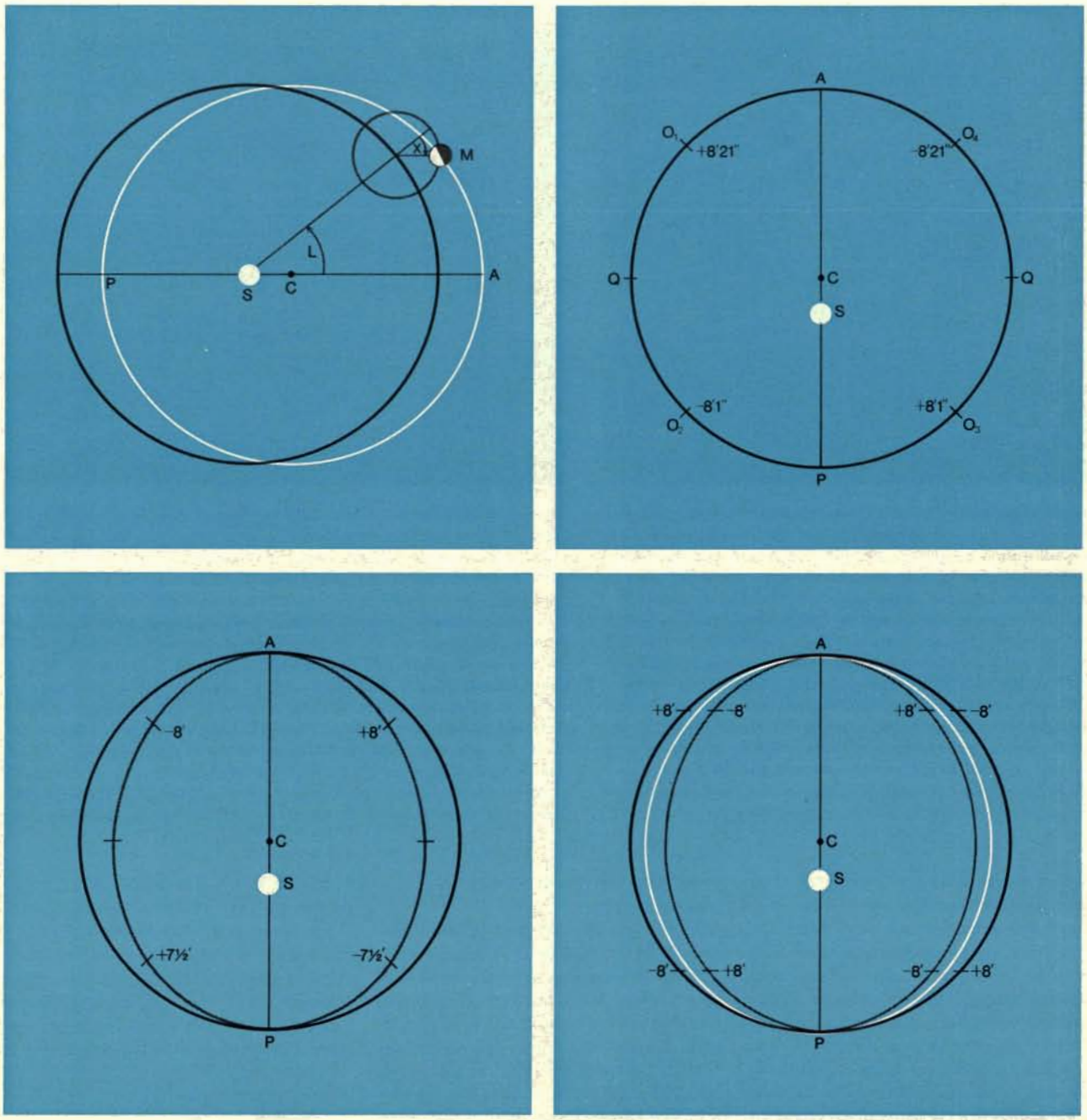
Kepler now had the correct ellipse and the area law: the first and second planetary laws that go under his name. More precisely, he knew that if the area law were right, then the orbit differed little if at all from a certain ellipse, the ellipse that would later be specified by the first Keplerian law. He was nonetheless in despair; he felt that his triumph over Mars was empty. He could not explain *why* the planet should go in this particular orbit.

In the seventh and final phase of the war on Mars, Kepler considered a diagram of the eccentric circle—not the correct orbit of Mars but the orbit of the fourth phase, before he discovered the oval shape. This circle circumscribed the correct elliptical orbit. If we assign 100,000 parts to the radius of the circle, then the ellipse comes within the circle at the quadrants by 429 parts, a number Kepler had calculated. He had been employing this circle constantly in his calculations of areas, and in the calculations a certain triangle played a role. One side of the triangle was drawn between the point of one of the quadrants and the off-center sun, a second side was between the sun and the center of the circular orbit and the third side ran from the center back to the quadrant point. The angle at the quadrant point, Kepler knew, was five degrees 18 minutes. Suddenly, quite accidentally, he happened on the secant of this angle, that is, the ratio of the quadrant-sun and the quadrant-center sides of the triangle. This ratio was 100,429 over 100,000. "It was as though," he said, "I had awakened from sleep, and seen a new light." At the quadrants the difference between the circular and the elliptical orbits is greatest, and in fact its width, 429, appeared to be exactly the excess of one side of the triangle over the other.

Immediately Kepler arrived at the notion that for any other point on the circle the distance of Mars from the sun should be given not by measuring the line between Mars and the sun in the circular orbit of the fourth phase of the war but by the perpendicular projection of this line onto the corresponding diameter of the circle.

Whence came Kepler's illumination? The course of his thought here is intricate, but the key factors may be indicated as follows. By substituting the perpendicular projection (a "diametral distance") for the sun-Mars distance (a "circumferential distance") Kepler believed he was on the way to the solution of two different problems.

The first problem had to do with why



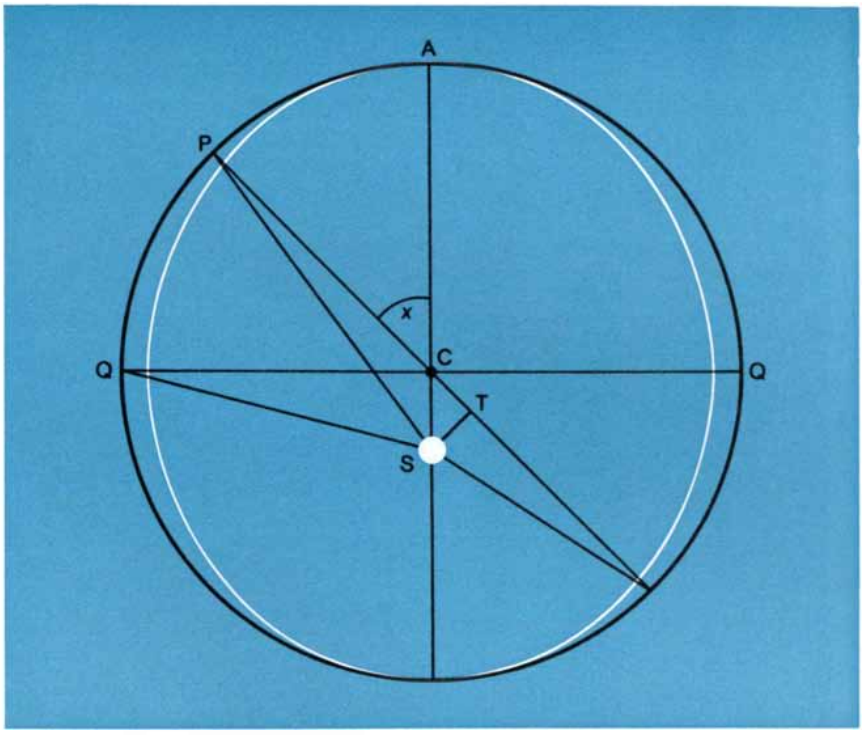
PTOLEMY'S EPICYCLIC SOLAR THEORY was the next theory Kepler tried to adapt to account for the eccentric orbit of Mars. In this theory the small circle (*gray*) of the epicycle (*top left*) moved uniformly counterclockwise around a circle (*black*) centered at the sun (*S*), whereas Mars (*M*) moved uniformly clockwise about the epicycle. The angular rate of the planet about the center of the epicycle was the same as that of the epicycle's center about *S*, so that angle *L* always equaled angle *X*. On these assumptions the path of Mars would be the white circle with center *C*. This Ptolemaic adaptation did not conform with the area law Kepler had previously derived. To do so and remain on a circle Mars would have to move nonuniformly about the epicycle. Kepler was skeptical, and he computed the heliocentric longitudes of Mars for different moments in the planet's orbital period (*top right*). The results agreed with the vicarious theory in the apsides (*A–P*) and the quadrants (*Q–Q*), but they showed a discrepancy of about eight minutes of arc in the octants (O_1, O_2, O_3, O_4). From the plus and minus signs of the errors it was evident that the planet was being made to move too rapidly about the apsides (aphelion and perihelion) and too slowly about the quadrants. Kepler concluded that the circular orbit was wrong; he brought the orbit within the circle around the quadrants so that the amount of area was reduced there and hence the times for given arcs were shortened. The new orbit was oval (*bottom left*). Kepler calculated the consequences of this hypothesis, simplifying the arduous calculations by substituting an "auxiliary ellipse" for the egg-shaped oval. The results again showed agreement with the vicarious theory in the apsides and quadrants and discrepancies in the octants. The errors, however, were equal in magnitude but opposite in sign to the errors in the circular hypothesis: Mars was made to go too slowly about the apsides and too rapidly at the quadrants. If the area law were right, the orbit would be less narrow (*bottom right*). Kepler chose an orbit (*white*) midway between the circle (*black*) of his war's fourth phase and the auxiliary ellipse (*gray*) of the fifth phase. The resulting elliptical orbit reduced the symmetrical errors to zero, showing that the right orbit could differ only negligibly from this final ellipse.

the planet's orbit was not a circle with the sun at its center. If the projected "diametral distances" are substituted for the "circumferential distances," the length of the radius vector extending from the sun to the planet varies in a way that agrees with observation. This variation can be calculated by multiplying the eccentricity by the cosine of the angle the diametral vector makes with the line of apsides at the time of the observation, and adding 1 (the radius of the circle) to the result. This specification of the sun-planet distances appealed to Kepler because he believed the sinusoidal variation was "the sort of thing that Nature does" and thus could be accounted for by means of a merely physical cause, a magnetic or quasi-magnetic force alternately attracting the planet to and repelling it from the sun. It must be admitted that Kepler's magnetic hypothesis, although it finally eliminated epicycles and planetary movers, remained a bag of tricks that had as its sole verifiable consequence the result that Kepler demanded of it, namely an elliptical orbit with the sun at one focus.

The second problem Kepler thought he could solve by substituting diametral for circumferential distances dealt with the relation between the area law and his earlier hypothesis that the times for equal arcs to be described along the orbit are inversely proportional to the distances of those arcs from the sun. Here Kepler's path is particularly confused and confusing, but he did come to understand at last that it is not the planet's orbital velocity but the component of the planet's velocity perpendicular to the radius vector that is inversely proportional to the sun-planet distance.

The substitution of diametral for circumferential distances led to a new problem requiring solution: the reconciliation of this new hypothesis with the ellipse arrived at in the sixth phase of the war on Mars. Kepler now had a rule for calculating the distances of Mars from the sun, but this rule was incomplete. At any given moment the correct distance is the "diametral distance" rather than the "circumferential distance," but how is this distance to be laid off? One end of it has to go at the sun; where does the other end go?

Kepler first verified his not quite defined formula by laying the diametral distance off along the radius vector from the sun to the planet in the old, discarded circular theory of the fourth phase. This theory gave errors in the heliocentric longitudes of eight minutes of arc in the octants. The orbit that resulted from laying off the distances in this way was not an ellipse. As Kepler pointed out, however, it did not matter: the angular positions of the new distances were not quite right, but they were never off by more than eight minutes. In an eight-minute shift at the upper octants, for example, the radius vector changed in length by about 25 parts. The distances determined by observation and trigonometrical calculation were uncertain by 100 or 200 parts. In other words, the observationally determined distances fitted orbits that differed slightly in shape from the ellipse.



CORRECT RULE FOR SUN-PLANET DISTANCES can be shown by circumscribing the correct elliptical orbit of Mars by Kepler's older circular orbit. If 100,000 parts are assigned to the radius of the circle, the ellipse comes within the circle at the quadrants *Q–Q* by 429 parts. A certain triangle had played a role in Kepler's calculations of areas: the triangle *CQS*, where *C* is the center of the circle, *Q* is the quadrant point and *S* is the sun. The angle *CQS*, Kepler knew, was five degrees 18 minutes. Quite accidentally he happened on the secant of this angle, that is, the ratio of *QS* over *QC*; it was 100,429 over 100,000. The width of the circumscribed circle over the ellipse at the quadrants, 429 parts, was exactly the excess of *QS* over *QC*. Kepler immediately arrived at the notion that when the angle the line *CP* makes with the line of apsides at the center of the circle is *x*, the sun-planet distance is not *SP* (as had been assumed in the circular theory) but *PT*, the projection of *SP* onto a diameter of the circle. Kepler called *PT* the "diametral distance." *PT* can be calculated by multiplying the cosine of angle *x* by the eccentricity *CS* and adding 1 to the result.

Kepler knew this. But he wanted to fit the distances of the new formula into the *ellipse*. The orbit *must* be elliptical, as he had convinced himself on the basis of the area law. After a time-consuming mistake Kepler discovered the right way. The new distances, laid off from the sun, were to be shifted in position from the radius vector in the discarded circular theory to a position such that the end point of the distance lay on a line perpendicular to the line of apsides [*see illustration above*]. Then the end points lie on an ellipse. Kepler here discovered a new piece of geometry, hitherto unknown.

Very well. The "war" was over. What the foregoing interpretation has aimed to show may be summarized as follows. The revolution Kepler brought about in astronomy—and the Keplerian revolution is the decisive one, because it is here rather than with Copernicus that man crossed the divide between ancient and modern astronomy—did not consist merely in the discovery of descriptive, empirical laws. If the first two Keplerian laws had been only descriptive, the ellipticity of the orbit would have had to be verified independently of the area law, and the area law would have had to be verified within the ellipse thus found. The actual process was the other way around. What Kepler verified with the

degree of precision he wanted was the proposition: If the area law is right, then the orbit is elliptical. It is true that he had satisfied himself on the basis of distance determinations, and thus independently of the area law, that the orbit was oval. But the distance determinations left an unsatisfying range of indeterminacy.

Newton later wrote: "Kepler knew ye Orb to be not circular but oval, & guest it to be Elliptical." This is correct, but Kepler's guess was no idle guess; it came out of a hunch actively pursued, in confrontation with all previous theories and with Tycho's new data.

If one were asked what was responsible for Kepler's discoveries, one undoubtedly would have to admit the role of chance or luck, or, as Kepler would have said, Providence. It was into his care that Tycho's observations were confided—the only observations that could have led him to his goal. The first task he was assigned was to construct a theory for Mars—the only planet whose elliptical path could have been discovered in the existing state of the observational art. The theories of the fourth and fifth phases of the war happened to err equally in opposite directions from the right result. He happened by accident on the secant of a certain angle, and so he emerged from his final perplexities.

What was just as important through all the accident and luck was Kepler's belief in the possibility of understanding, his devotion to his task that carried him through four years of reasoning and calculation, and finally, the rightness of his initial hunch and his ability to disentangle the confused state of things before him in the light of it. He proceeded indeed by a kind of Sherlock Holmesian logic, which presumes that in eliminating the impossible and the more implausible it has arrived at the true. Most of the physical hypotheses Kepler constructed had to be discarded later as being inconsistent with Newtonian theory. At the root of all his theorizing, however, was that initial sense of the significance of the inverse relation between velocity and distance—an anticipatory glimmer of what would one day be the law of the conservation of angular momentum. It was in the light of that hunch that he was guided through 900 pages of calculation to a planetary theory better than any that had been proposed before. No doubt Kepler's discoveries were a kind of miracle of chance and love, and also of inventive hypothesizing and detective logic.

Why Were Bankers Warned That This "New Book Could Upset the Savings Applecart"*

How to Make Up to 13½% or More on Your Savings—All Fully Insured

There are many things banks don't like to talk about. They don't like to talk about the fact that they do, indeed, pay interest rates of 8%, 10%, 13½%, and often more, to a select group of knowledgeable depositors!

What's the difference between these men and women and you? Simply this—*that they know certain "inside" techniques of depositing and withdrawing their savings (all perfectly legal, incidentally) that you don't! And—therefore they earn two to three times as much interest on those savings as you do!* Like this...

Would It Be Worth An Extra Ten Minutes A Month To You—To Bring Home Two Or Three Times As Many Dollars From Your Savings As You're Getting Today?

The procedure is simple. But to put it to work for you, tomorrow, you have to know these few "smart-money" facts:

1. Most thrifty people in this country today are actually *losing money* on their savings. The interest they get from a regular bank account is actually *far less* than the lost purchasing power that inflation cuts right out of their savings.

This is one of the great social tragedies of our time. It means that if you are thrifty and prudent in this country today, you are penalized. Either you are driven to speculate in the stock market, where you can be wiped out overnight. Or you try to secure safety for your hard-earned capital in a bank—*and watch inflation turn your dreams of early retirement and financial independence into dust!*

2. *But you just don't have to accept these two tragic choices any longer!* Now there is a Third Way to invest your money, that gives you the absolute safety you want, plus huge guaranteed returns that you may not even have dreamed possible before.

Read What The Banking Industry Itself Says About This Startling Volume!

NEW BOOK COULD UPSET THE SAVINGS APPLECART

TIGHT MONEY, Regulation Q, and the much-touted Age of the Consumer, are key ingredients in a flammable mixture about to be ignited by a book which could explode in the face of the commercial banking industry this year. This says:

"The millions of people who have saved a few dollars in the form of savings accounts and insurance have been prevented from gaining any profit from their investment—indeed they have been forced to accept real losses—by what amounts to government agency fiat. These depositors have contributed more, perhaps, to the growth of our economy than any other group, and it is unjust that controls apply only to interest rates to depositors, while there are no controls over the inflationary wage and price increases. Conditions permitting this 20 years of discrimination should be changed."

I am quoting from a book, titled, "Don't Bank On It! How To Make Up to 13½ percent and More on Your Savings—All Fully Insured."

The book is dedicated "to the members of the median income group, those truly forgotten men whose savings deposits make banking, as we know it, possible."

"Don't Bank On It" may be coming out at an auspicious time, as the general public is becoming more aware of high interest rates, and, thanks to truth in lending, is being conditioned to look at rates of 12 or 18 percent as low. No doubt he'll soon recognize that 4 or 5 percent is peanuts.

**Bank Marketing Management, Feb., 1970.*

About the Authors

Martin J. Meyer is president of the National Depositors Cooperative Association. He also serves as Vice President and Secretary of Intercept Tele-Communications, Inc., a new international cable and telegraphic interception and forwarding organization. Mr. Meyer has written numerous magazine articles on banking, thrift, and inflation.

Dr. Joseph M. McDaniel, Jr., recently elected President of the World Health Organization, was Secretary of the Ford Foundation from 1953 until his retirement in 1967 and Dean of the School of Commerce at Northwestern University. His distinguished career includes government service with the Economic Cooperative Association.

3. It is based on one simple fact: That most depositors are *completely passive* about where and how they save their money! They never take the one or two hours that are necessary to learn the "inside workings" of the banking system. Therefore, they never even hear about the "super-savings-accounts" that can yield them *far more* than ordinary interest on their money.

And, above all, they have never heard about the simple, ingenious techniques of "Loophole Depositing"! Active, precisely-timed deposits and withdrawals that take an average of ten minutes of your time per month—*and bring you back 8%... 13½%...even 19% on every dollar, with exactly the same total safety that you get on ordinary bank accounts today!*

You Do Only What The Banks And Savings And Loan Associations Themselves OPENLY PERMIT. Except Now—YOU DO IT A LITTLE SMARTER!

At this moment, in this entire country, out of the over one hundred million people who have savings accounts, *only about forty thousand of them use these techniques.* They are still brand-new—virtually-unknown. *Only now has a book been published that reveals them to every man and woman who is willing to risk an 8¢ stamp to learn them!*

The time required to read this book from cover to cover is approximately one weekend. Or, if you wish to skip the banking background at its beginning, it will take you about an hour or two to learn these "Active Depositing" techniques themselves. *And once you learn them, from that moment on, you will be able to exploit every legal loophole in the entire banking system, including:*

How to get more than 8% interest per year as *an absolute minimum,* with hardly any more work than filling out your deposit slip in a different way. And then go from there all the way up to as much as 19% to 25% in special situations, for limited periods!

How to protect yourself against the possibility of interest rates dropping in the future. So you're *guaranteed* the high interest rates available to you today, even if tomorrow your friends find *their* return on their savings cut in half!

How to make banks pay interest to you *on money you don't really even have on deposit—on non-existent money—on money you have already spent!* (And the bank loves you for it. Because, no matter how much you make, they make more! Page 143 shows you how.)

Yes, how you can even earn high interest on *your credit card!* So that you are now earning interest on other people's money—*and spending it at exactly the same time!*

Special Warning Section: Two common mistakes, that unknowingly trap thousands of depositors every year, *that could completely destroy your savings!*

And how to defer income tax on the interest you get. Two plans that offer marvelous tax-sheltered advantages.

And—let us repeat once again—*all completely protected by United States Government Insuring Agencies!* With *no* service charges—*no* minimum balances—no legal technicalities! Ready to go to work *right now* for the investor with $500—or $500,000!

Prove Every Penny Of It Yourself—Entirely At Our Risk!

The top money men in this country—the very banks themselves—use these techniques! Why shouldn't you? *All you risk to learn them all is an 8¢ stamp with the Coupon below! Why not send it in—and start your savings earning double and triple interest—8% to 13½% or more—TODAY!*

JAY NORRIS CORP.

25 W. Merrick Rd., Dept. BEG-2, Freeport, N.Y. 11520

MAIL NO RISK COUPON TODAY!

JAY NORRIS CORP. Dept. BEG-2, 25 W. Merrick Rd., Freeport, N.Y. 11520

Please rush me: _____ (quantity) copy(ies) of DON'T BANK ON IT. I understand the book is mine for $8.98 complete. In addition, I understand that I may examine this book for a full 30 days entirely at your risk. If at the end of that time, I am not satisfied, I will simply return the book to you for every cent of my money back.

☐ SAVE! Order TWO for $16.00 complete. Makes an ideal gift!

Enclosed is ☐ check or ☐ money order for total of $_____
New York residents add sales tax.

Name (print)_____

Address_____

City_____

State_____ Zip_____

MATHEMATICAL GAMES

The graceful graphs of Solomon Golomb, or how to number a graph parsimoniously

by Martin Gardner

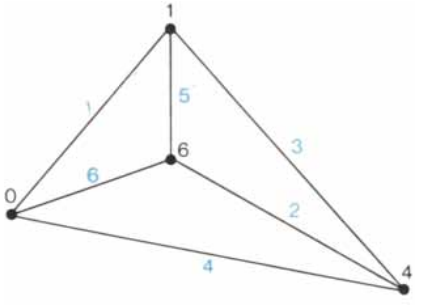
One of the least explored areas of modern mathematics is a class of problems that combine graph theory and arithmetic. Recreational problems of this type have been discussed before in this department, notably in December, 1965, when the topic was magic stars and their relation to the skeletons of polyhedrons. This month we take up a family of numbered-graph problems that has recently been defined and developed by Solomon W. Golomb, professor of engineering and mathematics at the University of Southern California. He is the author of *Polyominoes* (Scribner's, 1965), numerous articles on recreational topics and many technical papers. What follows is extracted from his correspondence and from an article titled "How to Number a Graph." The article will appear in the book *Graph Theory and Computing*, edited by Ronald C. Read, which will be published later this year by Academic Press.

Golomb has coined the term "graceful graph" for any graph that can be "gracefully numbered." He explains this terminology with a simple example: the graceful numbering of the graph shown in the illustration at left below. It is called the "complete graph for four points" because every pair of its four nodes is joined by a line called an edge. The graph is topologically equivalent to the skeleton of a tetrahedron. It is planar because it can be drawn on the plane without intersecting edges. A graph, as we shall see, need not be planar in order to be gracefully numbered, but it must be without loops (lines joining a node to itself) or multiple edges (more than one edge connecting the same pair of nodes).

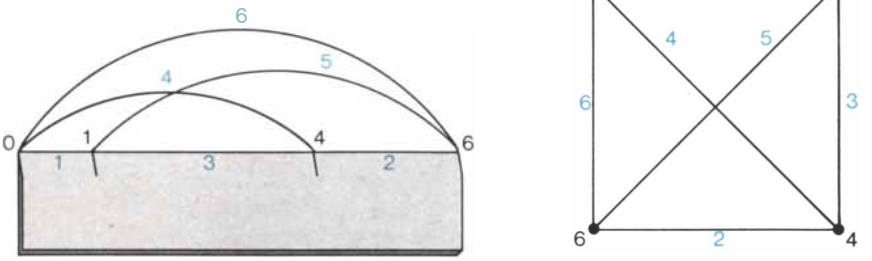
Each node is labeled with a positive integer. The lowest integer (by convention) is 0, and no two integers may be alike. After the nodes are numbered every edge is labeled with the difference between the numbers of its two end nodes. Like node numbers, all edge numbers also must be distinct (no two alike). The objective is to do all these things and keep the largest node number as small as possible. Obviously it cannot be smaller than the number of edges. If the largest node number equals the number of edges, e, the edge numbers will run consecutively from 0 through e, and we shall have achieved a graceful numbering. The number e will represent three values: the total number of edges, the highest node number and the highest edge number. Any graph that can be gracefully numbered is a graceful graph. Some graceful graphs have only one basic numbering, others more than one. (Trivial variations obtained by such symmetry operations as rotations and reflections, or by replacing each node number n by $e - n$, are not considered different.) A graph that cannot be numbered gracefully is called an ungraceful graph.

As Golomb points out, every complete graph can be drawn with all its nodes on a straight line and the remaining edges can be added as curved lines [*see left side of illustration at right below*]. Let us go further. Imagine that the straight line is the edge of a ruler with a length equal to the largest edge number of a numbered graph. The nodes of the graph are marks on the ruler at points that correspond to their numbers, each number indicating the mark's distance from the zero end of the ruler. Golomb calls such a ruler a "Euclidean model" of a numbered complete graph. The problem of gracefully labeling a complete graph of n nodes is equivalent to the problem of putting n marks on a ruler (always including the ruler's two ends as marks) so that every distance between a pair of marks is a distinct integer. In this example the ruler is marked at points 0, 1, 4 and 6, the node numbers of the complete graph for four points after it is gracefully numbered. Such a ruler clearly can measure lengths of one, two, three, four, five and six units. At the right of the ruler is shown another way of drawing the complete graph for four points: as a four-sided polygon with all its diagonals. (The intersection of the diagonals is not, of course, a node.) Note that the distances between adjacent marks on the ruler, together with the ruler's length, correspond to the perimeter numbers of the gracefully labeled square graph.

A closely related but less restricted ruler problem was discussed in Chapter 6 of my book *The Numerology of Dr. Matrix* (Simon and Schuster, 1967). Dr. Matrix' rulers measure all integral distances from zero to the length of the ruler, but the numbers of its "edges" (distances between any pair of marks) are not required to be different. With the added proviso that all such distances must be different, Dr. Matrix' ruler problem becomes identical with the problem considered here: that of finding a ruler with marks that correspond to the graceful numbering of a complete graph with n nodes. Golomb proves in his paper that this can be done only if n is 1, 2, 3 or 4. Expressed differently, no complete graph for n points, when n exceeds 4, can be gracefully numbered.



A graceful graph



Ruler version **(left)** ***of a complete graph*** **(right)**

If we keep the requirement that all distances between pairs of marks must be different, but we do not insist that they run consecutively from zero to the ruler's total length, we can still look for the shortest possible ruler of n marks (end points are included as marks) on which all distances between a pair of marks (which correspond to the edge numbers of the complete graph for n points) are different. In the chart of the shortest-known rulers when n is from 2 through 11 [*see top illustration on this page*], only the first three entries are solutions to Dr. Matrix' ruler problem. They correspond to the graceful numbering of complete graphs for two, three and four points. The other rulers do not have consecutive integral distances from zero to the ruler's length; they correspond to what Golomb calls the "best" numbering of complete graphs for more than four points. The numbers in each row give the distances between adjacent marks on rulers of two, three, four, ..., 11 marks. The chart, which extends downward to infinity, is called the Golomb triangle.

NODES	EDGES	DISTANCES BETWEEN ADJACENT MARKS										LENGTH
2	1	1										1
3	3	1	2									3
4	6	1	3	2								6
5	10	1	3	5	2							11
6	15	1	3	6	5	2						17
7	21	1	3	6	8	5	2					25
8	28	1	3	6	11	8	5	2				36
9	36	1	3	12	10	8	6	5	2			47
10	45	1	3	6	12	16	11	8	5	2		64
11	55	1	8	10	5	7	21	4	2	11	3	72

GOLOMB'S TRIANGLE

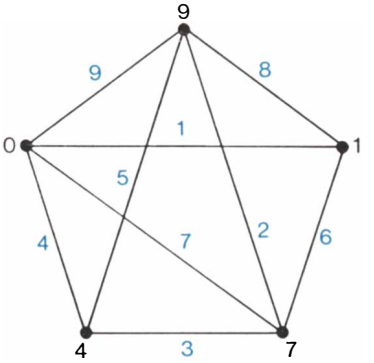
The shortest Golomb rulers with two through 11 marks

We can put the difference between Dr. Matrix' rulers and Golomb's rulers as follows. Dr. Matrix' rulers minimize the number of marks for a ruler of length k that can measure all integral distances from 1 through k. Golomb's rulers do not necessarily include all the integral distances from 1 through k; with Golomb's rulers, for a ruler with a given number of marks, the length of the ruler is minimized and all the integral distances the ruler does measure are different. If we draw a graph corresponding to a Dr. Matrix ruler, we may find two edges with the same edge number. By omitting all edges with duplicate numbers we can get a graceful graph that Golomb calls a "graceful approximation" of a complete graph. For example, by dropping one edge (the line between points 1 and 4) from a complete graph for five points [*see illustration at bottom left on this page*] the graph can be gracefully numbered. It is equivalent to Dr. Matrix' ruler with marks at points 0, 1, 4, 7 and 9.

With a yardstick, or 36-unit, ruler as an example, here is a quick way to prove that all distances measured by a Golomb ruler are distinct. The yardstick is the shortest-known ruler with eight marks. The top row [*see illustration at bottom right on this page*], taken from Golomb's triangle, gives the distances between adjacent marks on the ruler. These seven numbers, together with the ruler's total length, correspond to the eight edge numbers on the perimeter of an eight-sided polygon when it is made into a complete graph by drawing all its diagonals and then numbered as gracefully as possible. The second row of numbers is obtained by adding successive pairs of numbers in the row above it. The third row consists of adding successive triplets in the top row, the fourth row of adding successive quadruplets, and so on. The bottom number is the ruler's length. It is, of course, the sum of all the numbers in the top row. The 28 numbers of this triangle are the 28 edge numbers of the complete graph for eight points when it is given the best ungraceful numbering. If all these numbers are different, no two edge numbers of the complete graph will be alike and no two distances between pairs of marks on the corresponding ruler will be alike.

Golomb admits that for all rulers longer than six units the results were obtained (by himself and others) partly by trial and error. They have not yet been proved to be rulers of minimal length. (The ruler of length 47, for nine marks, was first found in 1965 by Matthew J. C. Hodgart of Brighton in England; the ruler of 72 lengths, for 11 marks, by Robert Reid of Miraflores in Argentina, also in 1965.) Perhaps readers can improve on these results or extend the triangle farther downward.

One of the many unusual properties for all graceful graphs discovered by Golomb is that the nodes of such graphs can always be divided into two sets—those with even numbers and those with odd—and the number of edges connecting the two sets will be $[(e+1)/2]$, where e is the total number of edges in the graph. The brackets mean that the expression is rounded down to the nearest integer. Golomb calls this a "binary



Graceful graph for a Dr. Matrix ruler

1 3 6 11 8 5 2

4 9 17 19 13 7

10 20 25 24 15

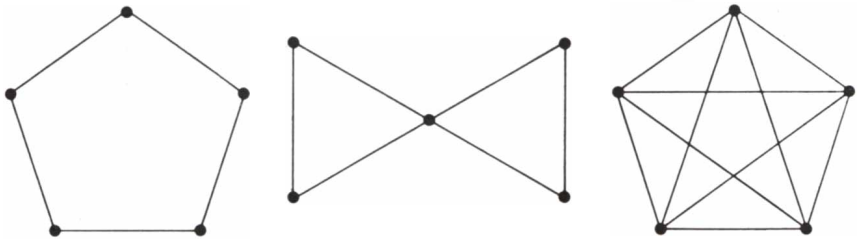
21 28 30 26

29 33 32

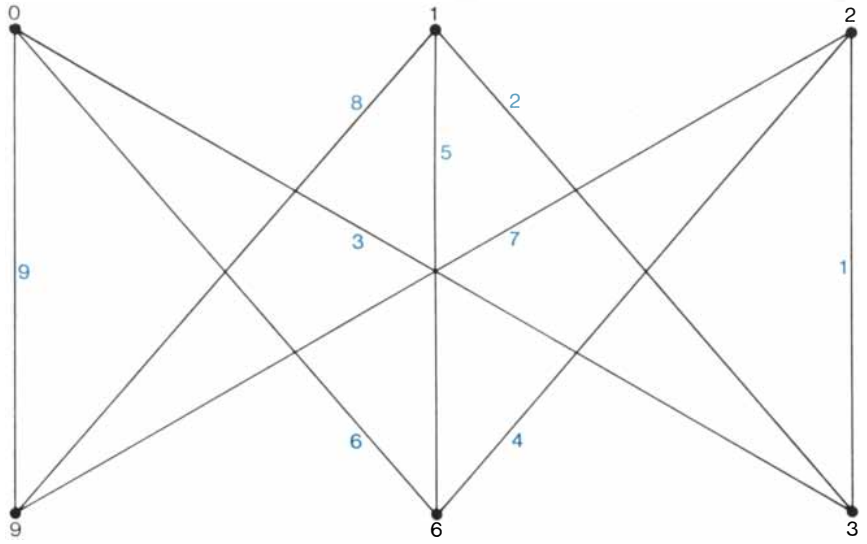
34 35

36

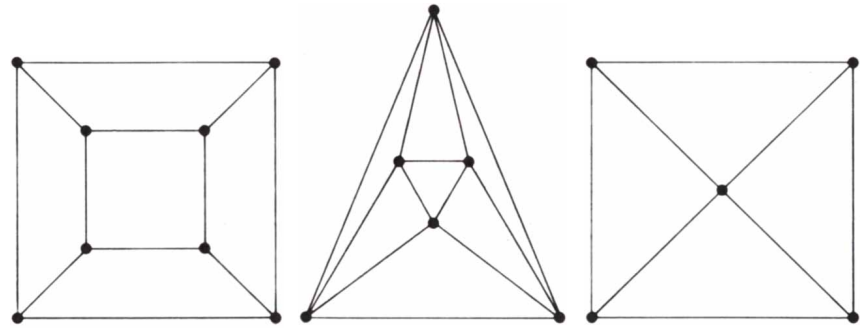
Proof for eight-mark Golomb ruler



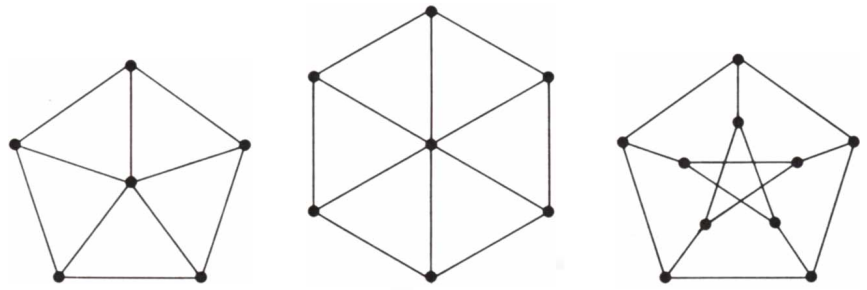
The only ungraceful graphs with fewer than six nodes



A graceful numbering of the Thomsen graph



Three graceful Schlegel graphs: cube (left), *octahedron* (center) *and Great Pyramid* (right)



Three graceful graphs by Golomb with six, seven and 10 nodes

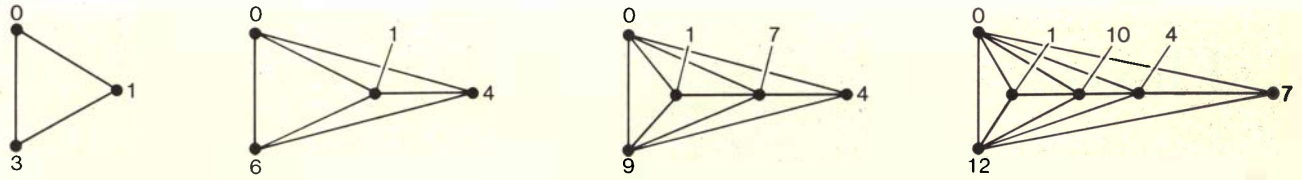
labeling." For example, the even set of nodes in the graph at the left on page 108 are numbered 0, 4 and 6, and the odd set has only the number 1. Inspection shows that the two sets are indeed joined by $[(6 + 1)/2] = 3$ edges.

Moreover, as Golomb proves, if all the nodes of a graph are of even order (attached to an even number of edges), the graph is graceful only if $[(e + 1)/2]$ is even. When this value is odd, binary labeling is impossible and therefore the graph cannot be gracefully numbered. Of the 29 topologically distinct graphs with five or fewer nodes, only three are ungraceful. All three have five nodes and all their nodes are of even order. The three graphs violate Golomb's theorem that $[(e + 1)/2]$ must be even [*see top illustration at left*]. Note that the first two graphs are planar whereas the third, the complete graph for five points, is not. This shows that not all planar graphs, and not all nonplanar graphs, are graceful. Can a nonplanar graph be graceful? Yes, as the graceful labeling of the Thomsen graph shows [*see second illustration from top at left*]. The Thomsen graph is sometimes called the utilities graph because it diagrams the well-known (and unsolvable) puzzle in which three houses are each to be connected to three utilities without any crossing of edges. The Thomsen graph is one of an infinite family of graphs, known as "complete bipartite graphs," in which every node in a set of a nodes is joined to every node in a set of b nodes, but nodes within each set are not connected. Golomb has established that all complete bipartite graphs are graceful.

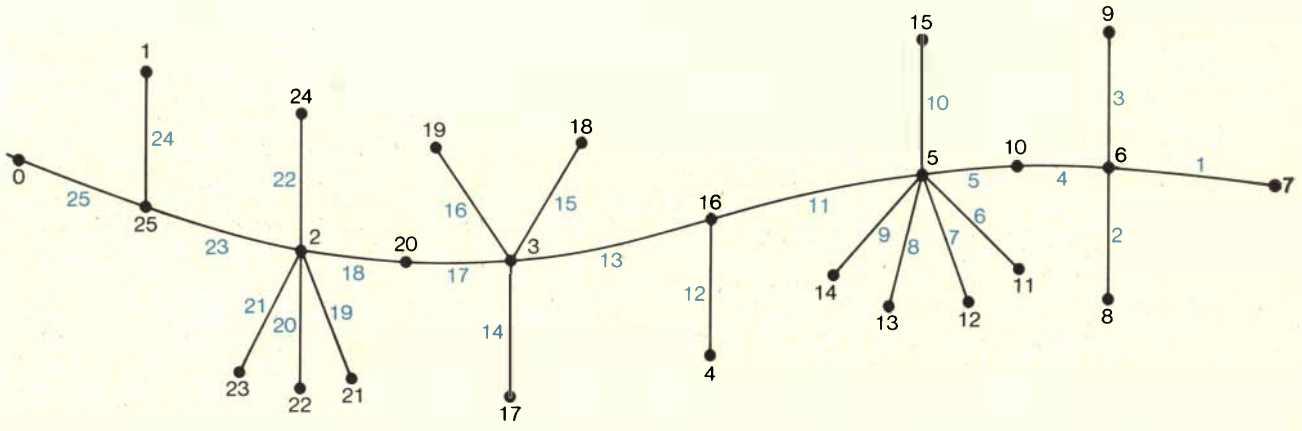
Skeletons of polyhedrons can be represented as planar graphs known as Schlegel diagrams. Of the five Platonic solids only the dodecahedron and icosahedron have not been shown to be graceful. We have seen how to gracefully number the tetrahedron. Can the reader gracefully number the Schlegel diagrams of the cube and octahedron [*see third illustration from top at left*] before Golomb's labelings are given next month? Can he do the same for the diagram of the skeleton of the Great Pyramid of Egypt? Can he discover graceful numberings for the dodecahedron or the icosahedron?

Three other graceful graphs by Golomb have six, seven and 10 nodes [*see bottom illustration at left*]. Can the reader number these also before the solutions are given next month?

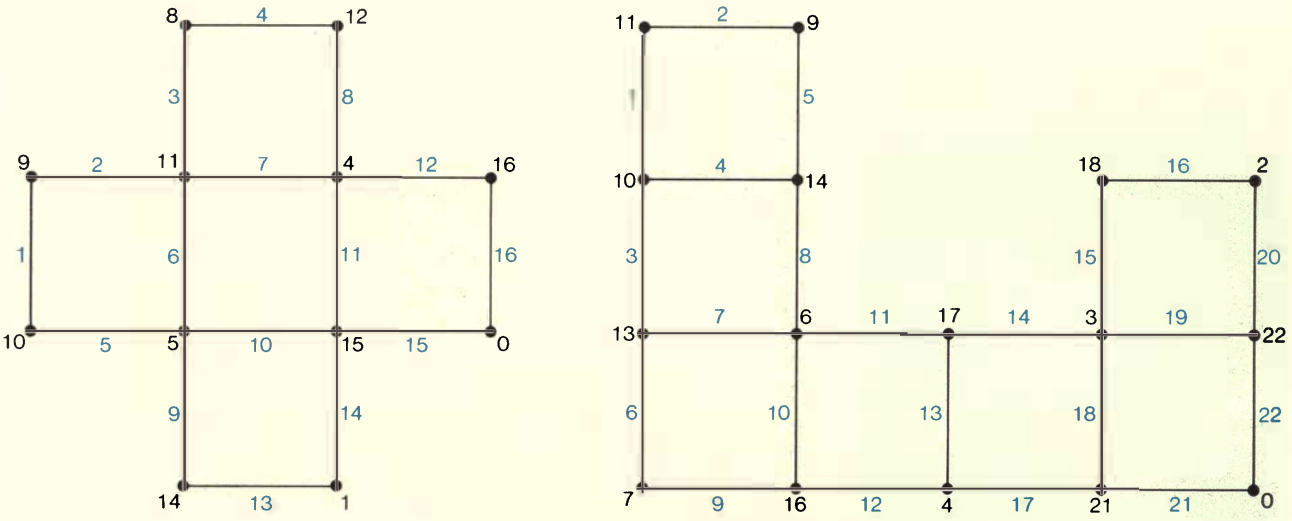
In addition to complete bipartite graphs there are other infinite families of graceful graphs. One found by Golomb



An infinite family of graceful graphs



A graceful caterpillar



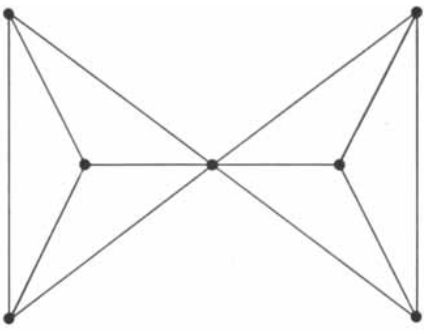
Graceful polyominoes

is shown in the top illustration above. The question arises: As the number of nodes approaches infinity, does the fraction of graceful graphs among all graphs of n nodes approach a limit? If so, what is the limit? For several years no fractional value from 0 through 1 was excluded, but recently Paul Erdös has been able to show that the limit is 0. His proof, not yet published, is difficult.

Although many unsolved problems about graceful graphs, some very technical, have now been cleared up by Golomb, Erdös and others, several major questions remain unanswered:

1. What are the necessary and sufficient conditions for a graph to be graceful? It is not even known if all tree graphs are graceful. (Tree graphs were discussed in this department in February, 1968.) Gerhard Ringel in 1963 apparently was the first to conjecture, in a different terminology and independently of Golomb's work, that all tree graphs can be gracefully numbered. This has been the subject of several papers by Alexander Rosa and other Czechoslovakian mathematicians. The conjecture has been established only for special kinds of trees such as "caterpillars": trees with every node on a central stalk or only one edge from the stalk [*see middle illustration above*]. In a typical gracefully numbered caterpillar the edge numbers run consecutively from one end of the tree to the other.

Golomb has discovered a similar algorithm for gracefully numbering an infinite class of polyomino graphs such as the pentomino and the heptomino [*see bottom illustration above*]. Note how the



A curiously ungraceful graph

consecutive numbers run diagonally upward, from left to right. Unfortunately there is an infinite class of polyominoes with a greater degree of concavity (the degree is not easy to define) for which this procedure fails even when they can be gracefully numbered.

A simple graph found by Golomb [*see illustration above*] is particularly ungraceful because it is not ruled out by any known general theorem.

2. What are the rules for forming Golomb's triangle? Put another way, is there a general algorithm for finding the shortest rulers that correspond to the best ungraceful numbering of a complete graph for more than four points?

3. Is there a graph that, when numbered as gracefully as possible, violates the conjecture that on all such graphs the highest node number and the highest edge number are equal? Golomb is now searching for a counterexample: a graph with the best numbering but with a highest node number that exceeds the highest edge number. (It cannot be the other way around.) "If I find one," Golomb writes in a letter, "the graph will not only be ungraceful but downright *disgraceful*."

The six questions raised last month have the following answers:

1. The year with the largest time span between itself and its inverse is 1066, the date of the Norman Conquest. The span: 9,901 − 1,066 = 8,835.

2. The only state capital that does not share a letter with its state is Pierre, S.D.

3. The one letter not in the name of any state is *Q*.

4. The only heteroliteral day-month pair, aside from Friday-June, is Sunday-October.

5. The single letter not in the name of any number from 0 through 100, but in all number names greater than 100 and less than 200, is *A*.

6.
$$1 + [(9 + 7)/2] = 9$$
$$-1 + [(9 + 7)/2] = 7$$
$$1 + [(9 - 7)/2] = 2$$

Vacation in the 18th Century.

The change will do you good.

It will take you back to a time as chaotic, yet far more courtly than our own. A time when so many new social, political, and scientific ideas burst forth on our world that calling those years "The Enlightenment" is clearly an understatement.

That's why we invite you to pick any three books about the 18th Century (or indeed any three listed on this page) as your introduction to The History Book Club.

You get all three for only 99¢ when you choose a fourth at the low member's price. Then you simply take four more new books of history or world affairs during the coming year from the 150 offered each month. You get these at savings of $2, $3, or $4 off bookstore prices. And you also earn other savings through bonus books you select from the Club's entire list.

Mail the coupon. Choose a vacation that gives you distance, without the cost that normally entails.

First price is publisher's list.
Boldface shows member's price.

857. The Rise and Fall of the Habsburg Monarchy
By Victor-L. Tapié. $15.00/**$8.95**

872. America at 1750: A Social Portrait
By Richard F. Hofstadter.
His last book. $6.95/**$5.80**

832. China & Russia: The 'Great Game'
By O. Edmund Clubb.
Three centuries of "massive contest" —always short of war. $12.95/**$8.75**

841. American Counterpoint: Slavery and Race in the North-South Dialogue
By C. Vann Woodward. $7.95/**$6.75**

837. The Age of Energy
By Howard Mumford Jones.
U.S. society from 1865-1915, a highly original interpretation by the noted cultural historian. $12.50/**$8.40**

775. The Rising Sun: The Decline and Fall of the Japanese Empire, 1936-45
By John Toland. $12.95/**$8.95**

708. Jefferson and the Rights of Man
By Dumas Malone. $10.00/**$7.95**

797. The War of American Independence: Military Attitudes, Policies and Practice
By Don Higginbotham. $12.95/**$8.95**

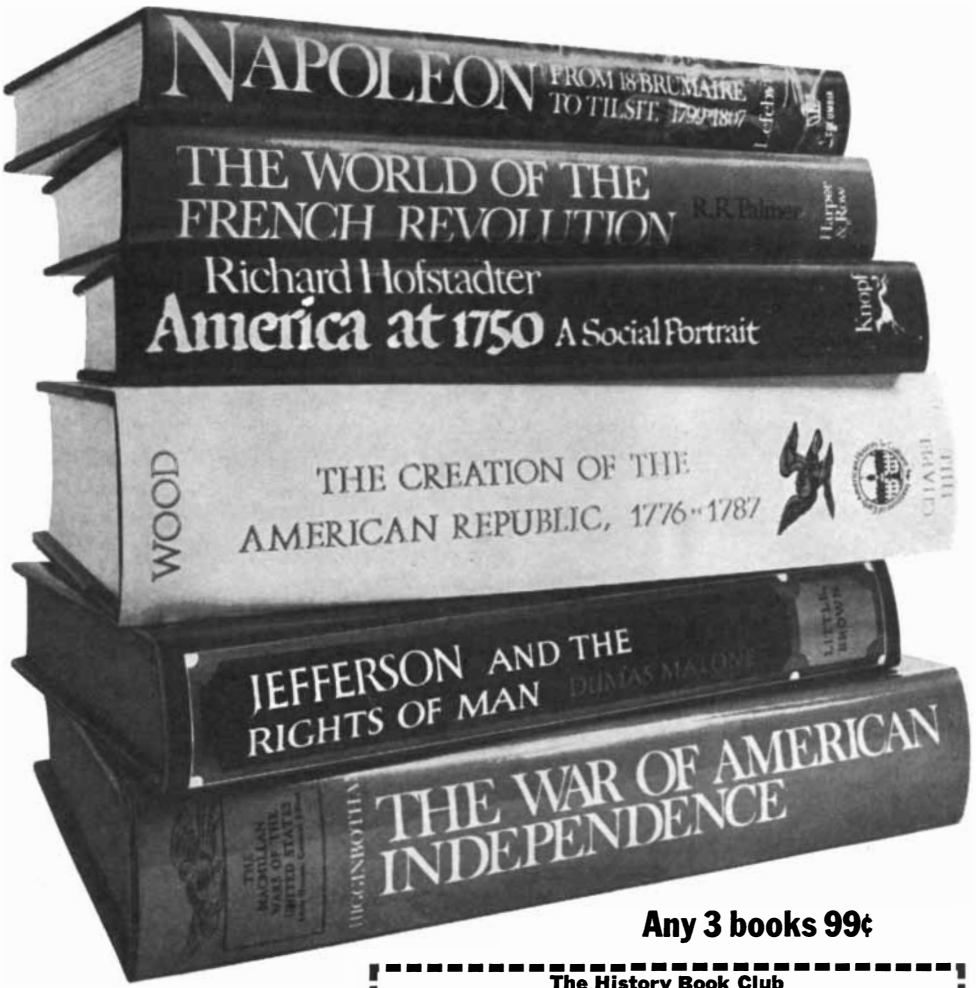
771. The World of the French Revolution
By R. R. Palmer. $8.50/**$6.75**

746. Empire of the Steppes: A History of Central Asia
By René Grousset. $17.50/**$9.95**

687. Anti-Intellectualism In American Life
By Richard Hofstadter.
Pulitzer Prize winner. $7.95/**$5.95**

677. Witchcraft at Salem
By Chadwick Hansen. $6.95/**$5.75**

671. Medieval History: The Life and Death of a Civilization (2nd Edition)
By Norman F. Cantor. $10.95/**$8.50**



Any 3 books 99¢

657. The Creation of the American Republic, 1776-1787
By Gordon S. Wood.
Bancroft Award winner. $15.00/**$9.95**

606. Henry VIII
By John J. Scarisbrick. $10.95/**$7.65**

476. The Life of Lenin
By Louis Fischer.
Winner of the National Book Award. $12.50/**$8.50**

791. Napoleon: From 18 Brumaire to Tilsit, 1799-1807
By Georges Lefebvre. $10.00/**$7.50**

369. The Spanish Civil War
By Hugh Thomas. $12.50/**$8.50**

767. Conquest of the Incas
By John Hemming. $12.50/**$8.75**

The History Book Club
Stamford, Connecticut 06904

Please enroll me as a trial member and send me, for only 99 cents, the three books whose numbers I have filled in:

[ ] [ ] [ ]

Also send me, at the special low member's price: [ ]

I enclose no money now. Send bill with books.

Within two weeks, I may return the books and owe nothing. Or keep them and agree to take only four more selections in the coming year from among more than 150 offered each month in the Club Review. I may choose any of these instead of the Editors' Choice, or no book at all, by returning the reply card accompanying the Review. SA 07S

Print Name _____

Address _____

City _____ State _____ Zip _____

A small shipping charge is added to each order. Residents of Canada: Please mail coupon to: The History Book Club, 16 Overlea Blvd., Toronto 17, Ontario

The History Book Club
Stamford, Connecticut 06904

THE AMATEUR SCIENTIST

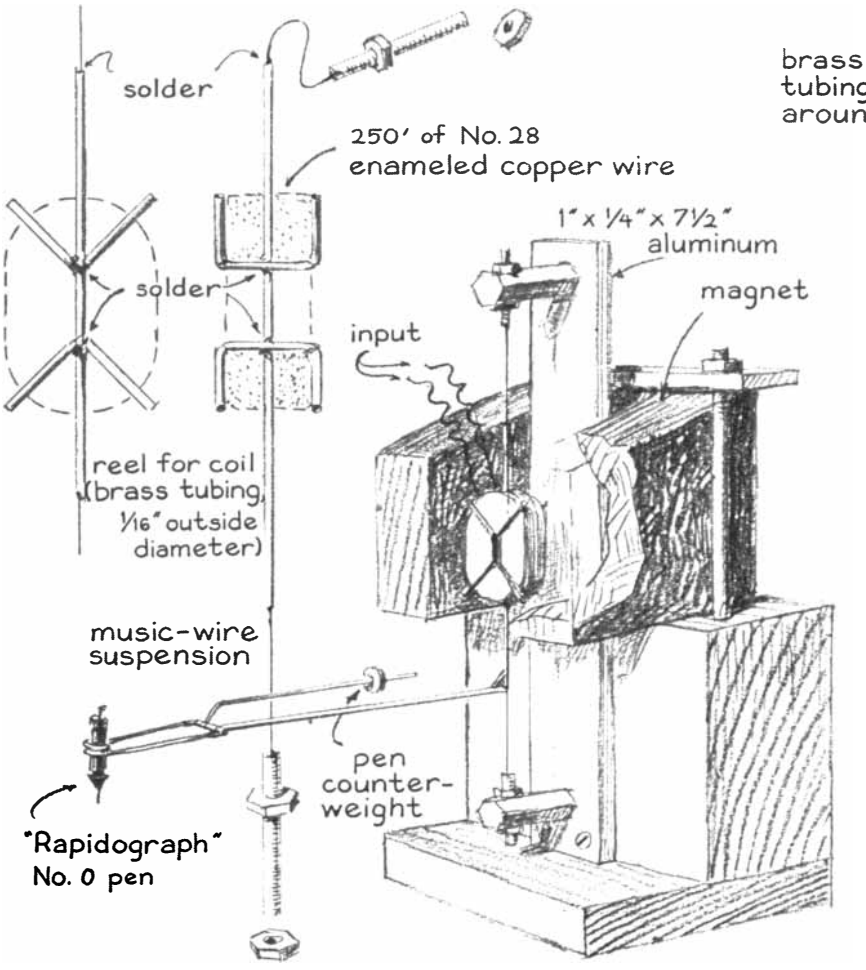
A simple motor with compensating devices is the key to a homemade chart recorder

Conducted by C. L. Stong

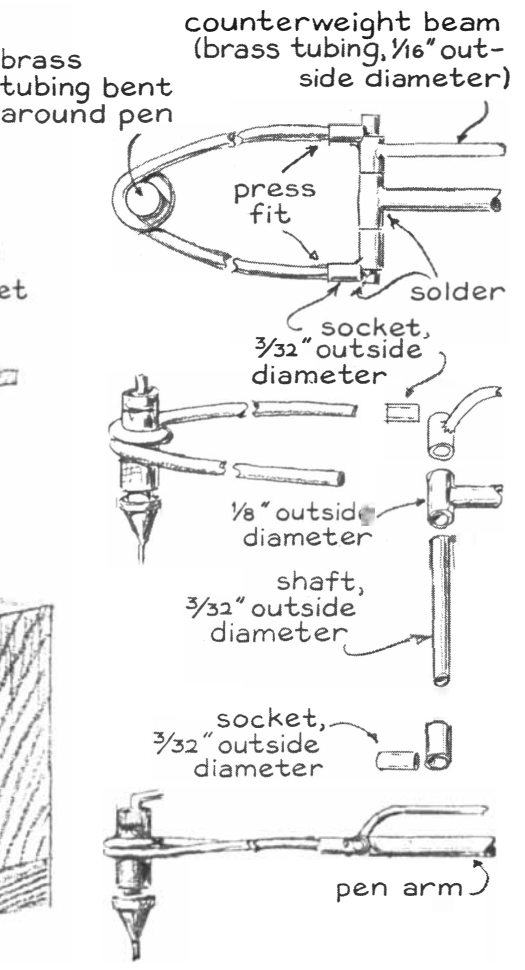
Chart recorders capable of registering signal frequencies of more than five cycles per second are difficult for amateurs to build. Such instruments are found in electrocardiographs, electroencephalographs and various kinds of machines that record telemetry signals. The response of recorders to frequencies of more than a few cycles per second has traditionally been achieved by minimizing both the mass and the friction of the moving parts. The parts of the motor that drive the pen weigh only a few grams, and the supporting shafts turn on accurately wrought pivots or on jewel bearings. Making such mechanisms requires the skills of a watchmaker.

J. Barry Shackleford, a simulation analyst with the Computer Sciences Corporation in Huntsville, Ala., has turned to an alternative scheme for achieving the desired high performance. He built a relatively crude pen motor that can be made inexpensively at home, and he applied power to it with transistor amplifiers that include feedback circuits to compensate for the mechanical deficiencies of the motor. Although the performance of Shackleford's recorder does not equal that of most commercial instruments, it is adequate for experiments of many kinds. Shackleford describes the construction as follows:

"Chart recorders consist of three principal subassemblies: a motor that operates the pen, a transport mechanism that advances the paper chart under the pen at a known rate and an electrical circuit that applies signal current to the pen mo-



J. Barry Shackleford's torsion-spring pen motor



Details of hinge assembly

tor. I made two motors that are structurally similar but employ different measuring devices: a torsion spring in one case and an electronic feedback mechanism in the other.

"Both motors have a coil of fine copper wire that is free to rotate about its vertical axis between the poles of a permanent magnet. The coil of one of the motors is supported at its top and bottom by a fine music wire in tension; the wire acts as a torsion spring [*see illustration at left on opposite page*]. An electron current applied to the coil generates a magnetic field.

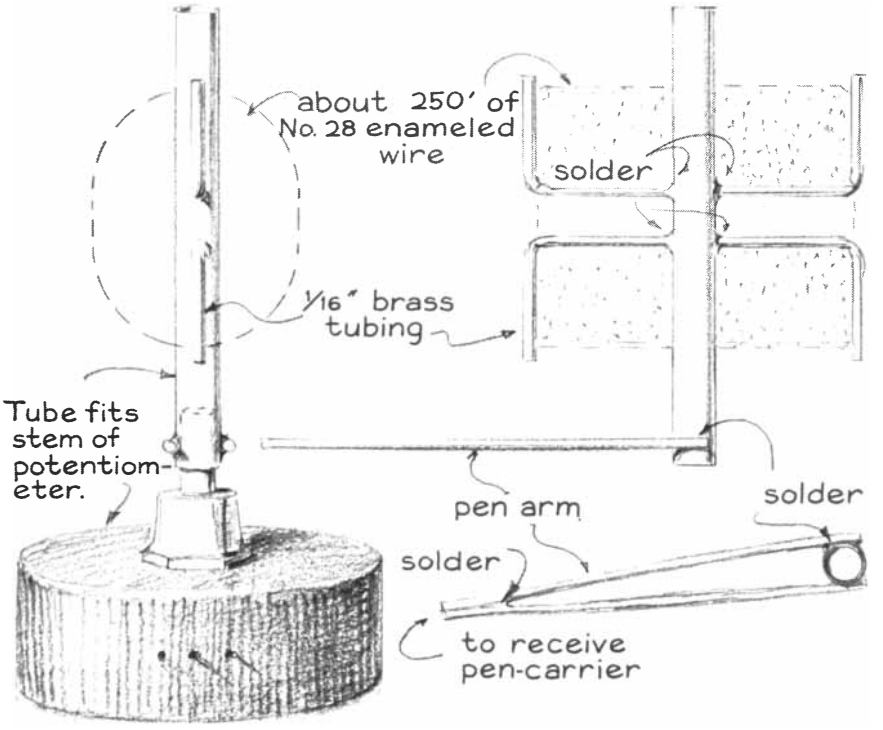
"The torsion fiber is adjusted initially to suspend the coil in the position in which the induced magnetic field makes a right angle with the field of the magnet. The induced field interacts with the permanent field to create a mechanical force that tends to rotate the coil to the position where the directions of the two fields coincide. The torque increases in proportion to the electron current in the coil. The torsion fiber exerts a countertorque that increases in proportion to the excursion of the coil from its zero, or initial, position. The motor is essentially a galvanometer of the type developed in the 19th century by the French physicist Arsène d'Arsonval. It is similar in principle to the motors in most voltmeters and ammeters except that a torsion fiber has been substituted for the pivots, shaft and hairspring of conventional meters.

"No advanced skills are needed for making the motor. The form on which the coil is wound consists of a framework of 1/16-inch brass tubing. The arm that supports the pen is also made of this tubing, which weighs about a third of a gram per inch. A two-inch length of tubing forms the spine of the framework.

"To the spine I solder a pair of *U*-shaped brackets that face in opposite directions. The coil is wound over the brackets. The brackets are an inch wide and 3/8 inch deep. Brass tubing of this diameter is available in hobby shops, particularly those that stock model-airplane supplies.

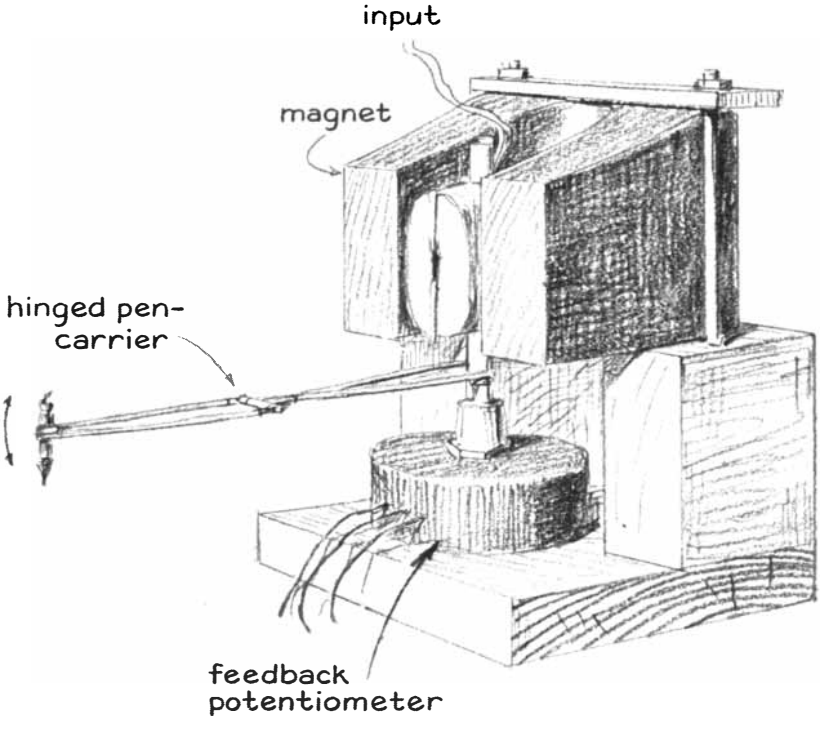
"Cut the tubing by placing it on a flat piece of cardboard and sawing it with the sharpest knife in the house. Let the tubing roll freely under the knife so that the resulting groove encircles the brass. The wall of the tubing is thin. Use moderate pressure to avoid collapsing the tube. You will have to resharpen the knife afterward. Burrs can be removed from the open ends of the tubing with a small twist drill.

"I join the parts with solder. To make the joint with a minimum of solder I first clean the brass at the joint and tin each part at this point with a small soldering iron. After the parts are assembled and clamped in the desired position I place a small bead of solder on the tinned parts and heat the joint with a propane torch adjusted to produce the smallest possible flame.



Arrangement of the coil for the servomotor

"Having assembled the framework, I add the torsion fiber, which is a straight length of music wire about .030 inch in diameter (Brown and Sharp gauge No. 20). The *E* string of a violin would do. Thread the wire through the spine of the framework and solder it in place at the



General view of the servomotor

ends of the spine. Use an acid flux and little solder to make the joint. Heat with a small soldering iron to avoid spoiling the temper of the wire. Remove excess acid flux with rubbing alcohol and apply a thin coat of plastic cement to the brass.

"The coil is wound with 250 feet of 28-gauge enameled copper magnet wire. I wrap the wire on the framework randomly, making no attempt to apply uniform layers. The resistance of the resulting coil is about 15 ohms. The ends of the coil terminate in flexible leads of 24-gauge wire. Solder the inner flexible lead to the end of the magnet wire, insulate the joint with a dab of plastic cement and tie it to the spine with silk thread before winding the coil. Terminate the outer end of the coil similarly and with thread lash both leads to the spine about 1/8 inch from the winding. The coil assembly will weigh some 25 grams.

"The arm that supports the pen is a length of 1/8-inch brass tubing. Solder the arm to the spine parallel to the axis of the coil. Add a small brace of the same tubing to stiffen the joint.

"Incidentally, the frequency response of the pen motor varies inversely with the length of the arm. Arms up to eight inches long can be used to record frequencies of up to two or three cycles per second, such as earthquake waves and microseisms. For recording heartbeats or brain waves the length of the arm from the center of the spine to the nib of the pen should not exceed three inches, including the hinge assembly that supports the pen.

"The hinge is made of brass tubing of two diameters: 1/8 inch and 3/32 inch [*see illustration at right on page 114*]. These sizes telescope to make a sliding fit. A 1/4-inch length of the 1/8-inch tubing is soldered at a right angle to the outer end of the pen arm. This tubing serves as the bearing of the hinge. A 5/8-inch length of the 3/32-inch tubing is slipped through the bearing and centered to form a shaft. Two collars, each 3/16 inch long, are cut from the 1/8-inch tubing and slid over the protruding ends of the shaft. A yoke for supporting the pen is bent from 1/16-inch tubing.

"The frequency response varies inversely with the inertia of the writing system. I prefer a light pen of the capillary type, such as the No. 0 or No. 1 Rapidograph. These pen tips are available from dealers in drafting supplies. The tip has a cylindrical ink reservoir that terminates in a cone from which a capillary tube extends to the paper. The flow of ink is regulated by a fine wire that extends through the capillary to the paper.

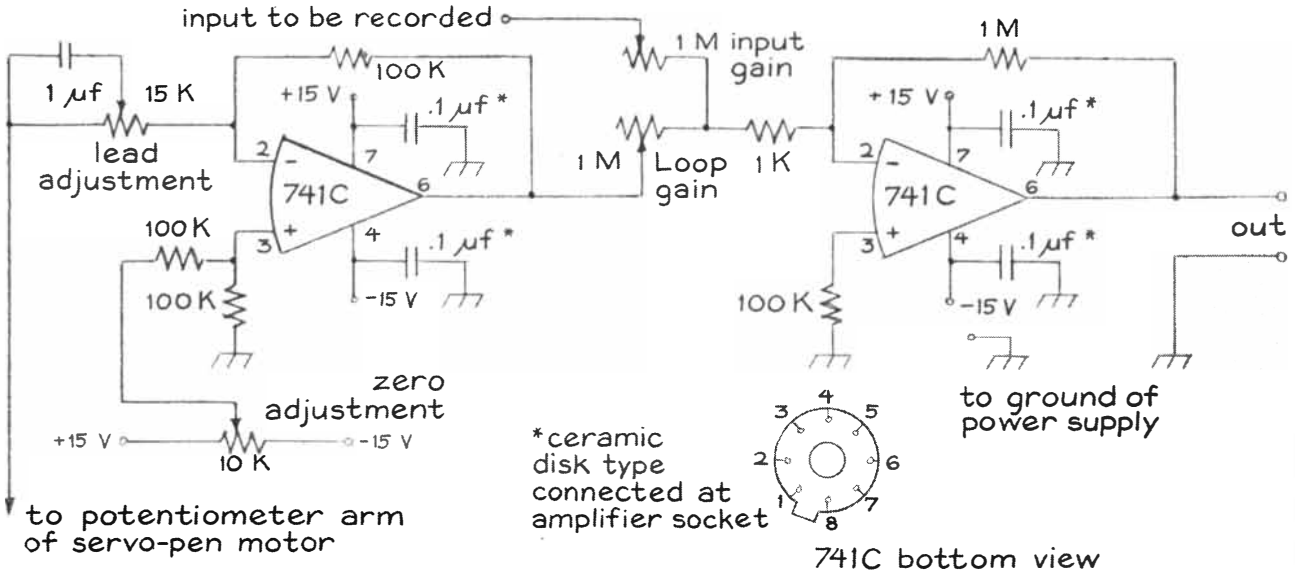
"The top of the wire terminates in a cylindrical weight that adds inertia to the system. I cut off and discard the weight, slip the wire into the end of a length of 1/16-inch brass tubing and fasten it by crimping the tubing. The upper end of the tubing is bent at a right angle to rest on the upper edge of the ink reservoir. The bend is made at a point such that the lower end of the wire extends to the tip of the capillary. The substitution of brass tubing for the cylindrical weight not only reduces the inertia of the system but also increases the ink capacity of the reservoir.

"The yoke that supports the pen is made by wrapping two turns of brass tubing around the reservoir. The ends of the tubing extend at an acute angle from the loop thus formed. At the point where they are separated about 1/2 inch I bend the ends parallel and slip them into the sockets of the hinge as shown in the illustration. Roger Hayward, who illustrates this department, suggests the substitution of a short strip of thin sheet metal, such as brass shim stock, for the hinge. The flexible strip would be lighter than the hinge and easier to make. I intend to try it.

"The coil and the pen arm are supported by a vertical bracket that has a facing pair of adjustment screws to which the torsion wire is attached. The wire passes through a small hole near the end of each screw and is soldered in place. The screws are tightened to exert a pull of about 10 pounds on the wire. The permanent magnet is clamped to a block of wood fixed to a wood base.

"The coil is centered between the poles at the most intense region of the magnetic field. The field intensity can be examined visually by covering the magnet with a sheet of cardboard and dusting the surface with iron filings. With the magnet clamped in position, rotate one of the screws attached to the torsion wire a half-turn. Rotate the other screw in the opposite direction until the axis of the coil makes a right angle with the magnetic field. Tighten the nuts. This adjustment stresses the wire in both torsion and tension.

"My instrument has a surplus magnetron magnet that is available from the Edmund Scientific Co., Barrington, N.J. 08007 (the catalogue number is 70,571). The gap between the poles is 1¼ inches and the width is 1¾ inches. The strength



Circuitry of the differential amplifier and the summing amplifier

of the field is rated at 2,075 gauss. The pen motor can be operated from any amplifier that develops 15 watts and has an output impedance of approximately 15 ohms.

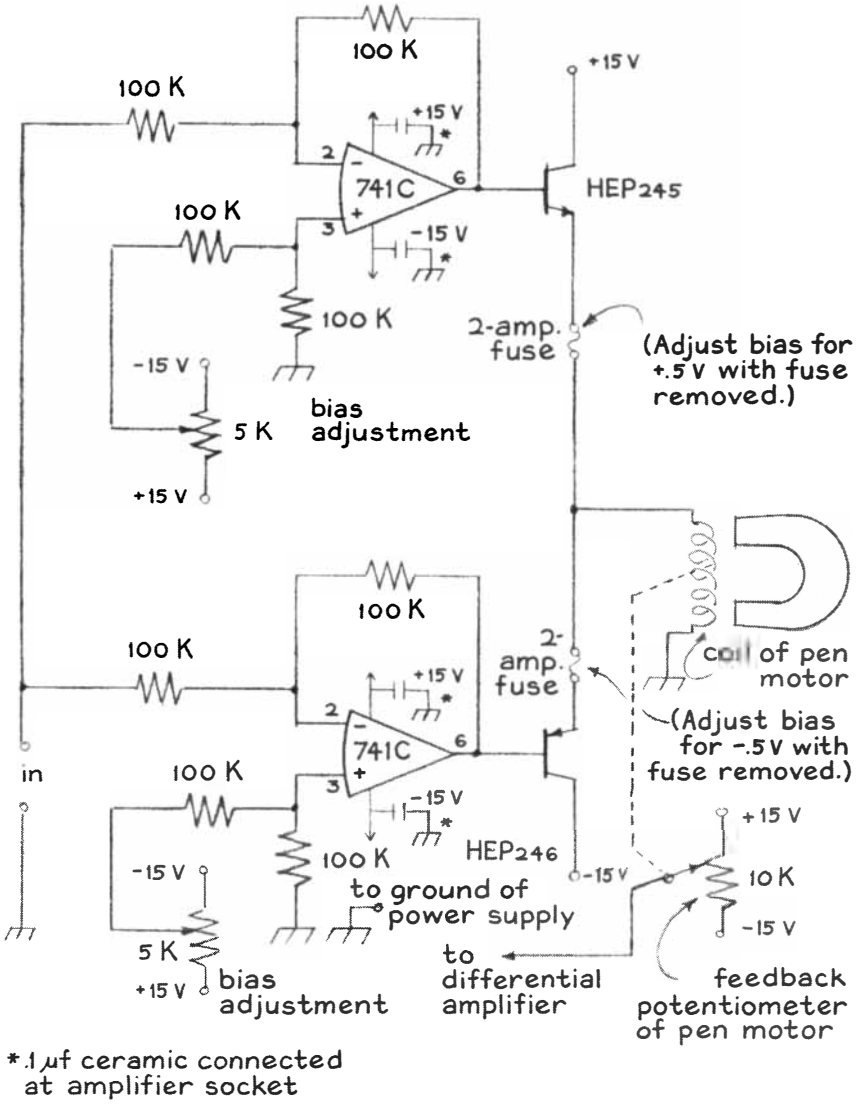
"My second motor is similar to the d'Arsonval type but develops much greater torque. It is of the servo type: the coil is mounted on the shaft of a potentiometer that develops a reference voltage. The reference voltage is fed back to the input of a differential amplifier. The feedback system replaces the torsion springs of the d'Arsonval instrument and stabilizes the performance. For example, variations in friction between the pen and the paper have little effect on the response of the system.

"The potentiometer must be of the low-friction type, preferably made with ball bearings. A suitable one is the Helipot 6502, which is rated at 10,000 ohms. The resistance of the potentiometer must not exceed 20,000 ohms. Precision single-turn potentiometers of the required kind are available from American Design Components, 39 Lispenard Street, New York, N.Y. 10013.

"A two-inch length of brass tubing that makes a tight fit with the shaft of the potentiometer becomes the spine of the framework on which the coil is wound [*see top illustration on page 115*]. *U*-shaped brackets of 1/16-inch brass tubing are soldered to the spine, as in the framework of the d'Arsonval pen motor. The framework is insulated with plastic cement, wound with an identical coil and supplied with a pen arm. The potentiometer is mounted shaft end up to a wood base. The permanent magnet can be clamped to a block of wood attached to the base.

"Connect the potentiometer to a power source of about 15 volts. Connect a voltmeter between a terminal of the source and the wiper arm of the potentiometer. Rotate the shaft of the potentiometer to the position where the meter indicates exactly half the source voltage. Without disturbing the position of the shaft, slide the spine of the coil assembly over the shaft so that the axis of the coil is at a right angle to the magnetic field.

"Although the pen servomotor is more powerful than the d'Arsonval type, it cannot drive a high-friction writing device such as a pencil or a crayon. Pen tips of hard felt or porous plastic work well. It pays in terms of frequency response, however, to minimize the mass of the assembly, particularly at the outer end of the pen arm. When I use a felt pen, I cut away most of the body and save only the tip and the ink reservoir.



Circuit diagram of the power amplifier

A counterweight is not ordinarily required in the pen servomotor.

"Functionally the electronic circuit of the recorder includes three amplifiers that are interconnected to form a closed loop. The signal to be recorded is applied to one input terminal of a summing amplifier [*see illustration on opposite page*]. The summing amplifier can accept two input signals. It multiplies their algebraic sum. The output of the summing amplifier is fed to the input of the power amplifier. The power amplifier drives the pen motor. The potentiometer, which is linked mechanically to the shaft of the motor, develops a voltage that varies in amplitude and polarity with the position of the shaft.

"The potentiometer voltage, which is known as the feedback voltage, is applied to one input terminal of a differential amplifier. The differential amplifier can accept two input signals. It multiplies their algebraic difference. The output of the differential amplifier is applied to the second input terminal of the summing amplifier, thereby closing the loop.

"An adjustable reference voltage developed by a potentiometer connected to the power supply is fed to the second terminal of the differential amplifier. In the absence of a signal to be recorded the amplified difference between the feedback voltage and the reference voltage is applied through the summing amplifier to the power amplifier. Energized in this way, the pen motor rotates in a direction that reduces the difference between the feedback voltage and the reference voltage to zero, thereby stopping the motor. If the reference voltage is altered manually, the resulting difference voltage again causes the motor to seek a position at which the difference is reduced to zero. Thus the pen of the recorder can be positioned at any desired

point on the chart by adjusting the reference voltage.

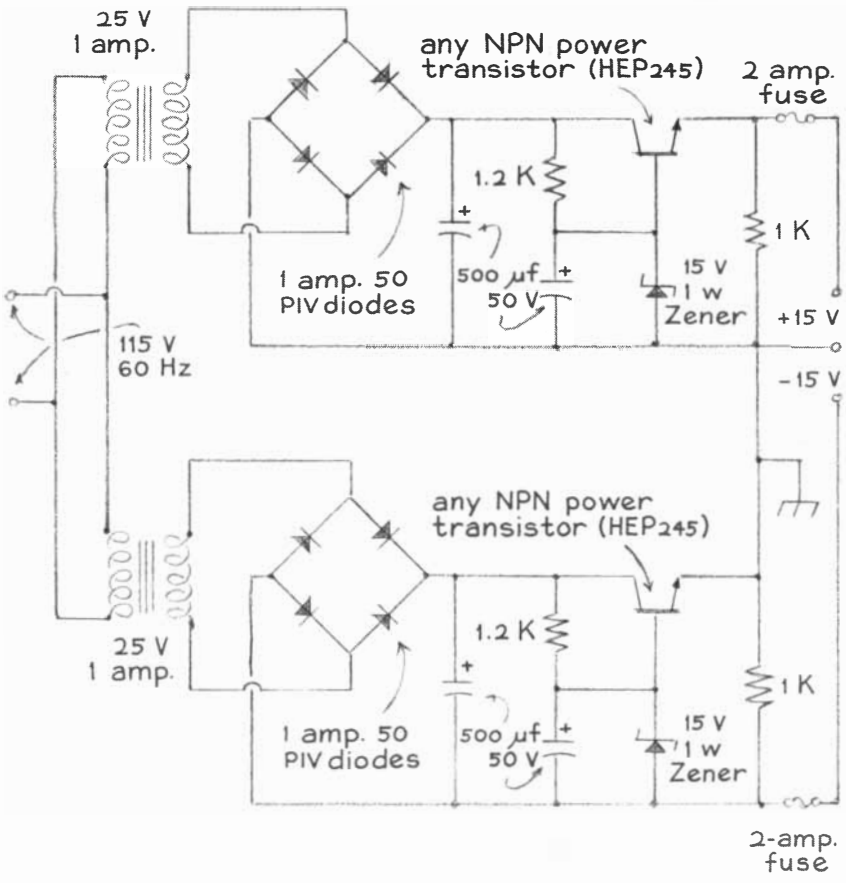
"A signal applied to the summing amplifier similarly energizes the power amplifier. The motor then seeks a position where the sum of the signal voltage and the difference voltage falls to zero. Voltages in the feedback loop are automatically amplified, with the result that the system tends to apply as much power as the motor requires to move the pen, regardless of variations of friction between the nib and the paper.

"Although the system has 82 transistors, the construction is fairly simple. All but two of the transistors are included in four operational amplifiers of the integrated-circuit type. The amplifiers are only slightly larger than the eraser of a pencil. They are available on the surplus market for about $1 each from suppliers such as Poly Paks, P.O. Box 942, Lynnfield, Mass. 01940.

"The electronic circuits should be made and tested one at a time. Make the regulated power supply first. *Do not omit the fuses.* The voltage of the power supply should not vary more than about 1 or 2 percent when a 30-ohm resistor is connected to the output. The resistor should be capable of dissipating about 30 watts.

"Next, build the power amplifier. Check the completed wiring against the schematic diagram [*preceding page*] at least twice. Remove one of the fuses, apply power and, with a voltmeter, adjust the corresponding bias potentiometer to deliver a potential of .5 volt with respect to ground. Replace the fuse and similarly adjust the bias of the complementary circuit.

"Make an operational test of the amplifier by connecting the output to a dummy load and applying power. I use for the load a small direct-current motor that is rated at 30 volts. Connect a 5,000-ohm potentiometer to the power supply and wire the arm of the potentiometer to the input of the amplifier. If the amplifier is operating properly, the speed of the motor and the direction of its rotation can be controlled by twisting the knob of the potentiometer. Incidentally, all power transistors in both the power supply and the power amplifier must be mounted on heat sinks that have large cooling fins. Without adequate heat sinks the power transistors will quickly burn out.



Arrangement of the power supply

"Finally, assemble the differential-amplifier circuit and the summing-amplifier circuit. Do not apply power to this unit until you have triple-checked the completed wiring against the schematic diagram. Failure to make the check can cost you amplifiers, as I have learned from sad experience.

"Test the amplifiers by connecting a voltmeter to the output and applying input voltage by means of a potentiometer wired as a voltage divider. After completing the checks connect the output of the summing amplifier to the input of the power amplifier, connect the pen motor to the output of the power amplifier and connect the arm of the feedback potentiometer to the lead adjustment of the differential amplifier. Center all potentiometers except the bias adjustments. All ground terminals of the system must be interconnected, including the power supply.

"Apply power to the system. Operate the zero-adjustment potentiometer that provides the reference voltage. If the pen cannot be centered on the chart by operating this control, it is probable that the polarity of the potentiometer is reversed. Interchange the connections of the potentiometer and the power supply. Center the pen. With the power on, push the pen from the zero position with your finger and release it. The pen should promptly return to the zero position and stop without overshooting. If it does not return promptly, increase the loop gain to the point at which the pen oscillates and then reduce the gain slightly. Flick the pen. It should overshoot the zero position and oscillate a few times.

"Next, operate the lead adjustment in the direction that arrests the oscillation. In effect, the lead adjustment gives the system the ability to second-guess what is going to happen next. Actually it has the effect of increasing the loop gain as the frequency of the feedback voltage increases and of altering the phase relations of currents in the network.

"The proper relative setting of the lead adjustment and the loop-gain adjustment must be found by experiment. The optimal settings may require considerable tinkering. When the adjustments have been made properly, the pen will snap back to zero and stop without overshooting when it is pushed from the zero position and released.

"This circuit was assembled from components that I happened to have at hand. The values of the resistors were determined experimentally. It is quite possible that the selected values are not optimal in every case. Doubtless the

system can be improved by further experimentation.

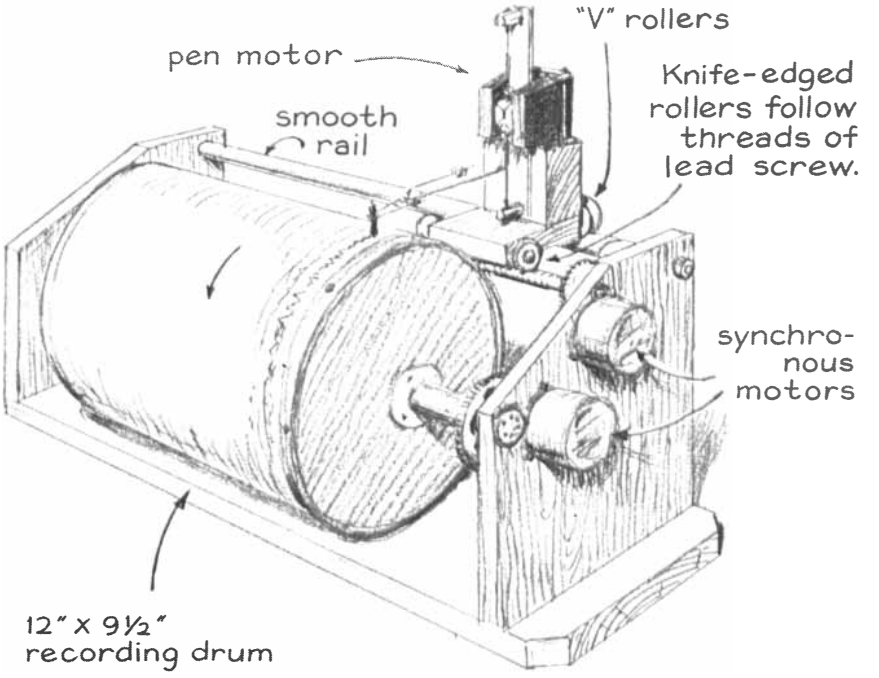
"Paper-chart transport mechanisms of two types have been built. One of the two carries a chart in the form of a paper cylinder. I use this unit for recording seismograms. The second mechanism pulls a strip of adding-machine tape under the pen. It is used primarily for recording transient signals of higher frequency such as bioelectric potentials. Both units can be made with ordinary hand tools.

"The drum that supports the cylindrical chart was made from a length of phenolic tubing of the kind used for telescope mountings. The material is 1/4 inch thick. It is available from the Edmund Scientific Co. (the catalogue number is 85,146). The ends were closed with disks of plywood. The axle on which the drum turns should be centered as accurately as possible, even though the hinged pen will follow departures from a true circle. Strips of chart paper are attached to the drum with adhesive tape.

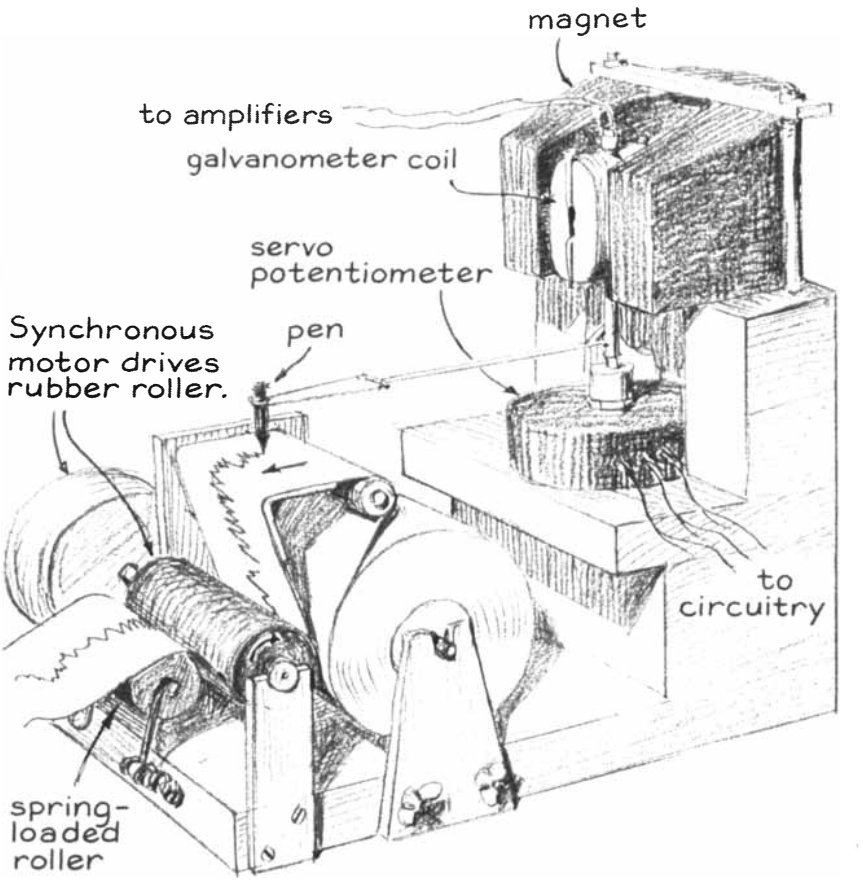
"The lead screw that advances the motor carriage across the drum is a long stud bolt of the kind stocked by most dealers in hardware. The drum and the lead screw can be rotated independently by synchronous motors or by belts or gears according to the requirements of the experimenter. For recording earthquakes I prefer synchronous motors [*see top illustration at right*].

"The transport mechanism for pulling adding-machine paper is somewhat smaller than the drum arrangement and is easier to make. The writing table consists of sheet steel bent to an acute angle and soldered at one edge to an upright plate attached to the base [*see bottom illustration at right*]. Paper is pulled across the writing table by a rubber roller. The roller is made of heavy-wall rubber tubing slipped over a shaft. A spring-loaded roller, also made of rubber tubing, increases friction between the paper and the driving roller.

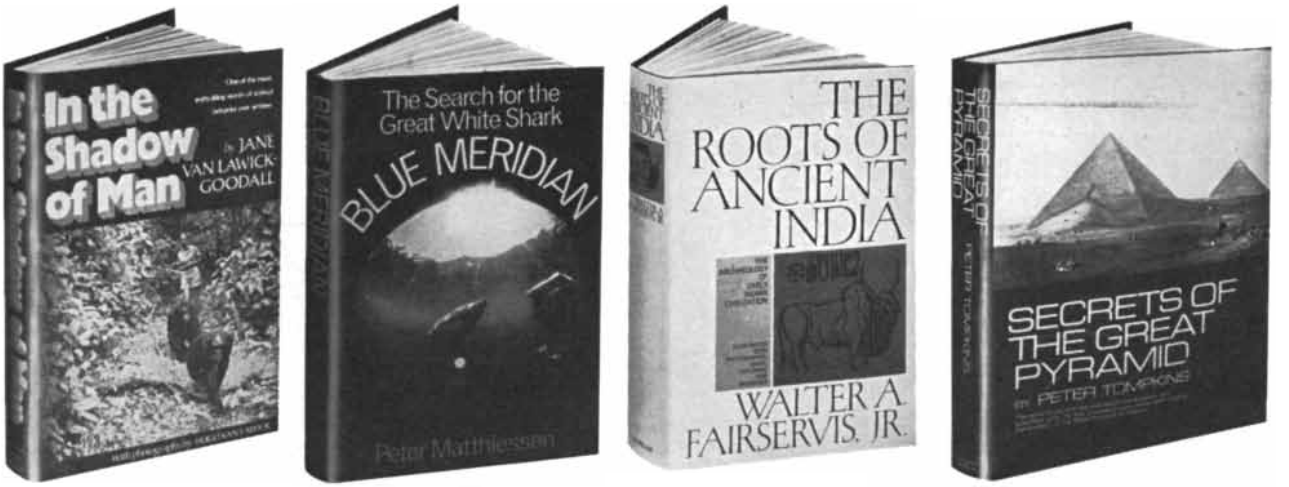
"Usually the experimenter is interested in timing the recorded signal. I accomplish this by driving the paper transport with a synchronous motor, an arrangement that automatically provides the graph with a known time base. The experimenter can substitute a variable-speed motor and record time marks and the signal simultaneously. This task can be done by feeding time signals to one input terminal of a summing amplifier and the data signal to the second input terminal. The combined signals would be fed to the input of the recorder."



The drum recorder



Recorder with chart in strip form



Join up now for the greatest reading adventure in the whole natural world.

Come. Spend a lively year with the wild chimpanzees. Follow the flashing trail of the great white shark. Walk the earth with the living Buddha. Explore the secrets of the Great Pyramid.

Just part of the whole, wide, wild, fantastic world of high adventure you will explore as a member of the Natural Science Book Club. Exciting reading. Rewarding reading. Important savings, too.

An 8¢ stamp is all it takes to see your first three books. If not delighted, return them and owe nothing. We're betting that once you start your fascinating journey, you won't want to stop!

Take any 3 books
(values to $42.45)
for only 99¢ each
with a trial membership in the
Natural Science Book Club
if you will agree to accept only 3 more selections or alternates in the next twelve months

(retail prices shown)

85060. ULYSSES AIRBORNE. *Mauricio Obregón.* The Odyssey relived by airplane. Filled with stunning pictures. **$8.50**

55440. INNOCENT KILLERS. *Hugo and Jane van Lawick-Goodall.* A tour in pictures and words into the little-known worlds of the African jackal, the wild dog and the spotted hyena. **$10.00**

69930. POPULATIONS, SPECIES, AND EVOLUTION. *Ernst Mayr.* The renowned zoologist and National Medal of Science winner has provided a sweeping exposition of the theory of evolution, complete with the latest refinements. **$10.00**

78410. SLAUGHTER THE ANIMALS, POISON THE EARTH. *Jack Olsen.* Unforgettable portrait of wildlife and its wanton destruction. **$6.95**

MEMBERSHIP APPLICATION 4-705

Natural Science Book Club
Riverside, New Jersey 08075

Upon acceptance of this order, please enroll me as a member and send the three books I have indicated. Bill me only 99c each for all three, plus postage and handling. If not delighted, I will return all books within ten days and this membership will be cancelled.

As a member, I need accept only three more selections during the next 12 months at reduced member prices, plus postage and handling. Savings range up to 30% and occasionally even more. I understand that I will receive free advance Reviews which fully describe each month's Main Selection and Alternates. If I wish the Main Selection, I need do nothing; it will come to me automatically. If I do not wish to receive the Main Selection, or I wish an Alternate Selection, I will advise you by using the convenient reply card provided with the monthly Reviews. I understand that I may choose a bonus book for every 4 selections purchased. (Introductory offer counts as first selection toward bonus book.)

Send no money. Members are billed when books arrive.

3 books for 99c each are: (write in numbers)

_____ _____ _____

Some expensive books count as 2 selections.

Name _____

Address _____

City _____ State _____ Zip _____

(Offer good in Continental U.S. and Canada only.)

36270. THE BIOLOGICAL IMPERATIVE. *Allan Chase.* A noted science writer examines present state of health services, medical education and national health in the U.S. **$8.95**

84900. THE TWILIGHT OF THE PRIMITIVE. *Lewis Cotlow.* What is happening to the Stone Age tribes in New Guinea, the primitives in the Mato Grosso, the Eskimos. A touching look at the changes taking place among the survivors of age-old cultures. **$10.00**

74340. THE ROOTS OF ANCIENT INDIA. *Walter A. Fairservis, Jr.* An incredibly rich development of Indian prehistory—from early Stone Age man to the days of the living Buddha—and the qualities than make it distinct from all the other ancient civilizations. *Counts as 2 of your 3 books.* **$17.50**

86510. THE WAY THINGS WORK. Vol. Two. Ski bindings, submarines, artificial kidneys—you name it—you'll find the inside, expert story about what makes everything tick. **$9.95**

78340. SHAPES, SPACE, AND SYMMETRY. *Alan Holden.* Explains, analyzes, and provides amazing insights into hundreds of mathematical solids ranging from the tetrahedron to the trapezoidal hexacontahedron. **$11.00**

37500. CHANCE AND NECESSITY. *Jacques Monod.* Presents a brilliant and revolutionary philosophy of life based on recent developments in molecular biology—in which the Nobel Laureate author played a major role. **$6.95**

64500. A NEW PHOTOGRAPHIC ATLAS OF THE MOON. *Zdenek Kopal.* Breathtakingly illustrated with more than 200 unique photographic plates, here is a lunar Baedeker that will allow you to explore the Moon first-hand. *Counts as 2 of your 3 books.* **$20.00**

37590. CHEROKEE TRAGEDY. *Thurman Wilkins.* The dramatic true story of the mysterious Ridge family—leaders of the Cherokee Nation; of their roles in the treaty that forced the Indians on the Trail of Tears—and of the brutal assassinations that followed. **$10.00**

85830. THE VIEW FROM SPACE. "Chock full of solid substance for the serious student of space photography yet has an historical vein and a pleasant conversational style. Timely, authoritative . . ."—James A. Van Allen. **$14.95**

36490. BLUE MERIDIAN. *Peter Matthiessen.* Powerful narrative of the search for the Great White Shark, the most savage creature on earth. **$8.95**

55620. IN THE SHADOW OF MAN. *Jane van Lawick-Goodall.* A fascinating account of some of the decade's most spectacular discoveries in animal research—adventures among wild chimpanzees. **$10.00**

62690. MICROBES AND MORALS. *Dr. Theodor Rosebury.* A fascinating and highly readable account of the cures, causes, history, and folklore of venereal disease. **$7.95**

68730. PHYSICS OF THE EARTH. *Frank D. Stacey.* A brilliant new volume in the science of geophysics incorporating the revolutionary changes in this most fundamental of sciences. **$11.95**

50110. GENETICS OF THE EVOLUTIONARY PROCESS. *Theodosius Dobzhansky.* The definitive work on the modern synthetic theory of evolution by the world's foremost geneticist. **$10.95**

45600. THE EROTIC OCEAN. *Jack Rudloe.* Everything marine enthusiasts need to locate, preserve, and identify oceanic specimens. This illustrated volume is an important contribution to conservation as well as a guide to scientific beachcombing. **$15.00**

77680. SECRETS OF THE GREAT PYRAMID. *Peter Tompkins.* Was the Great Pyramid of Cheops only a magnificent tomb? Or was it built to incorporate and hand down a mystical number system? Do its dimensions and structure hold the key to secrets about astronomy and the occult? **$12.50**

65210. OCEANOGRAPHY: An Introduction to the Marine Environment. *Peter K. Weyl.* Brings light to the mysteries of the ocean—the geology of the ocean basins, the chemistry of sea water, life beneath the sea and fundamental energy processes. Copiously illustrated. **$12.50**

52180. HAMLET'S MILL: An Essay on Myth and the Frame of Time. *Giorgio de Santillana and Hertha von Dechend.* A brilliant speculative inquiry into the origins of scientific thought in the preliterate world. **$10.00**

69270. A PICTORIAL HISTORY OF AMERICAN MINING. *Howard N. and Lucille L. Sloane.* The swindles and strikes, the colorful personalities and tragic disasters—are re-created in a meticulously researched, illustrated volume. **$12.50**

64580. THE NEW PROMETHEANS. *Robert S. de Ropp.* Perceives a continuing battle between the creative and destructive forces of science and applies this thesis to the great scientific developments of the recent past. **$7.95**

BOOKS

Standards and fundamental constants, and early European voyages to North America

by Philip Morrison

PRECISION MEASUREMENTS AND FUNDAMENTAL CONSTANTS: PROCEEDINGS OF THE INTERNATIONAL CONFERENCE OF THE NATIONAL BUREAU OF STANDARDS, 1970, edited by D. N. Langenberg and B. N. Taylor. National Bureau of Standards Special Publication 343, U.S. Government Printing Office ($6). Since the Enlightenment we have sought "philosophically true" bases for the units of measurement without which trade and manufacture could not grow. The savants of revolutionary Paris founded their meter not on the *pied du roi* but on the length of the earth's meridian. Pragmatism soon took over. Accurate comparisons were easier with a couple of scratches on a platinum bar than with any geodesy in the field; the grand idea behind the metric system was set aside for the humdrum utility of an arbitrary laboratory standard carefully made, painstakingly copied and safeguarded by international agreement. We live, however, in a world modularly constructed on the strictest rules; every atom is a storehouse of natural units safer than the Bureau International des Poids et Mésures in Sèvres. Building the world's units on atomic units is not a new idea; the rise of optical interferometry in the 1890's began a steady movement toward replacing the meter bar by the wavelength of some well-defined kind of light. Once single-isotope gases were available for spectral-line sources of ultimate reproducibility, the atomic standard could not be gainsaid. In 1960 the meter was internationally defined in terms of light waves emitted by a particular transition in the isotope krypton 86.

The past decade has seen the drift become a tide. The combined power, on the one hand, of experimental techniques such as atomic beams, radio-frequency spectroscopy and lasers and, on the other, of a detailed and deep theoretical understanding of atomic structure and of quantum electrodynamics has proved to be irresistible. The electronics of frequency division and our detailed knowledge of how atoms emit and absorb photons have given us frequency standards recognizably better than the earth's spin. With the invariant velocity of light we now have all but officially united the once independent standards of time and of length.

This large, timely volume—a great bargain from the Government Printing Office, although it is in a typeface few will enjoy—presents the papers of the first international conference of the new wave, held last year at the exurban site of our National Bureau of Standards. There are nearly 90 papers, plenty of verbatim discussion and even a shower of limericks (generally sophomoric). Although many of the papers are quite technical, demanding a sophisticated knowledge of quantum field theory, a score of them are reviews, with historical accounts, summaries of the state of the art, and projections. These alone will commend the volume to readers who enjoy the subtle and endless game of seeking precision and consistency in the magnitudes of nature. Particularly striking are the famous dimensionless numbers that are the modern sign of the Pythagorean insight that nature is ruled by number—say by 137.036, give or take a couple of parts per million. (There is even a neo-Pythagorean theory purporting to get that very number by a unique transcendental formula out of group theory, but it is not reported in this book.)

Frequency is still the quantity we measure best. The three techniques that contend for the palm of definition are the particle beam, its purified motion allowing the very detailed study of its own resonance absorptions; the storage vessel, holding the particles rattling around inside the resonance structure for a while, which reduces the fundamental time limitations on frequency definition, and the use of an absorber gas with an external beam of radiation. A jitter of less than a second in a million years is the 1970 performance of a laboratory-standard cesium-atomic-beam tube! With constant light velocity such frequency measurements imply length standards as well, and with Planck's constant knowing frequency means knowing energy too. Hence the standard volt is all but distributable by radio.

These physicists, a mixture of prudent metrologists who anxiously guard the meaning and the utility of the units and of refined students of atomic structure who use the hydrogen atom, the electron and the proton as the touchstone of the completeness of current theory, are happily contentious. Today there is no detectable inconsistency anywhere at the level of parts per million (although we do not know at all why the theory of photons and electrons is even finite). The most striking novelty (much discussed at a very advanced level) is the utility of macroscopic quantum fluids for precision measurement of Planck's constant. Two vessels of superfluid helium connected by a tiny hole will exchange an alternating flow of matter if they are irradiated by a supersonic vibration. At these low temperatures, however, the motion is quantized. The differences in liquid level between the two vessels can assume only a particular quantized set of values; the height of these steps in fluid level can be directly measured, and Planck's constant can be read off in terms of the frequency, the mass of a helium atom and the local acceleration of gravity. Exactly analogous effects are observed between superconductors, where the height differences are electrical voltage steps, the irradiating beam is microwave and the moving fluid is pairs of superconducting electrons.

There is plenty of interest here for the more old-fashioned physicist. Everything is not the Josephson effect and sixth-order diagrams. Consider modern X-ray interferometry, where these adepts count X-ray fringes directly, using epoxy glue to hold all the parts of the diffracting crystals together. They get

alignments and motions by smooth stressing of the glued blocks, pushing artfully here and there on cunning slotted blocks of crystal with piezoelectric elements that deflect just a little when they are signaled. Angles of a thousandth of a second of arc and distances of motion of a couple of angstroms are achieved in apparatus of tabletop size.

Good old *g*, the 980 centimeters per second squared acceleration of gravity, is still important. Just as you did it in freshman physics, so these witty specialists measure it by measuring how fast some object falls freely. They throw a fused-silica corner-cube reflector up a few inches in a vacuum and count optical interference fringes with fast electronics as the reflector rises and falls. In this way they can measure *g* absolutely to about one part in a billion. Sèvres is currently the virtuoso laboratory; their *g* does not change at about that level of precision either seasonally or from year to year, provided that they look out for (1) the barometric pressure, which pushes a bit on the local rocks, (2) the ocean waves pounding on the French coast, which shake the apparatus, (3) the position of desks and chairs around the laboratory and (4) small horizontal velocities, which change the measured value of *g* by centrifugal effects. After all, the earth does rotate and is not an inertial frame.

Not all readers but certainly physics buffs, students and their teachers, at the level of high school and higher, will find pleasure and insight somewhere in this experts' book.

THE EUROPEAN DISCOVERY OF AMERICA: THE NORTHERN VOYAGES A.D. 500–1600, by Samuel Eliot Morison. Oxford University Press ($15). That agreed-on myth, history, so constantly renegotiated as aspirations and nations grow and decay, has many pens to write it. History books are seldom reviewed in these columns, unless they center on science and technology. Historiography is mainly the evaluation of authority, of the written word and of sources, carefully checked and glossed for motive and meaning. It is essential scholarship, but it is not science, whatever *Wissenschaft* might mean. More and more we see a new trend, a kind of link between history and archaeology. Documents do not always mean what they plainly say. In statecraft or commerce the usual experience of the learned historian may be wide enough to grasp the reality behind the note, but for many other areas of human life a still wider experience is needed. Farming or sailing or science, like generalship, demands a special knowledge of the deed, not just the word.

Professor Morison is the dean and doge of American historians. He will document you documents and burrow in the sources with the best, and he has done that to general praise for close to six decades. He also does more. He is a durable prototype of the historian who knows more than the bookshelves. His peacetime métier is the sea under sail, and latterly also by light plane, flying low to make the same landfalls the voyagers he chronicles made so long ago from masthead and sterncastle. This thick, entertaining volume is a triumph of content, method, wit and assured and engaging style. (A similar volume on the southern voyages is to come—"God willing.")

Not all historians can find their way around outside the stacks. A late distinguished geographer, who "pontificated about the great navigators of history in a highly disparaging fashion," admitted "'I am absolutely terrified of the water and would not go on it for anything!'" Admiral Morison names names; his freely offered, blunt and well-informed judgment of his peers is one of the most characteristic and helpful parts of the work. It is evident that he is not always right; sometimes he passes a judgment too soon and sometimes his decisions founder on his own limitations. China is terra incognita to his eyes; he holds in spite of the bulk of the evidence that the magnetic compass originated in the Bay of Naples. He stoutly clears away, however, the vast sea wrack of fakes and wishful thinking and monomania that adheres to the stories of transatlantic voyages before Columbus. His view is explicit: Leif got to Vinland, which is Newfoundland and no grapier than Greenland is green. Otherwise Columbus was first. Yale's Vinland map is a fake, no more credible than a dozen other sailor's tales: Phoenician ore boats, the good St. Brendan, seaside towers and enigmatic rune stones. The long footnotes that support these firmly held views are among the most diverting and yet most useful parts of the book.

The metacenter of the book is the robust account of all the northern voyages Europeans made to America from A.D. 500 to 1600. The first of the voyagers after the Vikings was John Cabot, a Genoese probably, seeking the Indies by a shorter route, with his support and vessels out of the River Severn "shipshape and Bristol fashion." Then came the Anglo-Portuguese, the Florentine da Verrazzano sailing in French ships, Maître-Pilote Jacques Cartier and his associates and finally Elizabeth's mariners, from Martin Frobisher to those who sailed to the lost plantation of Roanoke. All these accounts are meticulous, lively, filled with detail and embedded in a context that is concerned with motive, means and men.

Two small tales will have to represent a compact volume of 700 pages of narrative. European sailing craft, drawn by the fisheries of cod, walrus and whale, had become "common every summer" around the Gulf of St. Lawrence by the 1560's. A brave English sailor, David Ingram, was put ashore with two others in 1567 on the Gulf coast of Florida. He tramped his way north and returned to Europe a couple of years later on a French ship he hailed at the mouth of the St. John River in New Brunswick. The Elizabethan chronicler Richard Hakluyt preserved Ingram's story. The doughty Ingram now swallowed the anchor to make his living thereafter by retelling his travels in taverns all across England. He finally elaborated a tall tale about the rich city of Norumbega on the Penobscot, where men wore pearls as big as your thumb and "hoopes" of gold and silver and dwelled with their gold-clad wives in houses with pillars of gold, silver and crystal! Norumbega had a long life on the maps; John Milton invoked the name. The Charles River (a little past Harvard) became the 19th-century site for that same city of legend—made into a Norse market town—and the place-name is Maine folklore today. The name is good Algonkian, but the place lies entirely within the imagination of Europeans.

When Frobisher sailed into the cold waters of Baffin Land in 1576, he fell in with Eskimo hunters in kayaks: "They be like to Tartars, with long blacke haire, broad faces, and flatte noses and tawnie in colour, wearing Seale skinnes." These men already knew European ships; they exercised in the rigging, showing themselves "verie strong of theyr armes, and nimble of theyr bodies." Five sailors went ashore to do some trading privately. No Englishman ever saw them again. In the 1860's an old woman of the Innuit told the American explorer Charles F. Hall that long, long ago five white men had been captured at a time when in three successive years ships had come to their country. (She gave almost exactly the number of ships, and in sequence, of Frobisher's three voyages.) The men were released, built a boat and sailed

away to perish at sea. That is 300 years of accurate oral tradition.

The docks and trenches Frobisher built are still to be found. He hauled home a couple of hundred tons of worthless black rock, held to be gold ore by assayers in London, who must have salted their samples. John Dee, the wizard crystal-scryer of Trinity College, was navigational consultant to the Frobisher voyages. John White, the artist, who became governor of Virginia in Raleigh's second planting, painted wonderfully well the three Eskimos whom Frobisher brought back as hostages to Bristol. There they are, the hunter and a woman with her baby in her hood, painted from life, real as Nanook, their furs the same as those Roald Amundsen found in 1905, to the very needlework. The three were a sensation in Bristol. The English used them as well as they could, feeding them meat and not forcing them to change their sealskins for woolens. All three sickened and died within the month.

Maps, index, notes, bibliography, facsimiles—almost all the apparatus of the book is splendidly generous. The photographs are interesting, and the technical deficiencies of some of the aerial views are insignificant, but the gray tones of the offset reproductions are not pleasing.

ANIMALS OF THE ARCTIC: THE ECOLOGY OF THE FAR NORTH, by Bernard Stonehouse. Holt, Rinehart & Winston ($10.95). Look down at the earth from above the North Pole. There is a landlocked sea, ringed by a wide continental belt. Through the forests not far north of Winnipeg, Oslo and roughly along the route of the trans-Siberian railroad runs a climatic boundary. It is the edge of the subarctic: a region of cold winters and cool summers. From that line northward the dark forests spread and dwindle until beyond a rather well-marked tree line close to the Arctic Sea no tree, a tree being defined as an individual of a forest species that grows tall enough to stand clear of winter snows, is to be found. The northernmost trees in the world are "stunted larch, spruce and alder" growing not far from the shores of the ice-covered Laptev Sea near the mouth of the great Lena River. The treeless world north of that boundary, all tundra, ice and sea, is the true Arctic, defined not by geometry but by biology. This beautifully illustrated and readable book, packed with paintings of birds and animals, with maps and with striking photographs from the ground and the air, presents the life of the two regions, with the focus on the animals of the land and the sea.

Arctic life tends to be new on the time scale of evolution. The tundra grows on the poorest soil, so recently freed from the grip of glacial ice that it is little more than rock dust and dead vegetation. There has been no time to evolve the close-woven network of mutual dependence seen in older abodes of life. There are plenty of plants in the tundra; Greenland alone has nearly 500 species of ferns and flowering plants. A Temperate Zone area of the same size, however, would have a much greater variety.

Indeed, never before in geologic history have there been real arctic conditions. The world has been cooling off, some 10 degrees Celsius in the mean since the days of the dinosaurs. It is continental wandering and the growth of mountains that must be the cause. When the poles are in the open ocean, the world is temperate; the ocean currents spread the heat, and there can be no glacial ages. Once the poles are landlocked, ice caps form. Today one pole is seated high among cold mountains in the south while the other remains in an oceanic basin where warming currents cannot penetrate. For a few million years, then, we have had a novel arctic world of life.

That life has had two dominant successes. At sea the phytoplankton of the subarctic marine zone prosper. The coldest waters hold dissolved gases well but are poor in nutrient salts; the Temperate Zone waters are low in gases but rich in nutrients. The mix, between the Atlantic and the Arctic, has the best of both regimes and the energy flow of continuous summer sun. The rich plant and zooplankton life feeds the great whales, the seals, the polar bear and the brave Eskimo hunter.

The other success is the warm-blooded branch of life: the birds and the mammals. They do well by land and by sea; the sea mammals, like the birds, were preadapted to the Arctic, since the sea is everywhere a drain on body heat. The birds too must fight for heat everywhere; they are light, all surface and no volume. The Arctic is easily settled by such immigrants: hardy, clever, mobile. Man is there, the wanderer from the tropical forest, wearing the furs of other animals. A photograph shows the whitened bones of a thousand walruses on the drift boulders of a Hudson Bay shore where every year for centuries the Eskimos have taken the big sea beasts. Another photograph shows a couple of dozen mosquitoes on a man's shirt; these tundra-pond successes feed the birds, taking in return small samples of vertebrate blood to nourish their egg-making. Great grazing

herds of caribou—the same species as the domesticated reindeer of the European Arctic—roam the margin between the forest and the tundra. A striking two-page aerial photograph shows a couple of thousand caribou white against the pale green moss tints and wintry-blue ponds that stretch across the Canadian Arctic. There are also orange lichens on old antlers and bloody red trails converging across the white pack ice to a lone sealing ship.

Altogether this is a book for many readers, for its visual richness and for its thoughtful text. It is one of the international enterprises of the modern book world, "designed and produced for the publisher by Eurobook Limited," and no doubt published by now in all the tongues of the marketplace. Eurobook has done a first-class job, but it is no cavil to enter one complaint. The numerous paintings and photographs each carry the surname of the artist responsible; for example, the single word Breummer runs in small capital letters along the edge of dozens of remarkable photographs of the Canadian Arctic and its life. We are told nothing else, however, of the sources or circumstances of the illustrations. They float all but anonymously out of the pool of the world's images. The text author is real enough—Bernard Stonehouse is a British ecologist—but the picturemakers elude us. For a volume that owes so much to its images, that presents so many arguments in visual form, this is not acceptable.

SHAPES, SPACE, AND SYMMETRY, by Alan Holden, with photographs by Doug Kendall. Columbia University Press ($11). The five solids of Plato and the four of Kepler and Poinsot make nine regular solids. In three-space, where we live, no more exist. Such a solid has all its corners alike and all its faces made of a single species of regular plane polygon. The Greeks knew the convex ones only; the other four, curiously puckered, were not exhibited until the 19th century. The frontispiece of this handsome, brief, closely argued book shows them in the life, crisply cut of photograph-mounting cardboard, lighted so well that the shadows are neither too deep for detailed study nor too dilute for clear perception of depth. They sit sharply on the page, providing the reader with the third-best way to learn about these objects. "The best way... is to make them, next best to handle them."

Of course the book, written in the precise yet informal style of the experienced author, goes well beyond so simple if so deep a display. A hundred and fifty pages later you can see the neatly packed group of 10 truncated octahedrons (six squares and eight hexagonal faces), capable of completely filling space while maintaining the full symmetry of the threefold rotation axes of a cube, so that they can be called isometric. In between you will find, for example, the figures named after Archimedes, with mixed faces but identical corners. With them, as with the solids of Plato, you can build the duals: replacing corners by faces and faces by corners. You can learn about stellation and faceting, about extending faces and facing off corners. There are "solids" of zero volume, interpenetrating solids, prisms built on convex polygons and on stars, antiprisms whose sides are triangles cunningly fitted to replace the rectangular prism faces. By the end of the book you have reached the solid whose boxlike members follow the course of a point executing an arabesque in the form of a trefoil knot among the points of a simple cubic lattice.

The book belongs to a well-defined genre, lying between first-rate recreational mathematics, formal higher geometry and the beginnings of crystallographic insight. Among the excellent and familiar members of its class it is distinguished by the coherence and lucidity of its arguments. The ideas and constructions almost always follow out of arguments based mainly on the logical continuity of construction. The entire text contains not one equals sign, yet it achieves many results explicitly and tabulates the full complexity of the world of polyhedrons. This exposition is a victory of clear, connected thinking over the theorematic method. You know why, as well as what, and you are very likely to have learned why not along several logical boundaries at once. This is not to say that the geometry, for instance, of the snub dodecahedron and its dual, the pentagonal hexecontahedron, is rendered effortlessly clear to the once-over view; not so. The reader can be sure, however, that neither jargon nor symbolism stands between his eye and mind and the limpid, if subtle, structure of the manifolds that flat planes can snatch from space. This is a book of mathematics but not of symbols.

A few pages show the tools and techniques for making all these objects out of white glue and cardboard cut by mounted shears. The photographer can be proud too: no page lacks its helpful picture, nearly all of them small works of still-life art. Indeed, one could well rec-

ommend the book for exercise in the skill of the geometric interpretation of the pictorial image.

THE GEM KINGDOM, by Paul E. Desautels. Special photography by Lee Boltin. Random House ($17.95). Aldous Huxley once observed that the human mind seems to delight in the experience of viewing colorful transparent volumes of space. The mystic vision, the effects of several hallucinogens and the real presence of gems induce in common that subtle pleasure. So does this volume aimed at the collector, or at least the admirer. Lee Boltin, whose photographs deserve the title-page epithet "special," has made the finest gem specimens glow in rich color from the pages of this stunningly illustrated volume without losing that sharpness of surface and hard mirror glint so characteristic of crystals. He has also succeeded in preserving the sheen of the microcrystalline jades and the resinous luster of amber. Turning the pages is a bit like wandering through the treasure room of the Smithsonian with an excellent dissecting microscope of eight or 10 power: most of the gems are as big as your hand across the page. A complaint might be murmured that the scale is lost, but it would be faint indeed. The photographs sometimes show a textured sandy background and sometimes a blankly mirrored one, doubling our view of the faceted gem. The pages are filled out by photographs of instruments, *objets d'art*, old prints and diagrams, and miners and lapidaries at work in many lands. The Italian printers in Verona share the credit with Boltin and the editors.

The text is concise but authoritative, the work of the curator of gems and minerals of the Smithsonian. It is written for the least trained of readers, not disdaining, for example, to present a clear paragraph distinguishing between weight and density. It opens almost every portion of the "gem world" carefully and intelligently, but it seeks to exhaust none. There are good annotated lists of precious and semiprecious stones, and material on cutting methods, on history, on sources, on royal jewels and even on techniques of counterfeiting and simulation, as well as on those of honest synthesis.

There is one stone beautifully cut from tektite glass, the child of some great meteorite or perhaps even of the moon. It is pale green and a little bubbly, but it gives pleasure to that mind which lies even deeper than the mysterious cortical region where glowing colorful volumes resonate.

INDEX OF ADVERTISERS

MARCH 1972

BIBLIOGRAPHY

Readers interested in further reading on the subjects covered by articles in this issue may find the lists below helpful.

FOOD ADDITIVES

FOOD STANDARDS COMMITTEE REPORTS. Great Britain Ministry of Agriculture, Fisheries and Food. Her Majesty's Stationery Office, London.

REPORTS OF THE CODEX COMMITTEE ON FOOD ADDITIVES. Food and Agriculture Organization of the United Nations, Rome.

REPORTS OF THE JOINT FAO/IAEA/WHO EXPERT COMMITTEE ON FOOD IRRADIATION. Food and Agriculture Organization of the United Nations, Rome.

REPORTS OF THE JOINT FAO/WHO EXPERT COMMITTEE ON FOOD ADDITIVES. Food and Agriculture Organization of the United Nations, Rome.

HANDBOOK OF FOOD ADDITIVES. Edited by Thomas E. Furia. The Chemical Rubber Co., 1968.

NONVISUAL LIGHT RECEPTION

ON TEMPORAL ORGANIZATION IN LIVING SYSTEMS. Colin S. Pittendrigh in *The Harvey Lectures, Series 56*. Academic Press, 1961.

THE ROLE OF THE EYE AND OF THE HYPOTHALAMUS IN THE PHOTOSTIMULATION OF GONADS IN THE DUCK. J. Benoit in *Annals of the New York Academy of Sciences*, Vol. 117, Art. 1, pages 204–216; September 10, 1964.

BIOCHRONOMETRY. Edited by M. Menaker. National Academy of Sciences, 1971.

RHYTHMS, REPRODUCTION, AND PHOTORECEPTION. Michael Menaker in *Biology of Reproduction*, Vol. 4, No. 3, pages 295–308; June, 1971.

GEOSYNCLINES, MOUNTAINS AND CONTINENT-BUILDING

NORTH AMERICAN GEOSYNCLINES: THE GEOLOGICAL SOCIETY OF AMERICA, MEMOIR 48. Marshall Kay. Geological Society of America, 1951.

COLLAPSING CONTINENTAL RISES: AN ACTUALISTIC CONCEPT OF GEOSYNCLINES AND MOUNTAIN BUILDING. Robert S. Dietz in *The Journal of Geology*, Vol. 71, No. 3, pages 314–333; May, 1963.

MIOGEOCLINES (MIOGEOSYNCLINES) IN SPACE AND TIME. Robert S. Dietz and John C. Holden in *The Journal of Geology*, Vol. 74, No. 5, Part 1, pages 566-583; September, 1966.

CONTINENTAL MARGINS, GEOSYNCLINES, AND OCEAN FLOOR SPREADING. Andrew H. Mitchell and Harold G. Reading in *The Journal of Geology*, Vol. 77, No. 6, pages 629–646; November, 1969.

MOUNTAIN BELTS AND THE NEW GLOBAL TECTONICS. John F. Dewey and John M. Bird in *Journal of Geophysical Research*, Vol. 75, No. 14, pages 2625–2647; May 10, 1970.

ELECTROSTATICS

ELECTRIC AND MAGNETIC FIELDS. Stephen S. Attwood. John Wiley & Sons, Inc., 1932.

INDUSTRIAL ELECTROSTATIC PRECIPITATION. Harry J. White. Addison-Wesley Publishing Company, Inc., 1963.

ELECTRICAL CORONAS. Leonard B. Loeb. University of California Press, 1965.

FROM RAINDROPS TO VOLCANOES: ADVENTURES WITH SEA SURFACE METEOROLOGY. Duncan C. Blanchard. Doubleday & Company, Inc., 1967.

ELECTROSTATICS: EXPLORING, CONTROLLING, AND USING STATIC ELECTRICITY. Arthur D. Moore. Doubleday & Company, Inc., 1968.

THE CAVE BEAR

DIE DRACHENHÖHLE BEI MIXNITZ. Edited by O. Abel and G. Kyrle. Speläologische Monographien, Vienna, 1931.

A REVIEW OF FOSSIL AND RECENT BEARS OF THE OLD WORLD. D. P. Erdbrink. Deventer, 1953.

PLEISTOCENE MAMMALS OF EUROPE. Björn Kurtén. Aldine Publishing Company, 1968.

CAVE BEARS. Björn Kurtén in *Studies in Speleology*, Vol. 2, Part 1, pages 13–24; July, 1969.

DO INFANTS THINK?

ATTENTION IN THE INFANT: EFFECTS OF COMPLEXITY, CONTOUR, PERIMETER, AND FAMILIARITY. Robert B. McCall and Jerome Kagan in *Child Development*, Vol. 38, No. 4, pages 939–952; December, 1967.

ATTACHMENT AND LOSS: VOL. I, ATTACHMENT. John Bowlby. The Hogarth Press, 1969.

CHANGE AND CONTINUITY IN INFANCY. Jerome Kagan. John Wiley, 1971.

SPEECH PERCEPTION IN INFANTS. Peter D. Eimas, Einar R. Siqueland, Peter Jusczyk and James Vigorito in *Science*, Vol. 171, No. 3968, pages 303–306; January 22, 1971.

AN EXPERIMENTAL TEST OF THE DISCREPANCY HYPOTHESIS. C. Super, J. Kagan, F. Morrison, M. Haith and J. Weiffenbach in *The Journal of Genetic Psychology*, in press.

THE SOURCES OF MUSCULAR ENERGY

THE REVOLUTION IN MUSCLE PHYSIOLOGY. A. V. Hill in *Physiological Reviews*, Vol. 12, No. 1, pages 56–67; January, 1932.

KINETICS AND MECHANISM OF OXYGEN DEBT CONTRACTION IN MAN. R. Margaria, P. Cerretelli, P. E. diPrampero, C. Massari and G. Torelli in *Journal of Applied Physiology*, Vol. 18, No. 2, pages 371–377; March, 1963.

EXERCISE AT ALTITUDE. R. Margaria. Excerpta Medica Foundation, 1967.

ENERGETICS OF MUSCULAR CONTRACTION. W. F. H. M. Mommaerts in *Physiological Reviews*, Vol. 49, No. 3, pages 427–508; July, 1969.

HOW DID KEPLER DISCOVER HIS FIRST TWO LAWS?

GESAMMELTE WERKE: VOL. III, ASTRONOMIA NOVA. Johannes Kepler. C. H. Beck, 1937.

KEPLER. Max Caspar. Abelard-Schuman, 1959.

KEPLER'S DERIVATION OF THE ELLIPTICAL PATH. Curtis Wilson in *Isis*, Vol. 59, Part 1, No. 196, pages 5–25; Spring, 1968.

FROM KEPLER'S LAWS, SO-CALLED, TO UNIVERSAL GRAVITATION: EMPIRICAL FACTORS. Curtis A. Wilson in *Archive for History of Exact Sciences*, Vol. 6, No. 2, pages 89–170; April 29, 1970.

MATHEMATICAL GAMES

GRAPH THEORY. Frank Harary. Addison-Wesley, 1969.

GRAPH THEORY AND ITS APPLICATIONS. Edited by Bernard Harris. Academic Press, 1970.

APPLIED GRAPH THEORY. Clifford W. Marshall. Wiley-Interscience, 1971.

THE AMATEUR SCIENTIST

HANDBOOK OF OPERATIONAL AMPLIFIER APPLICATIONS. Burr-Brown Research Corporation, Tucson, Ariz.

PRINCIPLES OF SERVOMECHANISMS: DYNAMICS AND SYNTHESIS OF CLOSED-LOOP CONTROL SYSTEMS. Gordon S. Brown and Donald P. Campbell. John Wiley & Sons, Inc., 1948.

When you buy a Gremlin, you get more than a great little economy car that's fun to drive.

You get a car that's been road-tested and checked over so thoroughly, we make this promise: If anything goes wrong and it's our fault, we'll fix it. Free.

And, if we have to keep your car overnight to fix it, over 1900 dealers will loan you a car. Free.

Finally, you get a name and toll-free number to call in Detroit if you have a problem. And you'll get action, not a runaround.

Nobody in the business does all this for you after you buy a car.

Which is probably why people who've never bought a car from us before are buying one now.

When you buy a new 1972 car from an American Motors dealer, American Motors Corporation guarantees to you that, except for tires, it will pay for the repair or replacement of any part it supplies that is defective in material or workmanship.

This guarantee is good for 12 months from the date the car is first used or 12,000 miles, whichever comes first.

All we require is that the car be properly maintained and cared for under normal use and service in the fifty United States or Canada and that guaranteed repairs or replacements be made by an American Motors dealer.

AMERICAN MOTORS BUYER PROTECTION PLAN