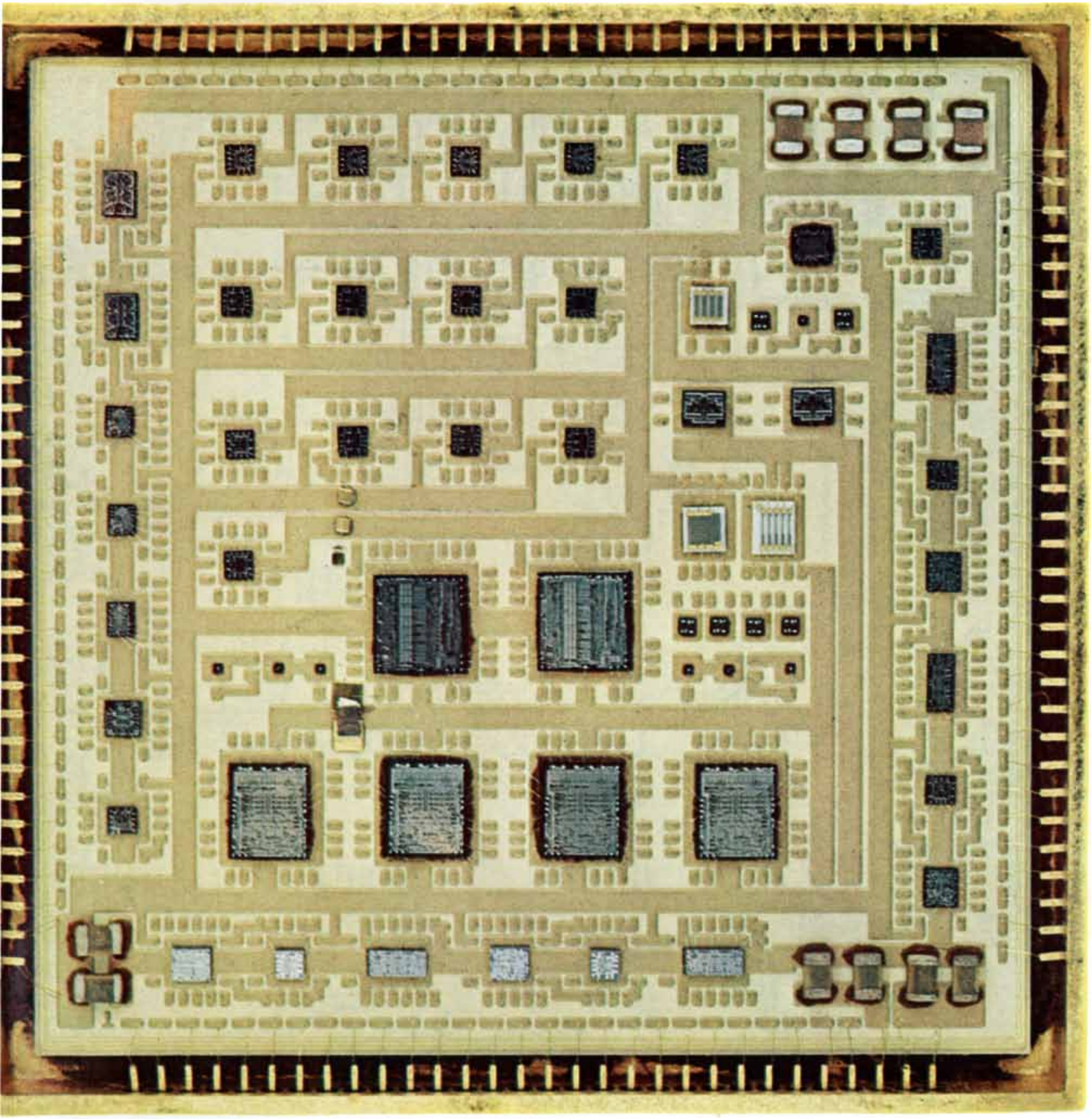
SCIENTIFIC AMERICAN



MICROCOMPUTERS

$1.25

May 1975

Announcing a new generation of International Paper Supertrees. And not a moment too soon.

Do you have any idea how much paper and wood the average American will use by the year 2000?

Over a ton a year.

Now multiply that by 264.4 million Americans, and ask yourself where on earth all that paper and lumber will come from.

At International Paper, we're working hard to come up with some answers.

Like Supertrees.

The average Supertree grows faster, taller, straighter, and healthier than ordinary pine trees.

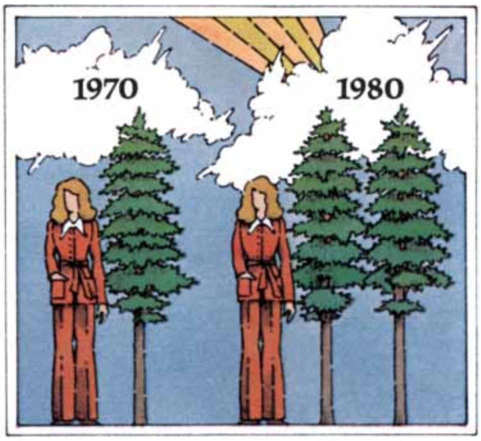
It produces more wood fiber for more paper. And like all trees, a Supertree is a renewable resource.

The best of the best.

How do you breed championship animals? Simply explained, you find two champions and mate them.

At International Paper, we've been doing the same thing with our Supertrees.

After years of testing, we've mated the best of our original Supertrees, and now we're beginning the second-generation selections. These new Supertrees will be even straighter. Even taller. Even healthier than their parents.



By 1980, International Paper will have planted two Supertrees for every person in America.

There will be more Supertrees, too. In fact, by 1980, International Paper will have planted two Supertrees for every man, woman and child in the country.

We've been working to build the world's fiber supply for decades. And not just with Supertrees.

Other things we're doing.

We've developed the Landowner Assistance Program, to help small landowners do a better job of managing their forests.

We're finding ways to get more wood fiber from our trees. For example, we're experimenting with a new machine that harvests tap roots that used to be left in the ground.

We're moving ahead on projects like cooperative nurseries. Tree farm programs. Forest research.

Will all this be enough to keep the world's fiber supply going strong?

It will help. But more must be done.

At International Paper, we believe forest products companies, private landowners and government should work together to develop more constructive policies for managing America's forests. The wrong policies can make tree farming impossible and force the sale of forest land for other purposes. The right policies can assure continuation of America's forest resources.

If you'd like more information about what has to be done to assure the world's fiber supply, write to: International Paper Company, Department 107-A, 220 East 42nd Street, New York, N.Y. 10017.

INTERNATIONAL PAPER COMPANY

220 EAST 42ND STREET NEW YORK, NEW YORK 10017


MCDONNELL
DOUGLAS

Established 1845 SCIENTIFIC AMERICAN May 1975 Volume 232 Number 5

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: $15 PER YEAR, U.S., POSSESSIONS AND CANADA; $18 PER YEAR, ALL OTHER COUNTRIES. EUROPEAN SUBSCRIPTION COPIES MAILED C/O PUBLICATIONS DISTRIBUTION SERVICE, ST. KATELIJNEVEST 14, B2000 ANTWERP, BELGIË.

THE COVER

The photograph on the cover shows a complete general-purpose computer enlarged 3½ diameters. The actual device is two inches square and .2 inch thick and weighs two ounces. One of a family of microcomputers designated TDY 52, it is made by the Teledyne Systems Company (see "Microcomputers," page 32). The unit depicted on the cover is a programmable, parallel 16-bit microcomputer. The term "16-bit" defines the length of the "word," consisting of 0's and 1's, that the machine can handle in its arithmetic operations. Conventional minicomputers, which occupy several cubic feet, typically process words of from 12 to 32 bits and are up to 10 times faster. Otherwise the tiny Teledyne device can do essentially everything that much larger machines can do. The two largest elements are highly complex integrated circuits, each containing some 6,000 transistors on a single chip of silicon. The two chips control the four slightly smaller chips, each a four-bit "slice," that incorporate the arithmetic and logic circuits and operate in parallel. These six principal integrated circuits and various smaller ones are made by the National Semiconductor Corporation. Teledyne assembles the chips on a ceramic wafer and interconnects them with six glass-insulated conducting layers. All told the system incorporates more than 100,000 transistors. It communicates with the outside world through 120 leads, 30 on a side, the ends of which are visible at the edges of the photograph. In quantities between 50 and 500 the Teledyne microcomputer, designed chiefly for military systems, sells for $1,295. Other companies package chips in less compact form in machines that sell for much less.

THE ILLUSTRATIONS

Cover photograph by Jon Brenneis

Page	Source	Page	Source
14	Dan Todd	54–55	Caroline Weaver, Bruce Coleman Inc.
15	U.S. Geological Survey	56–60	Bunji Tagawa
16–23	Dan Todd	65	Brian C. R. Bertram, University of Cambridge
24	Samuel Dales, Public Health Research Institute of the City of New York	67–78	Alan D. Iselin
26–31	Allen Beechel	80–81	George V. Kelvin
32–33	Motorola Semiconductor Products	82–87	Ilil Arbel
34	Jon Brenneis	88–89	Cornell University
35	Gabor Kiss	91	Henry S. Horn, Princeton University
36–37	Ben Rose	92–98	Lorelle M. Raboni
38–40	Gabor Kiss	103–106	Ilil Arbel
48–52	Jerome Kuhl	109–114	Jerome Kuhl

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

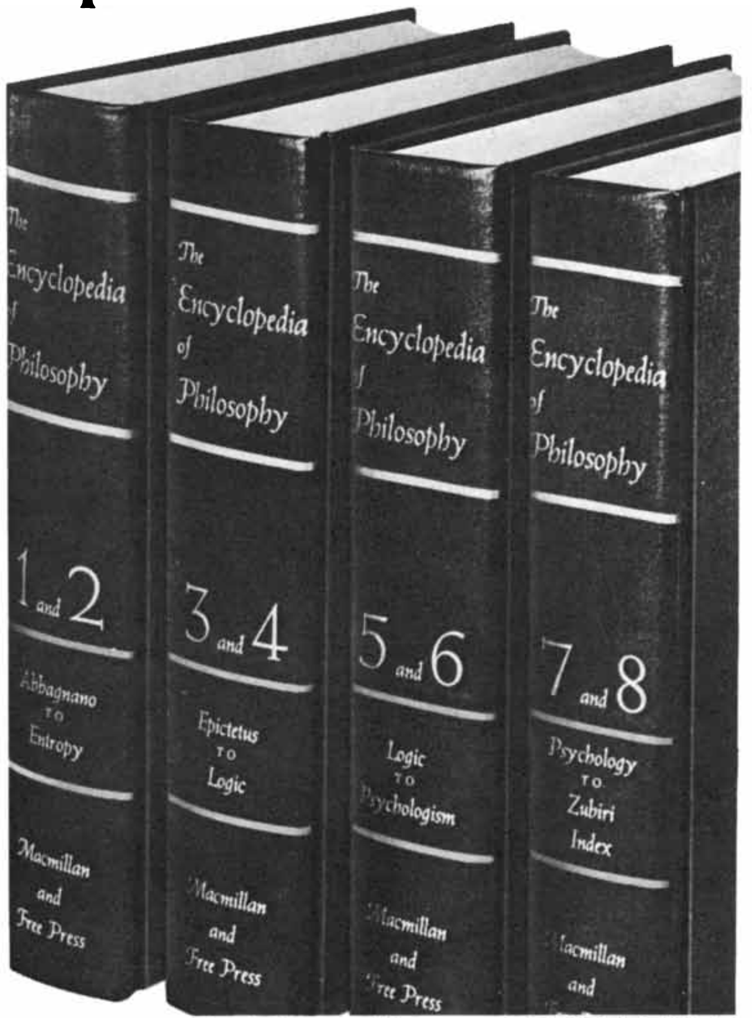
The most comprehensive encyclopedia of philosophy ever published...

FOR ONLY $17.50

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

PUBLISHER'S LIST PRICE: $99.50

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



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

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

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

SPECIAL FEATURES AND EDITORIAL FACTS

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

A SMALL SAMPLING OF THE IDEAS AND PHILOSOPHERS IN THESE VOLUMES

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

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

BOOK-OF-THE-MONTH CLUB, INC., Camp Hill, Pennsylvania, 17012 **5-A170-5**

Please enroll me as a member of the Book-of-the-Month Club and send me the four-volume reader's edition of *The Encyclopedia of Philosophy*, billing me $17.50. I agree to buy at least four Selections or Alternates during the first year I am a member, paying in most cases special members' prices. My membership is cancelable any time after I buy these four books. A shipping charge is added to all shipments.

65

MR.
MRS. } ....
MISS
(Please print plainly)

Address..*Apt*..................

City &
State..*Zip*..................

FACTS ABOUT MEMBERSHIP

- You will receive the Book-of-the-Month Club *News*, a literary magazine published by the Club fifteen times a year. The *News* describes the coming Selection and scores of Alternates, and will be sent to you approximately every three and a half weeks.
- If you wish to purchase the Selection, do nothing and it will be shipped to you automatically.
- If you do not want the Selection – or you would like one of the Alternates or no book at all – simply indicate your decision on the reply form always enclosed with the *News* and mail it so we receive it by the date specified.
- If, because of late mail delivery of the *News*, you should receive a Selection without having had at least 10 days to decide whether you want it, that Selection may be returned at Club expense.
- If you continue after your trial membership, you will earn at least one Book-Dividend® Credit for every Selection or Alternate you buy. These Credits enable you to acquire for your library a wide variety of books, called Book-Dividends, at astonishing savings – *at least 70% of publishers' list prices*. Under this unique system, more than 100 Book-Dividends will be available to choose from every year.
- All books distributed by the Book-of-the-Month Club are identical to the publishers' editions in content, format, size and quality.

LETTERS

Sirs:

In his article in the September 1974 *Scientific American* ["The Populations of the Underdeveloped Countries"] Paul Demeny expresses skepticism about the potential of family-planning programs for reducing fertility. His reasons are that "people can be expected to find the means to control fertility if they want to" and that "people tend to act in accordance with their interests as they best see them." It follows that couples have large numbers of children because they perceive it to be in their best interest to have large numbers of children or, more simply, because they want large numbers of children.

Two questions arise. First, do poor people actually want large families? Second, have family-planning programs been a failure? Demeny's reasoning is that the fact poor people have large families is itself convincing proof that they want large families. No other evidence is necessary. If we assume people always act in their own best interest, then anything they do is *ipso facto* what they want to do. Such reasoning is clearly tautological. Quite different explanations for such behavior are possible. In the case of fertility, couples may not know how to avoid having many children; or if they know, they may not have access to effective methods; or they may be deterred by religious or social taboos, or fear of health hazards, from even trying to control their fertility.

What empirical evidence we have by no means supports the contention that people in the underdeveloped countries want big families. A leading demographer who has done extensive fertility research in Latin America says: "In the past two decades more systematic information has been collected concerning Third World attitudes toward family size than on any other issue in the century. I would summarize the results of this research as follows: One common theme that has come out loud and clear is that most poor people do NOT want large families." (J. Mayone Stycos, *Family Planning Perspectives,* Vol. 6, No. 3; Summer, 1974.)

Nor does Demeny adduce convincing evidence that family-planning programs have failed. Studies by his colleagues at the Population Council show that in less-developed countries where programs employ at least four of the major current means of fertility control (conventional contraceptives, IUD's, pills, voluntary sterilization and abortion) fertility has been declining rapidly. (Parker Mauldin, "Assessment of National Family Planning Programs," a paper presented at the Population Tribune in Bucharest, August, 1974.)

Space forbids any attempt to survey the scores of family-planning programs operating throughout the world today. Two examples will suffice to suggest caution in accepting the sweeping verdict of the family-planning opponents. In South Korea the total fertility rate fell from 6.2 in 1960 to 4.2 in 1968. This is almost exactly the same decline as occurred in the U.S. from 1840 to 1885. The vigorous Korean family-planning program is credited by competent researchers with having had much to do with this greatly accelerated drop in fertility, particularly since fertility fell in rural areas by nearly the same amount as in urban areas. "Without the family-planning program, the rural fertility decline could have been only a fraction of the urban decline. Virtually no commercial contraceptive outlets existed in rural areas before the program began, and there are still few; abortion has become common only quite recently." (George Worth et al., "Korea/Taiwan 1970: Report on the National Family Planning Programs," *Studies in Family Planning,* Vol. 2, No. 3, page 60. The Population Council, 1971.)

The other example comes from Turkey. The total fertility rate in the Etimesgut district dropped from 4.9 in 1967 to 3.7 in 1973. Nusret Fisek concludes on the basis of an exhaustive study that the decline is evidence of "the effectiveness of a family-planning program in changing fertility behavior." (Nusret H. Fisek, "An Integrated Health/Family Planning Program in Etimesgut District, Turkey," *Studies in Family Planning,* Vol. 5, No. 7, page 210. The Population Council, 1974.)

Recent experience in the U.S. is also relevant. Many people in the U.S., as elsewhere, argued that the poor could not be expected to adopt effective methods of fertility regulation unless and until they became nonpoor. Proponents of this view opposed any direct program to make contraceptives available to low-income people. The skeptics have been proved wrong. Large numbers of low-income people have availed themselves of the birth-control services offered by a greatly expanded network of public and private clinics. Fertility rates—and the incidence of unwanted births—not only have declined but also have dropped more rapidly among low-income women than among higher-income women. (Frederick S. Jaffe, "Public Policy on Fertility Control," *Scientific American,* July, 1973.)

We do not know, of course, if family-planning programs will be sufficiently effective sufficiently fast to head off disastrous consequences of population growth. Even less do we know if the indirect effects of other development programs on fertility will have the desired effect. What we do know, from our brief experience so far, is that family-planning programs do contribute to the reduction of fertility and that they cost very little compared with other development efforts. The sensible strategy would seem to be to push ahead vigorously on the family-planning front as well as on the other development fronts. A negative attitude toward family planning, such as Demeny's, can only have the unfortunate effect of reducing the meager financial support for such programs and sapping the enthusiasm and energy with which they are carried on.

ALAN SWEEZY

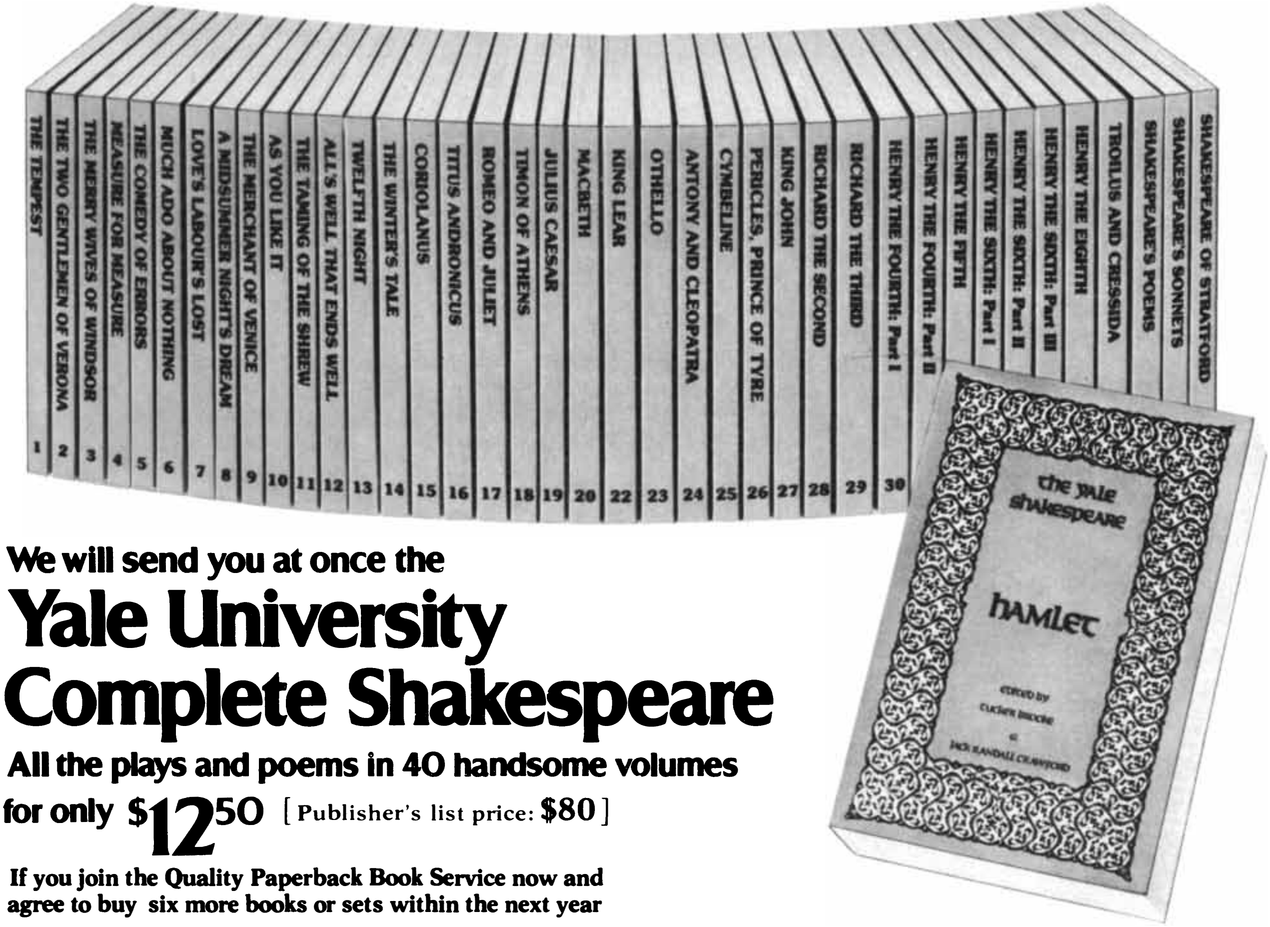
Professor of Economics
Associate Director, Population
Program
California Institute of Technology
Pasadena

Sirs:

When people say that they do not want to have large families but in fact keep having them, one must wonder about the reliability of the instruments that measure their wishes "loud and clear." More fundamentally, one must be mistrustful of the pertinence of a notion of wishes that are so neatly contradicted by actual behavior. If people want to have fewer children, why do they not have fewer children? To say that they do not know how to prevent them is hardly persuasive. Don't they know where babies come from? The question is neither callous nor irrelevant. There is ample historical evidence, including evidence from populations that were once poor, rural and illiterate, that fertility can be greatly reduced without access to modern methods of contraception. There is no reason to assume that the same thing would not happen anywhere in the contemporary world if that is what people really wanted.

Trying to answer the question of what people "really" want of course leads us straight into the domain of metaphysics. The need is for conceptual categories that make useful analysis and interpreta-

A dramatic introduction to the benefits you'll enjoy as a member of the
QUALITY PAPERBACK BOOK SERVICE



We will send you at once the

Yale University Complete Shakespeare

All the plays and poems in 40 handsome volumes

for only $12⁵⁰ [Publisher's list price: $80]

If you join the Quality Paperback Book Service now and agree to buy six more books or sets within the next year

HERE'S an extraordinary opportunity for those who collect the classics in fine paperback editions. You can add the complete forty-volume Yale University paperback edition of all Shakespeare's plays and poems—an $80 value—to your library for only $12.50, with membership.

This edition not only includes all the plays and poems; it also contains a wealth of background information prepared by distinguished Shakespeare scholars to help you gain a deeper understanding of Shakespeare and his works: copious notes, explanations of the sources and history of each play, useful glossaries of obscure words and phrases. Moreover, this paperback edition is one of the most readable Shakespeares in print.

As a QPB member, you'll save money on many of today's most significant paperbacks. Every Main Selection comes to you at a full 20% discount, and you'll also qualify immediately for our Free Book Plan.

If you've been frustrated by the difficulty—perhaps the impossibility—of locating good books in economical paperback editions, give the QPB the short trial outlined. You have everything to gain, including the forty-volume *Yale University Complete Shakespeare.*

HOW MEMBERSHIP WORKS

1 You will receive the *QPB Review* fifteen times a year—about every 3½ weeks. This informative catalog describes the Main Selection plus at least fifty Alternates. **The Main Selection is always offered at a 20% discount.**

2 If you want the Main Selection, do nothing. It will be shipped to you automatically If you want one or more of the Alternates—or no book at all—just indicate your decision on the reply form always enclosed with the *Review* and return it by the date specified.

3 **Free books.** For every book or set you buy (exclusive of the *Yale University Complete Shakespeare,* your introductory forty-volume set for $12.50) you earn at least one Bonus Point, and the Bonus Points you accumulate entitle you to free books and sets. (You pay only shipping charges.)

4 **Return privilege.** If your *Review* is delayed in the mail and therefore you receive the Main Selection without having had ten days to notify us if you did not want it, you may return that Selection at our expense.

QUALITY PAPERBACK
BOOK SERVICE, INC.
Middletown, Pennsylvania 17057

QUALITY PAPERBACK BOOK SERVICE, INC.
Middletown, Pennsylvania 17057

5-QB170-5

Please enroll me as a member of the Quality Paperback Book Service and send me all forty volumes of the paperback edition of the YALE UNIVERSITY COMPLETE SHAKESPEARE (publisher's list price: $80). Bill me only $12.50, plus shipping. My only obligation is to purchase six more books or sets during the coming year, receiving a 20% discount on every Main Selection I take. My membership is cancelable any time after I buy the six additional books or sets. A moderate shipping charge is added to all shipments.

9

Mr/Ms. ____________________ (Please print plainly)

Address ____________________ Apt ______

City ____________________

State ____________________ Zip ______

Q21

tion possible. The assumption that what people actually *do* indicates that they *prefer* that particular course of action to all alternatives available to them opens the door for a systematic investigation of how choices are affected when specific circumstances change. Since such a notion of preference is central to economic analysis, Professor Sweezy's complaint about applying such "tautological reasoning" to the understanding of fertility choices is quite puzzling. More seriously, reluctance to see high fertility in the case of an individual couple as a plausible expression of what is *best* for that couple *when all elements that influence fertility decisions are taken into account* has been and remains responsible for much wishful thinking and lopsided advice in designing population policy.

Knowledge of and access to contraceptives is only one of the many elements that condition fertility choices. It is hardly ever the dominant element. In the U.S., for instance, fertility (after more than a century of steady decline) reached below replacement levels in the 1930's. The postwar period brought a baby boom. Did people forget how to practice birth control? The boom ended and fertility resumed its downward course well before the advent of the pill and easy abortion.

The situation in less-developed countries defies easy generalization. That lowering the price of birth control increases the demand for it is not in question. The question is: By how much is the demand increased? One cannot tell without actual experience. The record thus far is at best mixed. Where socioeconomic change has been rapid, programs are quite successful, although success is perhaps best understood as demonstrating the emergence of demand rather than creating it. Elsewhere the accomplishments appear less promising. Professor Sweezy is clearly more inclined to describe the results of family-planning programs in optimistic terms than I am; it may simply be a matter of temperament. To ascribe the existence of persisting high fertility in much of the world primarily to inability to control fertility, however, is to take a patently untenable position.

It does not follow that efforts to provide better knowledge of and access to contraceptives are not worth while. I am certainly well aware of the "scores of family-planning programs operating throughout the world today," and I said in my article that their "potential is far from having been exhausted," a position certainly consistent with the advocacy of vigorous expansion of family-planning programs.

Paradoxically, to call for family-planning programs on the ground that people want to control their own fertility has provided less than adequate impetus to vigorous expansion of such programs in many developing countries. Indeed, a disproportionately heavy share of both advocacy and financing remains of foreign origin. Needs for all kinds of services are all too numerous and all too poorly satisfied; if the need for family planning is seen as being merely one of them, it is not unnaturally perceived as being a matter of low priority. To give action (including family-planning programs) in the population field greater urgency and better domestic support than it now has requires an analysis and identification of the population problem that cuts deeper than the facile diagnosis that couples would like to have lower fertility but simply do not know how to go about it.

PAUL DEMENY

The Population Council
New York

Scientific American, May, 1975; Vol. 232, No. 5. Published monthly by Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017; Gerard Piel, president; Dennis Flanagan, vice-president; Donald H. Miller, Jr., vice-president and secretary; George S. Conn, treasurer; Arlene Wright, assistant treasurer.

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

Advertising correspondence should be addressed to Harry T. Morris, Advertising Director, SCIENTIFIC AMERICAN, 415 Madison Avenue, New York, N.Y. 10017.

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

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.
U.S. and Canada:
☐ 1 year, $15 ☐ 2 years, $27 ☐ 3 years, $37
All other countries:
☐ 1 year, $18 ☐ 2 years, $33 ☐ 3 years, $45

NAME

NEW ADDRESS

OLD ADDRESS


PIONEER
1
PIONEER
1
PIONEER
3-MOTOR 3-HEAD STEREO TAPE DECK

50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

MAY, 1925: "In semidesert Bechuanaland in South Africa there has just been found the almost perfect fossil skull and jawbone of an extinct creature described by its sponsor, Professor Raymond A. Dart of the University of the Witwatersrand, as intermediate in type between the living anthropoid apes and man. The new fossil has been named *Australopithecus africanus,* or "South African ape," but it will probably also come to be known under the designation of the Taungs skull, from the name of the community near which the remarkable fossil was discovered."

"What is the strange process by which such a vast store of heat is created in the stars? Several answers have been given, of which the most radical of all now appears to be the best. Ever since the discovery of radioactivity it has been realized that matter possesses enormous internal energy by virtue of its own constitution. This relation, at first vague and speculative, was reduced to quantitative form by Einstein, who concluded that mass and energy should be interconvertible. If matter is transformed into heat in the stars, a tremendous amount of heat may be created at the expense of a slow decrease in mass."

"Thousands of universes, each similar to the lens-shaped galaxy of stars of which our sun is an insignificant member, with the nearest universe so distant that its light travels for a million years before it reaches us, have been shown to exist by Dr. Edwin P. Hubble. Through telescopes these universes appear to be clouds of light, more or less circular and with a spiral structure. By means of variable stars in these distant galaxies the distance of a few of them has been determined, and in the opinion of Dr. Harlow Shapley, director of the Harvard College Observatory, this proves the theory that they are 'island universes' lying entirely outside our own system."

"Trees that give weather reports for ages long past when there was no other weather bureau to record them are described by Professor A. E. Douglass of the University of Arizona. After discovering a correlation between known historical climatic conditions and comparatively recent tree rings, Professor Douglass began to carry his researches back into earlier centuries, using trees of the ages-old forests of the Southwest as his time sticks."

"William Jennings Bryan, the recognized leader of anti-evolutionary thought, is ignorant of the facts of evolution and has no legitimate claim to popular leadership in such an issue. This was the statement made by Professor Edward L. Rice before the American Association for the Advancement of Science. Although Mr. Bryan recently joined this great association himself, and it is hoped that he will now become letter-perfect in the scientist's point of view, an analysis of his writings against evolution shows that he has not taken the pains to inform himself on the subject. Professor Rice suggests that Mr. Bryan should get down to facts rather than opinions and concludes that Mr. Bryan advances no new evidence but ignores or denies the data collected by scientists."

SCIENTIFIC AMERICAN

MAY, 1875: "When Lyell and the rest of the uniformitarian school of geology began to attribute all geological changes to the protracted operation of the influences now remodeling the earth's surface, it was objected that time was too short for such proceedings. By Darwin's day such objections were worn out. Men had become accustomed to granting hundreds of millenniums for the periods of the geologist, yet they stood aghast at the demand for more. Geology had been modest in its demands compared with the rising science of biology. Darwin's theory called for an extension of time compared with which that of the geological record was small, and his opponents refused. Now we learn that, whatever objections may be urged against the evolution theory, lack of time for the slow development of creation is not one of them. The soundings of the *Challenger* expedition give a clue to ages of life whose duration dwarfs to insignificance that of the periods between the Lower Silurian and the present, the limits formerly set for the duration of life on the earth. Vast periods covered by the deposition of many thousands of feet of Cambrian and Laurentian rock must now be added, and even they do not bring us sensibly nearer to the beginning."

"Dr. W. A. Hammond, president of the Neurological Society, recently delivered an address before that body advocating the theory that the spinal cord shares with the brain the faculties of perception and volition. Dr. Hammond reviewed experiments showing that a frog from which the brain has been removed continues to perform those functions that are immediately connected with the maintenance of life. The heart beats and the stomach digests; if the web between the toes is pinched, the limb is immediately withdrawn; if the shoulder is scratched with a needle, the hind foot of the same side is raised to remove the instrument. Dr. Hammond said he did not contend that the spinal cord is, in the normal condition of the animal body, as important a center of mental influence as the brain. But it seems we are justified in concluding that the faculties of perception and volition are seated in the cord as well as in the central ganglia."

"The names of Croce-Spinelli and Sivel, two of the most daring and successful of French aeronauts, are now to be added to the long list of those who have laid down their lives in the cause of science. In company with M. Gaston Tissandier they attempted to ascend to a higher altitude than had ever before been reached. The balloon *Zenith* started on its voyage from Paris, shot directly upward and reached a height of 21,000 feet in a very few minutes. At this elevation the aeronauts inhaled a little oxygen from a respirator. Then three of the nine 80-pound bags of sand were emptied and the ascent continued. Sivel soon became intoxicated with repeated doses of oxygen and in his exhilaration threw over the respirator as well as the ballast and a number of the instruments. Again the *Zenith* soared aloft, and Tissandier, as he lapsed into a stupor, read from the barometer an altitude of 29,000 feet. When he awoke two hours later, the balloon was falling and both his companions were dead of suffocation."

"It is now accepted that storms are circular, and that most of them extend over a space hundreds of miles in extent, and often 1,000 or more. The storms are not only circular but also rotary, and they advance across the country at a rate varying from 200 or 300 to much more than 1,000 miles per day. Their average direction is a little north of east."

THE AUTHORS

FRANK PRESS ("Earthquake Prediction") is professor of geophysics and chairman of the department of earth and planetary sciences at the Massachusetts Institute of Technology. A graduate of the City College of New York, he obtained his M.A. and Ph.D. from Columbia University in 1946 and 1949 respectively. He taught geology at Columbia for several years before being appointed professor of geophysics and director of the seismological laboratory at the California Institute of Technology in 1955. He moved to M.I.T. in 1965. Press served as chairman of the Earthquake Prediction Panel of the U.S. Office of Science and Technology and has been chairman of the board of advisers of the Geological Survey's National Center for Earthquake Research since the center's inception in 1966. He is also currently a member of the National Science Board.

DEBORAH H. SPECTOR and DAVID BALTIMORE ("The Molecular Biology of Poliovirus") work together at the Center for Cancer Research of the Massachusetts Institute of Technology. Spector, who received her bachelor's degree from Smith College in 1971, began her Ph.D. training at M.I.T. at the same time that her husband entered medical school. "The combination has enhanced both of our approaches to science," she writes, noting that it has given her "the opportunity to discuss the implications and relevance of basic research for the medical sciences." Baltimore is American Cancer Society Professor of Microbiology at M.I.T. and consulting scientist at the Children's Hospital Medical Center and the Children's Cancer Research Foundation in Boston. He did his undergraduate work at Swarthmore College and his graduate work at M.I.T. and Rockefeller University, acquiring a Ph.D. from the latter institution in 1964. He was a research associate at the Salk Institute for Biological Studies for several years before joining the M.I.T. faculty in 1968. He comments that "I am perhaps best known outside the community of virologists for my work on reverse transcriptase, the enzyme that synthesizes a DNA product of the RNA genome of RNA tumor viruses."

ANDRÉ G. VACROUX ("Microcomputers") is a member of the technical staff of Bell Laboratories in Holmdel, N.J. A native of France, he studied modern languages, specializing in Russian, at the École Nationale Supérieure des Langues Orientales Vivantes in Paris at the same time that he attended the École Supérieure d'Électricité, receiving his Diplôme d'Ingénieur from the latter institution in 1959. He then came to the U.S., where he earned an M.S. in electrical engineering from the University of Notre Dame in 1961 and a Ph.D. in mechanical engineering from Purdue University in 1963. He was at the Illinois Institute of Technology until 1972, when he joined Bell Laboratories.

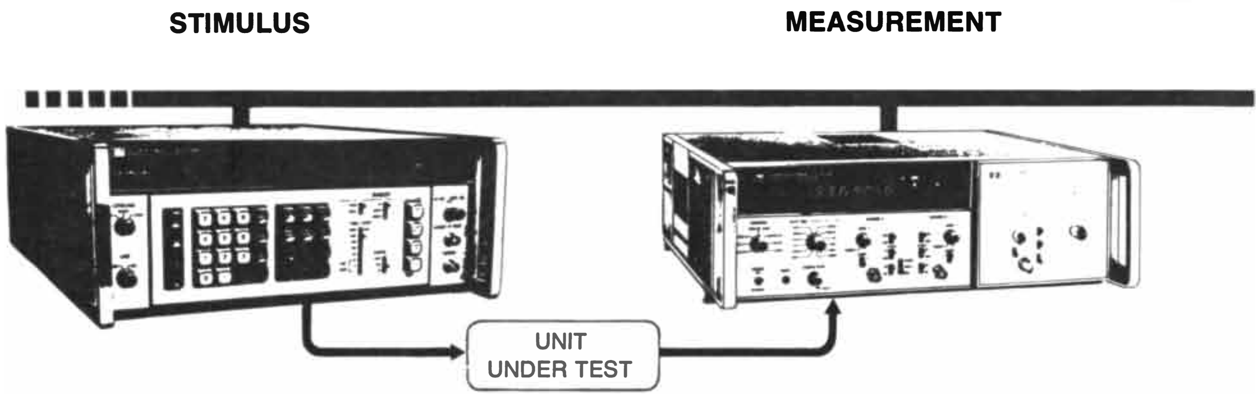
GREGORY J. CHAITIN ("Randomness and Mathematical Proof") lives and works in Buenos Aires, where he carries on his mathematical research "essentially as a hobby." Born in Chicago in 1947, Chaitin grew up in New York City, where he attended the Bronx High School of Science and the City College of the City University of New York. At the age of 11, he reports, "I won a first prize in the school science fair with a van de Graaff electrostatic generator based on a design in 'The Amateur Scientist' department of *Scientific American*. By junior high I began to concentrate my reading on mathematics and computers, and at the age of 15 I was able to actually play with computers, thanks to the Columbia University Science Honors Program." Chaitin developed his ideas on a definition of randomness based on the computer during his first year at City College. He left college without a degree in 1966 and moved with his parents to Buenos Aires, where he went to work for the IBM World Trade Corporation as a computer programmer. He also teaches courses on computers and mathematics at the University of Buenos Aires and has been invited to lecture on his research at a number of universities in the U.S.

BRIAN C. R. BERTRAM ("The Social System of Lions") teaches animal behavior at the University of Cambridge. Following his graduation from Cambridge in 1965, Bertram spent two years in India studying the Indian hill myna bird. He wrote up the results of this research for a Ph.D., which he received from Cambridge in 1969. His fieldwork on lions, which forms the basis of the present article, was done at the Serengeti Research Institute in Tanzania between 1969 and 1973. He writes: "I wish to thank the directors of the Tanzania National Parks and the Serengeti Research Institute for permission to work in the Serengeti; this is S.R.I. Publication No. 172. I am also grateful to the Natural Environment Research Council, U.K., to the Royal Society, London, and to the African Wildlife Leadership Foundation, Washington, D.C., for financial support."

STUART PLATTNER ("Rural Market Networks") teaches anthropology at the University of Missouri at St. Louis. After studying fine arts at Cooper Union in New York, Plattner went to work as a graphic designer. "After a few years," he relates, "I realized that I wanted to learn more about the world, and so I entered the School of General Studies at Columbia University. I majored in anthropology and received my B.S. at Columbia and my M.A. and Ph.D. from Stanford University. I wrote 'Rural Market Networks' while teaching anthropology with the Fulbright program at the University of San Antonio Abad in Peru. At present I am living in Chiapas in Mexico, continuing my studies of the economics of itinerant peddling in a developing area."

CARL SAGAN and FRANK DRAKE ("The Search for Extraterrestrial Intelligence") are professors of astronomy at Cornell University, where Sagan is director of the Laboratory for Planetary Studies and Drake is director of the National Astronomy and Ionosphere Center. "In returning from the International Astronomical Union meetings in Sydney in 1973," they write, "we spent some days skin-diving in Bora Bora in Tahiti, where our *Scientific American* article was first devised. Since Polynesia had been settled by voyagers crossing thousands of kilometers of ocean, we thought a two-kilometer journey by outrigger canoe would be a modest homage to those intrepid explorers, particularly since we were assured that such canoes are unsinkable. We discovered that this is true; when they are swamped, they only sink as far as the shoulders of the passenger, and the outrigger affords some discouragement to those sharks that are to starboard. The experience confirmed our belief that radio communication is easier than direct contact."

HENRY S. HORN ("Forest Succession") is a Princeton University professor of biology currently on leave at the Harvard Forest in Petersham, Mass. A 1962 graduate of Harvard College, he received his advanced training at the University of Washington, where he obtained a doctorate in zoology in 1966. He has been a member of the Princeton faculty ever since. Among his major interests apart from the theme of his article he lists "family, butterflies, carpentry and music."



A new standard communications link that facilitates conversation among instruments.

It wasn't long ago that all instruments were, in human terms, totally deaf and dumb. They could not hear instructions so you made them do their job by setting knobs and switches. And when the job was done, they could not tell you the results; the only way to find out was to read, and then analyze, their displays.

Many instruments have since learned to "talk." On command, they can output measurement results and transmit them remotely in code. More and more are being equipped to "listen": send them prearranged signals and they can program their own controls, remotely. Add a control function to such instruments—to tell them when to talk and when to listen—and they can communicate with each other automatically.

This sounds easy, but it hasn't been. Although the three basic elements for automatic instrumentation systems—talkers, listeners, and controllers—are readily available, one who sets out to design and assemble such a system quickly runs into severe frustrations. The different elements are rarely compatible; more often than not, they use different logic, speak a different language, and interconnect with different hardware.

Avoiding this electronic Tower of Babel is what the Hewlett-Packard Interface Bus (HP-IB) is all about. A standard interface system, the HP-IB forms a basic communications link that allows interconnected system components to communicate effectively, in an orderly and unambiguous manner. The interface system involves much more than the standardization of interconnecting cables; it also defines the interface logic capabilities within the system instruments, the scope of the data codes used on the interface, and the timing and control techniques for exchanging messages.

To talk or to listen: never a doubt.

In the HP Interface Bus, all system devices are exposed to all system communications. But a device can neither send nor receive a message unless told to do so by the system *controller*: at any given time, it can be either a *talker* or a *listener*, but not both. Listeners receive programming data from a controller or measurement data from talkers; talkers send measurement data to listeners. There can never be more than one active controller or one talker at the same time, but there can be as many as 14 concurrent listeners.

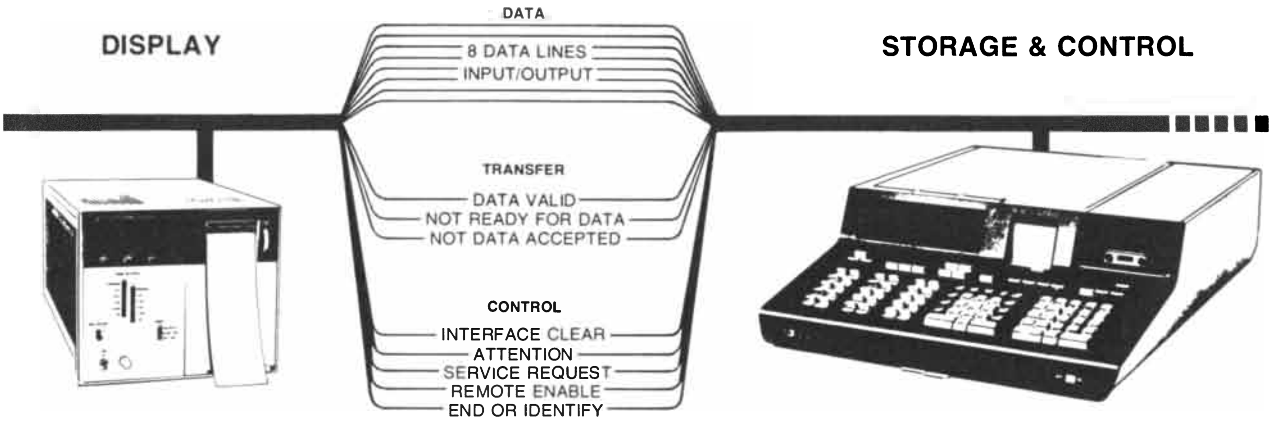
Depending on its capabilities, a device may play more than one role at different times. A calculator or computer, for example, can be talker, listener, or controller; a programmable digital voltmeter alternately talks when it outputs its measurement and listens when it's being programmed; a paper punch can only play the role of listener.

The bus: a common interconnection.

All system devices are interconnected on a common set of 16 signal lines. Eight of these lines form the *data bus* which carries all data messages bidirectionally between talkers and listeners, in bit-parallel byte-serial fashion. The *transfer bus* uses three lines to ensure that data is interchanged only from the intended talker to the designated listeners, through an interrogation and reply sequence. The remaining five signal lines constitute the *control bus*, by which the controller directs an orderly flow of information across the interface, sending commands to the devices and receiving service requests from them. Although system control is always delegated (never assumed), it may be shifted from one system device to another.

HP-IB simplifies systems, small or large.

An HP-IB system can consist of one talker, one listener, and *no* controller; for example, a counter and digital printer for semi-automatic data logging. At the other extreme, a completely auto-



matic system may include as many as 15 instruments possessing stimulus, measurement, display, storage, and control capabilities. Whether a calculator, computer, or the processor of a "smart" instrument, the controller operates the entire system through an interface connection (a single I/O card)—an obvious economy compared to non-bus systems that require one I/O card for each instrument.

System configuration: fundamental problems solved.

Although the HP-IB does not provide instant systems, it does solve the fundamental interface problems that have plagued instrumentation system designers and users until now. Designers no longer need to invent custom interfaces for each new product; users no longer need to familiarize themselves with an interface unique to each new product. Cable and connector problems are minimized by the use of a simple, passive cable interconnection system.

HP-IB protocol allows the designer to assign talk and listen addresses to each device to suit his purposes. Each address is set at the device to any desired value, through a switch on a rear panel, jumper wires on a PC board, or other convenient means.

The HP-IB imposes minimal functional restrictions on data transfer between a talker and a listener. For example, data bytes may consist of from one to eight bits. Once a device is addressed, data can be transferred using any coding and format convention appropriate to the application. The most commonly used codes are the printable characters of the ASCII code set, and the number representations are typically FORTRAN compatible.

Minimal timing restrictions are imposed on the data rates by the HP-IB. Data is transferred asynchronously at a rate that suits the devices involved; burst rates of 1 megabyte per second are possible over limited distances. Data may be transferred directly between devices, thus reducing message traffic on the bus.

More than a theory, HP-IB is a reality now.

Within Hewlett-Packard, the common interface concept has already been incorporated into a growing list of more than 25 instruments and accessory products as well as our computers and programmable desk-top calculators. Additionally, the HP-IB is our implementation of new IEEE Standard 488-1975—and it has served as a model for the IEC Recommendation recently released for ballot among member nations. Thus the possibility exists that this concept will become internationally applicable to the interfacing of instruments, without regard to manufacturer or nation of origin.

Obviously an idea whose time has come, the common interface is here now, still another aspect of the new measurement technology that is taking shape at Hewlett-Packard.

For more information, write to us. Hewlett-Packard, 1504 Page Mill Road, Palo Alto, California 94304.

Mail to: Hewlett-Packard, 1504 Page Mill Road, Palo Alto, CA 94304.

Please send me further information on the
Hewlett-Packard Interface Bus.

Name ________________ Title ________________
Company ________________
Address ________________
City ________________ State ________ Zip ________

00544

HEWLETT **hp** *PACKARD*

Sales and service from 172 offices in 65 countries.
Palo Alto, California 94304

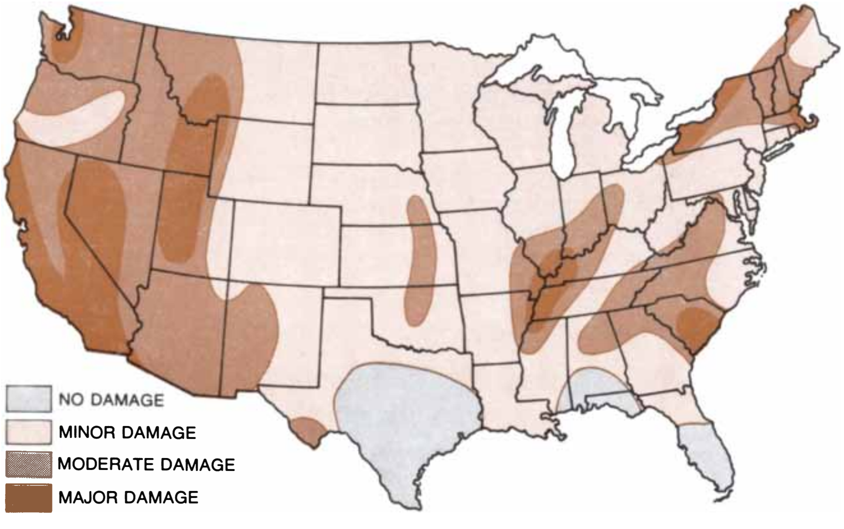
Established 1845 SCIENTIFIC AMERICAN May 1975 Volume 232 Number 5

Earthquake Prediction

Recent technical advances have brought this long-sought goal within reach. With adequate funding several countries, including the U.S., could achieve reliable long-term and short-term forecasts in a decade

by Frank Press

The forecasting of catastrophe is an ancient and respected occupation. It is only in recent years, however, that earthquake prediction has parted company with soothsaying and astrology to become a scientifically rigorous pursuit. At present hundreds of geophysicists and geologists, mainly in the U.S., the U.S.S.R., Japan and China, are engaged in research with earthquake prediction as the direct goal. Most of these investigators believe that the goal is attainable. Some are more pessimistic. A few actually think that the side effects of prediction might be worse than the benefits and that the goal should be abandoned. Research on earthquake prediction therefore exemplifies many of the problems that face modern society: technology assessment, the design and organization of a massive mission-oriented project, the competition for funds and the political niceties of an undertaking involving admittance to previously inaccessible regions of another country.



RISK OF DAMAGE FROM EARTHQUAKES is assessed in a broad, climatological sense in this map of the U.S., based on information compiled by the Coast and Geodetic Survey. The map is based primarily on historical records of destructive earthquakes and does not take into account the fact that earthquake tremors are much more frequent in the Western states than elsewhere. A third of the nation's population live in two darkest-colored regions.

I share the view of most of my colleagues that earthquake prediction is a highly desirable goal. Because of the large increase of population density in the earthquake-prone sections of the U.S., the potential loss from an earthquake as strong as the San Francisco shocks of 1906 could be as high as tens of thousands dead and hundreds of thousands injured, with property damage measured in the billions of dollars. A catastrophe on this scale would be unprecedented in the history of the country, yet it is an event that most seismologists expect to occur sooner or later. The seismic-risk map of the U.S. shows the most probable locations of strong earthquakes [*see illustration at left.*]. The map is based primarily on earthquake history; it does not take into account the frequency of occurrence. Hence Boston is shown to be as risky as Los Angeles (mainly because of a single great quake that occurred in the Boston area in 1755), even though tremors are 10 times less frequent on the East Coast than on the West Coast. It is a sobering thought that a third of the nation's population live in the two regions of highest risk.

Preliminary results of current investigations indicate that predictions of strong earthquakes could be made many years in advance. It also appears likely that a method for making short-term predictions, as short as weeks or even days, will be developed. With this dual capability it should become possible to devise a remedial strategy that could greatly reduce casualties and lower property damage. For example, the long-range prediction of a specific event could spur the strengthening of existing structures in the threatened area and motivate authorities there to enforce current building and land-use regulations and to revise such codes for new construction. A public-education campaign on safety procedures could also be instituted.

Short-term prediction could mobilize disaster-relief operations and set in motion procedures for the evacuation of weak structures or particularly flammable or otherwise hazardous areas. The shutdown of special facilities, such as nuclear power plants and gas pipelines, and the evacuation of low-lying coastal areas subject to tsunamis, or "tidal waves," could also follow a short-term forecast.

The problem of how one communicates an earthquake prediction to the public and the consequences that flow from such warnings (and from possible false alarms) are now being examined. Research into the social aspects of earthquake prediction will presumably advance along with progress toward a physical solution of the problem. For these reasons most experts consider the ability to predict earthquakes to be justifiable on both humanitarian and economic grounds.

With the advent of the theory of plate tectonics the distribution of earthquake belts around the world became understandable. According to this view, the earth's lithosphere, or outer shell, is divided into perhaps a dozen rigid plates that move with respect to one another. Most of the large-scale active processes of geology—vulcanism, mountain-building, the formation of oceanic trenches, earthquakes—are concentrated at or near plate boundaries [*see illustration on next two pages*]. It is easy to see why stresses build up along plate boundaries, where the relative motion of the plates is resisted by frictional forces. When the stress increases to the point where it exceeds the strength of the rocks of the lithosphere or overcomes the frictional forces at the boundary of a plate, fracturing occurs and an earthquake results. The plate-tectonic model combined with earthquake statistics already makes it possible to predict earthquakes in the climatological sense of identifying particularly dangerous areas

HOUSING TRACTS constructed within the San Andreas Fault zone near San Francisco appear in the aerial photograph at bottom, made by Robert E. Wallace of the Geological Survey in 1966; the photograph at top shows the same scene approximately 10 years earlier. Solid white line in each view traces the approximate position of fault along which the ground ruptured and slipped some two meters during the great earthquake of 1906. Broken white lines give approximate boundaries of main fault zone. Pacific Ocean is at lower left.

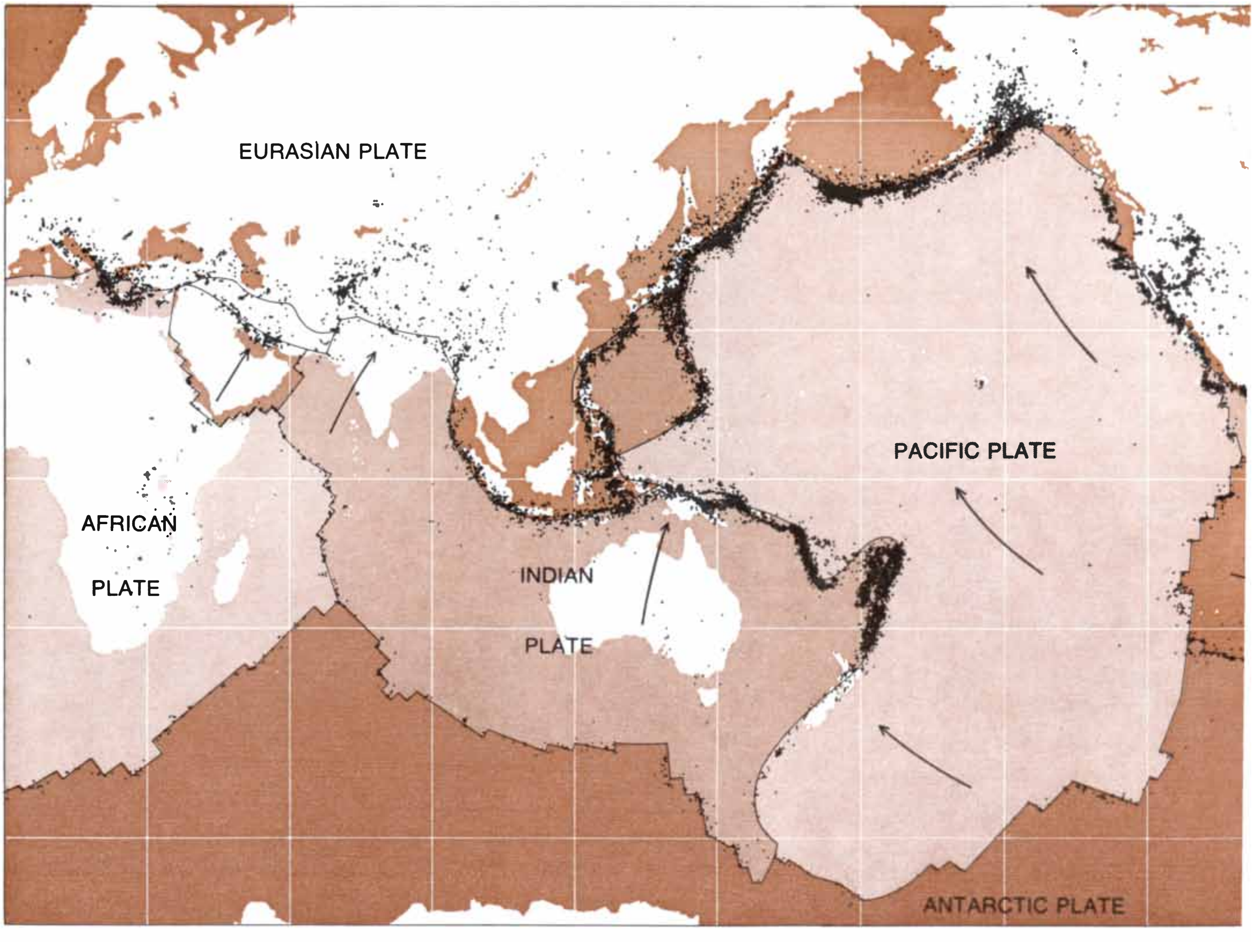
and estimating the relative degree of danger. What is needed, however, is prediction more akin to weather forecasting: Where and when is the next earthquake likely to take place?

A combination of laboratory and field experiments over the past five years has led to a breakthrough in thinking about the problem of earthquake prediction. When a rock is squeezed, it deforms and eventually breaks. Just before it breaks it swells, owing to the opening and extension of tiny cracks. This inelastic increase in volume, a phenomenon long known to laboratory experimenters as dilatancy, begins when the stress reaches about half the breaking strength of the rock. In the mid-1960's William F. Brace and his colleagues at the Massachusetts Institute of Technology showed that measurable physical changes accompany dilatancy in laboratory experiments; such effects include changes in the electrical resistivity of the rock and in the velocity at which elastic waves travel through the rock. Brace suggested that dilatancy and its effects might be detectable in the earth's crust and provide a basis for earthquake prediction; his suggestion generated much excitement at the time because it opened up the possibility that premonitory physical changes could be observed in advance of earthquakes.

In the late 1960's two Russian investigators, A. N. Semenov and I. L. Nersesov, startled the seismological world with a report that unusual variations in the velocity of seismic waves appeared just before earthquakes in the Garm region of Tadzhikistan. Subsequently the Russians announced that in earthquake-epicenter regions in Garm, Tashkent and Kamchatka they had detected changes both in electrical resistivity and in the content of the radioactive gas radon in the water of deep wells.

These reports triggered a flurry of activity in the U.S. American seismologists went to the U.S.S.R. to see the data firsthand. They also began arranging their own experiments in order to observe the precursory phenomena. Technical papers on these phenomena authored by Russian, American and Japanese workers began to be presented in increasing numbers at scientific meetings and in journals. Last year a group of American geologists and geophysicists visited China and found a large-scale earthquake-prediction program under way, with important results that had not yet been reported at international meetings or in publications.

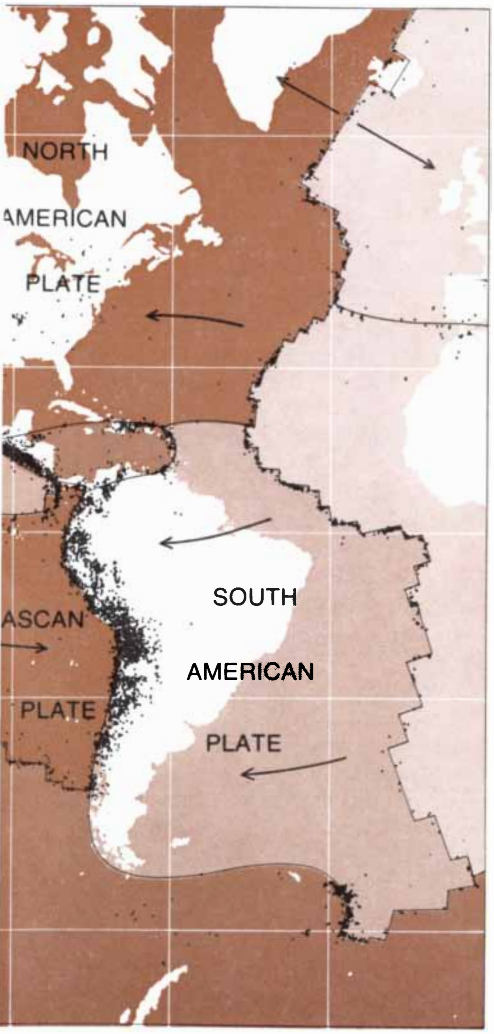
It is fortunate that a number of earthquake precursors have been found, each



EPICENTERS OF 30,000 EARTHQUAKES recorded between 1961 and 1967 are indicated on this world map plotted by M. Barazangi and H. J. Dorman of Columbia University on the basis of information supplied by the Coast and Geodetic Survey. Also shown are the dozen or so moving plates that, according to the modern theory of plate tectonics, comprise the earth's rigid outer shell. Most earthquakes take place at or near plate boundaries, where the relative motion of the plates is resisted by frictional sticking until the stress builds up to the point where the rock fractures, causing an earthquake. Intraplate earthquakes, such as those that appear in the based on a different physical measurement. Confidence in a prediction is enhanced when it is based on several independent lines of evidence, each with its distinctive "noise" history and its distinctive anomaly signaling an earthquake. How are the precursory anomalies observed?

An array of seismographs can be used to sense precursory changes in the velocity of compressional waves and shear waves in the focal region of an earthquake [*see illustration on next page*]. The seismic waves originate in smaller earthquakes within the focal region, in larger earthquakes outside the focal region or in artificial sources such as explosions or mechanical devices. Such anomalous changes have been observed in several parts of the U.S., the U.S.S.R. and China.

Seismically active regions have many more small earthquakes than large ones. This "background" of small tremors varies in time. Periods of calm before a strong shock are frequently observed; the background activity appears to go through a minimum and then to increase just before the main shock. The pattern of radiation of seismic waves reflects the stress field in the crust. In central Asia Russian investigators have found that the stress pattern shown by the small tremors is random during the calm period but becomes highly organized beginning three or four months before the main shock. The compressional stresses become aligned in the same direction as that of the forthcoming main shock.



eastern U.S., are rare but can be destructive. China, squeezed by large plates on the south and the east, has a high level of seismic activity, which may be attributable to the existence of "miniplates" in central Asia.

Another approach is to measure anomalous changes in the volume of crustal rock in the focal region. The changes can be observed by tiltmeters, by devices for monitoring changes in sea level (corrected for oceanographic and meteorological effects) and by repeated surveying. In parts of Japan and China historical records of precursory changes in the level of lakes, rivers or the sea, sometimes dating back hundreds of years, may be related to the same phenomena [*see illustration on page 19*].

Precursory changes in water level, water turbidity and temperature in deep wells can be observed visually or with instruments. Observing the radon content of well water, a technique used extensively in the U.S.S.R. and China, also seems to be a sensitive indicator of forthcoming seismic activity [*see top illustration on page 21*].

If an electric current is fed into the earth's crust between two points several kilometers apart, voltage changes between two other points will show up if the resistivity of the intervening crustal rocks changes. Such precursory fluctuations have been reported in the U.S., the U.S.S.R. and China [*see bottom illustration on page 21*].

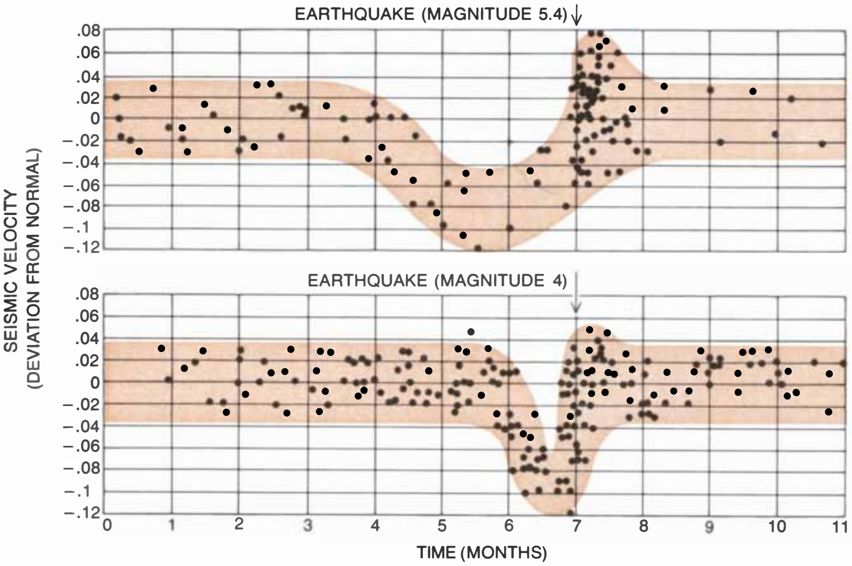
Magnetometers on the earth's surface can detect changes in magnetic field with a strength of about a hundred-thousandth of the earth's natural field. By subtracting the changes sensed by "standard" instruments removed from the epicentral region, noise introduced by fluctuations in the stream of electrically charged particles from the sun (the "solar wind") can be reduced and anomalous changes in the focal region can be detected. Precursory magnetic signals have also been observed in the U.S., the U.S.S.R. and China.

Although one can conceive of an earthquake-prediction strategy based purely on empirical observations such as these, it is highly desirable to have a physical model that explains the observations. A model not only enhances confidence in the basic notion of predictability but also makes for more efficient research procedures.

Two principal models have been proposed, both growing out of laboratory experiments. The dilatancy-diffusion theory, proposed by Amos M. Nur of Stanford University in 1972 and extended by Christopher H. Scholz, Lynn R. Sykes and Y. P. Aggarwal of Columbia University in 1973, is supported by most American specialists. Another model, which might be called the dilatancy-instability theory, was proposed in 1971 by workers at the Institute of Physics of the Earth in Moscow. It also has a few American and Japanese adherents. The models have a common feature: the growth of cracks as stress builds up in the crust just before an earthquake [*see illustration on page 23*].

Both models begin with a stage in which elastic strain builds up in the earth's crust. In the next stage small cracks open in the strained portion of the crust and dilatancy becomes a dominant factor. In the Russian view the development of cracks "avalanches" in this stage. In both models it is the second stage that marks the real beginning of precursory phenomena, since the open cracks change the physical properties of the rock. Seismic velocity (the ratio of compressional-wave velocity to shear-wave velocity) drops. Electrical resistivity increases if the rock is dry and decreases if it is wet. Water flow through the rock increases (and therefore more radon enters the water from the rock). Volume in the dilatant zone increases. In the American model the number of small tremors decreases in this stage because the cracks become undersaturated in water as they increase in number; as a result sliding friction increases and inhibits faulting.

The two models differ markedly in the third stage. In the American model water diffuses into the undersaturated dilatant region. The main effect of this inflow is to increase the seismic velocity and to raise the pore pressure in the cracks, weakening the rock to the point where small earthquakes increase in number and the main shock follows. In the Russian model water plays no role in the third stage. Instead the avalanche-like growth of cracks leads to instability and rapid deformation in the vicinity of the main fault. The stress load drops partially in the region surrounding the zone of unstable deformation, cracks partially close and the rock recovers some of



PREMONITORY CHANGES in seismic velocity (the ratio of compressional-wave velocity to shear-wave velocity) were observed in the late 1960's just before two fairly large earthquakes in the Garm region of Tadzhikistan by A. N. Semenov and I. L. Nersesov of the Institute of Physics of the Earth in Moscow. These composite diagrams, drawn from their work, are based on a number of smaller earthquakes in the region. Each point represents a deviation of the seismic velocity from the normal regional value and is derived by measuring the travel times of compressional waves and shear waves from each small earthquake to a local network of seismograph stations. The colored bands indicate the statistical scatter of the observations. The duration of the calm period preceding the main earthquake appears to increase with the magnitude of the forthcoming event. (The two earthquakes shown measured 5.4 and 4 on the Richter scale.) Seismic-velocity anomalies of this type have been observed about 18 times in U.S.S.R., 10 times in U.S. and several times in China.

its original characteristics. This sequence of events accounts for the increase in seismic velocity, the decrease in volume and the other changes typically observed in the third stage. The developing instability finally gives way to faulting, and the main shock ensues. In both models stress is released by the earthquake, and the crustal rock recovers most of its original properties.

An empirical formula, derived by James H. Whitcomb, J. D. Garmany and Don L. Anderson of the California Institute of Technology, connects the duration of the precursory anomaly with the magnitude of the predicted earthquake. For example, an event with a magnitude of 5 on the Richter scale has an anomaly lasting for about four months, whereas a major earthquake, with a magnitude of 7, say, would be preceded by an anomaly beginning some 14 years before the event. The formula is still rough, particularly in the high-magnitude range, but it appears that the large earthquakes will provide warning times on the order of 10 years. The discovery that the size of an earthquake, as well as its location and timing, is predictable should hold important implications for the design of an earthquake-mitigation strategy. Fortunately the larger the magnitude of the forthcoming quake, the longer the lead time available for making plans to combat its effects.

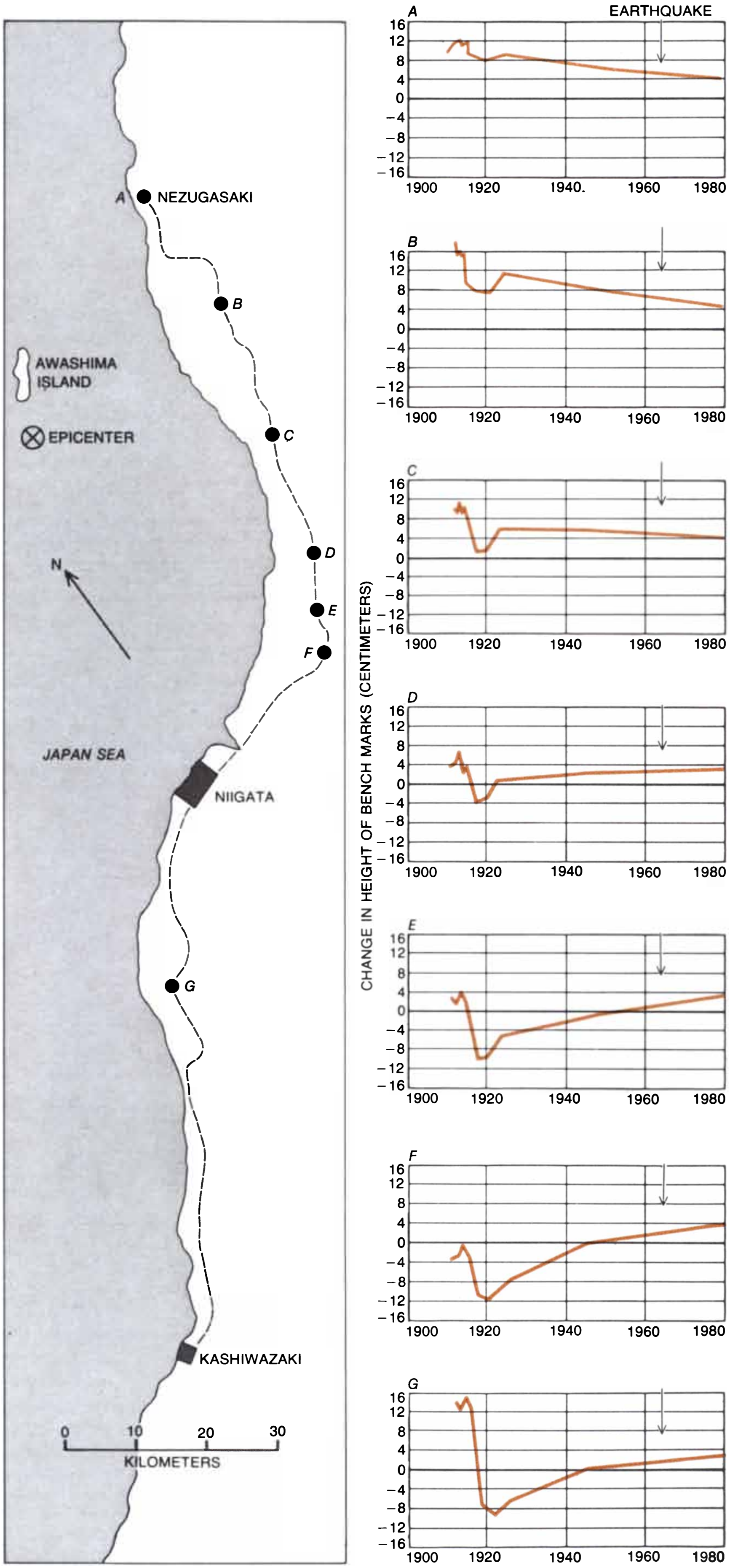
What is most needed now to bring earthquake-prediction technology to the point of implementation is a larger number of examples of successful prediction. So far only about 10 earthquakes have been predicted before the fact. Perhaps three times as many have been "predicted" after the fact by going back to the data and finding precursory signals. It is difficult to know how many formal predictions, based on the methods described above, have failed. The number is probably less than 10, which is not bad for the rudimentary research networks now in operation. That is still too small a sample to eliminate unreliable methods and to design a comprehensive, operating prediction system. Although the major earthquake belts extend for tens of thousands of kilometers, only a small fraction of that distance is instrumented adequately to test prediction methods. With the pooling of the data being gathered in various countries, however, the number of case histories should grow rapidly in the next few years, and statistically valid tests of prediction methods should be forthcoming. Seismologists of different nationalities, like workers in any other field of science, need to combine their results in order to advance toward a common goal.

The leading agency for earthquake-prediction research in the U.S. is the Geological Survey, which runs a strong program centered in California and supports a research program in several universities. In central California, the region where the San Andreas Fault is most active, the Geological Survey has installed a network of stations equipped with seismometers and tiltmeters. Magnetic and electrical observations are also conducted but to a much lesser degree. In southern California a large number of instruments are being installed in a joint effort involving the Geological Survey and Cal Tech. Data from these arrays are mostly telemetered into Menlo Park and Pasadena on telephone and microwave circuits. This growing ability to pinpoint earthquake locations and monitor precursory velocity changes, tilts, magnetic fluctuations and changes in electrical resistivity is beginning to pay out. Recently workers associated with the Geological Survey found that 10 California earthquakes were preceded by tilt changes in the vicinity of the epicenter [*see illustrations on page 20*]. Precursory changes in seismic velocity have been reported for about 10 earthquakes in California and New York.

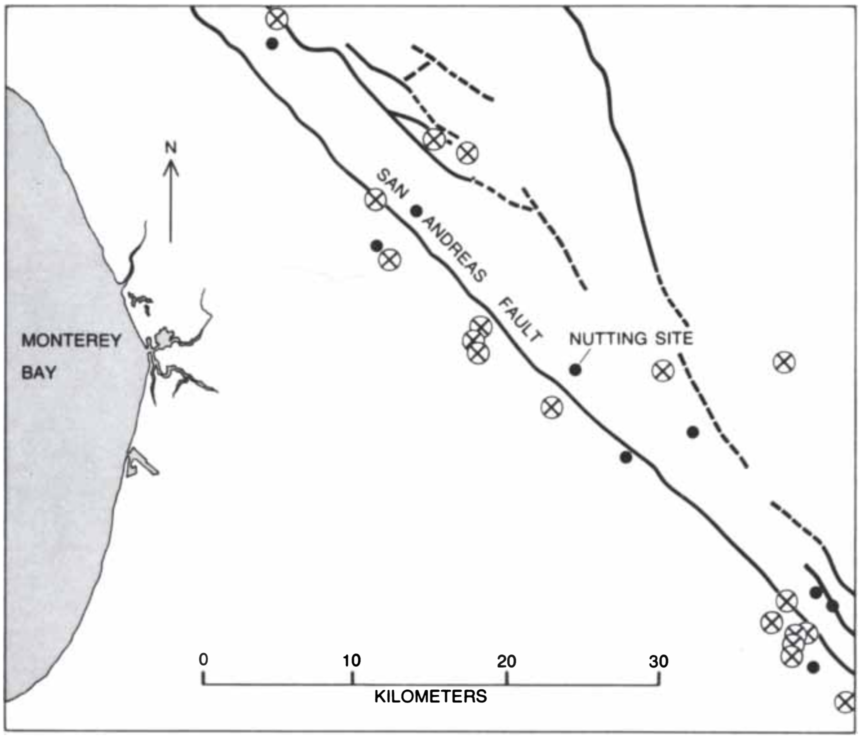
Perhaps the most significant new data were gathered on November 28, 1974, when a magnitude-5 earthquake struck about 10 miles north of Hollister in central California. The tremor was preceded by distinct tilt changes and magnetic fluctuations convincingly above the noise level, and with indications of seismic-velocity changes. John H. Healy of the Geological Survey chided his colleagues the night before the quake for not publicly announcing the forthcoming event.

In spite of these interesting results the U.S. program is still not sufficiently supported to make prediction a reality within the next decade. It is simply a matter of too few methods being tested in too few places. With the present level of support many potentially important methods cannot be tested, such as arrays of wells monitoring water level and radon content, networks of resistivity sensors, sea-level gauges, advanced surveying techniques and so forth. Even now more data are being accumulated than can be digested, a situation that could easily be rectified if a large computer were provided to scan and automatically analyze the incoming stream of data. Universities and industries with much research talent are insufficiently involved because of the lack of funds. Few studies are being conducted outside California. An additional $30 million per year could make prediction within a decade a realistic goal. The cost-effectiveness of such an investment is obvious when one remembers that the relatively modest San Fernando tremor (magnitude 6.6) that struck just north of Los Angeles in 1971 resulted in damage of more than $500 million.

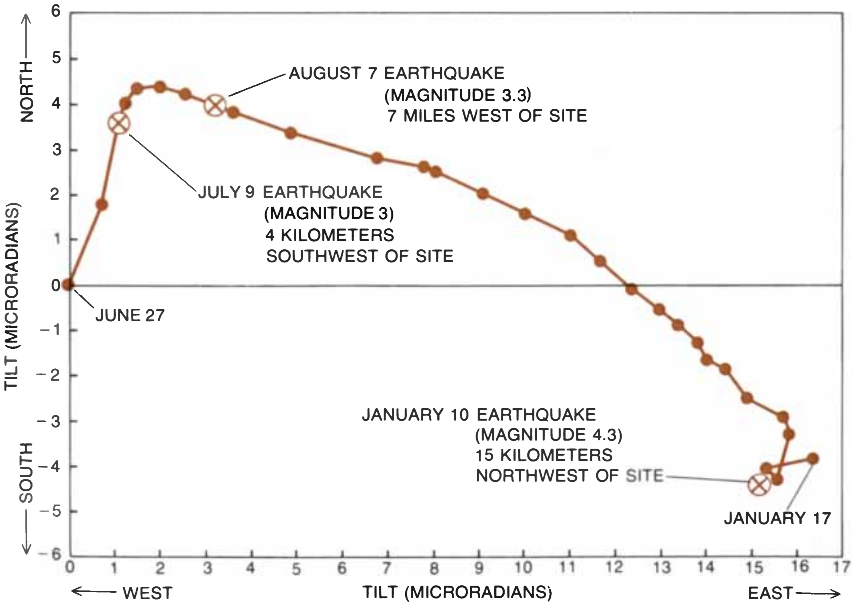
The earthquake-prediction program of the U.S.S.R. is centered in the Institute of Physics of the Earth in Moscow. A program involving laboratory and field measurements, comparable in level to our own, is being carried out. The Russian field experiments form the longest series so far, having been started nearly 20 years ago. The impressive discovery of anomalous precursors stems from these efforts. The strategy of the Russian investigators is somewhat different from our own in that several experimental sites are being monitored in central Asia and Kamchatka with a lower density of instruments compared with our heavier emphasis on specific areas in California. Moreover, the Russians are exploring more methods than we are. Nevertheless, it appears that in the absence of a major new initiative, operat-



ANOMALOUS UPLIFT of the earth's crust in the vicinity of Niigata in Japan was observed for about 10 years before the disastrous 7.5-magnitude earthquake there in 1964, according to the Japanese investigator T. Dambara. The uplift was detected by plotting changes in the height of bench marks measured in repeated land surveys. The graphs at right correspond to the lettered bench-mark sites (*black dots*) shown on the detailed map. Evidence of the crustal uplift was also obtained from records showing a precursory drop in mean sea level observed by a tide-gauge station at Nezugasaki.



TILTING OF EARTH'S CRUST just before earthquakes has been observed by investigators associated with the Geological Survey using an array of sensitive tiltmeters (*black dots*) installed along 85 kilometers of the San Andreas Fault east of Monterey Bay. Circled crosses denote epicenters of all earthquakes with a magnitude greater than 2.5 recorded in the region between July, 1973, and March, 1974. Data summarized in illustration below were obtained at Nutting tiltmeter site, seven kilometers southwest of the town of Hollister.



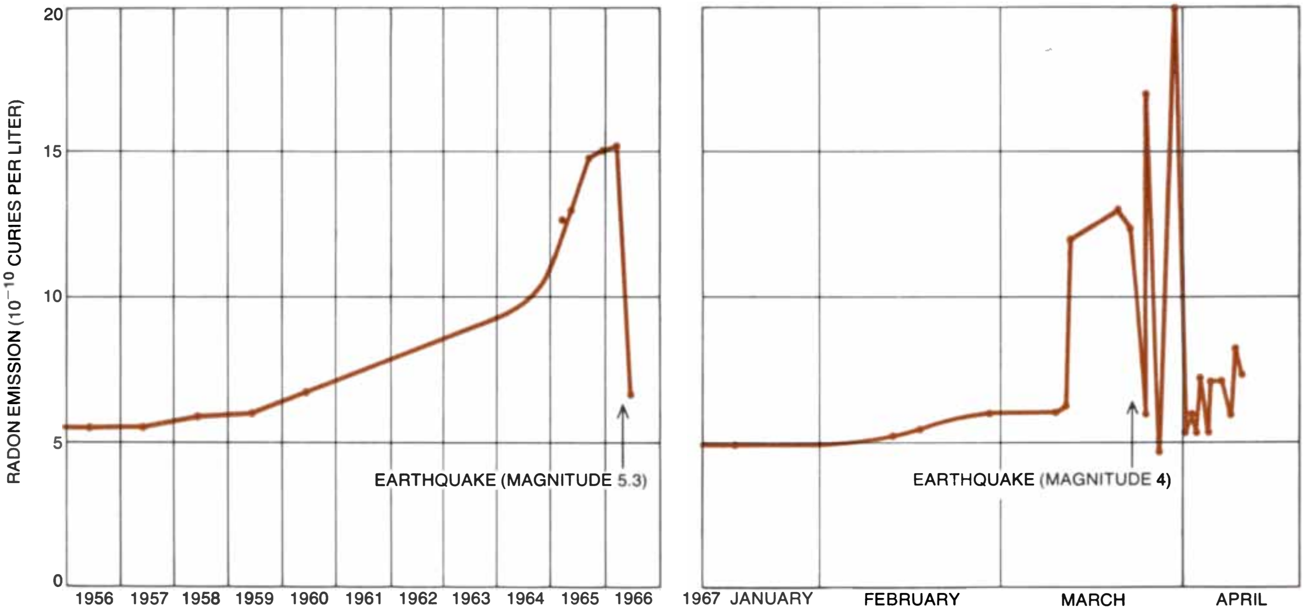
SEVEN-MONTH RECORD of crustal tilting was made during parts of 1973 and 1974 with the aid of a tiltmeter located in a shallow hole at the Nutting site. The colored dots represent the weekly mean tilt direction and magnitude. Several major local earthquakes are indicated; each is preceded by clear precursory change in tilt direction. M. J. S. Johnston and C. F. Mortensen of the Geological Survey report that precursory anomalies of this type have been detected on at least 10 occasions; the largest such event was on November 28, 1974, when a magnitude-5 earthquake struck about 10 miles north of Hollister. So far no comparable tilt change has been recorded that has not been followed by an earthquake.

ing prediction systems covering large areas will not be forthcoming in the U.S.S.R. either in this decade. As part of the environmental treaty between the U.S. and the U.S.S.R. there has been a rewarding exchange of ideas and personnel in the fields of earthquake prediction and seismic engineering. A formal bilateral working group has been established. In this way both American and Russian workers are kept informed of the latest unpublished developments; joint experiments are under way, and there is healthy criticism of each side's efforts by the other. This kind of close cooperation would have been unthinkable a few years ago.

Although Japanese earth scientists have been devoted to the notion of earthquake prediction since the turn of the century, a formal research program dedicated to this goal did not get under way until 1965. For years reports of anomalous sea-level changes and tilts prior to earthquakes have emanated from Japan, but the data were sparse and of uneven quality, and the world community of geophysicists was unimpressed. It now seems that some of these reports must have described true precursory phenomena. In any case the Japanese workers include some of the world's best geophysicists. It is therefore a tragedy that a strike has crippled the Earthquake Research Institute in Tokyo for several years.

The Japanese are currently emphasizing surveys every five years extending more than 20,000 kilometers. So far 17 observatories have been equipped with strain detectors and tiltmeters. Observations of the level of seismicity, of changes in the velocity of seismic waves and of magnetic and electrical phenomena are also under way. Cooperation between the U.S. and Japan in this field is quite close.

This past October I had the good fortune to participate in a month-long trip to China as a member of a group of 10 American earthquake specialists. This tour of Chinese research facilities followed a visit by 10 Chinese earthquake experts to the U.S. earlier last year. Since scholarly publication in China was suspended during the "cultural revolution," almost everything we saw in China was new to us. Following the destructive Hsing-t'ai earthquake of 1966 the Chinese embarked on a major effort in the field of earthquake prediction. Chairman Mao and Premier Chou En-lai issued statements charging Chinese scientists with achieving this goal. At present some 10,000 scientists, engi-

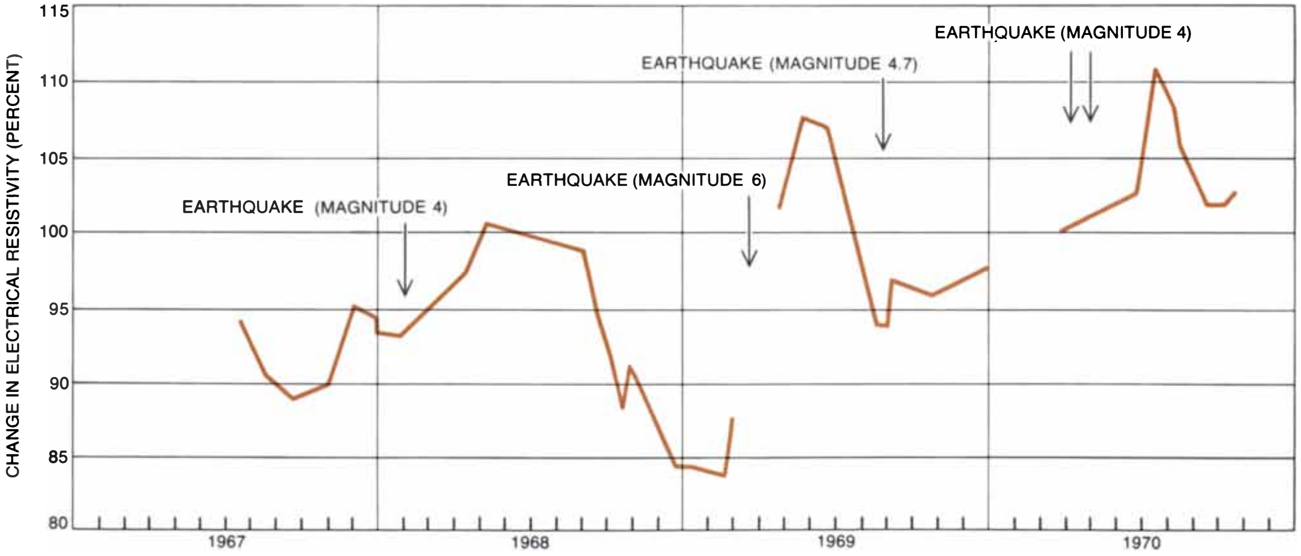


AMOUNT OF RADIOACTIVE GAS RADON dissolved in the water of deep wells has been found by Russian researchers to increase significantly in the period preceding an earthquake. The two examples shown here were recorded before two major earthquakes in the vicinity of Tashkent. The 1966 event (*left*) had a magnitude of 5.3; the 1967 aftershock (*right*) had a magnitude of 4. This promising observational technique is used extensively in both the U.S.S.R. and China, but it has not yet been tried in U.S.

neers, technicians and other workers are engaged in the program—more than 10 times the number of such workers in the U.S.

A unique feature of the Chinese approach is the use of an even larger number of amateurs, mostly students and peasants, who build their own equipment, operate professional instruments in remote areas and educate the local people about earthquakes. So far 17 fully equipped seismograph stations and 250 auxiliary stations have been installed. Data pertinent to earthquake prediction are being obtained at a total of 5,000 points. Every method described in this article is being tested in China. The Chinese say that they have made successful predictions, involving the evacuation of people from their homes and a consequent saving of lives. They also admit to false alarms and failures, chalking these up to the fact that their program is new and they are still in a learning phase. The motivation for success is strong. The high population density, the nature of rural construction and the high degree of seismicity make China particularly vulnerable to earthquakes.

Although it is difficult to gauge the quality of the Chinese program from a brief visit, there is no question that the potential is great. In a few years the Chinese will probably be gathering more data than anyone else, owing to the size of their program and the more frequent



CHANGES IN ELECTRICAL RESISTIVITY of the earth's crust prior to earthquakes have been reported in the U.S.S.R., China and the U.S. The data for this graph were obtained by G. A. Sobolev and O. M. Barsukov for a series of earthquakes monitored in the U.S.S.R. between 1967 and 1970. Measurements of this type are made by feeding an electric current into the ground and observing voltage changes a few kilometers away. In general it has been found that earthquakes are preceded by a decrease in crustal resistivity.

incidence of earthquakes in their country. It may well be that the first statistical validation of prediction methods will come from China. It would be a pity if political considerations were to inhibit close international cooperation on this score, because joint projects with China could pay out in a more rapid achievement of a mutually desirable goal.

Although the prediction of earthquakes has been emphasized in this account, a comprehensive program to reduce vulnerability to destructive earthquakes includes progress in other areas: earthquake engineering, risk analysis, land-use regulation, building codes and disaster preparedness. Unlike earthquake prediction, about which there is a good deal of optimism but (so far at least) no guarantee of success, research and development in these other areas is bound to result in reduced casualties and lowered economic losses. Earthquake engineering deals with the efficient and economic design of structures that may have to withstand the shaking of earthquakes. The alteration of existing structures to improve their performance is included. Not only are residences, commercial buildings, schools, hospitals, dams, bridges and power plants examined individually but also the interaction of all these elements in the system we call a community is considered. This developing technology can serve its purpose only if it is transferred from the investigators to the professional practitioners and to the regulatory bodies that draft building codes.

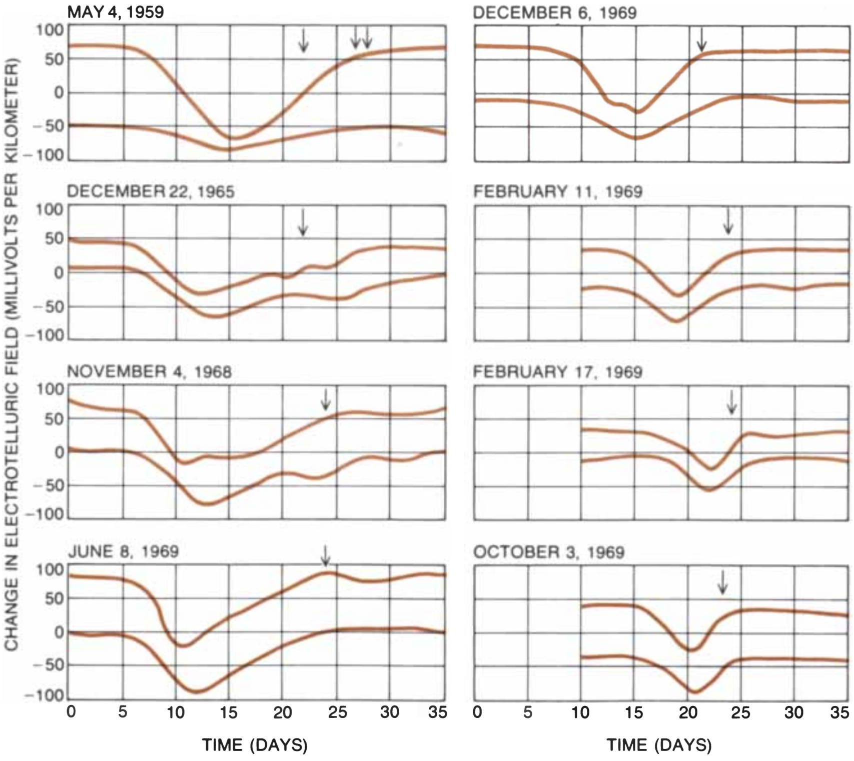
The damage caused by recent earthquakes in Japan and Alaska dramatized the fact that structures that could withstand the shaking were nevertheless toppled by foundation failure. Severe ground-shaking can cause soils to settle or liquefy and thereby lose their ability to support structures. Research on this poorly understood phenomenon is an important aspect of earthquake engineering. When it is better understood, it might be possible to take countermeasures or to institute land-use regulations that would limit construction on vulnerable soils as well as along active faults, in potential landslide areas or in coastal zones subject to tsunamis.

Some regions suffer major earthquakes frequently, others suffer them infrequently. In some places the potential for severe ground-shaking is higher because seismic waves propagate with less attenuation or because the soil resonates and amplifies the ground motion. In one city the problem following a quake is fire; in another it is flooding. Construction practices differ from region to region. Some of these factors are known explicitly; some can only be described in probabilistic terms. All of them, and other factors as well, must be combined into an overall assessment of risk on which decisions must be based. Risk assessment is a new and important part of earthquake research. Also helping to provide a rational basis for decision making about land use and construction in earthquake regions is economic analysis of such questions as to what degree, in an area with a given probability of strong earthquakes, the added costs of safer construction are offset by the potential saving of life, property and productivity.

The possibility of controlling or modifying earthquakes arose a few years ago as the result of a chance discovery. The injection of wastewater into a deep well near Denver was found to have triggered small earthquakes. Since that time both laboratory and field experiments have shown that the injection of a fluid in a fault zone reduces frictional resistance by decreasing the effective normal stress across the fault. In a sense fluid injection serves to weaken the fault, whereas fluid withdrawal can strengthen it. If a preexisting stress is present, an earthquake could result if a fault were unlocked by fluid injection. In a remarkable field test of these ideas workers from the Geological Survey injected and withdrew fluids in a water-injection well of the Rangeley oil fields in Colorado and found that in this way they could switch seismicity on and off.

The extension of these results to the control of a major active fault such as the San Andreas Fault is unlikely in the near future. Some future generation, however, may be able to modify earthquakes by the injection of fluid and the controlled, gradual release of crustal strain. Science often advances more rapidly than is expected, of course, and in any case research on the possibility of modifying earthquakes should be encouraged for the sake of the next generation, if not of our own.

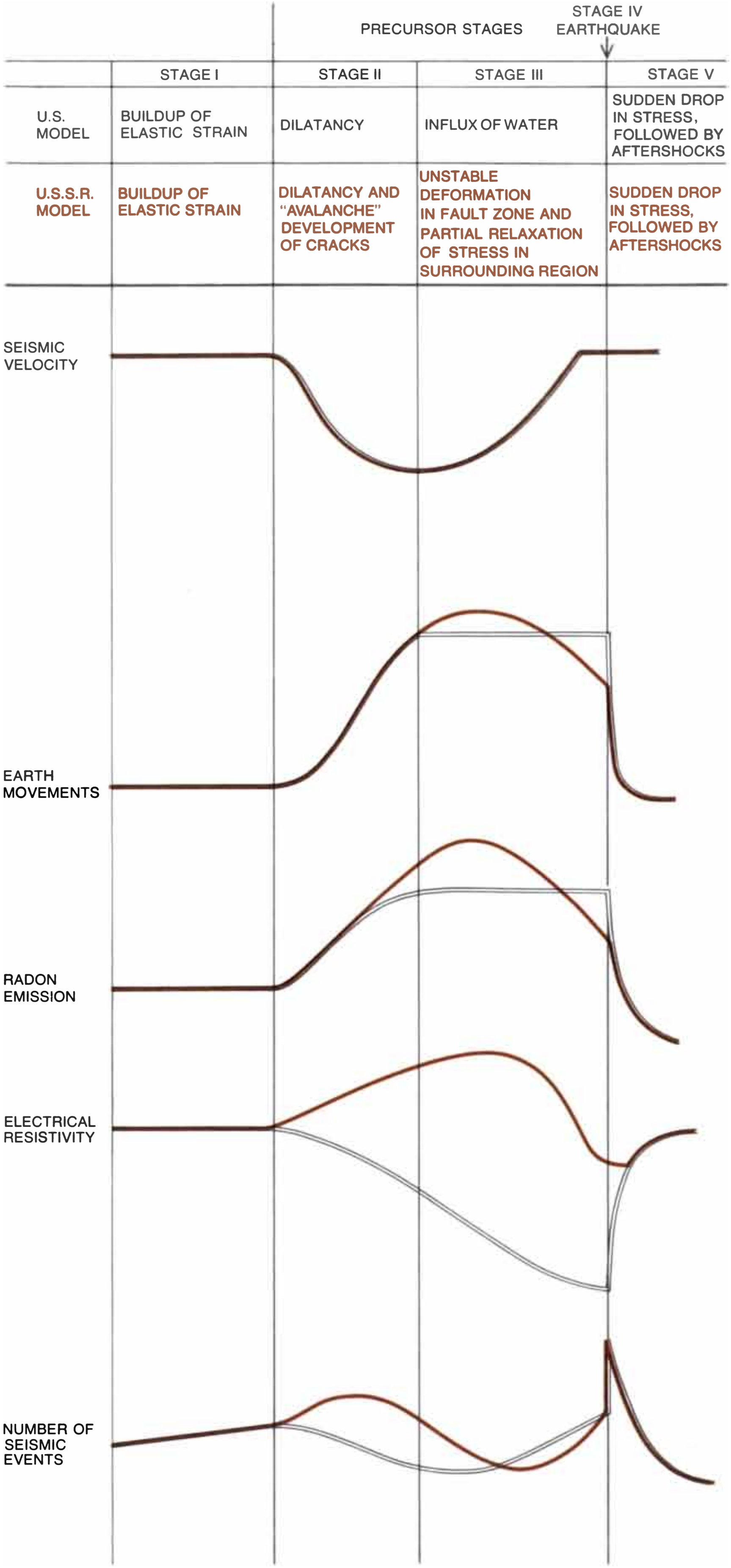
Although the number of case histories is still too small to make a positive statement about the feasibility of earthquake prediction, most seismologists would agree that prediction is an achievable goal in the not too distant future. Unfortunately the level of present effort in the U.S. is below that required to move rapidly to an operating prediction system. If a major earthquake were to



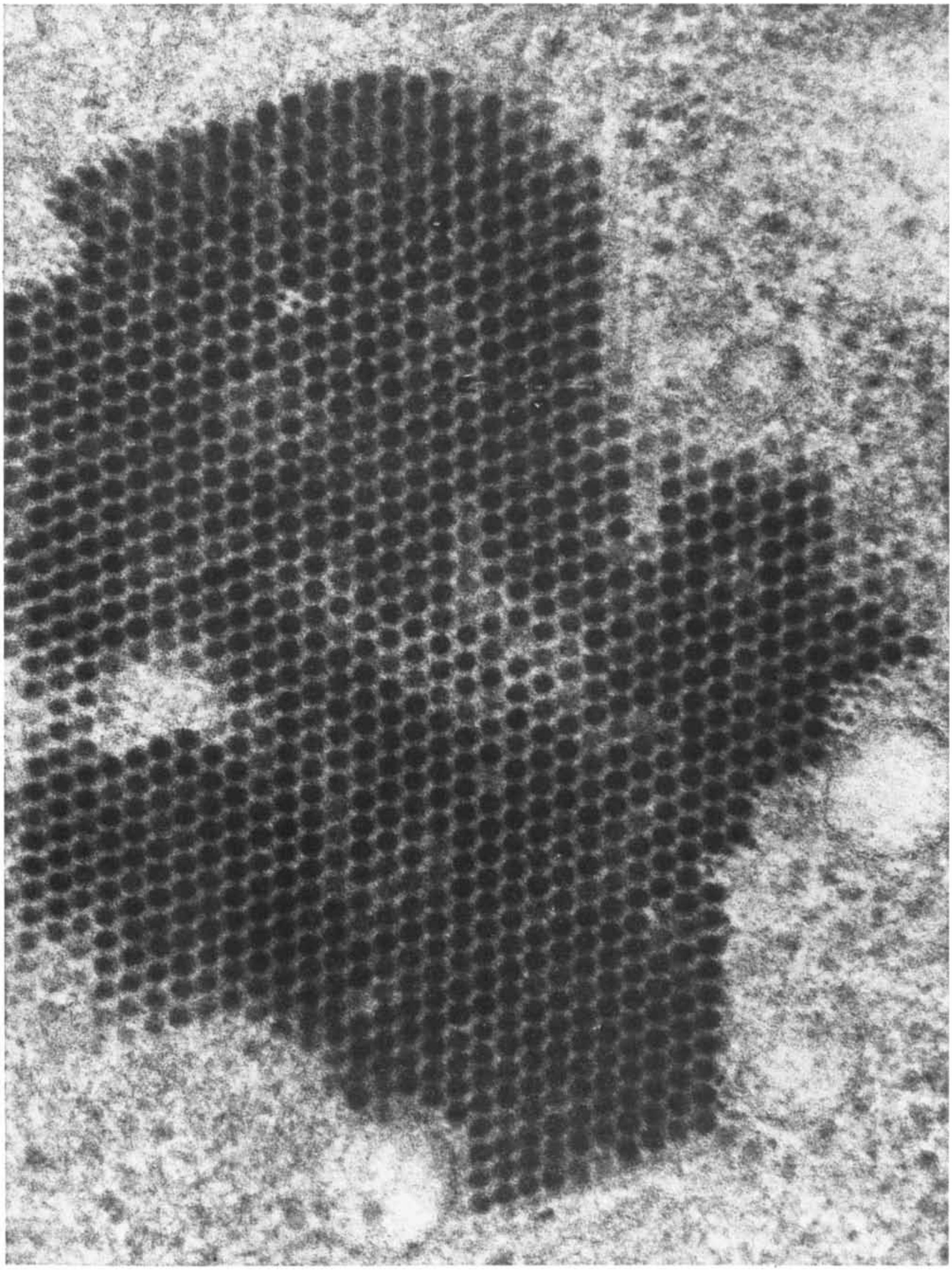
REDUCTION IN NATURAL ELECTRIC CURRENTS inside the earth just before an earthquake has also been observed by the Russian investigators. Data were obtained by recording voltage changes between points a few kilometers apart. Arrows denote earthquakes.

strike the U.S., the following day would almost certainly see abundant resources made available for a large-scale earthquake-mitigation program. (The earthquake-prediction programs of China and the U.S.S.R. were launched after severe earthquakes in each of those countries.) How does one sell preventive medicine for a future affliction to government agencies beleaguered with current illnesses?

It is proper, I believe, for scientists to assume an advocate role when they perceive an inadequate government response to some new opportunity or to some future danger. Earth scientists have a case to make. They can point to housing tracts placed in fault zones or on unstable hillside slopes. They can cite a newly built hospital that collapsed when shaken by the moderate San Fernando earthquake. The same tremor caused a dam to be stressed to near the failure point. A slightly larger shock would have resulted in casualties in the tens of thousands in the floodplain below the dam. Scientists can question the policy of a government that spends billions in construction but is unable to support research that would safeguard its own investment. They can question the wisdom of budgeting less than a tenth of a percent of the total construction investment for research on possible hazards. They can show how a research dollar invested today can yield an enormous return in lives saved and property preserved tomorrow. At a time when basic research budgets have not kept pace with the growth of the economy as a whole, earth scientists can point up the practical value to society of their new comprehension of the forces that have shaped the earth.



TWO MODELS of the mechanism responsible for earthquakes have been proposed in an attempt to put earthquake prediction on a sound theoretical basis. One view, called the dilatancy-diffusion model, was developed mainly in the U.S. The alternative, sometimes known as the dilatancy-instability model, was formulated in the U.S.S.R. The black-outlined curves show the expected precursory signals according to the American model; the colored curves show the expected precursory signals according to the Russian model. (Dilatancy is the technical term used to describe the inelastic increase in volume that begins when the stress on a rock reaches half the breaking strength of the rock.) The illustration is based on the work of Christopher H. Scholz, Lynn R. Sykes and Y. P. Aggarwal of the U.S. and V. I. Myachkin and G. A. Sobolev of the U.S.S.R.



POLIOVIRUS PARTICLES lie closely packed in the cytoplasm of a cell. About 1,000 virions, or virus particles, form a crystalline array in the plane of the electron micrograph; other virions are out of the plane. The viruses are stained with lead and enlarged about 200,000 diameters in the micrograph, made by Samuel Dales of the Public Health Research Institute of the City of New York.

The Molecular Biology of Poliovirus

The agent of a once dreaded disease has become a tame laboratory organism, an excellent instrument for studying the multiplication of a virus and the cellular machinery it converts to that purpose

by Deborah H. Spector and David Baltimore

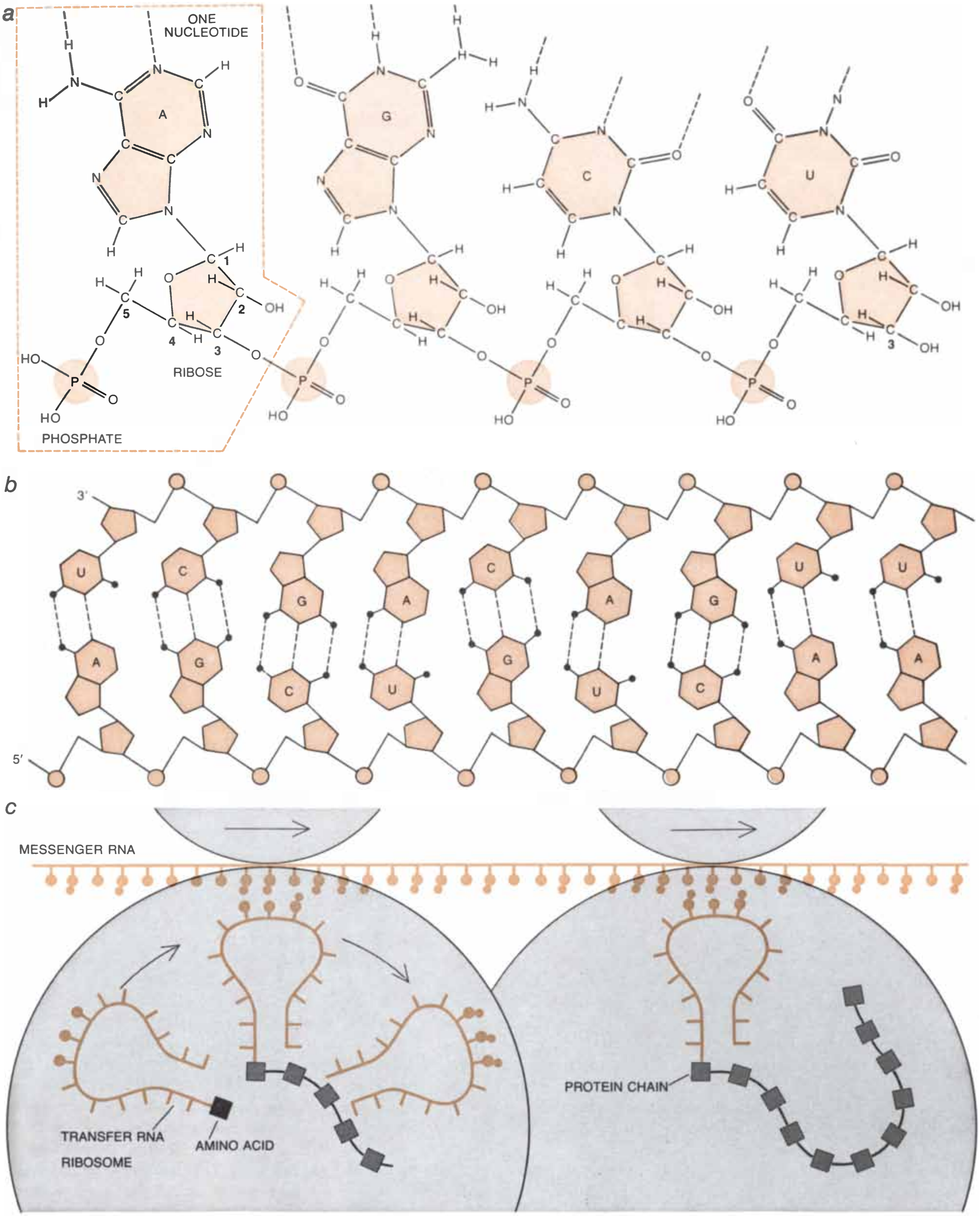
A virus is the most minimal of organisms, a protein-packaged snippet of genetic information that takes over the mechanisms of a living cell and converts them to a new purpose, which is simply the manufacture of new viruses. In the process viruses often kill cells and thus destroy tissue and cause disease. On the other hand, viruses are good laboratory animals, as it were, and even good laboratory instruments. Because they contain little genetic information and give rise to a limited number of recognizable products, they are ideal organisms in which to study the basic life processes: the replication of genetic information and its translation into protein. And because they reproduce only by appropriating particular elements of a cell's machinery, they are also excellent tools with which to probe into cells and learn how those cellular elements function.

One of the most effective viruses for these purposes is the once dreaded poliovirus. It has been intensively investigated since its discovery in 1909, originally of course because it attacks the nerve cells of the gray matter of the spinal cord and is the agent of poliomyelitis. It grows well and to high concentrations in laboratory cultures of the human tumor-cell line called HeLa. It is stable and easy to handle, and the development of the Salk and the Sabin vaccines in the 1950's rendered it safe. Poliovirus is also small even by viral standards; it seems to contain just enough genetic information to accomplish its own reproduction, with little extra genetic material to complicate matters. In the past 15 years a great deal has been learned about the multiplication of poliovirus and thus about certain cellular mechanisms. Here we shall give an account of some of those findings, after first describing the virus and its parasitic way of life.

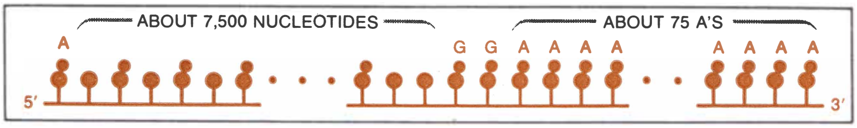
An electron micrograph of the poliovirus virion (the virus particle) reveals a sphere 27 nanometers (millionths of a millimeter) in diameter. There are indications that the sphere is actually an icosahedron, a polyhedron with 20 faces, but that shape has not yet been resolved in micrographs. The virion consists only of protein and the nucleic acid RNA. The protein is formed into a capsid, a coat enclosing the nucleic acid that is constructed of 60 identical subunits arranged in icosahedral symmetry; it incorporates 60 copies each of four proteins that are designated VP 1, VP 2, VP 3 and VP 4.

Poliovirus RNA is a single chain of about 7,500 of the subunits called nucleotides, each of which consists of a ribose sugar component and one of four organic bases: adenine, uracil, guanine and cytosine. The nucleotides are linked by phosphate groups joining the carbon atom at position No. 3 on one sugar to the carbon at position No. 5 on the adjacent sugar. The two ends of the poliovirus RNA molecule are chemically distinct: at one end there is a free No. 3 position on the sugar and at the other end there is a No. 5 position with a terminal phosphate group. The two ends of the RNA molecule are therefore designated the 5′ and the 3′ ends [*see illustration on next page*]. There are four kinds of nucleotide, named for their bases: adenylic acid, uridylic acid, guanylic acid and cytidylic acid, better known as *A*, *U*, *G* and *C*. For the most part they are distributed in what looks like random order, but their sequence is actually the code that specifies the genetic message. At the 3′ end, however, there is a unique sequence: a string of about 75 adenylic acids in a row. Such a sequence is called polyadenylic acid, or poly-*A*. Its presence in poliovirus was discovered in 1972 by John A. Armstrong, Mary P. Edmonds, Hiroshi Nakazato, Bruce A. Phillips and Maurice H. Vaughan, Jr., of the University of Pittsburgh; later Yoshiaki Yogo and Eckard Wimmer, then at the Saint Louis University School of Medicine, showed that the poly-*A* constituted the 3′ end of the molecule. Similar stretches of poly-*A* have been found on a number of cellular and viral RNA's.

In cells and in some viruses the genetic material is DNA, and RNA is the substance of various intermediary structures in the translation process; messenger RNA, for example, is transcribed from the original DNA and is thereupon translated into protein. In the poliovirus and certain other viruses that contain no DNA, the RNA is itself the carrier of the genetic information, which is translated into molecules of protein by the infected cell's machinery. Poliovirus RNA is both a genetic RNA and a messenger RNA: the information in the viral RNA is translated directly into protein, with no intermediate step of transcription into another RNA. The sequences of amino acids that make up the poliovirus proteins are specified by the 7,500 nucleotides of poliovirus RNA, and so the proteins must consist overall of about 2,500 amino acids (because it takes a code "word" of three nucleotide "letters" to specify one amino acid). That number of amino acids would make about 10 average-sized proteins. In a cell there are thousands of different proteins, and so poliovirus has only a tiny fraction of the information possessed by a cell. Since a virus is such a simple organism, however, and since it turns the cell's complex machinery to its own purposes, this small amount of information can be devastating: six



RNA, the genetic material of the poliovirus, is a chain of nucleotides (*a*), each consisting of a ribose sugar, a phosphate group and one of four organic bases: adenine (*A*), uracil (*U*), guanine (*G*) and cytosine (*C*). The sequence in which the nucleotides appear constitutes the genetic code. The phosphates join the carbon atom at position No. 3 on one sugar to carbon No. 5 on the next sugar; an RNA molecule therefore has distinct 3′ and 5′ ends. RNA replicates (forms copies of itself) through assembly of complementary nucleotides to form a complementary strand (*b*) according to base-pairing rules: *A* pairs with *U* and *G* pairs with *C*. Hydrogen bonds (*broken lines*) link complementary bases. RNA, in the form known as messenger RNA, is the medium whereby the genetic code is translated into protein (*c*) in the cellular structures called ribosomes. The RNA is "read" by small transfer RNA's that recognize three-letter codons (sequences of three bases) as codes for specific amino acids. The elongating chain of amino acids folds to become a protein.



POLIOVIRUS RNA is a single chain of nucleotides. There is an *A* at the 5′ end, followed by roughly 7,500 nucleotides that code for the viral proteins. Then, after two *G*'s, there is a distinctive sequence of about 75 *A*'s, or what is called poly-*A*, at the 3′ end of the strand.

hours after a single poliovirus attaches itself to a cell there are 100,000 new virions and the cell is dead.

The details of the events that accomplish that end—the replication of the viral RNA, its translation into proteins and the assembly of the proteins and the RNA into virions—are now starting to become clear in our laboratory at the Massachusetts Institute of Technology and in other laboratories. Beyond that, studies with poliovirus RNA and related RNA's have provided detailed information about the translation mechanisms in the eukaryotic (nucleated) cells of mammals and other organisms, and have at least narrowed down the possible functions of poly-*A*.

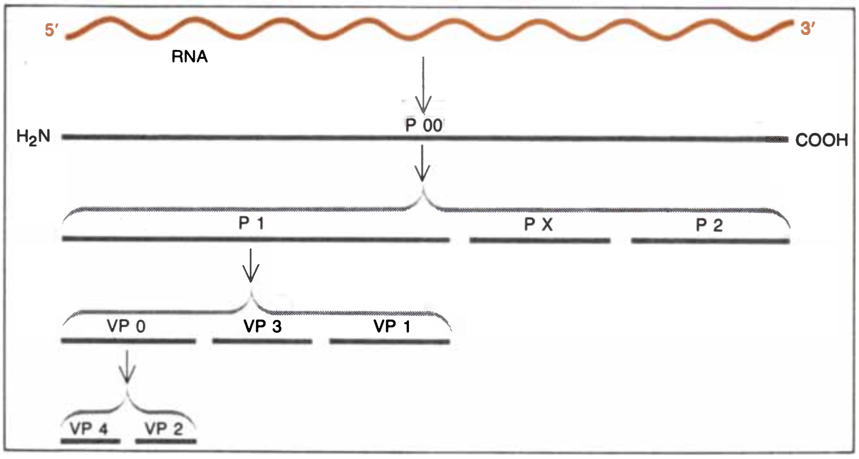
The infection of a cell in man or some other primate by poliovirus depends in the first instance on the virion's adsorption to the cell. There is an attraction between capsid protein and a cellular receptor so specific that even viruses closely related to the poliovirus bind to different receptors. (Naked infectious poliovirus RNA can infect cells that lack specific receptors, but in that case the proliferating virions are unable to bind to the cells, and so the infection is confined to a single cycle.) The binding of the virion is followed by the release of RNA from its protein coat and its entry into the cytoplasm of the cell; just how the transfer is accomplished is still not known.

Poliovirus replicates in the cytoplasm of the infected cell, not in the nucleus; even cells whose nucleus has been removed will grow the virus. The time course of viral multiplication varies considerably, depending on such factors as the specific virus strain, the host cell, the nutritional state of the cell and the multiplicity of infection. When HeLa cells are infected under optimal conditions, no new virus is detectable until about three hours after infection; during the intervening "eclipse" period the viral functions are getting under way. During this phase the virus also inhibits the synthesis of cellular RNA and protein, thus freeing most of the cell's machinery for the virus's purposes.

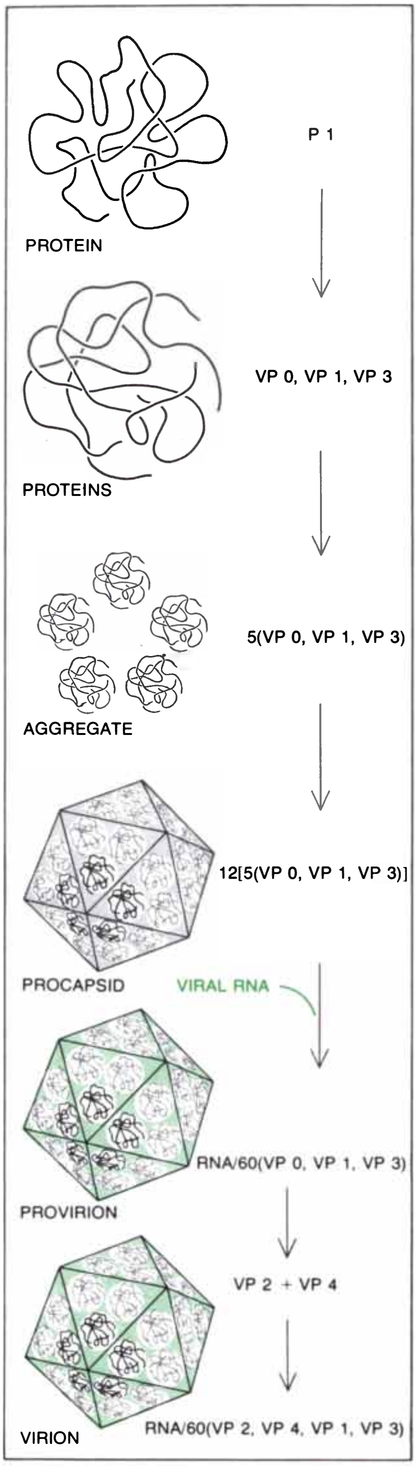
As we have indicated, the injected poliovirus RNA must serve two basic functions in order to initiate an infection. First, it must act as the messenger RNA to be translated into viral protein. More specifically, if the viral RNA is to be replicated, there must be an active replicase (a replicating enzyme, in this case RNA-dependent RNA polymerase). Since there is little evidence of replicase activity in uninfected cells, it appears that the synthesis of at least some of the enzyme is directed by the viral RNA. Second, the viral RNA must act as a template for the synthesis by the replicase of a new molecule of RNA, one that has a sequence of nucleotides that is the mirror image of the viral-RNA sequence. This is accomplished through the fundamental complementarity in the structure of nucleic acids: according to the base-pairing rules, *A* always pairs with *U* and *G* always pairs with *C*. The virion RNA is called plus RNA and the complementary RNA that is synthesized in the cell is called minus RNA. The minus RNA serves in turn as a template for the synthesis of new copies of plus RNA.

Let us follow the translation process step by step. The viral-RNA molecule becomes attached to cellular ribosomes, the major constituents of any cell's protein-synthesizing system. The RNA and several ribosomes form a polyribosome, the structure on which protein synthesis is carried out. Here the viral RNA acts as messenger RNA to specify the amino acid sequence of the various viral proteins. The first viral RNA to enter the cell has to compete in the translation process with cellular RNA. Subsequent molecules of viral RNA have an easier time because the virus somehow interferes progressively with the ability of the host cell's ribosomes and messenger RNA to interact and thus makes the ribosomes available for synthesizing viral proteins. HeLa-cell polyribosomes, for example, begin to disintegrate immediately after poliovirus infection; they are replaced by larger polyribosomes, and it is in these larger structures that the synthesis of viral proteins proceeds. One of the most interesting unanswered questions of virology is how poliovirus and many other viruses selectively inhibit host-cell protein synthesis and take over the ribosomes to make viral proteins.

Many nucleic acid molecules appear to have "punctuation points" along their length that signify the beginning and end of a gene, which can be defined as a sequence of nucleotides that codes for a single kind of protein. The poliovirus RNA molecule appears to have only two punctuation points: one "start" signal and one "stop" signal. Ribosomes therefore attach themselves near the 5′ end of the RNA and proceed along the entire length of the molecule, forming a giant chain of amino acids with a molecular weight of about 250,000. This huge protein, really a polyprotein, is then systematically cleaved by proteolytic enzymes. (The early stages of the cleavage process are very fast; we can only identify the complete polyprotein by incor-



TRANSLATION of poliovirus RNA results in the synthesis of a single very large protein designated P 00, which is immediately cleaved to yield three proteins: P 1, P X and P 2. P X and P 2 are probably forms of replicase, the enzyme that polymerizes new RNA strands. P 1 is the precursor of structural proteins that form the coat of new virus particles. P 1 is cleaved to form VP 0, VP 3 and VP 1; finally VP 0 is cleaved into VP 4 and VP 2.



CAPSID IS FORMED in a series of steps in which the capsid proteins are cleaved and aggregated. A likely pathway (details are still being studied) is shown here. P 1 is cut into three smaller molecules that remain assembled in a structural subunit, five of which aggregate. Twelve aggregates are assembled (possibly as shown here, at the vertexes of an icosahedron) into the empty procapsid. The provirion encloses a viral-RNA molecule. Finally the protein VP 0 is cleaved and the new virion is complete.

porating into it chemical analogues of some of the normal amino acids, which prevents cleavage.) The polyprotein, which we call P 00, is cleaved at two sites to make three products, P 1, P X and P 2. Probably either P X or P 2 represents the replicase enzyme; perhaps both of them do.

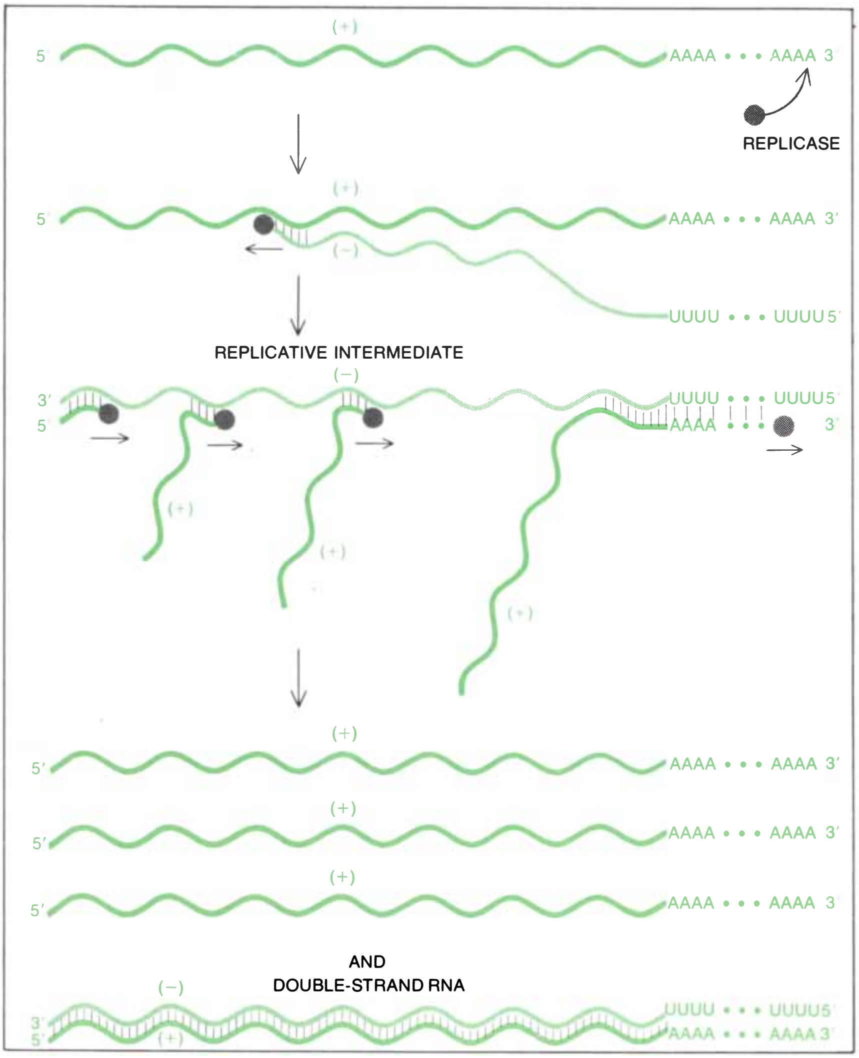
P 1 is the precursor molecule for the viral capsid proteins, which are successively cut from it and aggregated into structures of increasing size. First P 1 is cleaved to yield three direct products, VP 0, VP 1 and VP 3, that seem to remain aggregated as a single structural subunit. Five of these subunits aggregate into a fivefold molecule and then 12 of these pentamers come together to form an empty shell, the procapsid. A procapsid combines with a molecule of viral RNA, making a provirion. Finally the VP 0 molecules of the provirion are cleaved into two distinct proteins, VP 2 and VP 4. We have distinguished each of these steps in the laboratory by separating the different structures and particles on the basis of their sedimentation rates and then analyzing their protein and RNA content.

Whereas most poliovirus research has been conducted with HeLa cells, in the past four years we and our associates Lydia Villa-Kamaroff and Harvey F. Lodish have developed systems that make it possible to investigate the translation of poliovirus RNA outside the cell. When we add poliovirus to extracts of various cells (including HeLa cells) under the proper conditions, we can detect translation, at a low but measurable efficiency, into a protein that is recognizable as P 00, the largest poliovirus protein. The synthesis is initiated at only one site on the RNA molecule, which is an indication of correct translation in a cell-free system; the protein chains do tend, however, to stop growing before they are completed. We are now exploiting these cell-free systems to study various aspects of the translation process in greater detail and also to try to learn how the virus inhibits cellular protein synthesis.

Once some of the infecting RNA molecules have been translated, the enzyme replicase becomes available for the transcription of complementary RNA (the minus strand) from viral RNA (the plus strand). Presumably the replicase attaches itself to the poly-*A* tail at the 3′ end of the viral RNA and proceeds to synthesize a minus strand, which begins with a sequence of complementary poly-*U* at its 5′ end [*see illustration on opposite page*]. As each minus strand is completed it serves as a template for the simultaneous transcription of several new viral (plus) molecules. We have isolated a structure reflecting that stage: several plus strands of graduated lengths partially bonded to the minus strand on which they are being synthesized; we call such a structure the replicative intermediate. (We still do not know whether two different replicases are required for transcription from plus to minus and from minus to plus or whether a single enzyme performs both functions.) Each replicative intermediate transcribes plus strands for only a short time. Apparently a plus strand then fails to peel off and remains bonded to its minus strand, so that what was a functioning replicative intermediate becomes a double-strand RNA molecule. Such double-strand molecules accumulate until at the end of the infection cycle they account for a fairly large proportion of the total viral RNA in the infected cell.

Experiments in our laboratory and in that of Wimmer, who is now at the State University of New York at Stony Brook, suggest that the poly-*A* tail at the 3′ end of viral-RNA molecules is transcribed from the poly-*U* sequence at the 5′ end of the minus strand. The poly-*U* sequence is heterogeneous, ranging in length from about 50 to more than 200 nucleotides, whereas the poly-*A* sequence runs about 75 nucleotides. It is possible that a plus strand can be pushed off its minus-strand template before all the *U*'s are transcribed by the arrival of the next replicase molecule with its plus strand. We have noted that late in infection, when the rate of RNA synthesis has slowed down and there may be less such pushing, there are poly-*A* tails with more than 75 *A*'s on many newly synthesized viral-RNA molecules, suggesting that the length of the tail is determined by the period during which the viral-RNA molecule remains associated with the replicative intermediate.

The synthesis of poliovirus RNA can be divided into two periods. During the eclipse phase (the first three hours after infection) RNA synthesis proceeds at an exponential rate; toward the end of that period protein synthesis reaches a peak. Then for about an hour the rate of synthesis remains constant and viral RNA accumulates at a linear rate. The cause of the switch from exponential to linear synthesis is not known, but it comes just about when the cell is beginning to manufacture new virions from the newly synthesized RNA and proteins. At the end of the hour-long second phase the rate of



REPLICATION of poliovirus RNA (the "plus" strand) begins with the attachment of a molecule of the enzyme replicase at the 3′ poly-*A* end. Moving along the plus strand, the replicase assembles a complementary minus strand, which then serves as the template for the synthesis of a number of new plus strands. The minus strand has anywhere from 50 to 200 *U* nucleotides at the 5′ end, but most of the new plus strands end up with only about 75 of the complementary *A*'s; possibly they are "pushed off" the minus strand before all the *U*'s have been transcribed into *A*'s. After a number of new plus strands have peeled off, the replicative intermediate ceases to function and ends up as a double strand of RNA.

RNA synthesis falls off rapidly, as does the synthesis of viral protein; the cell is by now functionally dead, although some viral activity may continue for a while before the cell bursts. The changing events of the viral multiplication cycle, such as the phases of RNA synthesis, appear not to be the result of direct temporal control of various genes. In some other viruses particular genes are turned on and off at different times; there are "early" enzymes, for example, that are synthesized at the direction of genes activated early in the infective process. In the case of poliovirus the same set of viral-gene products is synthesized throughout the infection, albeit at different rates; something other than genetic control of protein synthesis must regulate the changes in rate.

To take one example, every newly synthesized RNA molecule has three possible destinies. It can serve as a template for the transcription of a minus strand, it can serve as a template for translation into protein or it can become associated with capsid proteins and form a new virion [*see illustration on next page*]. We know that as the infection proceeds more and more of the RNA goes to form virions, presumably because more capsid protein is available for the RNA to associate with. For a time it appeared that the poly-*A* sequence might be involved in regulating the destiny of viral RNA, but that has not proved to be the case.

Does the poly-*A* have any important role in the viral growth cycle, then? That is what we have been trying to determine for the past year. We began by analyzing the poly-*A* of RNA molecules involved in replication, translation and the formation of virions; there was no difference in the average size of the poly-*A* tail in the three conditions. The only consistent difference in size we could find was the temporal one to which we have alluded: viral-RNA molecules with longer than normal tails are produced late in the infection cycle. These long-tailed molecules are never found in virions; perhaps it is the excess poly-*A* that keeps them out. That still would not explain the normal role of normal amounts of poly-*A*, however.

In order to learn whether the poly-*A* mattered at all to the virus we asked what would happen to the infectivity of naked poliovirus RNA if it lacked the poly-*A* tail. The poly-*A* could be removed by an enzyme, ribonuclease *H*, that has a special property: it breaks down only that part of a single-strand RNA molecule which is bonded to a complementary strand of DNA. In DNA thymidylic acid (*T*) takes the place of the *U* in RNA and therefore bonds to *A*. By hybridizing a DNA molecule consisting of poly-*T* to the poly-*A* of poliovirus RNA molecules we were able to digest away more than 80 percent of the poly-*A* with ribonuclease *H*. The resulting RNA, now deficient in poly-*A*, turned out to be less than 5 percent as infectious as poliovirus RNA with a normal complement of poly-*A*. That could be because the RNA needs poly-*A* in order to get into the cell. We cannot test that possibility, however, because one can detect RNA entry only through its infectious effects. In any case it seems unlikely. We have assumed, therefore, that the poly-*A* does not prevent the entry of RNA but rather plays a critical role in some intracellular step in the multiplication of poliovirus.

When we examine cells exposed to RNA that is deficient in poly-*A*, we find no evidence that the RNA can multiply. In view of the RNA's double function in multiplication, as a template for protein synthesis and as a template for its own replication, the poly-*A* might be necessary either for translation or for replication. One experiment argues against a role for poliovirus poly-*A* in the translation of viral proteins: at least in two cell-free protein-synthesizing systems, poliovirus RNA that is deficient in poly-*A* is translated as efficiently as RNA with a normal complement of poly-*A*. That leaves us with the likelihood that poly-*A* plays some necessary role in the replication of poliovirus RNA, but the nature of that role is obscure.

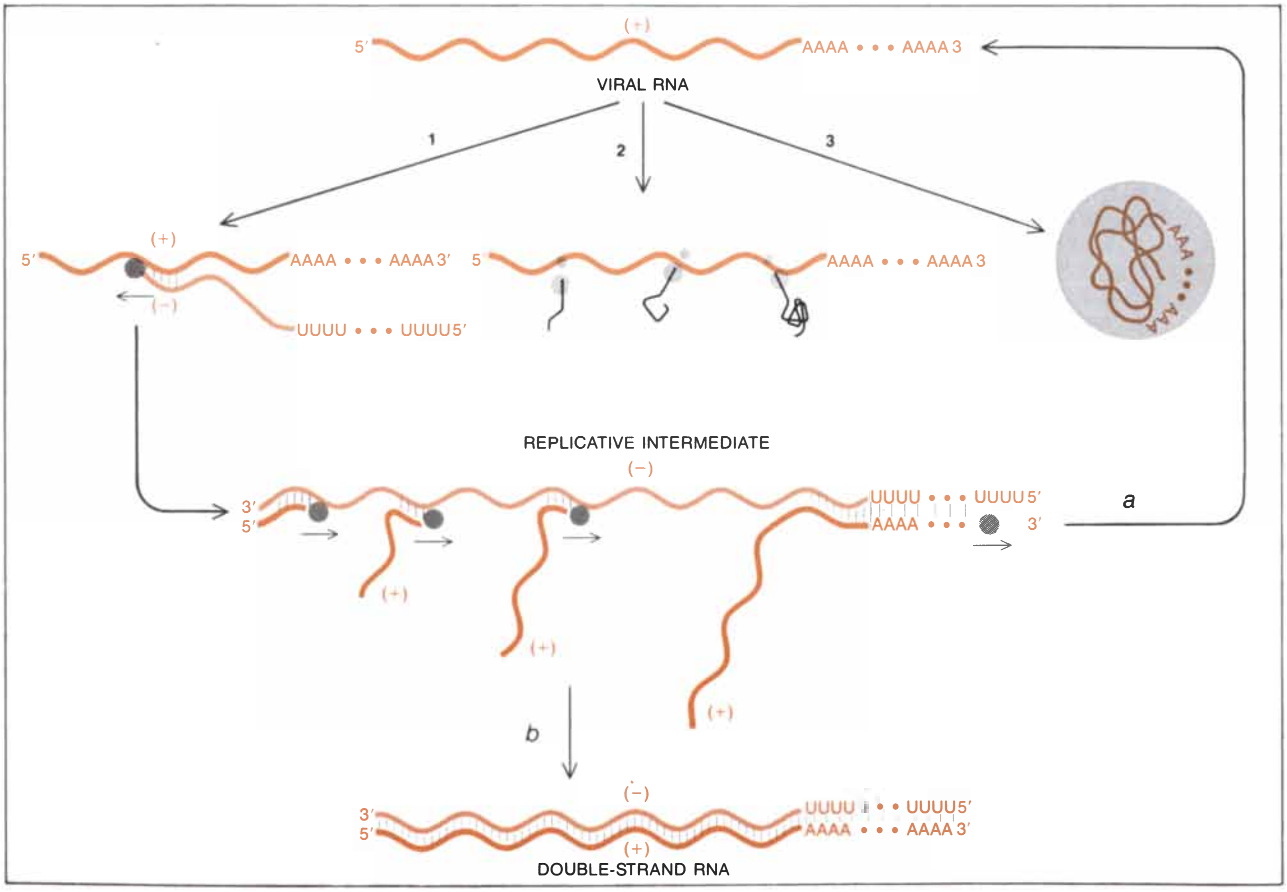
Sequences of poly-*A* are not peculiar to poliovirus RNA. They are also found as integral elements at the 3′ ends of a number of other RNA's, notably the messenger RNA of nucleated cells. In spite of investigations by a number of laboratories the biological function of these sequences is still not known.

The mechanism of messenger-RNA synthesis in nucleated cells is still an unsettled question. One widely accepted hypothesis is that the messenger RNA is derived from longer molecules called heterogeneous nuclear RNA, a collection of RNA molecules of different lengths that are transcribed from the cell's DNA and are found in the nucleus. A sequence of about 200 nucleotides of poly-*A* is somehow added to the 3′ end of some of the heterogeneous RNA molecules. This posttranscriptional addition of poly-*A* differs from the transcriptional addition we have described for poliovirus RNA, which utilizes poly-*U* as a template. For one thing, in cellular DNA there are no long sequences of poly-*T*, the DNA analogue of RNA's poly-*U*, that could serve as templates for poly-*A* transcription. Furthermore, there is a lapse of about 10 minutes between the transcription of the heterogeneous nuclear RNA molecules and the addition of poly-*A* to some of them. It appears that pieces of the heterogeneous RNA that have poly-*A* tails are cut down to the size of messenger RNA and then transported through the nuclear membrane into the cytoplasm of the cell, there to become messenger-RNA molecules that attach themselves to ribosomes and are translated into protein [*see bottom illustration on opposite page*]. While the messenger-RNA molecules are in the cytoplasm their poly-*A* tails gradually become shorter, until they are only about 100 nucleotides long. (On the other hand, there is evidence that poly-*A* is sometimes added to messenger RNA in the cytoplasm.)

Several enzymes that act to add poly-*A* to RNA have been found in both the nucleus and the cytoplasm of many kinds of cells. The ubiquity of these enzymes implies not only that the addition of poly-*A* could take place in either the nucleus or the cytoplasm but also that the poly-*A* plays some significant role in the genesis or functioning of messenger RNA. The role is presumably not in replication (as it apparently is in the case of poliovirus RNA) since no evidence has been found for the replication of cellular RNA. And the role is presumably not in the transcription of RNA from DNA because most of the poly-*A* is added some time after transcription.

It has been suggested that the primary function of poly-*A* has to do with the processing of messenger RNA and its transport from the nucleus to the cytoplasm. This hypothesis seems to be negated by the finding that there are perfectly good messenger RNA's that are made in the nucleus of cells and transported to the cytoplasm and that do not contain poly-*A*.

It seems more likely that poly-*A* is

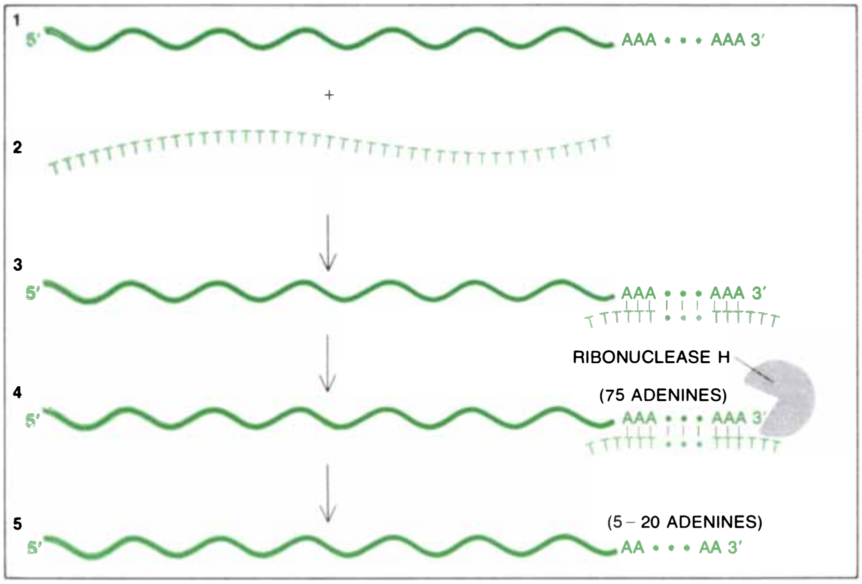


NEWLY SYNTHESIZED VIRAL RNA can go one of three ways. It can be transcribed into a minus strand (*1*), be translated into protein (*2*) or be associated with capsid proteins to form a new virion (*3*). If it is transcribed into a minus strand, the minus strand serves as a template for the transcription of more viral RNA that either is recycled (*a*) or ends up in a double strand of RNA (*b*).

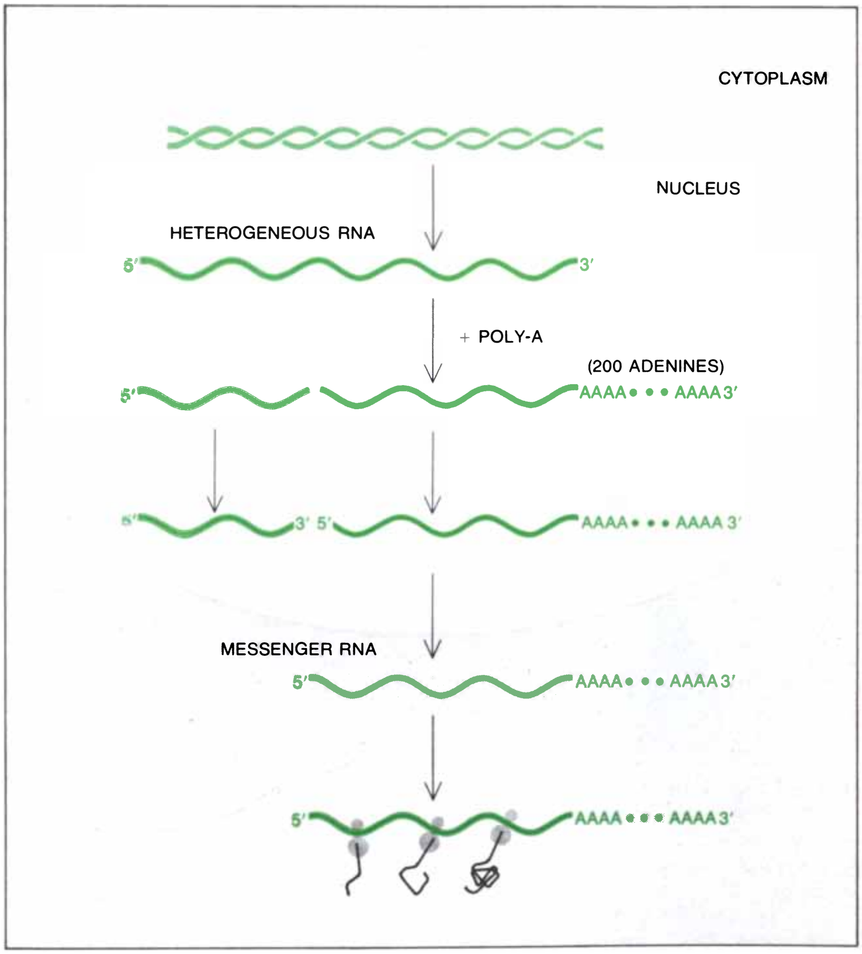
involved in the translation of messenger RNA into protein. Here the evidence appears somewhat contradictory. There are some indications, first of all, that poly-*A* is not absolutely necessary for translation. That was the implication of our findings with poliovirus RNA in cell-free systems. Similarly, the efficiency of translation of rabbit messenger RNA coding for hemoglobin or of mixed cellular messenger RNA has been shown to be only slightly lower than normal when the RNA is deficient in poly-*A*, again in cell-free systems.

Protein synthesis is less efficient in most cell-free systems, however, than it is in cells. When all the cellular machinery is present, the results are somewhat different. Investigators at the Free University of Brussels and at the Weizmann Institute of Science in Israel recently injected rabbit-hemoglobin messenger RNA with and without poly-*A* tails into the egg cells of the toad *Xenopus laevis*. (The egg's protein-synthesizing system had previously been shown to be capable of translating most messenger RNA from other species.) During the first hour after infection the two kinds of RNA supported identical rates of hemoglobin synthesis. After that, however, the rate of synthesis began to decline in the cells injected with RNA that lacked poly-*A*, and by the fifth hour it was only half the original rate. Translation of the RNA that contained poly-*A*, on the other hand, continued at the original rate for at least 48 hours. Although it is difficult to formulate any general conclusion based on the translation of a specific RNA in a very specialized cell, the result suggests that poly-*A* may help to stabilize cell messenger RNA's in some way and thus keep them operating at maximum efficiency.

The poly-*A* story, incomplete as it is, is typical of the kinds of knowledge that emerge from the study of viruses and the cells they infect. By studying poliovirus multiplication we have learned some things about HeLa cells. The work has defined specific receptors on the cell surface, has explained some elements of the cellular machinery for protein synthesis and has shown how newly made proteins are processed. In the case of poly-*A* we have been able to demonstrate a role in viral metabolism for a structure that is ubiquitous in cells but has no clearly defined role in the cell. In the future we can expect that more detailed knowledge of how a number of mammalian viruses grow will help to unravel the complex mechanisms that enable cells to live and fulfill their varied functions.



POLY-A is removed from poliovirus RNA with an enzyme, ribonuclease *H*, that digests only RNA bonded to complementary DNA. Viral RNA (*1*) is combined with poly-*T*, the DNA analogue of poly-*U* (*2*). The poly-*T* bonds to poly-*A* (*3*), and the ribonuclease *H* degrades most *A* nucleotides (*4*), leaving a viral-RNA molecule deficient in poly-*A* (*5*).



CELLULAR RNA may also have poly-*A* sequences. They are found, for example, on messenger RNA and on what is called heterogeneous RNA. The poly-*A* is apparently added to the heterogeneous RNA after the RNA has been transcribed from DNA. Then the RNA is cleaved and the strands with poly-*A* at the 3′ end are transported through the nuclear membrane into the cytoplasm, where they serve as messenger RNA that is translated into protein.

MICROCOMPUTERS

Evolutionary successor of the minicomputer, the microcomputer is a set of microelectronic "chips" serving the various computer functions. It has opened up new realms of computer applications

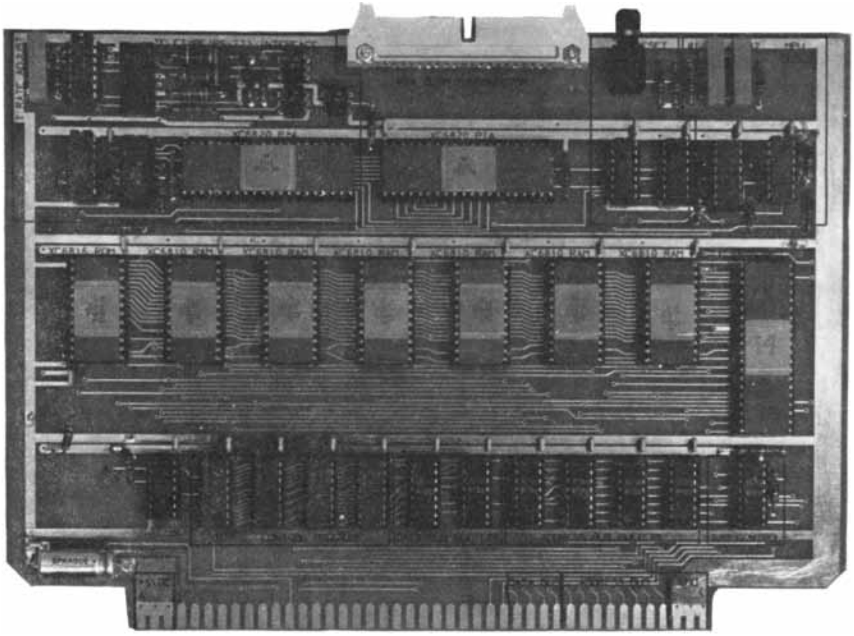
by André G. Vacroux

It is now about 15 years since the electronics industry learned to make miniature electronic circuits on a "chip" of silicon substrate by alternating processes of masked etching and diffusion. In the early 1960's the commercially available integrated circuits incorporated at most a score of components such as diodes, transistors and resistors. Production yields (the fraction of circuits that worked) were low, and packaging technology did not allow the realization of practical devices with more than a dozen leads, or connections. The basic technology, however, was so amenable to improvement and the rivalry among manufacturers was so keen that every year since then the number of components that could be economically placed on a single chip has doubled. Today chips less than a quarter of an inch on an edge can incorporate well over 20,000 components. As a result the cost per component has in 10 years dropped by a factor of more than 100, from about 20 cents to a small fraction of a cent.

The steady increase in component density, combined with parallel advances in circuit organization and complexity, has predictably led to the microcomputer, a full-fledged general-purpose machine whose logic and memory circuits can be mounted on a single plastic card that would fit comfortably inside a cigar box. Where space is at a special premium the complete microcomputer can be squeezed onto a substrate two inches square [*see illustration on the cover of this issue*].

The microcomputer is a direct descendant of the minicomputer, whose first embodiments were the PDP-5 and PDP-8, small parallel data processors introduced in 1963 and 1965 by the Digital Equipment Corporation. The PDP-8, roughly the size of a two-drawer legal file cabinet, became known as a minicomputer primarily because of its physical size, not because of limitations in its performance. Nearly as powerful as much larger computers costing several times more, it was soon widely imitated. Within a decade it had given rise to an entire industry concerned not only with "hardware" (the computers themselves) but also with "software" (the programs) and with innumerable peripheral devices and auxiliary services.

Minicomputers rapidly found their way into existing systems. To them they brought, in addition to cost reductions, the flexibility and simplified design of stored-program machines. More important, they made possible a wide range of

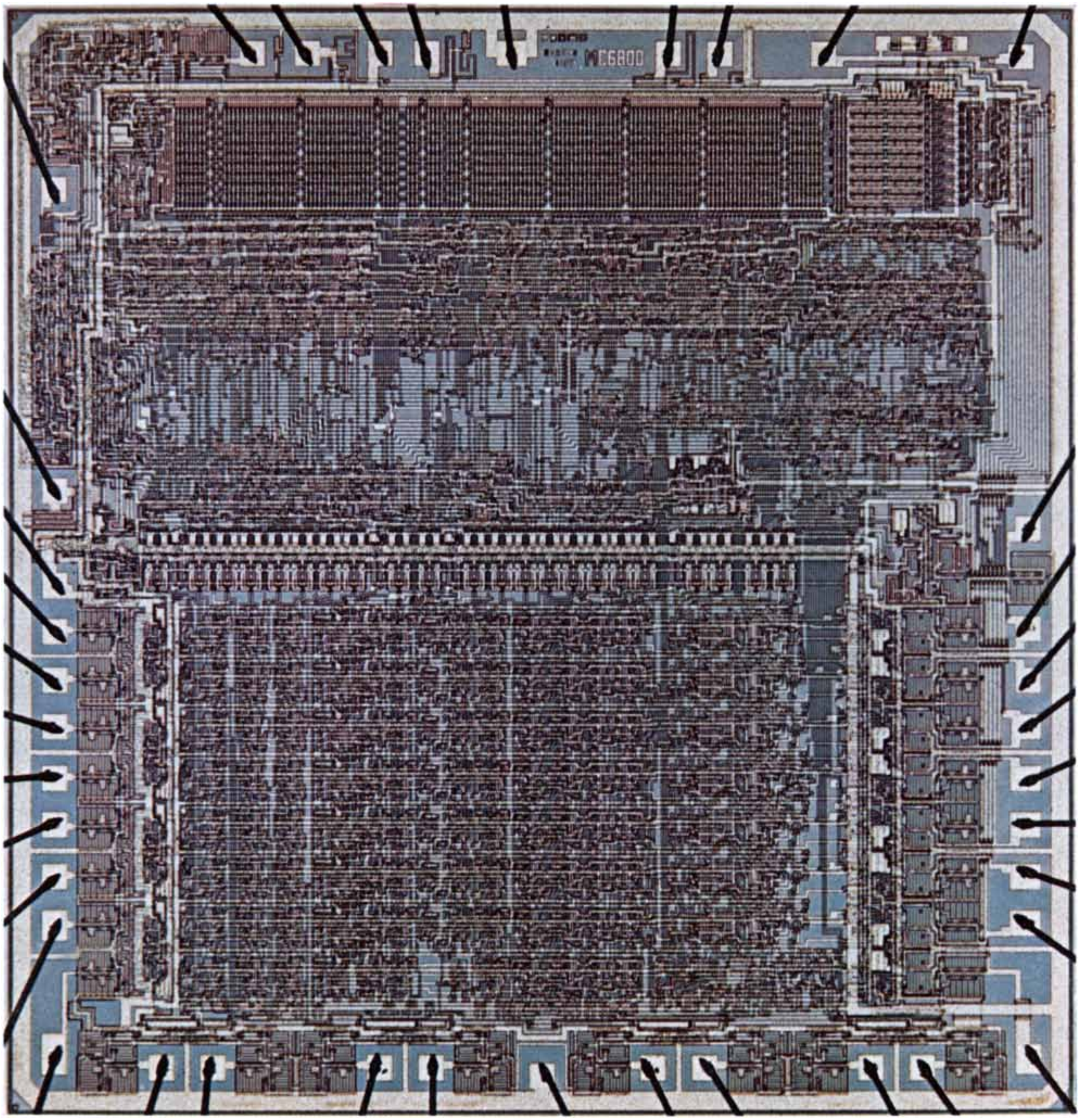


COMPLETE MICROCOMPUTER, manufactured by Motorola Semiconductor Products, is contained on a single plastic card 9¾ by 5¾ inches. More than a dozen manufacturers make similar machines. The large rectangular package with 40 pins, or leads, at the extreme right is the microprocessor, or central processing unit, which performs all arithmetic and logic functions and supervises the operation of the entire system. The microprocessor is a complex integrated circuit fabricated on a chip of silicon less than a quarter of an inch on an edge (*see illustration on opposite page*). The Motorola microcomputer also has seven memory packages: one read-only memory (ROM) and six random-access, read/write memories (RAM's). Two other large integrated-circuit packages, known as peripheral interface adapters (PIA's), manage the flow of information to and from peripheral equipment (such as keyboards and tape units) that can be connected to the microcomputer. The entire unit contains some 60,000 transistors. It operates on five volts and performs more than 100,000 operations per second. It sells for $975. Simpler machines can be assembled for much less.

new applications that called for an inexpensive resident computer.

Much of the success of the minicomputer industry was due to the dramatic advances being made in microelectronics throughout the 1960's. These advances had a twofold effect. The performance of minicomputers improved steadily, and the size and cost of the systems decreased at a rate averaging 30 percent per year. (The PDP-5 sold in 1963 for $27,000; only two years later the more powerful and significantly smaller PDP-8 was available for only $18,000.) It was clearly just a question of time until the further integration of microscopic components would lead to a microcomputer, a machine that would have tens of thousands of components on a single chip, or at most a few chips, and that would require no more than a few hundred milliwatts of power.

By the late 1960's it was still difficult to predict when single-chip computers would become a reality because semiconductor manufacturers had been concentrating on the development of what are termed bipolar devices. In such devices the currents in the transistors are carried by both electrons and "holes." (A hole is a site in a semiconductor where an electron is absent, so that whereas the electron is a negative cur-



MICROPROCESSING UNIT of the microcomputer on the opposite page is an integrated circuit laid down on a single-crystal silicon chip .210 by .217 inch, enlarged about 33 times in this photomicrograph. The chip bears about 5,400 transistors fabricated by the metal-oxide-semiconductor (MOS) technology. Since current is carried by electrons it is known as an *N*-channel device (for negative channel). Black lines projecting outward are leads bonded to pins of finished package. In small lots the unit sells for $360.

rent carrier the hole is a positive current carrier.) With bipolar devices high component densities were difficult to achieve, and substrate sizes were limited because of low production yields and problems with the dissipation of heat. At about this time, however, older fabrication problems that had plagued the metal-oxide-semiconductor (MOS) technology were solved, and it became possible to economically manufacture large unipolar devices with high component densities and low heat production [see "Metal-Oxide-Semiconductor Technology," by William C. Hittinger; SCIENTIFIC AMERICAN, August, 1973]. Initially these devices were exclusively of the *P*-channel type (PMOS): their transistors operated with positive current carriers (holes). Although PMOS devices are an order of magnitude slower than bipolar devices, they have higher component densities, lower power requirements and better production yields that make them ideally suited for applications such as electronic calculators and computer memories.

The PMOS technology contributed to a major conceptual advance in 1971, when the Intel Corporation, which had undertaken to develop a calculator chip, chose to design it as a more versatile, programmable, single-chip microprocessor. A microprocessor is equivalent to the central processing unit of a larger computer. Known as the Intel 4004, the device processes blocks consisting of four bits, or binary digits, at a time. It has 2,250 transistors on a silicon substrate measuring .117 by .159 inch, and it comes in a package with 16 pins, or leads. All that is needed to make the microprocessor into a minimal general-purpose microcomputer that could sell for less than $50 are two additional devices (a control memory and a temporary storage memory) and a master clock to time the system's operation [*see illustration on page 38*].

WORKING MICROCOMPUTER is contained in a small desk-top unit to which various input and output devices can be attached. The Intellec-8 unit shown here, built by the Intel Corporation, is coupled to a teletypewriter and a paper-tape punch for recording the machine's output. For the most part microcomputers are used not as a substitute for a larger computer but rather as specially programmed subsystems that serve as "resident" computers in a complex device or system, such as a process-control instrument or a traffic-light-control system. In developing programs that will eventually be placed in read-only memories of a resident microcomputer, it is convenient to use a specially equipped version of microcomputer itself, such as the one in Intellec 8. The price of the Intellec 8 is $3,540.

A few months later Intel introduced an eight-bit microprocessor chip, the 8008, that had more computing power and flexibility than the 4004 and was more suitable for control applications and data handling. In spite of limitations that resulted mainly from a packaging constraint of 18 pins, the 8008 was to remain for two years the only available eight-bit machine.

The two Intel microprocessors and two other early PMOS models (made by Rockwell International and National Semiconductor) were quickly incorporated into a wide range of applications, from specialized laboratory instruments to sales terminals and electronic games. As more engineers became aware of the microprocessor technology, other new uses evolved rapidly. At the same time more semiconductor manufacturers sought to establish a foothold in the rapidly expanding microprocessor market.

Although there have been a few failures, more than 20 different microprocessors have appeared within the four years since the introduction of the Intel 4004. Most of the machines still use the PMOS technology, which has been developed to such a level that one 12-bit microprocessor, made in Japan by Toshiba, has 11,000 transistors on a single chip .22 by .24 inch. Most of the newer microprocessors, however, use either the faster NMOS technology, in which the transistors operate by means of negative current carriers (electrons), or the complementary MOS (CMOS) technology. CMOS combines the PMOS and NMOS technologies to achieve a reduction in power requirements and to improve resistance to extraneous noise. The first bipolar microprocessors began to appear late last year.

Coming three decades after the first electronic computers, microprocessors benefited from the experience accumulated in system organization and computer architecture. Many advanced concepts and features, frequently unavailable in machines several orders of magnitude larger and more expensive, are standard in almost every microprocessor. One such feature, known as a stack, is a set of electronic registers organized so that the subroutines called up by a program are handled on a last-in-first-out basis: after executing one subroutine or more the microprocessor can return quickly to the main program sequence.

The parallel development of many microcomputer systems has given rise to a number of original designs and architectures, a term that refers to the organization of a computer. The everyday meaning of architecture, referring to both a style of construction and a particular way of assembling structural materials to achieve a functional goal, carries over into the field of computer design. Computer architecture describes the arrangement of the central processing unit (CPU), the memory elements for the storage of programs and data, the input and output devices and the master clock. Thus one architecture may emphasize facility of arithmetic operations and another may stress convenience of input and output operations. Whereas both have a CPU, a memory and input and output ports, the first is more suited for lengthy numerical analysis and the second is more suited for control applications and the monitoring of external equipment.

MACHINE LANGUAGE	ASSEMBLY LANGUAGE	EXPLANATION
00 100 001 00 000 010 00 000 111	ABS LXIH 2 7	Load into the memory-address "pointer" (a special register in the CPU) the address, defined by 2 and 7, of the location where X is to be found in memory.
01 111 110	MOVA, M	Move the contents of memory X into the accumulator.
00 100 011	INXH	"Increment" the pointer (that is, add 1 to the address of X to locate the address of Y).
10 010 110	SUBM	Subtract the contents of memory Y from the accumulator (so that it now contains $X - Y$).
11 110 010 00 001 011 00 000 011	JP LOC	If the accumulator is positive ($X > Y$), jump to LOC. (The address of LOC is given by the next two eight-bit words, the binary representations of 11 and 3.)
00 101 111	CMA	Since the accumulator represents a negative number ($X < Y$), the corresponding positive number is obtained, in binary arithmetic, by "complementing" the accumulator (that is, exchanging all 1's and 0's) and . . .
00 111 100	INRA	. . . by incrementing the accumulator (adding 1).
11 000 110 00 000 101	LOC ADI 5	Add 5 to the accumulator (which gives either $X - Y + 5$ or $Y - X + 5$).
00 100 011	INXH	Increment the pointer (which means adding another 1 to the address of Y to locate the address of Z).
01 110 111	MOVM,A	Move the contents of the accumulator into memory (Z).
11 001 001	RET	The subroutine has been completely executed. Return to the main program.

HIGH-LEVEL LANGUAGE (PL/M)

```
ABS:   PROCEDURE;
IF X > Y, THEN Z = X - Y + 5; OR ELSE Z = Y - X + 5;
RETURN;
END ABS;
```

HIERARCHY OF LANGUAGES originally developed for programming large computers is now used with microcomputers. This example shows the structure of three equivalent languages used with the Intel 8080 microcomputer: machine language, assembly language and a high-level (or compiler) language, in this case PL/M, a language comparable to Fortran in sophistication. The microcomputer is being instructed in a subroutine to subtract the smaller of two positive numbers, either X or Y, from the larger, to add 5 and to make Z equal to the result. The machine language consists of a collection of eight-bit words stored in the machine's read-only memory. (The division of the eight bits shown here is strictly for the convenience of the programmer.) The assembly language employs a dictionary of words, letter sequences that serve as mnemonic aids and correspond to one or more eight-bit "words" of machine language. The instructions conveyed by the two equivalent languages are explained in the third column. In this example it is assumed that X, Y and Z are stored in sequentially numbered memory locations starting at 1794 (00000111 00000010) and that both X and Y are positive integers smaller than 127 $(2^{7} - 1)$. ABS and LOC are symbolic addresses arbitrarily labeled by the assembly-language programmer. Note that one of the instructions, ADI 5, calls for two eight-bit words of memory and that two others, LXIH 2 7 and JP LOC, call for three words. The central processing unit contains a 16-bit "address pointer" capable of designating some 65,000 different memory locations. The compiler language assumes that variables X, Y and Z have been defined earlier in the program.

The CPU, or microprocessor, is the most expensive component (or group of components) of a microcomputer. It fetches the control instructions stored in the memory and then decodes, interprets and implements them. The CPU manages the temporary storage and retrieval of data and regulates the exchange of information with the outside world through the microcomputer's input and output ports. It incorporates the arithmetic and logic unit (ALU), in which all operations are performed, and a certain number of registers. Finally, it synchronizes the operation of the various components.

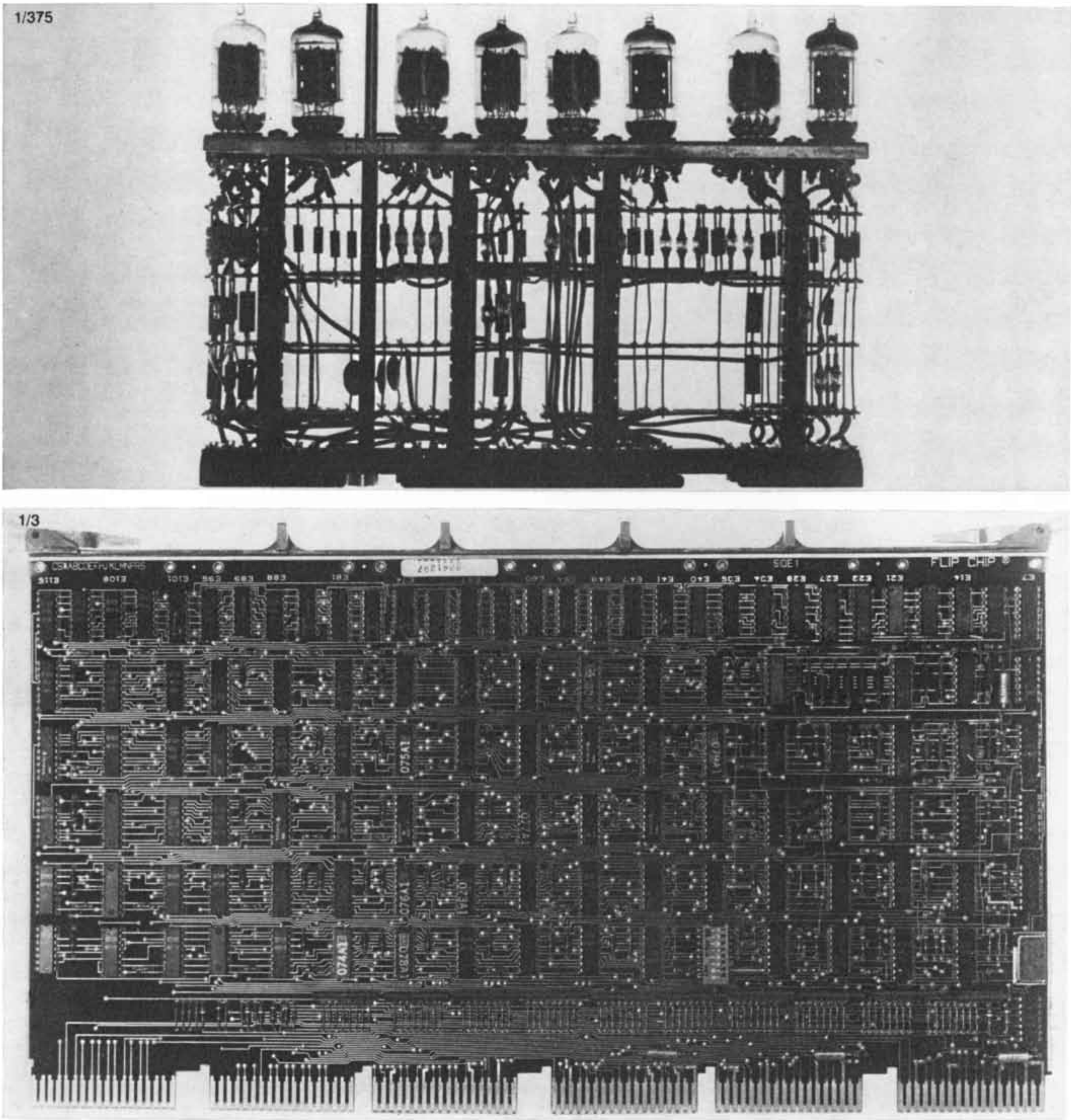
Microcomputers are usually classified according to the number of bits that can be handled by their CPU. Their performance is judged by the richness of their instruction set, by the bit efficiency of their program (the number of bits that need to be stored in the program for the implementation of a given set of tasks) and by the speed with which they execute typical programs. Such distinctions have mainly to do with their capacity for operating in "real time." If speed of operation is not a consideration, almost any microcomputer can serve in a given application. Some machines, however, may be more economical than others for particular jobs.

Four-bit single-chip microprocessors such as the Intel 4004 or the PPS-4 made by Rockwell International are particularly economical. They are well suited to systems designed for decimal arithmetic, or to systems that do not have to deal with "words" consisting of many bits. Four-bit microprocessors are the most natural choice for products such as electronic scales (which are now common in supermarkets) and sales terminals. For such applications a microcomputer control system consisting of three or four chips costing less than $50 can be placed on a printed-circuit board between four and six inches square.

Eight-bit microprocessors are the most popular at the present time. Their word length makes them a natural choice for all applications that involve communications equipment, which commonly works with eight-bit encoded characters. They have more complete instruction sets and more computing power than the four-bit units and have many of the features found in larger machines. Although many of them require more supporting devices, some of the most recent eight-bit CPU's can make a complete minimal microcomputer with as few as five devices. A few 12- and 16-bit machines have been introduced, mostly for process control and other complex tasks. Some of them are highly integrated versions of previously available minicomputers, for which they are an economical substitute when speed is not critical.

Although the design concept of putting an entire microprocessor on a single chip is an attractive one, a system consisting of several chips, each with a smaller number of components, has its advantages: higher production yields, fewer pins per device and better heat-dissipation characteristics. Accordingly some manufacturers have brought out multichip microprocessors. The usual approach is to design one common control section and several identical two- or four-bit "slices" for the arithmetic and logic unit and the registers. The slices are connected in parallel, making it possible to put together machines that have several different word lengths. This approach was first taken by National Semiconductor, which introduced a four-bit PMOS slice that could be used as a modular unit in the design of machines ranging from four to 32 bits. Bit-slice architecture, as it is called, is now the rule in bipolar microprocessors because of the necessity of limiting component densities to facilitate the dissipation of heat.

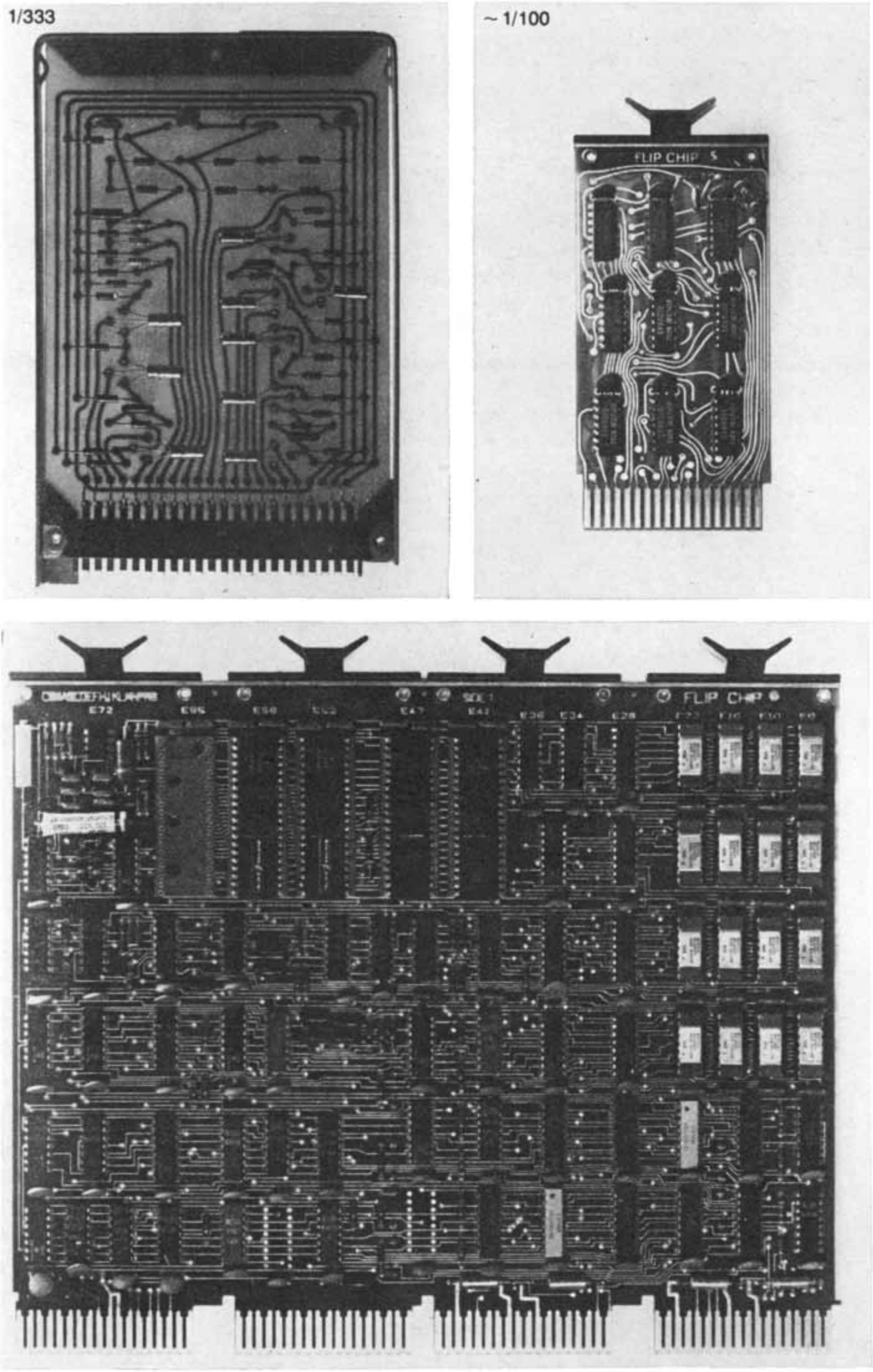
Some multichip machines have the advantage that they can be "microprogrammed" by the user. Microprogramming, which is not the same thing as programming a microcomputer, is a technique by which the functional architecture of some machines can be redefined



EVOLUTION OF COMPUTER CIRCUIT TECHNOLOGY is epitomized in this sequence of five modules from computers built over the past 20-odd years. All are depicted 45 percent actual size. The numbers at the upper left corner of each photograph show approximately what portion of the computer's total circuitry is represented by the module. The first picture (*top left*) shows a module from one of the first large commercial computers, announced in 1952. The entire computer contained some 3,000 vacuum tubes. The remaining four modules were all built by the Digital Equipment Corporation. The first (*top middle*) is from the earliest minicomputer, the PDP-5, which was introduced in 1963. The module contains nine transistors in individual "cans" and other discrete com-

By working at the microprogram level one can speed up the creation of new instructions with which certain programs are executed, and even modify the microcomputer so that it can emulate the operation of another type of machine altogether.

Although the microprocessor is the most complicated and expensive single function of a microcomputer system, it is in fact completely controlled by the memory that surrounds it. In the evaluation of a CPU attention must be given to the variety of modes made available for accessing instructions and manipulating data, since these operations significantly influence the ease of programming, the speed of execution and eventually the size of the memory itself. In many cases the most costly part of a system's hardware is the memory subsystem. For many years the memory requirements of digital systems were met with ferrite cores: tiny ceramic rings densely strung on a mesh of fine wires. Core memories are still widely used in many large computers and minicomputers, but an ever increasing number of users are turning to semiconductor memories. Fabricated on silicon chips by the same technology used to make microprocessors, semiconductor memories require less power than core memories, are easier to use, take up less space and have recently become less expensive. Since MOS memories are also far easier to integrate into an MOS microcomputer system, they are the overwhelming choice of microprocessor users.



ponents. In the next module (*top right*), from the PDP-8/I of 1968, five to 10 transistors are in each of nine integrated-circuit packages. The next module (*bottom left*), from the PDP-8/A of 1974, incorporates more than 110 integrated-circuit packages, each containing between 25 and 50 transistors. The last module, Digital's LSI-11 (*bottom right*) is a complete minicomputer, with 65,000 bits of memory, containing more than 110,000 transistors. Its list price is less than $1,000, and it is more powerful than PDP-5, which sold for $27,000.

Most semiconductor memories found in microcomputers are of the random-access type. Access to any memory location can be gained in a uniform amount of time. The two main types of random-access memory differ in their "volatility," that is, in their ability to retain their contents under various operating conditions. The computer program and tables of fixed data are mainly stored in "read-only memory," or ROM, devices. ROM's are nonvolatile: their contents cannot be altered during the operation of the computer, and the retention of stored data does not depend on a supply of power. The contents of an ROM are simply binary patterns of 1's and 0's that are programmed in advance by the user. When large volumes of identical ROM's are required, the programming is most economically done by the semiconductor manufacturer's custom-making one of the masks employed during fabrication. Such an ROM is said to be "mask-programmed." Mask-programmed ROM's can store up to 2^{14} (16,384) bits, and the cost can fall to less than a tenth of a cent per bit.

When only a few ROM's are to be programmed, as in laboratory development work, one can use "field-programmable" ROM's (PROM's). A popular type of PROM, introduced almost concurrently with the first microprocessors, can be programmed and then erased. It incorporates arrays of floating-gate avalanche-injection transistors, which are capable of trapping a charge when a pulse in excess of 40 volts is applied to them. (Microcomputers, depending on the technology used in the chip, operate at from five to 17 volts.) The charges can subsequently be removed by exposing the device to intense radiation (ultraviolet or X rays). The trapping of charge, of course, is what makes it possible to program the device; the removal of charge is equivalent to erasure. The cost per bit for PROM's is about 10 times more than the cost for mask-programmed ROM's, and the capacity of the devices is at present limited to 2^{13} (8,192) bits.

The temporary storage of data calls for memories that can be modified while the microcomputer is operating. This capability is found in the semiconductor "read/write" memories called RAM's. (Some engineers prefer the designation R/WM to avoid possible confusion with ROM.) Read/write memories are currently one of the most competitive provinces of the semiconductor industry. Typical devices can store as many as 2^{12} (4,096) bits at costs that are dropping toward a tenth of a cent per bit.

A typical microcomputer system incorporates both volatile and nonvolatile memories. Some applications, however, do not need temporary data storage beyond what can be handled by the microprocessor's internal registers, so that the read/write memory can sometimes be omitted. In other applications permanent storage is omitted and read/write memories serve to store both the program and the data. When only read/write memories are employed, the program must be reloaded each time the power is shut off (either on purpose or accidentally) unless a battery backup system has been included in the system to maintain the contents of the memory. With some of the newer CMOS semiconductor memories such a backup system can operate for weeks on as few as three or four "penlight" batteries.

In applications that require the bulk storage of large amounts of data other storage mediums such as tape cassettes or flexible (floppy) disks can be used. The peripheral devices for recording in these storage mediums were initially developed for use with minicomputers, but their cost has now decreased to justify their use in some minicomputer systems. Currently under development are all-electronic bulk-storage systems that will use charge-coupled devices or magnetic-bubble memories. They should be cheaper, faster, more capacious and more reliable than present systems.

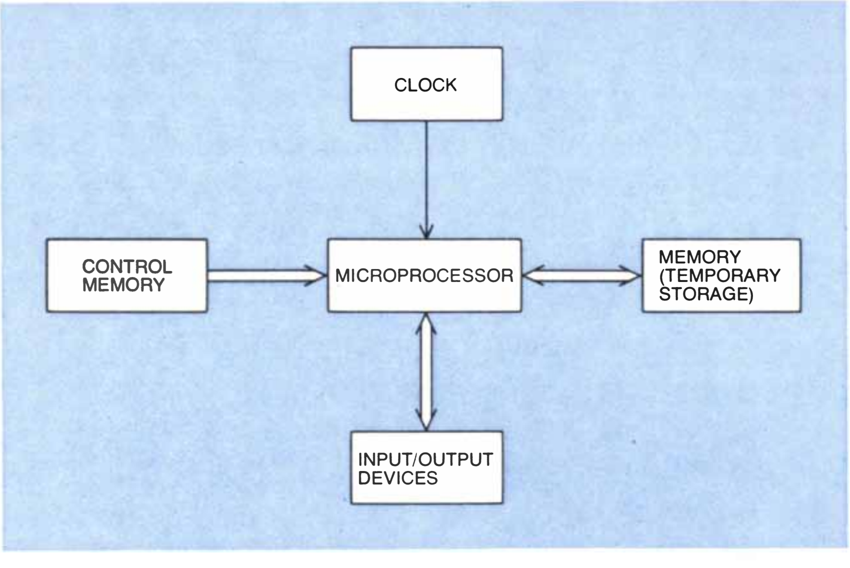
Semiconductor manufacturers saw in microprocessors an opportunity both to introduce new products and to increase their memory business. What they did not foresee was the strong demand that would arise from their customers for engineering support to exploit the new technology. The manufacturers soon discovered it was necessary to supply not only devices but also complete systems, including hardware and software of the kind available from the manufacturers of larger computers.

Hardware support was quick to come, from outside venders as well as from the semiconductor manufacturers, in the form of completely assembled subsystems. Fabricated as boards or modules, the subsystems can be directly incorporated into larger systems of many kinds. The availability of standardized but programmable modules is particularly useful for prototype development or for equipment manufactured in small series, which do not justify a custom-made microcomputer design.

It is well known that in any computer system the software is likely to account for the largest fraction of the development cost. This is no less true for microcomputers, and it is indicative of the importance of giving the user comprehensive support. Although few microcomputer software packages compare favorably with those available for larger and more expensive machines, they considerably simplify the task of putting together a microcomputer system.

A microcomputer program consists of a sequence of binary words stored in a control memory. The instructions thus defined are said to be written in machine language [*see illustration on page 35*]. Although a programmer can elect to write his program directly in this form, the process is time-consuming and prone to error. Programming is made considerably simpler by assembly languages, which are available for all microcomputers. These languages allow the substitution of mnemonic words such as ADD, SUB and JUMP for the binary words of the machine language; they also simplify the task of putting program data into a memory by giving the memory "addresses" arbitrary labels instead of absolute locations. An assembly-language program must be translated into machine language before it is committed to a memory; this conversion is accomplished by an assembler, which checks the assembly-language program for certain types of errors and, if none are found, produces the desired machine-language code. The assembler is a program that sometimes can be executed by the microcomputer itself.

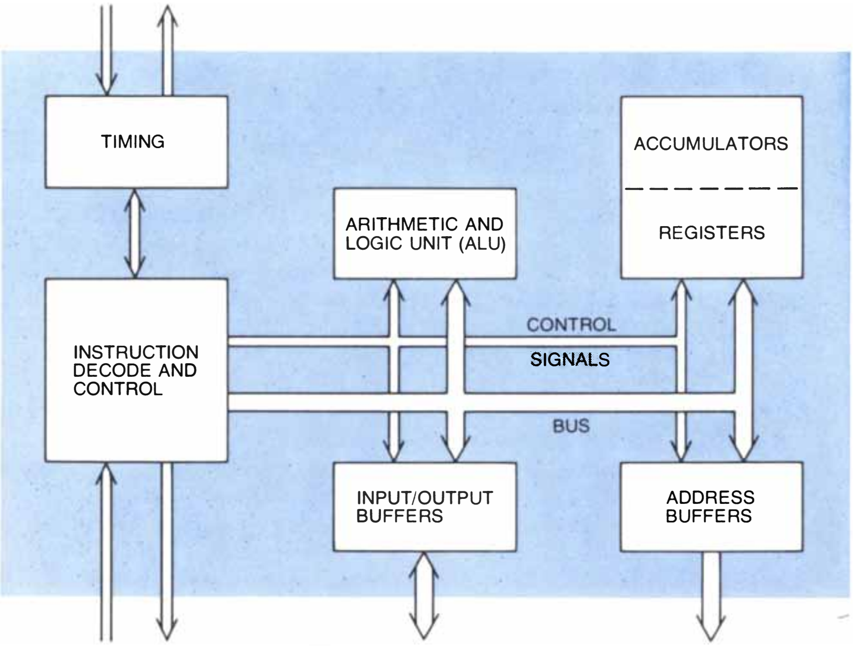
A higher level in the hierarchy of programming languages is represented by procedure-oriented languages, such as Fortran or PL/M. To translate statements written in these languages into machine language one uses a compiler. When a compiler is available, it speeds up programming, and the resulting easier to understand programs simplify the problems of documentation and mainte-



FUNCTIONAL DIAGRAM OF MICROCOMPUTER is similar to a diagram of a large conventional machine. The microprocessor performs all arithmetic operations, logic operations and data manipulation, following instructions stored in the control memory. A read/write memory provides temporary storage for the data generated during a computation. An electronic clock provides timing pulses for overall synchronization of all operations. Finally, input and output devices enable the microprocessor to communicate with the outside world.

nance. Unfortunately compilers, because of their general nature, tend to generate machine-language programs that are not highly efficient in speed of execution or in number of instructions. Typically they require from 10 to 100 percent more control memory than would have been needed if the programmer had worked at the assembly-language level. The decision whether or not to use a compiler is based on trade-offs involving the experience of the programmer, the time available for software development and, probably most important, the expected production volume of the system. Economies in memory achieved by more efficient programs can obviously justify greater software-development costs.

The differences between microcomputer software and minicomputer software go beyond the support level. In general, microcomputers have a significantly less extensive instruction set. Their shorter word lengths (four to 16 bits against the minicomputer's eight to 32) restrict the number of binary combinations that can be used in the instructions and the number of ways memories can be addressed, which increases the number of steps needed to execute an operation. This fact combines with the inherently slower performance of MOS devices to make present microcomputers between three and 10 times slower than minicomputers.



SINGLE-CHIP MICROPROCESSOR, such as the chip on page 33, contains the basic components shown in this block diagram. Instructions from the stored program are decoded by the decode-and-control unit and implemented by the circuits in one or more of the other blocks. The arithmetic and logic unit (ALU) performs arithmetic and logic operations. The registers serve as an easily accessible memory and can be used for data that are frequently manipulated. An accumulator is a special register closely associated with the ALU. It is one of the sources of data for the ALU and is the immediate destination of all its results. The address buffers supply the control memory with the address from which to fetch the next instruction. They also supply the read/write memory with the address from which the next datum is to be read or into which it is to be written. Input/output buffers are bidirectional: they can read instructions or data into microprocessor or send data out.

Microcomputers should be judged, however, not by how they compare with machines many times bigger and more expensive but by the impact they have had on the design of electronic systems. In many cases they are replacing systems based on custom-designed large-scale integrated circuits (LSI's). These specially made integrated circuits can be justified only when they are produced in volumes large enough (50,000 units, say) to amortize their development costs. They combine the advantages of high density and low power dissipation, but they cannot be modified or adapted for an application for which they were not designed. One can readily see the broad appeal of high-volume versatile devices such as microprocessors incorporating no custom-made features except the contents of a separate and easily programmed read-only memory. Not the least of the microprocessor's attractions is that it eliminates the lengthy effort required to design, "debug" and manufacture a special LSI circuit.

Until recently digital circuits were designed almost exclusively with "hard-wired" logic. Integrated circuits with a low or medium level of integration (typically with from 20 to 200 transistors per device) were interconnected by wires or by metallized paths on printed-circuit boards. Although this approach is very powerful (indeed, it is used in most of the electronic equipment manufactured today), it yields an inflexible product that is difficult to modify or improve without major redesign. Most of the microcomputers in service today were adopted as a replacement for inflexible circuitry. Their use would be justified by their advantages in number of components, fabrication costs and power requirements, but it is to their flexibility, inherent in all stored-program systems, that they owe most of their popularity. Development time is significantly decreased. Modifications, improvements or selection of options can be implemented by simple program changes even after the machine has been installed in the field; all it takes is the substitution or the reprogramming of a PROM device.

A good example of an application for which microcomputers are ideally suited is systems designed to control automobile traffic lights. Although these control systems have a constant objective (the regulation of vehicle and pedestrian flow), the requirements and constraints can change significantly, depending on state and city regulations, the intersection configuration, the traffic patterns and densities, the seasons and even the time of day. In earlier designs, which incorporated a great many separate circuits and relays, sections of each controller had to be custom-wired to meet the requirements of a specific intersection. With the substitution of microcomputers, traffic-light controllers can use common equipment throughout; the only custom-designed features are at the program level and are included in the read-only memories. Thousands of such controllers are currently being installed.

Other typical microcomputer applications are desk-top computers, compact business machines, bank terminals, check processors, payroll systems, process controllers and chemical analyzers. Over the next few years microcomputers should penetrate strongly into such areas as communications (with "smart" terminals), biological and medical research (with better monitoring and diagnostic instruments) and education (with more practical and economical teaching machines). In the 1980's microcomputers will be commonplace in the home in consumer electronic products, appliances, security devices and innumerable gadgets and toys.

The automotive industry will probably be the largest single user of microcomputers. They will replace the custom-designed LSI devices already installed by many manufacturers for monitoring brakes, lights and battery, for seat-belt interlocks and for skid control. Together with appropriate transducers and sensors, they will take up many new functions such as the control of ignition, the optimization of the fuel-air mixture, the reduction of exhaust emissions and even the continuous indication of fuel consumption in miles per gallon.

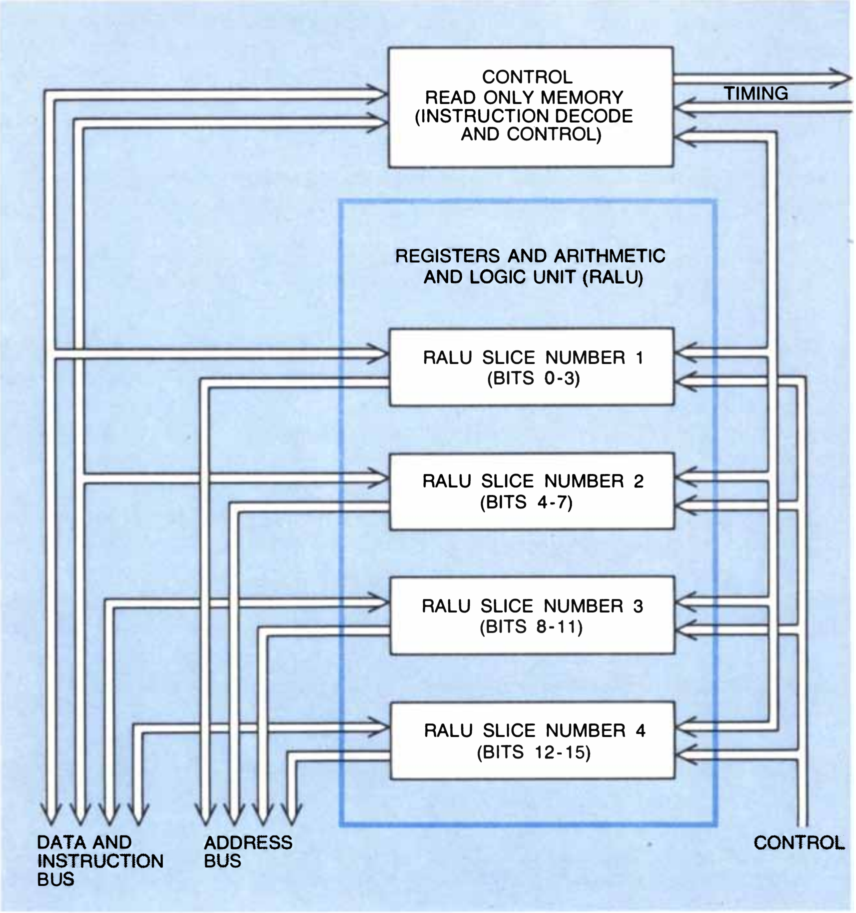
The telephone system also offers a large market for microcomputers. Already used for some time at Bell Laboratories in instruments specially designed for testing equipment, microcomputers are now being included in some of the newer Bell System central-office equipment for automatic maintenance and protection. They are also at the heart of the new family of "transaction" telephones that enable merchants and banks to check a customer's credit within a few seconds. Under microcomputer control the new telephones automatically read the customer's magnetically encoded credit card, consult a computer data base over the ordinary lines of the direct-dialing network and authorize the transaction.

The demand for microcomputers and the continuing evolution of solid-state technology will reduce the cost of microprocessors and memories, will improve production yields and will lead to higher levels of integration. Particularly promising at the present time is one of the newest bipolar technologies: integrated injection logic (IIL). This technology is characterized by a greater density of components on the substrate, higher speed of operation and lower power requirements. Microprocessors will be among the first devices to exploit these advantages.

As chips with larger numbers of components on them become more economic, more sophisticated microcomputer architectures can be expected. For example, larger read/write memories will be available on the microprocessor chips themselves, and entire portions of central processing units will be duplicated or triplicated in order to simplify the handling of multiple processes or to provide self-checking.

As microcomputers find wider ranges of applications, relatively simple systems built around a single microprocessor will continue to account for the largest fraction of the production volume. It is predicted that factory sales of microcomputers will reach $500 million by 1980 and $1 billion by 1983. One can expect that an increasing amount of attention will be given to systems incorporating more than one microprocessor, whether it is for multiprocessing (in which a certain kind of task can be shared by several machines) or for distributed intelligence (in which a large number of microcomputers with assigned roles operate under the supervision of a central machine). This trend will shift the developmental center of gravity away from the architecture and performance of single microcomputers to the organization of hierarchical systems incorporating many computers. Such systems will find applications in process control, data acquisition and inventory control. They will be a small factor in the total microcomputer market, but they will be unique in that they will bring together microcomputers, minicomputers and full-sized computers.

Any attempt to speculate on the potential applications of a new device as powerful and versatile as the microcomputer is bound to appear naïve within a very few years. The early user of a new technology is constrained by his past experience to direct his thinking along familiar channels. Moving beyond its currently obvious applications (such as traffic-control systems), the microcomputer will soon be in the hands of a new generation of designers who will be trained to regard it as a simple device, much as today's engineers look at transistors or even moderately complex integrated circuits.



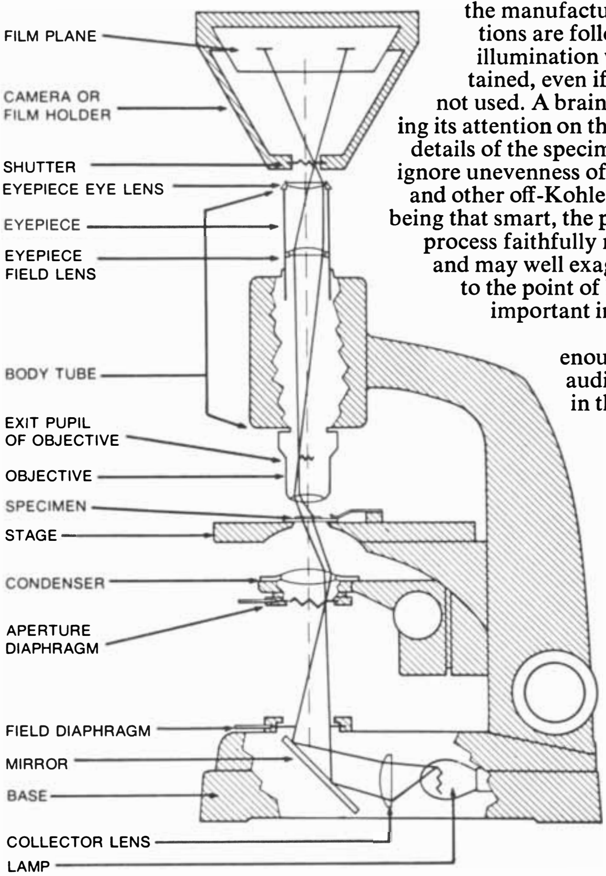
"BIT SLICE" ARCHITECTURE, represented in this functional block diagram, provides a means for achieving variable word length. In this scheme the microprocessor consists of one or two common sections that control several parallel "slices" incorporating the registers and the arithmetic and logic unit (RALU). If each slice is capable, as is shown here, of processing four bits of information, an assembly of four slices can simultaneously process 16 bits. Microcomputer on the cover of this issue follows this bit-slice architecture.

We want to be useful...and even interesting

Communication by microscope

Besides looking through a microscope, you can take pictures with it. You may want the pictures only as a record of what you saw. Or you might want them for communicating to others your thoughts on what you saw. If so – unless you feel there is no room for improvement in your photomicrography – Kodak Publication P-2,* "Photography through the Microscope," may prove helpful. In the course of telling how to get the most out of our products, it discusses things we don't even provide. For example, Kohler illumination:

There are microscope illuminators (in 10th-grade biology labs, perhaps) that do not use lenses to direct the light into the substage condenser. In that case, don't worry about Kohler illumination because you can't have it, nor can you expect as much resolving power from a high-aperture objective as theory predicts. If you have progressed beyond such equipment, carefully read the manual that comes with the microscope. Hope that if the manufacturer's suggestions are followed Kohler illumination will be maintained, even if that name is not used. A brain, concentrating its attention on the fascinating details of the specimen, learns to ignore unevenness of illumination and other off-Kohler effects. Not being that smart, the photographic process faithfully renders them and may well exaggerate them to the point of losing what's important in the subject. It can be bad enough to rob the audience of faith in the cleverness of the communicator.



THE KOHLER ILLUMINATION PRINCIPLE

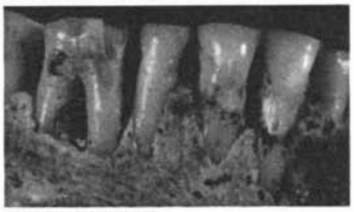
The collector lens gathers light from the emitting source, whether lamp filament or arc. Source should be in sharp focus at the aperture diaphragm. Its image should be centered with that diaphragm, and it should be large enough to extend just a bit beyond the widest opening. If it is much larger, you are wasting light. If it is much smaller, you may be losing some of the resolving power the objective could be delivering. Two further images of the light source will be formed, one by the condenser and objective in combination and the second by the eyepiece. If an image of the aperture diaphragm and within it the relayed image of the light source coincide with the plane of the shutter and are no larger than the shutter opening, the film is getting the full benefit of the light to keep down problems from long exposures, *and it is seeing the most of the specimen's structure and the least of the light source's structure.*

She must have gone to the dentist


Panchromatic


Reflected UV only


UV-excited visible fluorescence


Reflected IR only

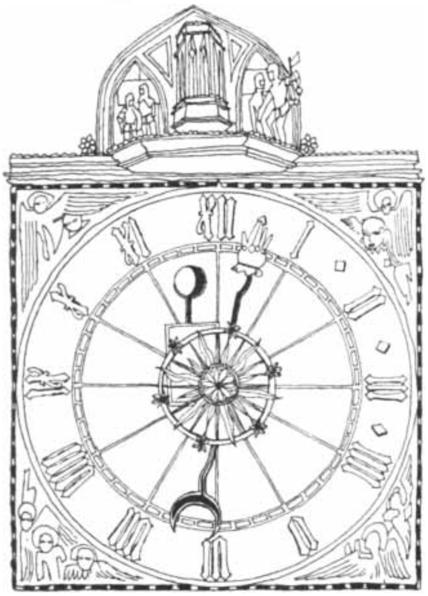

Visible-excited IR luminescence

Since the funeral took place as long ago as 600 A.D. in what the palefaces later named Jersey County, Illinois, the archaeologists took the liberty of disturbing the lady's rest. Their investigations bring new respect for the culture of her people. Therapeutic dentistry, it seems, was not beyond their technological capability. To evaluate the evidence for that, send for the article "Multi-spectrum Investigation of Prehistoric Teeth."* If more interested in the photographic methods than in the findings, get Kodak Publication M-27,* "Ultraviolet and Fluorescence Photography."

Kodak

* **KODAK Directory of Products and Services for the Health Sciences** *is available free from Dept. 55M, Kodak, Rochester, N.Y. 14650, as is the article on the evidence for prehistoric dentistry. Kodak Publications P-2 and M-27 can be ordered from photo dealers, some bookstores, or at $3.25 and $2.25, respectively, from our Dept. 454. (Prices subject to change without notice. Please add applicable taxes.)*

SCIENCE AND THE CITIZEN



No Reasonable Alternative

The unprecedented price of imported oil and the uncertainty of deliveries from the Middle East have stimulated a fractious debate about the most desirable way for the U.S. to meet its future energy needs. On various grounds, often coupling a concern for the environment with a concern for human welfare and safety, critics have objected to each of the principal avenues for expanding the U.S. energy supply: the strip-mining of coal, the mining and retorting of oil shale, offshore drilling for oil and gas and the construction of nuclear power plants. The last of these options has now been vigorously defended in a statement signed by a group of 32 scientists headed by Hans Bethe of Cornell University. Eleven members of the group, including Bethe, are Nobel prizewinners. Some excerpts from the statement follow:

"In the next three to five years conservation is essentially the only energy option.... But there must also be long-range realistic plans and we deplore the fact that they are developing so slowly.... There are many interesting proposals for alternative energy sources which deserve vigorous research effort, but none of them is likely to contribute significantly to our energy supply in this century.... If we look at each possible energy source separately, we can easily find fault with each of them, and rule out each one. Clearly, this would mean the end of our civilization as we know it.

"Our domestic oil reserves are running down. [Thus] we must...permit off-shore exploration.... We shall have to make much greater use of solid fuels, [of which] coal and uranium are the most important options.... The U.S. choice is not coal *or* uranium; we need both. Coal is irreplaceable as the basis of new synthetic fuels to replace oil and natural gas. However, we see the primary use of solid fuels, especially of uranium, as a source of electricity. Uranium power, the culmination of basic discoveries in physics, is an engineered reality generating electricity today. Nuclear power has its critics, but we believe they lack perspective as to the feasibility of non-nuclear power sources and the gravity of the fuel crisis.

"All energy release involves risks, and nuclear power is certainly no exception.... We have confidence that technical ingenuity and care in operation can continue to improve the safety in all phases of the nuclear power program, including the difficult areas of transportation and nuclear-waste disposal.... On any scale the benefits of a clean, inexpensive and inexhaustible domestic fuel far outweigh the possible risks. We can see no reasonable alternative to an increased use of nuclear power to satisfy our energy needs."

The Nobel prizewinners who signed the statement, in addition to Bethe, were Luis Alvarez, John Bardeen, Felix Bloch, Joshua Lederberg, Willard F. Libby, Edwin M. McMillan, Edward M. Purcell, I. I. Rabi, Glenn T. Seaborg and Eugene Wigner.

The SALT-free Weapons

The publicity surrounding the bilateral arms-control negotiations carried on by the U.S. and the U.S.S.R. over the past few years under the general heading of SALT (Strategic Arms Limitation Talks) has tended to focus attention on only one component of each side's arsenal of nuclear weapons: the "strategic" component, which by definition includes only those long-range nuclear weapons capable of hitting targets inside one nation from launching sites inside the other nation (or from missile-launching submarines on the high seas). At present the U.S. has a wide numerical lead over the U.S.S.R. in independently targetable strategic nuclear warheads. It is estimated that by the middle of this year the U.S. will have emplaced approximately 8,500 such strategic weapons (deliverable by land-based missiles, sea-based missiles and bombers), compared with the Russian total of perhaps 2,800. For the past four years the U.S. has been producing strategic nuclear weapons at the rate of about three per day. Under the "ceilings" set by the two superpowers in their November 1974 Vladivostok understanding the U.S. could have as many as 21,000 strategic nuclear weapons deployed by 1985.

What is ignored in this strategic arithmetic is the vast stockpiling of "tactical," or comparatively short-range, nuclear weapons by both sides. According to a recent report published in *The Defense Monitor,* the newsletter of the Center for Defense Information, the U.S. alone has approximately 22,000 tactical nuclear weapons distributed around the world. This total includes about 7,000 on land in Europe, 1,700 on land in Asia, 2,500 aboard U.S. Navy combat ships and the remainder (10,800) assigned to bases in the U.S. The total number of tactical nuclear weapons deployed by the U.S.S.R. is not public knowledge, but it is believed the Russian forces in Europe have some 3,000 to 3,500 tactical nuclear weapons at their command, giving the U.S. forces in Europe a roughly 2 : 1 advantage in this category.

Tactical nuclear weapons were first introduced in Europe by the U.S. in 1954, three years before the Russians began their deployment. At present the U.S. tactical nuclear weapons in Europe include at least four different kinds of surface-to-surface missiles (Lance, Sergeant, Honest John and Pershing), two sizes of nuclear artillery shells (155-millimeter and 203-millimeter), nuclear surface-to-air missiles (Nike-Hercules), nuclear air-to-surface missiles (Walleye), nuclear demolition munitions and nuclear depth bombs, as well as large numbers of nuclear air-to-surface bombs capable of being delivered by more than 500 U.S. fighter-bombers. The U.S. tactical nuclear forces are stationed in a number of European countries but are most heavily concentrated in West Germany. (France also maintains some of its tactical forces in West Germany.)

The researchers at the Center for Defense Information point out that their estimate of 2,500 for U.S. tactical nuclear weapons at sea is "conservative." There are currently some 284 ships and submarines in the U.S. Navy that can carry nuclear weapons, and the authors state that "the maximum loading of nuclear weapons would result in a number four times larger than the Center estimate."

The Defense Monitor report concludes by calling for a "national debate on U.S. tactical nuclear weapons," arguing that "the excessive secrecy surrounding tactical nuclear weapons hinders oversight by Congress and is unnecessary to preserve U.S. security." The very presence of U.S. tactical nuclear weapons abroad in their view invites preemptive nuclear strikes, the likely result of which would be "the destruction of the country in which they were used." In addition to the risk that "an exchange of tactical nuclear weapons would escalate into a full-scale nuclear war," the researchers contend that "the dispersion of so many tactical nuclear weapons around the world greatly increases the danger of theft, terrorism and accidents." The security of the U.S. would be enhanced, they believe, by the removal of most land-based U.S. tactical nuclear weapons from Europe, all land-based tactical nuclear weapons from Asia and all nuclear bombs and nuclear air-to-surface weapons from U.S. aircraft carriers.

What Are the New Particles?

The experiments in high-energy physics that culminated last November in the discovery of two massive subnuclear particles were undertaken in the hope of choosing from numerous possibilities a single valid theory of particle structure. The result has been quite the opposite: a rich crop of theories proposing to explain the new particles has sprouted in *Physical Review Letters,* and as yet there is no way to winnow the harvest.

The particles can be observed only as "resonances," or enhancements of the probability of an interaction between other particles; they cannot be detected directly. They were discovered at about the same time but by entirely different techniques at the Stanford Linear Accelerator Center (SLAC) and the Brookhaven National Laboratory. Estimates of their mass have been revised since the initial reports and are now given as 3.095 GeV (billion electron volts), or slightly more than three times the mass of a proton, and 3.684 GeV. At SLAC they have been named psi(3095) and psi(3684); at Brookhaven only the less massive particle has been observed, and it is called *J.*

The most intriguing characteristic of the particles is their long lifetime. They decay in 10^{-20} second, and although by ordinary standards that is a very brief interval, it is about 1,000 times as long as was expected. Other resonance particles decay by the "strong" interaction in about 10^{-23} second. Virtually all speculations on the nature of the new particles are in essence attempts to account for their curious persistence.

One large group of theories involves revisions of or elaborations on the "quark model," which postulates pointlike entities that can be combined in pairs or triplets to make up known particles such as protons and pions. Even though quarks have never been observed in isolation, the theory that predicts them is firmly established, and there is good reason to believe that the psi particles, like other kinds of matter, are made up of them. The psi particles might not, however, be made of the same kinds of quarks as ordinary matter.

The original formulation of the quark model required three quarks; a decade ago another formulation was proposed, incorporating a fourth quark with a new property arbitrarily called charm. One of the first suggested explanations of the new particles was that psi(3095) might be a charmed quark and a charmed antiquark, bound into an "atom" of charmonium. The delayed decay of psi(3095) into ordinary particles could then be explained by a tendency to conserve charm. In this scheme and similar ones discussed by several theorists psi(3684) is merely an excited state of charmonium. At least two additional excited states are predicted, and it should be possible to detect the radiation emitted in transitions between them.

Another modification of the quark hypothesis postulates a property of matter arbitrarily called color. Each of the three original quarks is assumed to come in three colors, so that there are nine quarks all together; ordinary matter does not display color because it consists of a quark and an antiquark or of three quarks with all the colors in equal proportion. It has been suggested that the psi particles are the first observed states of colored matter, that is, they may consist of a combination of quarks having net color. Their lifetime would be explained by a tendency to conserve color; once again a multiplicity of other particles is predicted.

The charm and the color hypotheses have in common the supposition that the decay of the psi particles into ordinary matter is inhibited by the conservation of a previously unobserved property. Several theorists have offered explanations based on other hypothetical properties. One group has postulated a new property called paracharge, which would be conserved in the strong interactions but not in electromagnetic interactions, which generally take about 10^{-21} second. Another theory proposes the new property gentleness, which likewise would be conserved in strong interactions but not in electromagnetic interactions. This theory assumes an ensemble of six quarks: the three usual ones and three gentle ones.

Another hypothesis that attributes the psi lifetime to conservation predicts that the new particle is a chimeron, a particle that exhibits certain properties some of the time but different properties at other times. If this formulation should prove correct, interactions involving the psi particle would violate the conservation of electric charge and parity.

A quite different approach to the problem of explaining the psi is taken by two groups of investigators who suggest that it is merely an aggregate of previously observed particles. The most likely candidate is the omega particle, which has a mass of 1.672 GeV; the psi(3095), they propose, is a bound state of an omega and an antiomega particle. The decay of the psi is then equivalent to the annihilation of the bound pair, which might be inhibited by the angular momentum of the system. No new conserved properties are required by this theory, but many similar bound states of other particles would be expected.

Perhaps the most sweeping theories of all are those that require not a new property of matter but a new kind of force or interaction. Two groups of physicists have made such proposals. These explanations have the advantage that the strength of the interaction can be adjusted to exactly the value required to account for the lifetime of the psi particle. On the other hand, they demand a fundamental revision of the doctrines of physics merely to explain the existence of two anomalous particles.

Experiments capable of testing at least some of these hypotheses are under way at SLAC and at several other laboratories. Not all the proposals are immediately subject to test, but it is hoped that this time the experimental work will at least reduce the number of viable theories rather than breeding more.

Acetaldehydism

It is clear enough that prolonged, heavy consumption of grain alcohol damages the liver, the heart and other tissues—but why? The mechanism of these familiar effects has never been clearly established; indeed, much of the blame has often been put on the nutritional deficiencies that frequently accompany serious alcoholism. Now it appears that acetaldehyde, into which alcohol is converted in the liver, may be largely to blame. Alcoholics may actually suffer from acetaldehydism.

Acetaldehyde was known to be a potent cell poison and so has been under suspicion for some time. The primary enzyme system by which alcohol is oxidized to form acetaldehyde is quickly saturated, however, so that the acetaldehyde level does not keep climbing as alcohol is consumed. That seemed to argue against a major role for acetaldehyde in causing damage that is clearly associated with the ingestion of alcohol. Charles S. Lieber and his associates at the Bronx Veterans Administration Hospital and the Mount Sinai School of Medicine undertook a precise study of the relation of alcohol and acetaldehyde blood levels. Their results were reported in *The New England Journal of Medicine* by Mark A. Korsten, Shohei Matsuzaki, Lawrence Feinman and Lieber.

The investigators infused into the veins of six patients with chronic alcoholism and five control patients enough pure grain alcohol (ethanol) to attain a concentration in the blood normally associated with moderate intoxication, and then monitored the alcohol and acetaldehyde levels continually for eight to 10 hours. Both levels were predictably high after the infusion. The alcohol level fell off steadily as the alcohol was metabolized, but the acetaldehyde level did not vary for some time; it remained on a plateau until a certain alcohol level (remarkably similar in all patients) was reached, at which point it fell abruptly. The acetaldehyde plateau was significantly higher in the alcoholics than in the control patients, however.

Lieber and his associates interpret their results as indicating, first of all, that there is a logical dose-response relation between alcohol and acetaldehyde, albeit a complex one. The abrupt end of the acetaldehyde plateau (at an alcohol level at which the primary oxidative mechanism for alcohol, mediated by the enzyme alcohol dehydrogenase, is still saturated) could indicate the presence of a second enzyme system that handles high concentrations of alcohol. Most important, the high level of the plateau in alcoholics suggests that they are indeed different: that they either metabolize alcohol more quickly or fail to break down acetaldehyde as quickly. The latter seems to have been the case in the patients under study. There may be a vicious circle at work. Acetaldehyde is known to cause liver damage; the damage may interfere with acetaldehyde metabolism in the liver, leading in turn to still higher acetaldehyde levels and more liver damage.

Shang China and Olmec Mexico

How is it that, no later than 1,000 years before the Christian Era, high cultures flourished in both the Old World and the New? Did the peculiar blend of technology and social organization that we call civilization arise only once and then diffuse worldwide, or did each major civilization develop independently? Writing in *American Anthropologist,* Betty J. Meggers of the Smithsonian Institution asks this still unanswered question while presenting data strongly suggestive of a transpacific contact in the second millennium B.C. between the first civilization in China, the Shang, and the first in the New World, the Olmec.

Meggers suggests that unprejudiced analysis of the parallels between Shang and Olmec may allow a decision as to which of three leading hypotheses is correct. These are, first, that civilization arose only once, so that all examples found anywhere in the world are related; second, that civilization arose independently in the Old World and the New, and third, that civilizations have arisen repeatedly. Meggers herself is an advocate of the first hypothesis; she and her husband, Clifford Evans, have presented persuasive evidence of involuntary prehistoric traffic between Japan and Ecuador (see "A Transpacific Contact in 3,000 B.C.," by Betty J. Meggers and Clifford Evans; SCIENTIFIC AMERICAN, January, 1966).

The Olmec civilization arose abruptly and without apparent preliminary stages among the agricultural villages of southern Mexico in about 1200 B.C. The most spectacular Olmec sites lie along the coast of the Gulf of Mexico in the states of Veracruz and Tabasco. Olmec influence and trade in exotic materials such as jadeite, hematite, basalt and obsidian, however, extended to the Pacific coast states of Guerrero, Oaxaca and Chiapas and even into neighboring Guatemala. The principal Olmec monuments are both flat-topped and pyramidal earth platforms oriented on a north-south axis, inscribed altars and stelae made of basalt, colossal basalt heads and great quantities of polished jadeite, some in the form of axe blades and adze blades, some in the form of small figurines and some in apparently ceremonial forms, including elongated tablets. Olmec religious sculpture represents, among other deities, a feline form (the "were-jaguar") and a precursor of the better-known Toltec and Aztec deity, the feathered serpent.

The Shang civilization arose abruptly among agricultural villages in the lower Yellow River basin around 1750 B.C. Its best-known sites are in northern Honan, but Shang remains are found from Shantung in the east to Shensi in the west and as far south as Hunan. Anyang, the Shang "capital" in the 14th century B.C., contains numerous earth platforms oriented on a north-south axis. Shang craftsmen worked in jade and bone and were expert bronze casters. Shang officials evidently carried jade tablets of various shapes and sizes as badges of rank; most of the jade was imported from Sinkiang, some 1,400 miles to the west. The principal Shang religious practice appears to have been ancestor worship, but spirits of the sky and the earth were also recognized. The earth spirit was often depicted as a tiger and sometimes as a snake.

Meggers summarizes the principal Shang-Olmec parallels as follows:

Writing: Divinatory Shang inscriptions are the earliest-known form of Chinese ideographic writing. No significant Olmec inscriptions are known, but certain designs that are frequently repeated in Olmec carvings may have symbolic meaning; three of the Olmec designs resemble Shang ideographs.

Symbols of rank: Certain Olmec jade tablets have the same shape as certain Shang jade rank badges; they may have had a similar function.

Architecture: Both Shang and Olmec centers feature north-south-oriented earth platforms and buried drains.

Settlement pattern: Neither the Shang nor the Olmec centers seem to have housed large resident populations. They appear to have been occupied mainly by administrators, religious leaders and craft specialists.

Iconography: A feline deity is featured in both Shang and Olmec religious art; in both, many of the representations omit the animal's lower jaw. The parallel between the Olmec snake with feathers and the Shang flying snake, or dragon, may also be significant.

If the rise of the Olmec civilization was a consequence of transpacific contact with China, Meggers writes, the implications are far-reaching. She notes that many authorities "already accept the interrelationship of all civilizations in the Old World." If diffusion from the Old World is responsible for the appearance of civilization in the New World, Meggers points out, all civilizations "would be traceable to a single origin."

Videograph

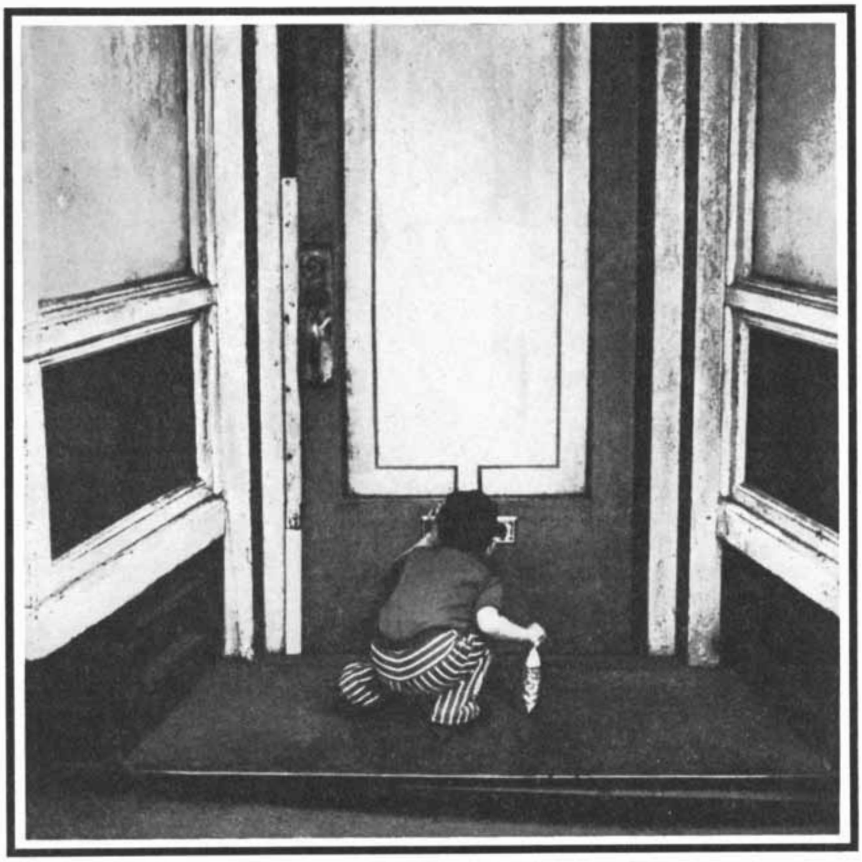
A "videodisc" system that presents recorded pictures and sound on a standard television set will be put on the home-appliance market next year by North American Philips Corporation and MCA Inc. (formerly the Music Corporation of America). The player looks somewhat like a phonograph and the record resembles a phonograph record. The distinctive feature of the system is an optical "stylus," which replaces the needle of the phonograph. The stylus is a scanner employing a one-milliwatt helium-neon laser as its sensor.

Because of this feature the record is not touched by the playing device and therefore will not wear out or deteriorate. The optical stylus also makes it possible for the viewer to pick out randomly part of a record for display, to run a scene in reverse or in slow motion, to look at single frames and to view a film frame by frame. The sponsors plan also to make available a device resembling a pocket calculator, whereby the viewer can enter numbers from a catalogue and summon up individual pictures, such as photographs of famous paintings.

The two companies plan to sell the player for about $500 and the records for from $2 to $10 each, depending on their length and content. Among the recorded programs that will be available are commercial motion-picture films, instructional films and such nontheatrical subjects as sports events and newsreels.

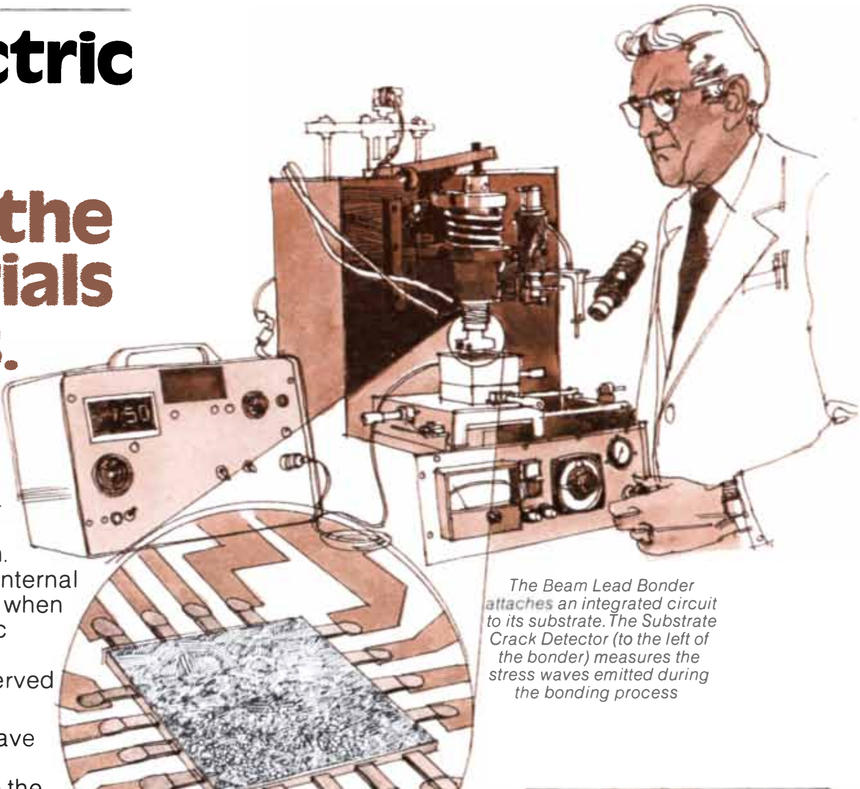
Featherburgers

In February this department carried a brief account of investigations into the breeding of featherless chickens, which would have the advantage of not wasting feed by converting it into inedible feathers. Harvey C. McCaleb of Columbia, Mo., writes to point out that feathers are not necessarily inedible. He cites the work of John Cherry of the University of Georgia, who is studying methods of isolating the protein of feathers so that it can be used as a food additive. According to Cherry, chicken-feather protein is already being used as a supplement to feeds, and there is no reason why it cannot be added to dog, cat and even human food. McCaleb comments: "I think 'Science and the Citizen' should be fair to our feathered friends, as there are those who think there is a great future for feathers."



Western Electric Reports:

Listening to the cry of materials under stress.



The Beam Lead Bonder attaches an integrated circuit to its substrate. The Substrate Crack Detector (to the left of the bonder) measures the stress waves emitted during the bonding process

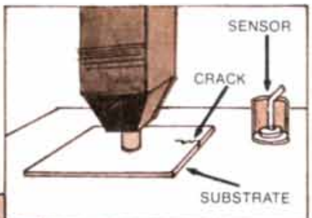
Opera stars have shattered a glass by singing at its resonant frequency. The phenomenon: Acoustic standing waves in the glass can exceed its breaking strength. Acoustic waves can also be set up by internal metallurgical processes. For example, when a material is strained beyond its elastic limit, it emits a "cry." This characteristic signal, called a "stress wave," was observed by Bell Labs' scientists in 1948.

Now Western Electric engineers have applied acoustic techniques to non-destructive testing of materials used in the manufacture of Bell System telephone equipment. Many manufacturing operations, such as drilling, welding and thermocompression bonding, place materials under a great deal of stress, which may lead to cracking.

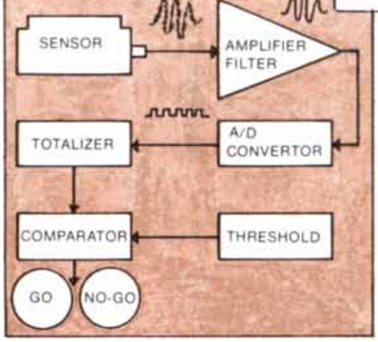
For example, when silicon integrated circuits are bonded to their ceramic mounts, the ceramic may sustain very small "microcracks." While that may be of no immediate consequence, it can lead to failure years later, when moisture accumulated in the microcracks may disrupt circuitry. Detecting microcracks with a microscope or by destructive sampling is both costly and tedious.

Western Electric's Engineering Research Center has developed the bold idea of detecting microcracks by *listening to the sound of the bonding process*. The "cry" of the unit being bonded is compared to sounds normally emitted during a satisfactory bond. Anything outside the normal range indicates microcracking. The process can detect microcracks smaller than a thousandth of an inch.

Benefit: Monitoring stress wave sounds has made possible real-time, non-destructive testing of ceramic substrates on the production line, to help ensure that each circuit will give years of uninterrupted service to Bell System users.



An acoustic sensor, attached to the base of the bonder, detects the sound of a microcrack



Thresholds were determined experimentally at the Engineering Research Center. A go/no-go signal tells the operator whether to accept or reject the unit.

Western Electric

We're part of the Bell System.
We make things that bring people closer.

RANDOMNESS AND MATHEMATICAL PROOF

Although randomness can be precisely defined and can even be measured, a given number cannot be proved to be random. This enigma establishes a limit to what is possible in mathematics

by Gregory J. Chaitin

Almost everyone has an intuitive notion of what a random number is. For example, consider these two series of binary digits:

01010101010101010101
01101100110111100010

The first is obviously constructed according to a simple rule; it consists of the number 01 repeated ten times. If one were asked to speculate on how the series might continue, one could predict with considerable confidence that the next two digits would be 0 and 1. Inspection of the second series of digits yields no such comprehensive pattern. There is no obvious rule governing the formation of the number, and there is no rational way to guess the succeeding digits. The arrangement seems haphazard; in other words, the sequence appears to be a random assortment of 0's and 1's.

The second series of binary digits was generated by flipping a coin 20 times and writing a 1 if the outcome was heads and a 0 if it was tails. Tossing a coin is a classical procedure for producing a random number, and one might think at first that the provenance of the series alone would certify that it is random. This is not so. Tossing a coin 20 times can produce any one of 2^{20} (or a little more than a million) binary series, and each of them has exactly the same probability. Thus it should be no more surprising to obtain the series with an obvious pattern than to obtain the one that seems to be random; each represents an event with a probability of 2^{-20}. If origin in a probabilistic event were made the sole criterion of randomness, then both series would have to be considered random, and indeed so would all others, since the same mechanism can generate all the possible series. The conclusion is singularly unhelpful in distinguishing the random from the orderly.

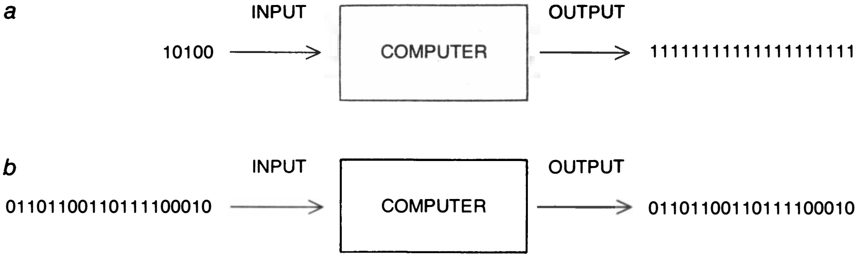
Clearly a more sensible definition of randomness is required, one that does not contradict the intuitive concept of a "patternless" number. Such a definition has been devised only in the past 10 years. It does not consider the origin of a number but depends entirely on the characteristics of the sequence of digits. The new definition enables us to describe the properties of a random number more precisely than was formerly possible, and it establishes a hierarchy of degrees of randomness. Of perhaps even greater interest than the capabilities of the definition, however, are its limitations. In particular the definition cannot help to determine, except in very special cases, whether or not a given series of digits, such as the second one above, is in fact random or only seems to be random. This limitation is not a flaw in the definition; it is a consequence of a subtle but fundamental anomaly in the foundation of mathematics. It is closely related to a famous theorem devised and proved in 1931 by Kurt Gödel, which has come to be known as Gödel's incompleteness theorem. Both the theorem and the recent discoveries concerning the nature of randomness help to define the boundaries that constrain certain mathematical methods.

Algorithmic Definition

The new definition of randomness has its heritage in information theory, the science, developed mainly since World War II, that studies the transmission of messages. Suppose you have a friend who is visiting a planet in another galaxy, and that sending him telegrams is very expensive. He forgot to take along his tables of trigonometric functions, and he has asked you to supply them. You could simply translate the numbers into an appropriate code (such as the binary numbers) and transmit them directly, but even the most modest tables of the six functions have a few thousand digits, so that the cost would be high. A much cheaper way to convey the same information would be to transmit instructions for calculating the tables from the underlying trigonometric formulas, such as Euler's equation $e^{ix} = \cos x + i \sin x$. Such a message could be relatively brief, yet inherent in it is all the information contained in even the largest tables.

Suppose, on the other hand, your friend is interested not in trigonometry but in baseball. He would like to know the scores of all the major-league games played since he left the earth some thousands of years before. In this case it is most unlikely that a formula could be found for compressing the information into a short message; in such a series of numbers each digit is essentially an independent item of information, and it cannot be predicted from its neighbors or from some underlying rule. There is no alternative to transmitting the entire list of scores.

In this pair of whimsical messages is the germ of a new definition of randomness. It is based on the observation that the information embodied in a random series of numbers cannot be "compressed," or reduced to a more compact form. In formulating the actual definition it is preferable to consider communication not with a distant friend but with a digital computer. The friend might have the wit to make inferences about numbers or to construct a series from partial information or from vague instructions. The computer does not have that capacity, and for our purposes that deficiency is an advantage. Instructions



ALGORITHMIC DEFINITION of randomness relies on the capabilities and limitations of the digital computer. In order to produce a particular output, such as a series of binary digits, the computer must be given a set of explicit instructions that can be followed without making intellectual judgments. Such a program of instructions is an algorithm. If the desired output is highly ordered (*a*), a relatively small algorithm will suffice; a series of twenty 1's, for example, might be generated by some hypothetical computer from the program 10100, which is the binary notation for the decimal number 20. For a random series of digits (*b*) the most concise program possible consists of the series itself. The smallest programs capable of generating a particular series are called the minimal programs of the series; the size of these programs, measured in bits, or binary digits, is the complexity of the series. A series of digits is defined as random if series' complexity approaches its size in bits.

given the computer must be complete and explicit, and they must enable it to proceed step by step without requiring that it comprehend the result of any part of the operations it performs. Such a program of instructions is an algorithm. It can demand any finite number of mechanical manipulations of numbers, but it cannot ask for judgments about their meaning.

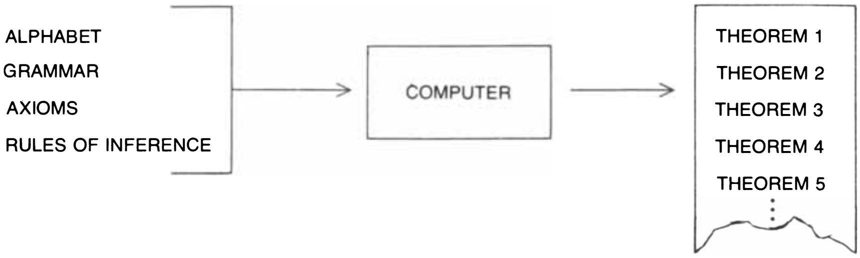
The definition also requires that we be able to measure the information content of a message in some more precise way than by the cost of sending it as a telegram. The fundamental unit of information is the "bit," defined as the smallest item of information capable of indicating a choice between two equally likely things. In binary notation one bit is equivalent to one digit, either a 0 or a 1.

We are now able to describe more precisely the differences between the two series of digits presented at the beginning of this article:

01010101010101010101
01101100110111100010

The first could be specified to a computer by a very simple algorithm, such as "Print 01 ten times." If the series were extended according to the same rule, the algorithm would have to be only slightly larger; it might be made to read, for example, "Print 01 a million times." The number of bits in such an algorithm is a small fraction of the number of bits in the series it specifies, and as the series grows larger the size of the program increases at a much lower rate.

For the second series of digits there is no corresponding shortcut. The most economical way to express the series is to write it out in full, and the shortest algorithm for introducing the series into a computer would be "Print 01101100110111100010." If the series were much larger (but still apparently patternless), the algorithm would have to be expanded to corresponding size. This "incompressibility" is a property of all random numbers; indeed, we can proceed directly to define randomness in terms of incompressibility: A series of numbers is random if the smallest algorithm capable of specifying it to a computer has about the same number of bits of information as the series itself.



FORMAL SYSTEMS devised by David Hilbert contain an algorithm that mechanically checks the validity of all proofs that can be formulated in the system. The formal system consists of an alphabet of symbols in which all statements are to be written; a grammar that specifies how the symbols are to be combined; a set of axioms, or principles accepted without proof, and rules of inference for deriving theorems from the axioms. Theorems are found by writing all the possible grammatical statements in the system and testing them to determine which ones are in accord with the rules of inference and are therefore valid proofs. Since this operation is performed by an algorithm it could be done by a digital computer. In 1931 Kurt Gödel demonstrated that virtually all formal systems are incomplete: in each of them there is at least one statement that is true but that cannot be proved.

This definition was independently proposed about 1965 by A. N. Kolmogorov of the Academy of Sciences of the U.S.S.R. and by me, when I was an undergraduate at the City College of the City University of New York. Both Kolmogorov and I were then unaware of related proposals made in 1960 by Ray J. Solomonoff of the Zator Company in an endeavor to measure the simplicity of scientific theories. During the past decade we and others have continued to explore the meaning of randomness. The original formulations have been improved and the feasibility of the approach has been amply confirmed.

Model of Inductive Method

The algorithmic definition of randomness provides a new foundation for the theory of probability. By no means does it supersede classical probability theory, which is based on an ensemble of possibilities, each of which is assigned a probability. Rather, the algorithmic approach complements the ensemble method by giving precise meaning to concepts that had been intuitively appealing but that could not be formally adopted.

The ensemble theory of probability, which originated in the 17th century, remains today of great practical importance. It is the foundation of statistics, and it is applied to a wide range of problems in science and engineering. The algorithmic theory also has important implications, but they are primarily theoretical. The area of broadest interest is its amplification of Gödel's incompleteness theorem. Another application (which actually preceded the formulation of the theory itself) is in Solomonoff's model of scientific induction.

Solomonoff represented a scientist's observations as a series of binary digits. The scientist seeks to explain these observations through a theory, which can be regarded as an algorithm capable of

OBSERVATIONS	PREDICTIONS	THEORY	SIZE OF THEORY
0101010101	01010101010101010101	TEN REPETITIONS OF 01	23 CHARACTERS
	01010101010000000000	FIVE REPETITIONS OF 01 FOLLOWED BY TEN 0's	44 CHARACTERS

INDUCTIVE REASONING as it is employed in science was analyzed mathematically by Ray J. Solomonoff. He represented a scientist's observations as a series of binary digits; the observations are to be explained and new ones are to be predicted by theories, which are regarded as algorithms instructing a computer to reproduce the observations. (The programs would not be English sentences but binary series, and their size would be measured not in characters but in bits.) Here two competing theories explain the existing data; Occam's razor demands that the simpler, or smaller, theory be preferred. The task of the scientist is to search for minimal programs. If the data are random, the minimal programs are no more concise than the observations and no theory can be formulated.

generating the series and extending it, that is, predicting future observations. For any given series of observations there are always several competing theories, and the scientist must choose among them. The model demands that the smallest algorithm, the one consisting of the fewest bits, be selected. Stated another way, this rule is the familiar formulation of Occam's razor: Given differing theories of apparently equal merit, the simplest is to be preferred.

Thus in the Solomonoff model a theory that enables one to understand a series of observations is seen as a small computer program that reproduces the observations and makes predictions about possible future observations. The smaller the program, the more comprehensive the theory and the greater the degree of understanding. Observations that are random cannot be reproduced by a small program and therefore cannot be explained by a theory. In addition the future behavior of a random system cannot be predicted. For random data the most compact way for the scientist to communicate his observations is for him to publish them in their entirety.

Defining randomness or the simplicity of theories through the capabilities of the digital computer would seem to introduce a spurious element into these essentially abstract notions: the peculiarities of the particular computing machine employed. Different machines communicate through different computer languages, and a set of instructions expressed in one of those languages might require more or fewer bits when the instructions are translated into another language. Actually, however, the choice of computer matters very little. The problem can be avoided entirely simply by insisting that the randomness of all numbers be tested on the same machine. Even when different machines are employed, the idiosyncrasies of various languages can readily be compensated for. Suppose, for example, someone has a program written in English and wishes to utilize it with a computer that reads only French. Instead of translating the algorithm itself he could preface the program with a complete English course written in French. Another mathematician with a French program and an English machine would follow the opposite procedure. In this way only a fixed number of bits need be added to the program, and that number grows less significant as the size of the series specified by the program increases. In practice a device called a compiler often makes it possible to ignore the differences between languages when one is addressing a computer.

Since the choice of a particular machine is largely irrelevant, we can choose for our calculations an ideal computer. It is assumed to have unlimited storage capacity and unlimited time to complete its calculations. Input to and output from the machine are both in the form of binary digits. The machine begins to operate as soon as the program is given it, and it continues until it has finished printing the binary series that is the result. The machine then halts. Unless an error is made in the program, the computer will produce exactly one output for any given program.

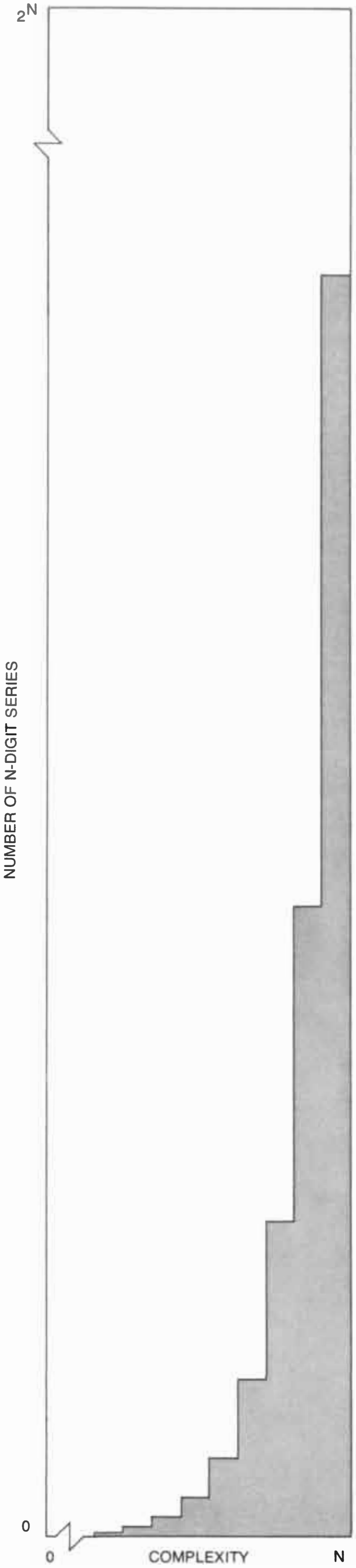
Minimal Programs and Complexity

Any specified series of numbers can be generated by an infinite number of algorithms. Consider, for example, the three-digit decimal series 123. It could be produced by an algorithm such as "Subtract 1 from 124 and print the result," or "Subtract 2 from 125 and print the result," or an infinity of other programs formed on the same model. The programs of greatest interest, however, are the smallest ones that will yield a given numerical series. The smallest programs are called minimal programs; for a given series there may be only one minimal program or there may be many.

Any minimal program is necessarily random, whether or not the series it generates is random. This conclusion is a direct result of the way we have defined randomness. Consider the program *P*, which is a minimal program for the series of digits S. If we assume that *P* is not random, then by definition there must be another program, *P′*, substantially smaller than *P* that will generate it. We can then produce S by the following algorithm: "From *P′* calculate *P*, then from *P* calculate S." This program is only a few bits longer than *P′*, and thus it must be substantially shorter than *P*. *P* is therefore not a minimal program.

The minimal program is closely related to another fundamental concept in the algorithmic theory of randomness: the concept of complexity. The complexity of a series of digits is the number of bits that must be put into a computing machine in order to obtain the original series as output. The complexity is therefore equal to the size in bits of the minimal programs of the series. Having introduced this concept, we can now restate our definition of randomness in more rigorous terms: A random series of digits is one whose complexity is approximately equal to its size in bits.

The notion of complexity serves not only to define randomness but also to measure it. Given several series of numbers each having n digits, it is theoretically possible to identify all those of complexity $n-1$, $n-10$, $n-100$ and so forth and thereby to rank the series in decreasing order of randomness. The exact value of complexity below which a series is no longer considered random remains somewhat arbitrary. The value ought to be set low enough for numbers with obviously random properties not to be excluded and high enough for numbers with a conspicuous pattern to be disqualified, but to set a particular numerical value is to judge what degree of randomness constitutes actual randomness. It is this uncertainty that is re-



RANDOM SEQUENCES of binary digits make up the majority of all such sequences. Of the 2^n series of n digits, most are of a complexity that is within a few bits of n. As complexity decreases, the number of series diminishes in a roughly exponential manner. Orderly series are rare; there is only one, for example, that consists of n 1's.

flected in the qualified statement that the complexity of a random series is *approximately* equal to the size of the series.

Properties of Random Numbers

The methods of the algorithmic theory of probability can illuminate many of the properties of both random and nonrandom numbers. The frequency distribution of digits in a series, for example, can be shown to have an important influence on the randomness of the series. Simple inspection suggests that a series consisting entirely of either 0's or 1's is far from random, and the algorithmic approach confirms that conclusion. If such a series is n digits long, its complexity is approximately equal to the logarithm to the base 2 of n. (The exact value depends on the machine language employed.) The series can be produced by a simple algorithm such as "Print 0 n times," in which virtually all the information needed is contained in the binary numeral for n. The size of this number is about $\log_2 n$ bits. Since for even a moderately long series the logarithm of n is much smaller than n itself, such numbers are of low complexity; their intuitively perceived pattern is mathematically confirmed.

Another binary series that can be profitably analyzed in this way is one where 0's and 1's are present with relative frequencies of three-fourths and one-fourth. If the series is of size n, it can be demonstrated that its complexity is no greater than four-fifths n, that is, a program that will produce the series can be written in $4/5n$ bits. This maximum applies regardless of the sequence of the digits, so that no series with such a frequency distribution can be considered very random. In fact, it can be proved that in any long binary series that is random the relative frequencies of 0's and 1's must be very close to one-half. (In a random decimal series the relative frequency of each digit is, of course, one-tenth.)

Numbers having a nonrandom frequency distribution are exceptional. Of all the possible n-digit binary numbers there is only one, for example, that consists entirely of 0's and only one that is all 1's. All the rest are less orderly, and the great majority must, by any reasonable standard, be called random. To choose an arbitrary limit, we can calculate the fraction of all n-digit binary numbers that have a complexity of less than $n - 10$. There are 2^1 programs one digit long that might generate an n-digit series; there are 2^2 programs two digits long that could yield such a series, 2^3 programs three digits long and so forth, up to the largest programs permitted within the allowed complexity; of these there are 2^{n-11}. The sum of this series $(2^1 + 2^2 + 2^3 + \ldots + 2^{n-11})$ is equal to $2^{n-10} - 2$. Hence there are fewer than 2^{n-10} programs of size less than $n - 10$, and since each of those programs can specify no more than one series of digits, fewer than 2^{n-10} of the 2^n numbers have a complexity less than $n - 10$. Since $2^{n-10}/2^n = 1/1{,}024$, it follows that of all n-digit binary numbers only about one in 1,000 have a complexity less than $n - 10$. In other words, only about one series in 1,000 can be compressed into a computer program more than 10 digits smaller than itself.

A necessary corollary of this calculation is that more than 999 of every 1,000 n-digit binary numbers have a complexity equal to or greater than $n - 10$. If that degree of complexity can be taken as an appropriate test of randomness, then almost all n-digit numbers are in fact random. If a fair coin is tossed n times, the probability is greater than .999 that the result will be random to this extent. It would therefore seem easy to exhibit a specimen of a long series of random digits; actually it is impossible to do so.

Formal Systems

It can readily be shown that a specific series of digits is not random; it is sufficient to find a program that will generate the series and that is substantially smaller than the series itself. The program need not be a minimal program for the series; it need only be a small one. To demonstrate that a particular series of digits is random, on the other hand, one must prove that no small program for calculating it exists.

It is in the realm of mathematical proof that Gödel's incompleteness theorem is such a conspicuous landmark; my version of the theorem predicts that the required proof of randomness cannot be found. The consequences of this fact are just as interesting for what they reveal about Gödel's theorem as they are for what they indicate about the nature of random numbers.

Gödel's theorem represents the resolution of a controversy that preoccupied mathematics during the early years of the 20th century. The question at issue was: "What constitutes a valid proof in mathematics and how is such a proof to be recognized?" David Hilbert had attempted to resolve the controversy by devising an artificial language in which valid proofs could be found mechanically, without any need for human insight or judgment. Gödel showed that there is no such perfect language.

Hilbert established a finite alphabet of symbols, an unambiguous grammar specifying how a meaningful statement could be formed, a finite list of axioms, or initial assumptions, and a finite list of rules of inference for deducing theorems from the axioms or from other theorems. Such a language, with its rules, is called a formal system.

A formal system is defined so precisely that a proof can be evaluated by a recursive procedure involving only simple logical and arithmetical manipulations. In other words, in the formal system there is an algorithm for testing the validity of proofs. Today, although not in Hilbert's time, the algorithm could be executed on a digital computer and the machine could be asked to "judge" the merits of the proof.

Because of Hilbert's requirement that a formal system have a proof-checking algorithm, it is possible in theory to list one by one all the theorems that can be proved in a particular system. One first lists in alphabetical order all sequences of symbols one character long and applies the proof-testing algorithm to each of them, thereby finding all theorems (if any) whose proofs consist of a single character. One then tests all the two-character sequences of symbols, and so on. In this way all potential proofs can be checked, and eventually all theorems can be discovered in the order of the size of their proofs. (The method is, of course, only a theoretical one; the procedure is too lengthy to be practical.)

Unprovable Statements

Gödel showed in his 1931 proof that Hilbert's plan for a completely systematic mathematics cannot be fulfilled. He did so by constructing an assertion about the positive integers in the language of the formal system that is true but that cannot be proved in the system. The formal system, no matter how large or how carefully constructed it is, cannot encompass all true theorems and is therefore incomplete. Gödel's technique can be applied to virtually any formal system, and it therefore demands the surprising and, for many, discomforting conclusion that there can be no definitive answer to the question "What is a valid proof?"

RUSSELL PARADOX

Consider the set of all sets that are not members of themselves.
Is this set a member of itself?

EPIMENIDES PARADOX

Consider this statement: "This statement is false."
Is this statement true?

BERRY PARADOX

Consider this sentence: "Find the smallest positive integer which to be specified requires more characters than there are in this sentence."
Does this sentence specify a positive integer?

THREE PARADOXES delimit what can be proved. The first, devised by Bertrand Russell, indicated that informal reasoning in mathematics can yield contradictions, and it led to the creation of formal systems. The second, attributed to Epimenides, was adapted by Gödel to show that even within a formal system there are true statements that are unprovable. The third leads to the demonstration that a specific number cannot be proved random.

Gödel's proof of the incompleteness theorem is based on the paradox of Epimenides the Cretan, who is said to have averred, "All Cretans are liars" [see "Paradox," by W. V. Quine; Scientific American, April, 1962]. The paradox can be rephrased in more general terms as "This statement is false," an assertion that is true if and only if it is false and that is therefore neither true nor false. Gödel replaced the concept of truth with that of provability and thereby constructed the sentence "This statement is unprovable," an assertion that, in a specific formal system, is provable if and only if it is false. Thus either a falsehood is provable, which is forbidden, or a true statement is unprovable, and hence the formal system is incomplete. Gödel then applied a technique that uniquely numbers all statements and proofs in the formal system and thereby converted the sentence "This statement is unprovable" into an assertion about the properties of the positive integers. Because this transformation is possible, the incompleteness theorem applies with equal cogency to all formal systems in which it is possible to deal with the positive integers [see "Gödel's Proof," by Ernest Nagel and James R. Newman; Scientific American, June, 1956].

The intimate association between Gödel's proof and the theory of random numbers can be made plain through another paradox, similar in form to the paradox of Epimenides. It is a variant of the Berry paradox, first published in 1908 by Bertrand Russell. It reads: "Find the smallest positive integer which to be specified requires more characters than there are in this sentence." The sentence has 114 characters (counting spaces between words and the period but not the quotation marks), yet it supposedly specifies an integer that, by definition, requires more than 114 characters to be specified.

As before, in order to apply the paradox to the incompleteness theorem it is necessary to remove it from the realm of truth to the realm of provability. The phrase "which requires" must be replaced by "which can be proved to require," it being understood that all statements will be expressed in a particular formal system. In addition the vague notion of "the number of characters required to specify" an integer can be replaced by the precisely defined concept of complexity, which is measured in bits rather than in characters.

The result of these transformations is the following computer program: "Find a series of binary digits that can be proved to be of a complexity greater than the number of bits in this program." The program tests all possible proofs in the formal system in order of their size until it encounters the first one proving that a specific binary sequence is of a complexity greater than the number of bits in the program. Then it prints the series it has found and halts. Of course, the paradox in the statement from which the program was derived has not been eliminated. The program supposedly calculates a number that no program its size should be able to calculate. In fact, *a* This statement is unprovable.

b The complexity of 01101100110111100010 is greater than 15 bits.

c The series of digits 01101100110111100010 is random.

d 10100 is a minimal program for the series 11111111111111111111.

UNPROVABLE STATEMENTS can be shown to be false, if they are false, but they cannot be shown to be true. A proof that "This statement is unprovable" (*a*) reveals a self-contradiction in a formal system. The assignment of a numerical value to the complexity of a particular number (*b*) requires a proof that no smaller algorithm for generating the number exists; the proof could be supplied only if the formal system itself were more complex than the number. Statements labeled *c* and *d* are subject to the same limitation, since the identification of a random number or a minimal program requires the determination of complexity.

the program finds the first number that it can be proved incapable of finding.

The absurdity of this conclusion merely demonstrates that the program will never find the number it is designed to look for. In a formal system one cannot prove that a particular series of digits is of a complexity greater than the number of bits in the program employed to specify the series.

A further generalization can be made about this paradox. It is not the number of bits in the program itself that is the limiting factor but the number of bits in the formal system as a whole. Hidden in the program are the axioms and rules of inference that determine the behavior of the system and provide the algorithm for testing proofs. The information content of these axioms and rules can be measured and can be designated the complexity of the formal system. The size of the entire program therefore exceeds the complexity of the formal system by a fixed number of bits c. (The actual value of c depends on the machine language employed.) The theorem proved by the paradox can therefore be stated as follows: In a formal system of complexity n it is impossible to prove that a particular series of binary digits is of a complexity greater than $n + c$, where c is a constant that is independent of the particular system employed.

Limits of Formal Systems

Since complexity has been defined as a measure of randomness, this theorem implies that in a formal system no number can be proved to be random unless the complexity of the number is less than that of the system itself. Because all minimal programs are random the theorem also implies that a system of greater complexity is required in order to prove that a program is a minimal one for a particular series of digits.

The complexity of the formal system has such an important bearing on the proof of randomness because it is a measure of the amount of information the system contains, and hence of the amount of information that can be derived from it. The formal system rests on axioms: fundamental statements that are irreducible in the same sense that a minimal program is. (If an axiom could be expressed more compactly, then the briefer statement would become a new axiom and the old one would become a derived theorem.) The information embodied in the axioms is thus in itself random, and it can be employed to test the randomness of other data. The randomness of some numbers can therefore be proved, but only if they are smaller than the formal system. Moreover, any formal system is of necessity finite, whereas any series of digits can be made arbitrarily large. Hence there will always be numbers whose randomness cannot be proved.

The endeavor to define and measure randomness has greatly clarified the significance and the implications of Gödel's incompleteness theorem. That theorem can now be seen not as an isolated paradox but as a natural consequence of the constraints imposed by information theory. In 1946 Hermann Weyl said that the doubt induced by such discoveries as Gödel's theorem had been "a constant drain on the enthusiasm and determination with which I pursued my research work." From the point of view of information theory, however, Gödel's theorem does not appear to give cause for pessimism. Instead it seems simply to suggest that in order to progress, mathematicians, like investigators in other sciences, must search for new axioms.

Ours is not the only way to build a car. But it's the only way to build a Jaguar.

There are very few people today who care about doing a thing as well as it can be done. Among those few who care are the people who build Jaguars.

The results of this caring are evident in our Jaguar XJ6L and our XJ12L*, motorcars with a level of luxury and high performance characteristics that make them stand out from other luxury sedans.

For just one example, the facia on every Jaguar XJ sedan is carved from one piece of burled Belgian walnut. The grain is matched by hand and the wood is finished by hand, as are the prime British hides that cover Jaguar's orthopedically designed seats.

But the interior refinements of the Jaguar XJs are only a preliminary to driving them.

No full-sized luxury car handles quite like a Jaguar XJ sedan. The feel is solid and very comfortable, yet the response is quick and precise like a sports car because the handling characteristics of the XJs are derived from Jaguar's extensive and heroic racing experience. Standard equipment includes independent suspension and power disc brakes on all four wheels and the precision of rack and pinion steering.

The XJ12L is powered by Jaguar's innovative, turbine-like aluminum alloy V-12. It is remarkably smooth and develops great reserves of torque even at low speeds. Yet the Jaguar V-12 displaces only 326 cubic inches as compared to popular American luxury V-8s which displace up to 500 cubic inches.

The XJ6L engine is a refinement of the overhead-cam Six that made history at LeMans.

And every Jaguar is individually road-tested over no less than 14 miles of Coventry roads.

Yes, there are other ways to build cars. But they wouldn't be Jaguars. We invite you to drive these uncommon luxury cars. For the name of the dealer nearest you, call these numbers toll-free: (800) 447-4700, or, in Illinois (800) 322-4400. *NOT AVAILABLE IN CALIFORNIA

Jaguar

BRITISH LEYLAND MOTORS INC., LEONIA, N.J. 07605

BRITISH
LEYLAND

The Social System of Lions

The organization of the social unit of lions, the pride, and the behavior of the animals in it are a sensitive reflection of the lion's adaptation to its environment

by Brian C. R. Bertram

One of the principal habitats of the lion (*Panthera leo*) is the Serengeti National Park in Tanzania, an area of 5,000 square miles where about 1,500 lions live with some two million other large mammals. Here other investigators and I, working at the Serengeti Research Institute, have been able to make detailed observations of the lion's habitat and way of life over a period of years. From these observations is beginning to emerge an overall picture of the integrated system consisting of the lion's environment, its prey, its hunting behavior, its social organization, its sexual behavior and its reproduction.

Lions are distinctive as hunters and as social animals. No other animal that hunts in groups on land approaches the lion in size. No other species of cat is social.

In the Serengeti National Park lions are the largest of the predators and, except for hyenas, the most numerous. Their prey includes a variety of large mammals; zebras and wildebeests are taken most often, but the lions also kill buffaloes, gazelles, hartebeests, impalas, warthogs and other animals. Lions are particularly adept at catching such prey in fairly open country, mainly because they tend to hunt cooperatively.

Even for lions prey animals are not easy to catch, since the prey animal can usually run faster than a lion. Therefore most hunting efforts by lions fail, even at night, when they do most of their hunting and when they most often succeed at it. When lions hunt in a group, they spread out and each one stalks toward the prey, which is therefore likely to be surrounded. A prey animal fleeing from a lion it has detected may run within range of one it has not seen and thus be caught. And when a prey animal is caught, all the lions in the hunting group feed on it.

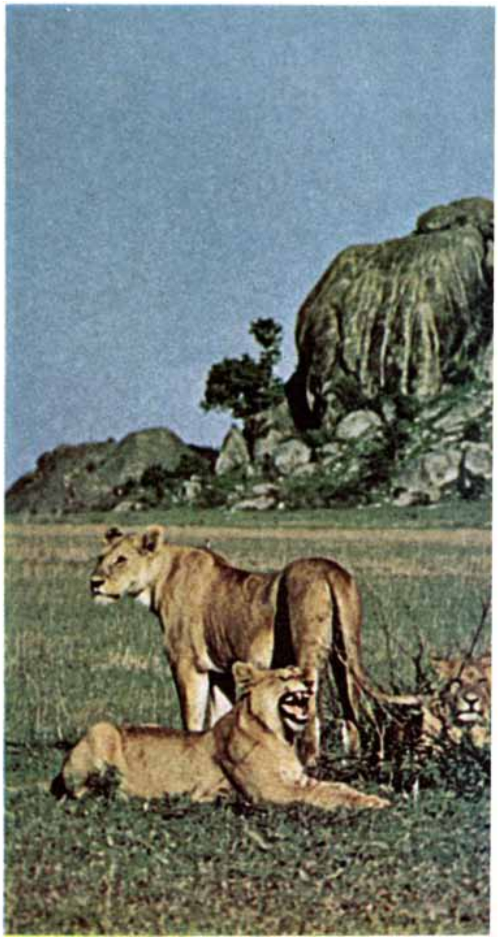
Cooperative hunting tends to be more successful for the lions than solitary hunting, although solitary hunting is not unusual among them. The females do most of the hunting. One can speculate that they are more effective at it than the males because they do not have an unwieldy large mane and are lighter. (A female weighs about 120 kilograms, or 265 pounds, and a male about 180 kilograms, or 395 pounds.) In any case the males, being stronger than the females, can gain access to any kill made by the females.

The communal hunting and communal living of lions are clearly interrelated. The social unit of the lion—the pride—is a long-lasting entity. Its nucleus is a group of from three to 12 females of breeding age. They are accompanied by a smaller group of males of breeding age; most often a pride will have two such males, but the number can range from one to six. The pride also includes cubs of various ages.

The members of a pride are not likely to be together all the time or even most of the time. Individual lions sometimes go off on their own or in the company of one or two other lions. Some lions tend to be with certain individuals much more frequently than they are with others. All members of the pride encounter one another fairly often, however, and at such times they interact peaceably.

Each pride occupies a territory that is a few miles in diameter. Intruding lions are kept away, particularly by the males. The territories do not have sharply defined boundaries, so that between adjoining territories both overlapping and gaps are found. A territory may shift with the seasons, although if prey animals can be found in it throughout the year, such shifts are small. Most of the woodland area of the park can be visualized as being subdivided into a mosaic of separate, adjoining pride territories with little overlap.

In addition to the lions living in territorial prides one finds lions that are nomadic. They tend to follow the migratory herds of prey animals, and so they wander through the territory of prides and also through areas not permanently



PRIDE OF LIONS lives in the Serengeti National Park in Tanzania. This group, which is part of a pride, includes an adult occupied by other lions. About 15 percent of the lion population live in this way. The nomads are not a different population but represent the surplus from the resident breeding population. Many of the nomads are males at a particular stage in their life cycle; some nomads are females that have been expelled from their pride. Since the nomads make up only a small part of the total population and are much less successful in reproducing, I shall not discuss them further and shall concentrate on the more typical resident lions, as exemplified by two adjacent Serengeti prides.

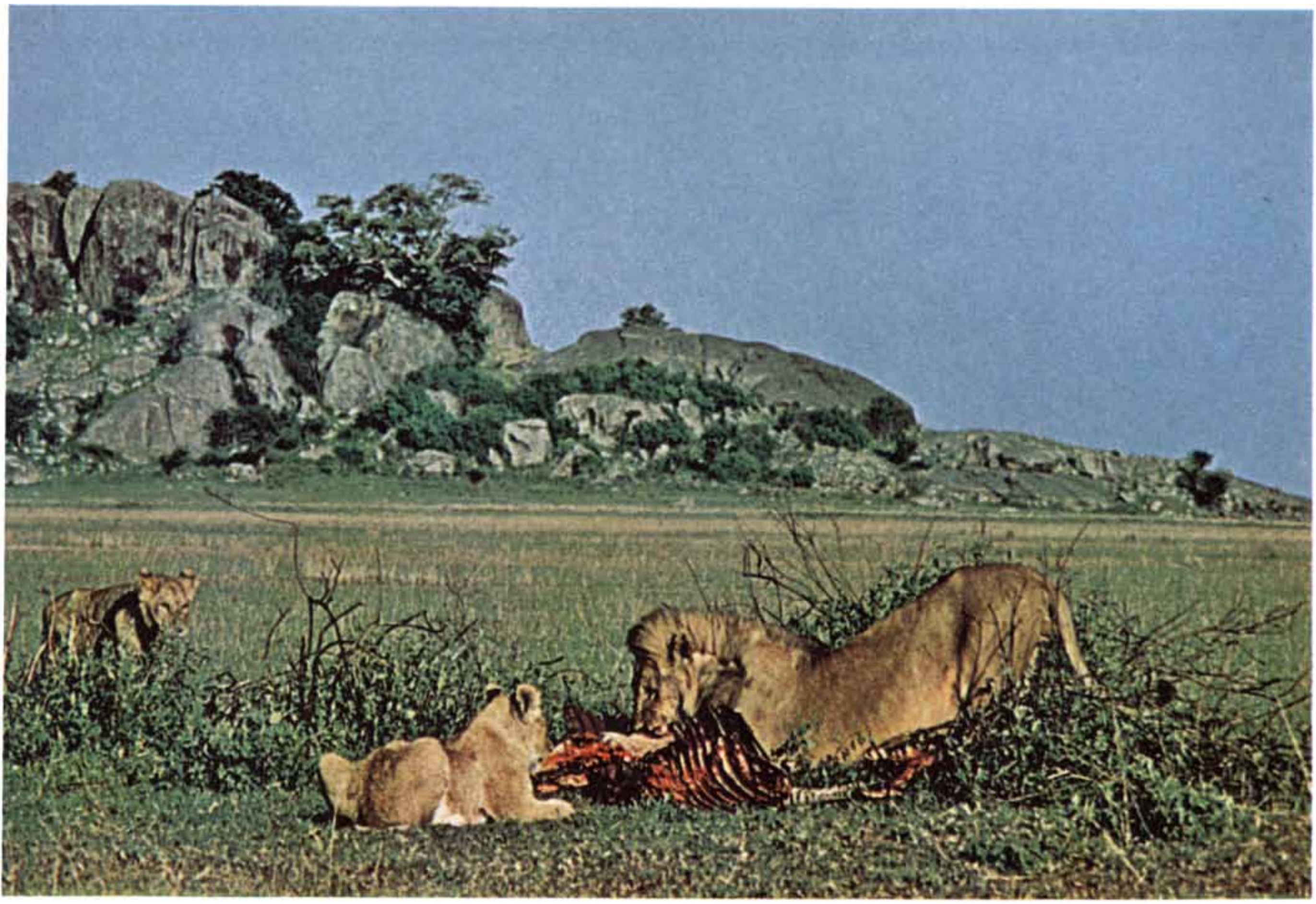
The data on the two prides, which occupy territories near Seronera in the middle of the park, were gathered over seven years of observation. George B. Schaller of the New York Zoological Society had started keeping records on the same prides in 1966. I continued the records from 1969 to 1973, and my successors at the Serengeti Research Institute are maintaining them now.

Since the prides are adjacent, the lions in the two different social groups are usually within four or five miles of each other. The climatic conditions in the areas are similar, and the densities of prey change seasonally in much the same way. Therefore one can reasonably assume that differences between the two prides are caused by factors operating within the prides rather than by gross seasonal factors common to both of them.

The prides are in an area frequented by tourists, and so the lions have become accustomed to vehicles. In an automobile one can get close to the animals without disturbing them. They are easy to observe at ranges of from 10 to 40 yards. Moreover, after watching a pride for a period of time one finds that the lions are recognizable individually by such markings as nicks in the ears (caused by squabbling at kills), scars, missing teeth and the arrangement of whisker spots. I made an identity card for each lion; it included photographs of the animal and notations on the animal's characteristic features.

Finding the lions was a bigger problem than recognizing them. In another area I had been locating the animals by tracking the radio signal emitted by a small transmitter in a collar fitted around each animal's neck. That, however, could not be done in the tourist area. Therefore I had to drive around in the territories of the prides in the hope of finding some of their members, and often I could find only a few. As a result my observations of individuals were necessarily intermittent and irregular. When I did find a group, I made a note of which individuals were there and also kept records of newborn cubs and lions that were pregnant, feeding, mating or sick.

Even though the observations were intermittent, the fact that they were conductedover a period of seven years made it possible to amass a great deal of information. (Observations must be made over a long period of time in any case because lions are long-lived and have a long generation time.) Where necessary I have drawn on data from other prides to supplement the records. Let us now examine what this information reveals



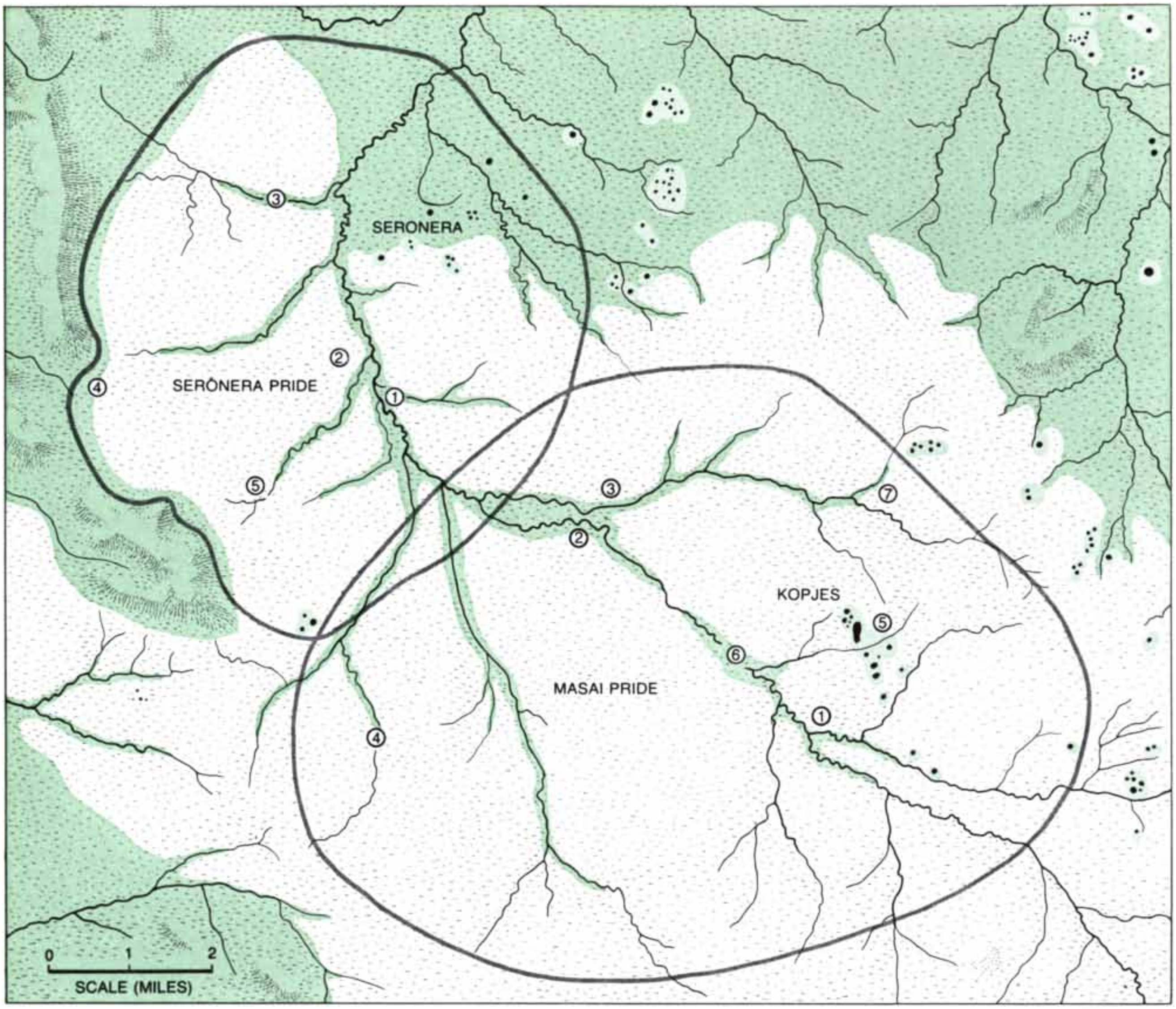
male eating a prey animal at right and an adult female standing at left. The other lions are cubs or subadults. A typical pride usually includes two or three adult males, from five to 10 adult females and a number of cubs. Lions are termed cubs up to the age of two years and subadults until they are four years old. In the background is a kopje, an outcropping of rock that is characteristic of the region.

about the interaction of the social system of lions and their reproduction.

Considering the females first, it is significant that no strange female joined any of the prides observed. Therefore every female in a pride was born and reared in that pride. No prides were observed to die out, and no new pride was seen to form. (It is likely that both events occur but that they are uncommon.) If a pride lasts for some decades at least, it follows that all the females in it are likely to be related to one another. They are sisters, mothers, grandmothers, half-sisters, cousins and so on. The permanent nucleus of a pride is therefore a group of related females with a range of ages.

To be born in a pride, however, is not necessarily to remain in it. At the age of about three years a subadult female is either recruited into the pride or driven out of it. (Lions are termed cubs until they are two years old and subadults between the ages of two and four.) A lioness is more likely to be recruited if the pride has relatively few adult females than if it has a large number.

A lioness that is expelled becomes nomadic. She leaves the territory of the pride and roams over large distances in search of a living. Such females do not reproduce nearly as effectively as the resident females. Schaller has shown that they do not survive as long and that they produce smaller litters, whose members are less likely to survive.



TERRITORIES OF TWO PRIDES in the Serengeti National Park show where the members might be on a typical day. In the Seronera pride's area are (*1*) a male, two females and five small cubs; (*2*) a male and female mating; (*3*) three females and seven large cubs; (*4*) a peripheral female, and (*5*) two subadult males, nomadic intruders. Masai pride has (*1*) a male alone; (*2*) two males with a kill; (*3*) four females and 11 cubs; (*4*) a female and four subadults; (*5 and 6*) females with new cubs, and (*7*) a female alone.

It would seem that if a group of nomadic lionesses were to come on a suitable area for a territory, they would settle there and establish a pride. It is likely, however, that any suitable area would already be occupied by a pride and that the members of it would not allow the newcomers to settle there or to join it.

A female that is recruited into a pride has an easier and more productive life. She will generally produce her first litter when she is about four years old and will continue bearing litters until within a couple of years before her death at the age of about 18. A recruited female therefore has a reproductive period of some 13 years.

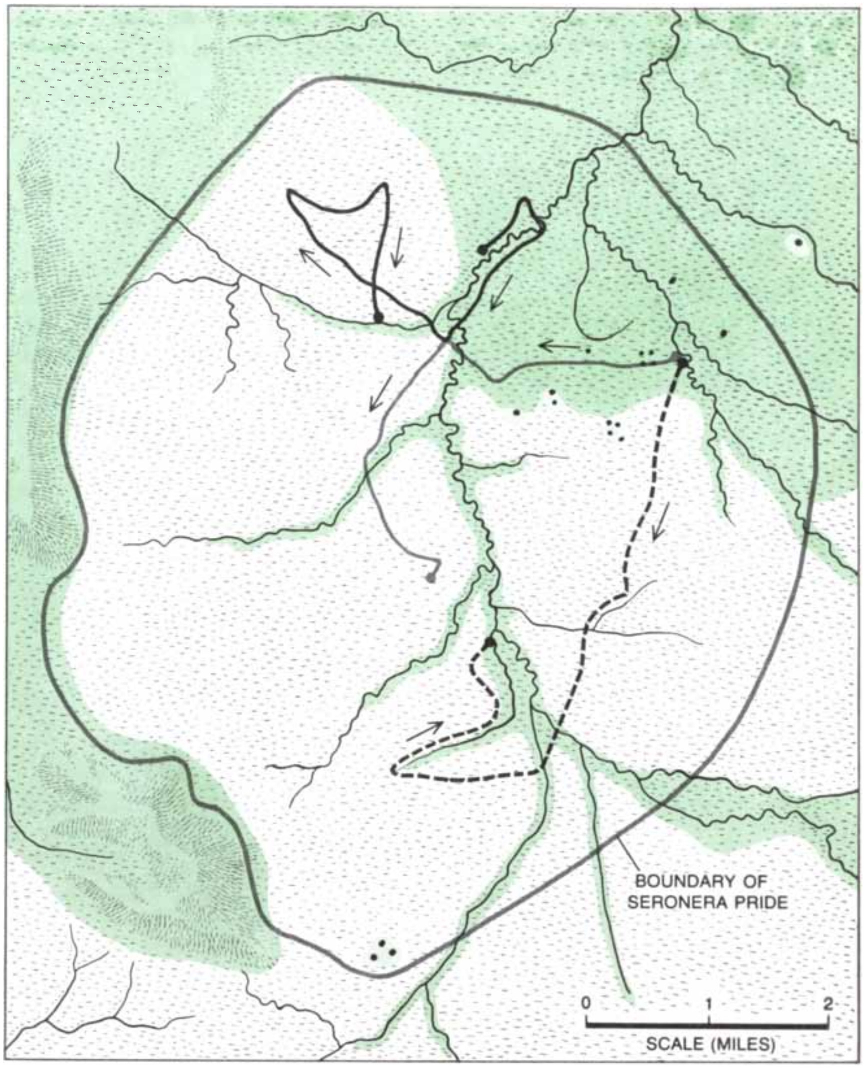
For young males the ordinary course of events is quite different. At the age of about three they leave the pride in which they were born. They may leave voluntarily or they may be expelled. Males depart in small groups of up to half a dozen members each. Staying together as a group, they join the nomadic portion of the population.

At that age a young male has very little mane. He is sexually active, however, and will mate if he encounters an unattended estrous female. During the next two years or so the male gradually reaches adult size, and his mane develops from a scruffy fringe around the neck to a darkening cape from head to shoulders. The young males hunt for themselves and also scavenge what they can from the carcasses of animals that have died of natural causes or have been killed by other predators.

After about two years of this kind of life the group of males is likely to find a pride it can take over. It may be a pride whose males have left or died. If it is a pride with males still in residence, the new group may take over by driving them out.

The driving out of resident males is sometimes a slow process. The newcomers settle at the edge of the pride's territory and gradually expand their area of operation. In other cases the takeover is sudden and dramatic, being brought about by a fight that the resident males lose. In either case the new males come into possession of the pride and its territory.

The pride they have taken over is unlikely to be the pride they grew up in. It might be any of the prides in the park. As a result the males are unlikely to be genetically related to the females. On the other hand, the males are closely related to one another, because they are the collective offspring of a group of related females and have stayed together since they left those females. This distinction accounts for the apparently unique social structure of lions: a permanent group of closely interrelated females and a smaller group of separately interrelated coequal males, which are with the females for a short period.



DAY'S MOVEMENTS by several members of the Seronera pride are plotted for a group of three adult females and seven cubs (*black*), an adult male (*gray*) and another adult male (*broken line*). Most activities, such as hunting, feeding, finding companions and patrolling the territory, take place at night. During most of the day the animals rest in the shade.

The length of time during which the males stay in possession of the pride varies. On the average it is two or three years before they in turn are expelled by a new group of younger, stronger or more numerous males. The lions that are expelled are unlikely to find another pride they can take over, because they are older now and also may have been injured. Moreover, they are probably less efficient at hunting than they were as young lions. In the pride they could depend on the females to do most of the hunting. Meanwhile they have become bulkier and have acquired a mane. Therefore the conditions of life for males that have been expelled from a pride are harsh. For these reasons the effective reproductive period of the males is much shorter than that of the females, being only as long as the few years during which the males are in possession of a pride.

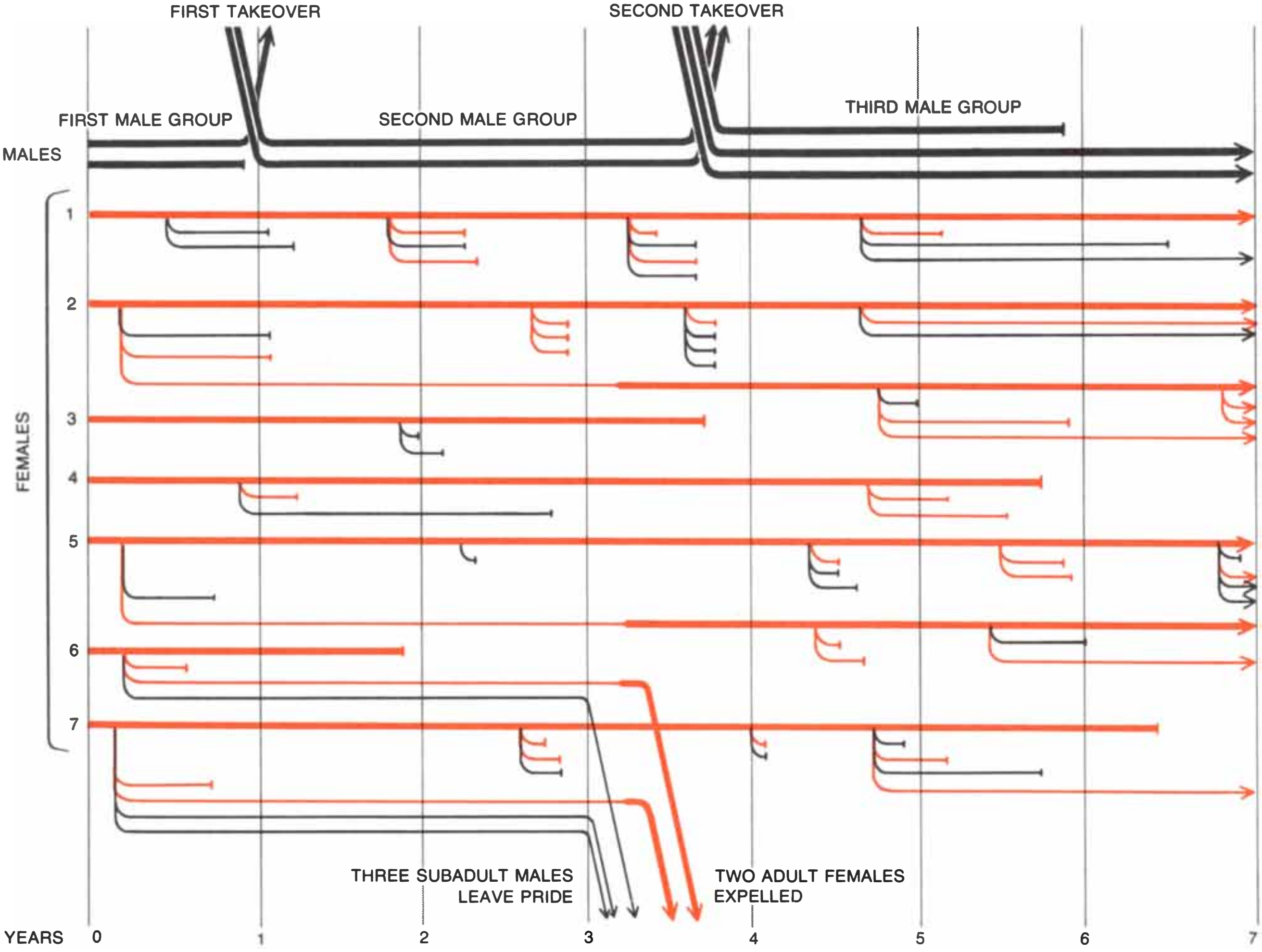
The brevity of the male reproductive period is a result of the strong competition between male groups. Males are likely to be expelled from a pride soon after they are past their prime or are reduced in number by injury, disease or death. A single male can almost never retain possession of a pride in the Serengeti because almost any pair of males, fighting in tandem as they do, can defeat him. Similarly, larger groups of males can defeat smaller groups, other things being equal, and so the larger groups have the longest tenure of prides. In addition large groups (of from three to six males) sometimes manage after a while to take over a second adjacent pride too and to keep possession of both prides concurrently. Thus it is to the selective advantage of a male to have male companions.

Lionesses do not appear to have a regular estrous cycle in the wild: they come into heat at variable intervals of between three weeks and a few months. A lioness usually remains in heat for a few days, and during that time she will mate on the average every 15 minutes. The first male to encounter a female coming into estrus temporarily gains dominance over other males. He mates with her and by his presence keeps other males 20 yards or more away. Competition among males for an estrous female is rare. A female sometimes changes males while she is in heat, but she seldom does so more than once a day. Occasionally, if several females are in heat at the same time, a male may mate with more than one female.

The females of a pride are often in heat synchronously. The synchrony is unrelated in time to the synchrony in adjacent prides, so that it must be caused by some factor operating within the pride. Possibly the females are responding to signals from one another, such as pheromones in the urine, or to signals from the males. Another possibility is that they are responding to particular characteristics of their food intake, such as a series of kills of large prey animals. The data are scanty, and it is difficult with lions in the wild to obtain information on the mechanisms involved.

Notwithstanding the energetic mating performance of lions, most mating periods do not result in the birth of cubs. It is difficult to tell why, again for want of adequate information. The male almost always ejaculates and probably produces viable sperm. Whether the low level of births has to do with ovulation, fertilization or abortion in the female is not known.

When a female does conceive, the gestation period is some 14 or 15 weeks, which is a remarkably short time for such a large animal. As a result the cubs are tiny at birth, weighing less than 1 percent of the adult weight. A litter usually consists of two or three cubs; the range is from one through five. I could not find



REPRODUCTIVE EVENTS in a pride of lions are portrayed for males (*black*), females (*color*) and cubs (*thin lines*). At the end of the first year the original group of males is driven out by a new group, which in turn is driven out some three years later. Below the events involving adult males one sees what happened to the pride's permanent group of interrelated females and their cubs. A horizontal line that stops means that the individual died. Mortality among cubs is high, often because males on taking over a pride will kill the cubs already there. One also sees, notably after the second takeover in this case, a tendency toward synchrony of births and enhanced survival of cubs at times when the pride has few or no older cubs. Cub survival also affects timing of a female's successive litters.

any factors that could be shown to influence the size of the litter.

Cubs can be born in any month of the year. The timing of births appears to be influenced by social factors rather than by the climate or the food supply. A lioness generally produces a litter when the cubs of her preceding litter are from 20 to 30 months old, provided that they have survived. If they have died, her next litter will be born some six to 12 months after the death of the last cub.

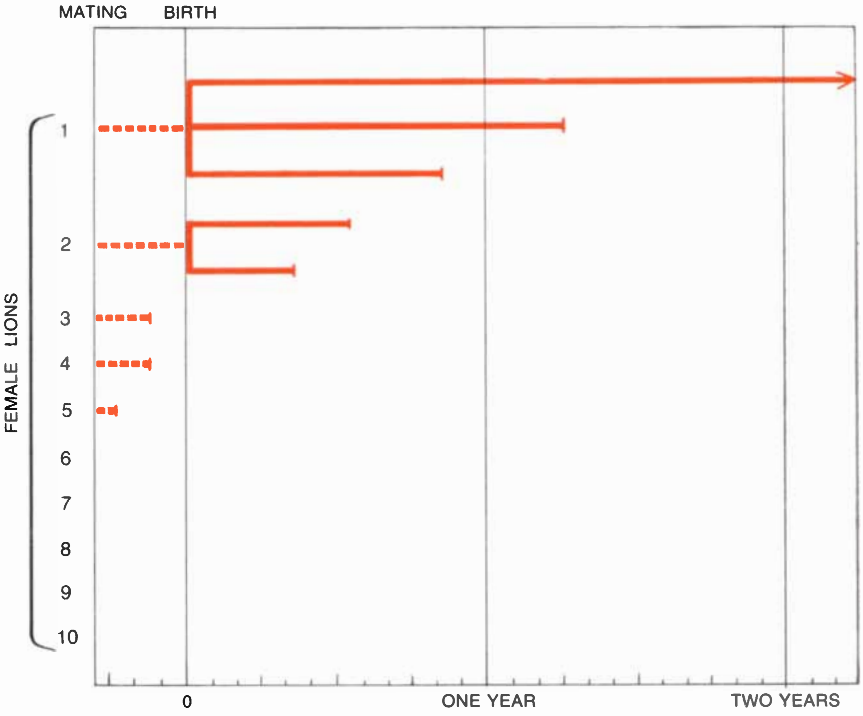
The females of a pride tend to produce litters at about the same time. The synchrony of estrus contributes to the synchrony of births, of course, but it cannot be the only factor because the synchrony of estrus is much more precise and because most periods of estrus do not result in cubs. The physiological causes, possibly pheromonal ones, for the synchrony of births remain to be determined, and it will not be easy to determine them with lions in the wild.

Evidence can be found for the influence of further social factors on the timing of births. For example, if one examines the timing of births in relation to the arrival of new adult males, one finds that for about six months after the new males take over few litters are born. Then the production of litters rises sharply.

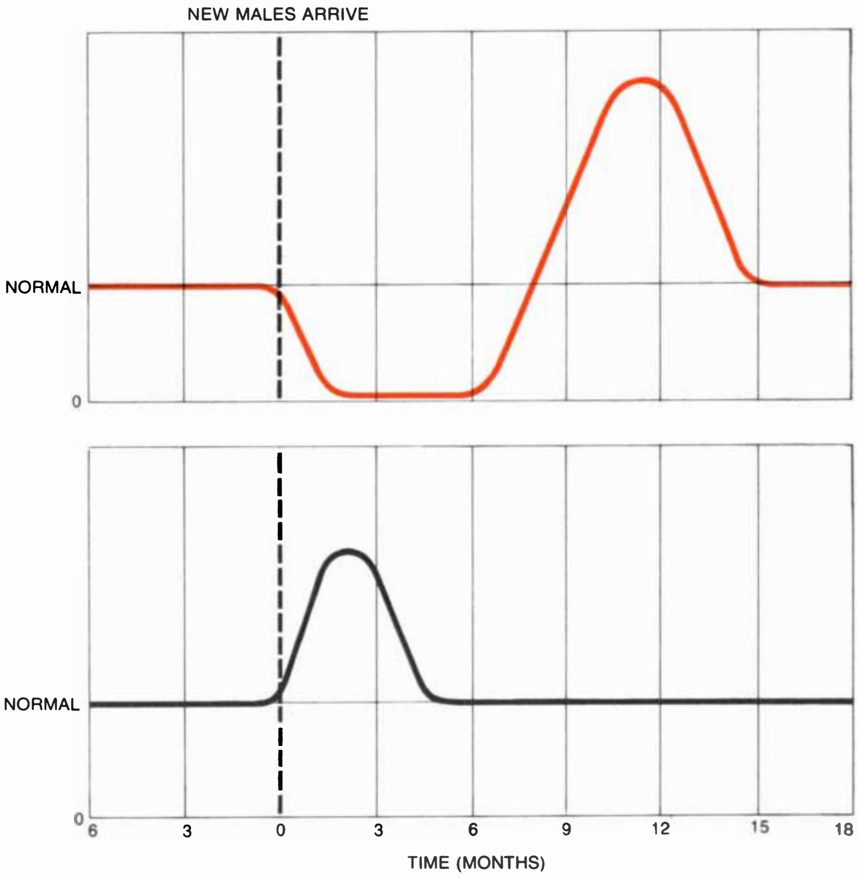
Why should it be that males do not succeed in siring offspring until they have been in the pride for about three months? Possibly stress associated with their presence makes the females less likely to conceive. Their presence may also cause abortion among females that were pregnant when the new males arrived.

Lion cubs suckle as long as they can, which is from six to eight months, when the mother stops lactating. Long before that, at the age of two or three months, the cubs have started to eat meat, so that they undergo a period of gradual weaning. The cubs are totally dependent on the adults to get food for them until they are at least two years old.

Only a small proportion of cubs attain that age; the mortality among cubs is about 80 percent. Schaller has concluded that about a fourth of the deaths are violent, being brought about mainly by other lions but also by hyenas, buffaloes and accidents. Another fourth of the deaths are attributable to starvation, since cubs are the least able to get and keep even small pieces of meat at times when food is in short supply. Half of the deaths are of unknown origin, because



RESULTS OF MATING are indicated for 10 female lions. Usually a large number of matings is needed for a female to produce cubs. Because of this and the mortality rate of about 80 percent among cubs, number of matings per offspring reared to adulthood is about 3,000.



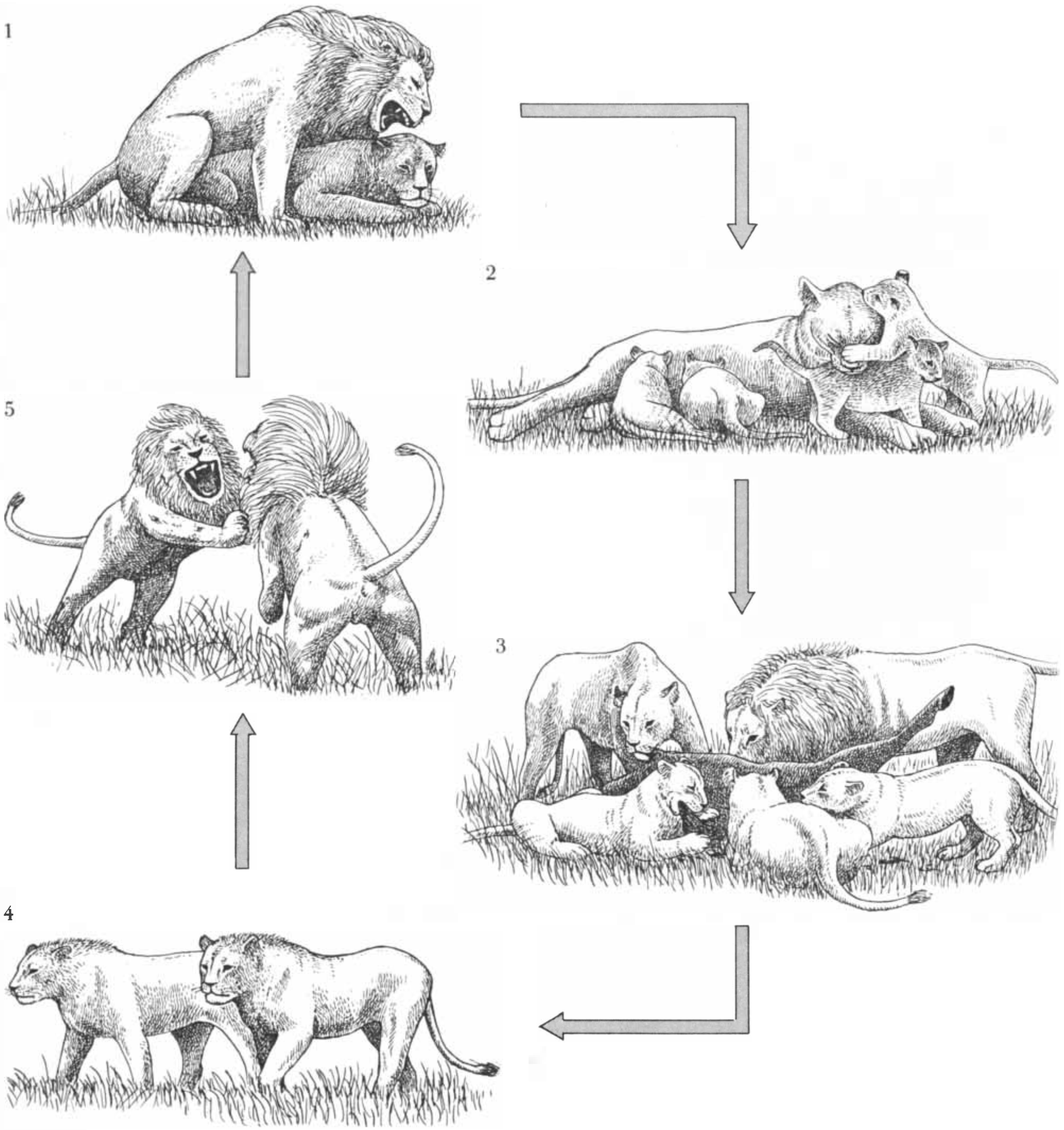
EFFECT OF TAKEOVER on the birth rate (*color*) and the death rate (*black*) of cubs is charted. At the point indicated by the broken line a pride is taken over by a new group of males. Mortality of cubs rises immediately, and few litters are born for several months.

the carcass is not found and the animal was not previously seen to be ailing.

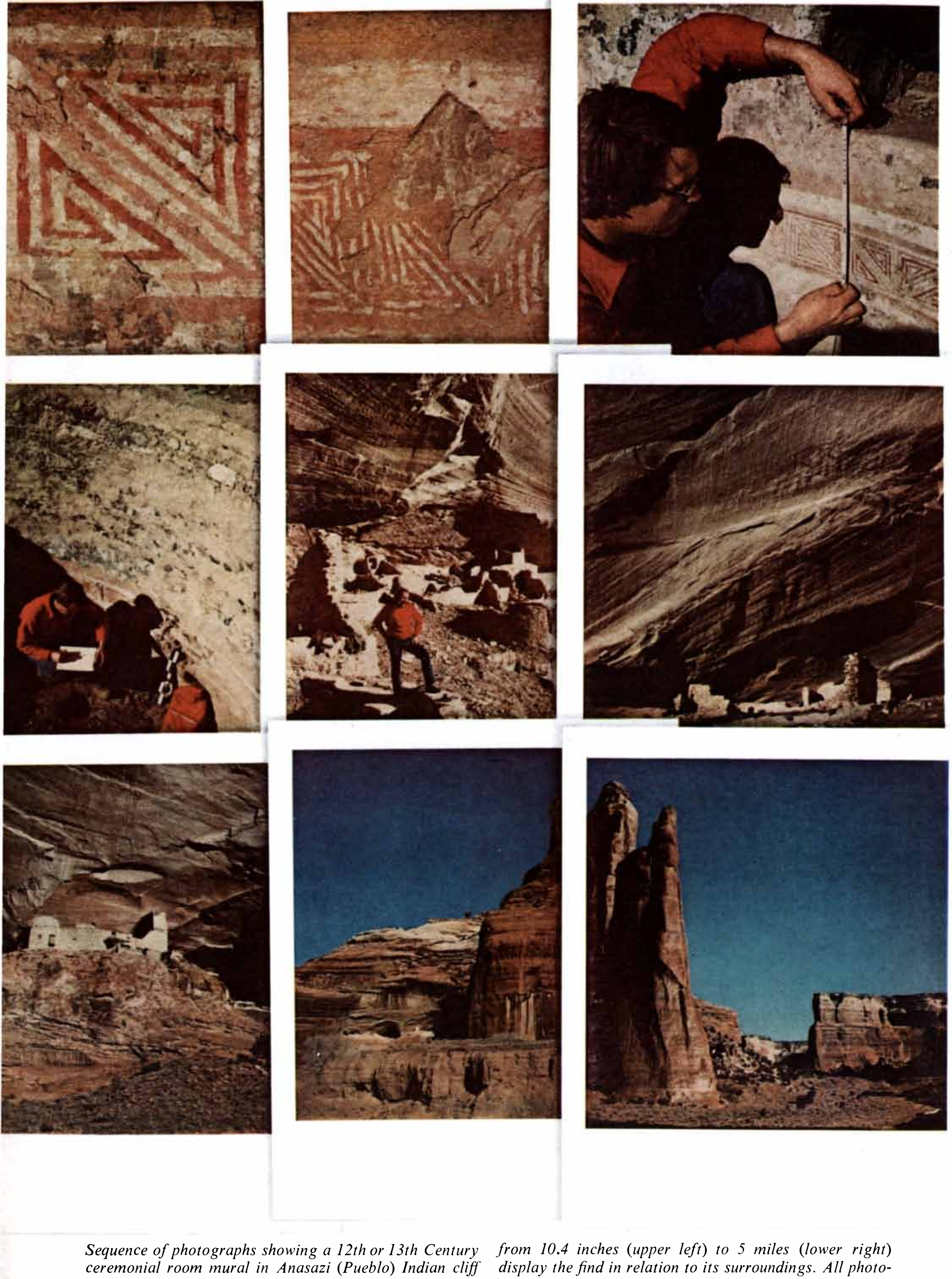
I have looked for factors that might influence the survival of cubs. The mortality was higher at times when few prey animals were available. In the two prides near Seronera the level of mortality tended to be higher from October to March, but the times of food shortage differ greatly from year to year and from pride to pride, depending on the movements of the migratory prey species.

The time of year when cubs were born did not influence their chance of survival, but the time of birth in relation to the birth of other cubs in the pride was important. Litters tended to show poorer survival if appreciably older cubs were present in the pride. The reason is likely to be that at kills the stronger animals can plunder the share of the weaker ones.

Litters that were born synchronously survived better than litters that were not. Synchrony of births makes communal suckling and rearing possible, so that cubs have a more regular supply of milk



EVENTS IN LIFE CYCLE of a pride are portrayed, beginning with mating (*1*). Cubs suckle for several months (*2*) and until the age of two are dependent on adults to provide food by catching prey animals (*3*). Subadult males leave at about age three and become nomadic (*4*). Females usually remain with the pride but may be expelled at about age three if the pride has an excess of females. Expelled females become nomads. Young males from a pride tend to stay together. At about age five two or more or them will take over a pride by driving the resident males out (*5*). The pride is unlikely to be the pride in which the young males were reared.



Sequence of photographs showing a 12th or 13th Century ceremonial room mural in Anasazi (Pueblo) Indian cliff ruin in Canyon del Muerto, Arizona. Photographs ranging from 10.4 inches (upper left) to 5 miles (lower right) display the find in relation to its surroundings. All photographs are reproduced at their actual size.

ONE CAMERA: 10.4 INCHES TO INFINITY.

Slip an SX-70 Land camera in one jacket pocket, and some film packs in another, and you're ready to document your field trip.

Your field may be archeology, paleontology or criminology. It doesn't matter. What matters is the SX-70's versatility and portability.

Consider these unique virtues:

With a touch of the finger and a glance through the split-image rangefinder, you can focus to get a one-half life-size image of a subject 10.4 inches from your lens. Or a panoramic view at infinity. Or anything in between. (You can even get a *life*-size image, if you need it, by attaching our optional close-up lens.)

When you're working in the field, you have to concentrate on your surroundings. So it's a blessing to have a camera that doesn't compete for your attention. With the SX-70 you just focus on your subject, frame and press the electronic button. Period.

There's no worry that part of the subject will be cut out of your print. (Our single-lens reflex viewing system delivers what you see in the viewfinder.)

No preoccupation with your light conditions. (Our automatic electronic exposure control system takes care of *that.*)

No need to time the development of your prints as they emerge from the camera into broad daylight. And nothing to peel or throw away. (Our film takes care of all *that.*)

With the SX-70, you can take a sequence of up to ten photographs at intervals as short as 1.5 seconds.

The color values of those photographs are incredibly true-to-life. And our revolutionary metallized dyes will keep those colors from fading for an incredibly long time.

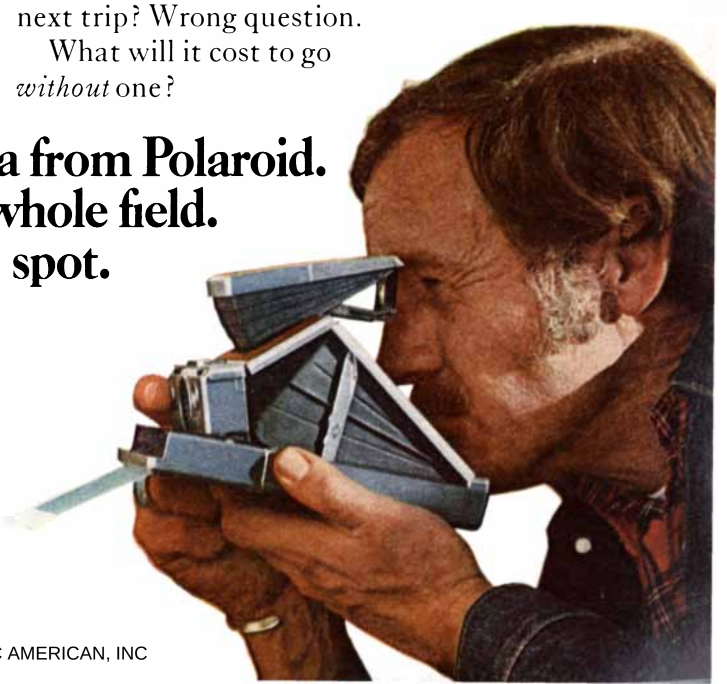
Furthermore, since SX-70 prints are virtually grainless, you could photograph a medieval castle (for example) at infinity, and study its gate hinges under a magnifying lens back at your desk.

But before you leave to return to your desk, you'll know that each of your photographs contains exactly what it's supposed to. Because the print develops itself completely in minutes. So, if the camera moved, or you want to change the angle of a particular shot, you can reshoot then and there.

What will it cost to have an SX-70 on your next trip? Wrong question.

What will it cost to go *without* one?

**The SX-70 camera from Polaroid.
Covers the whole field.
On the spot.**



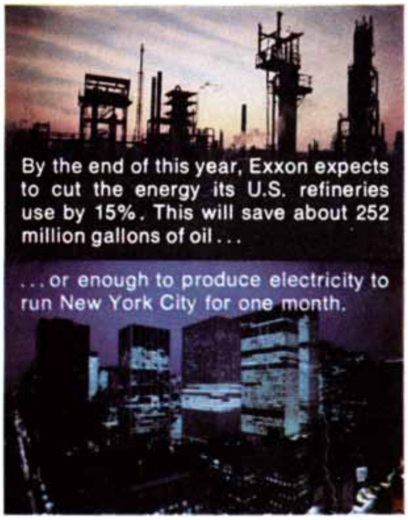
Saving energy

American industry offers a tremendous potential for energy savings right now.

One way to reduce our dependence on foreign oil is to use energy more efficiently. This means using less electricity and heating oil in our homes and saving gasoline in our cars.

American industry also offers a great potential for using energy efficiently. Why? Because industry uses at least one-third of all the energy consumed in the U.S. today.

Exxon will save enough energy to run New York City for 30 days.
All major industries require large amounts of energy. In fact, 25 percent of all our energy is consumed by just six industries: farming and food processing, aluminum, chemical, iron and steel, paper, and petroleum refining.


By the end of this year, Exxon expects to cut the energy its U.S. refineries use by 15%. This will save about 252 million gallons of oil...

...or enough to produce electricity to run New York City for one month.

In the case of Exxon, we use energy to make energy. But, by the end of this year Exxon expects to cut energy usage at our U.S. refineries by 15 percent of what we used in 1972. The energy we save could heat the homes in Pittsburgh for one year or provide enough electricity to run New York City for one month.

No more "full speed ahead."
A ship captain can save fuel the same way you save gasoline in your car. By slowing down our U.S. tankers and towboats and by cutting nonessential power demands, Exxon saved 5.5 million gallons of fuel last year. That is enough to power 5700 farm tractors for a year.


Exxon's U.S. tankers reduced fuel consumption by 5.5 million gallons last year...

...or enough to fuel 5700 average-sized farm tractors for one year.

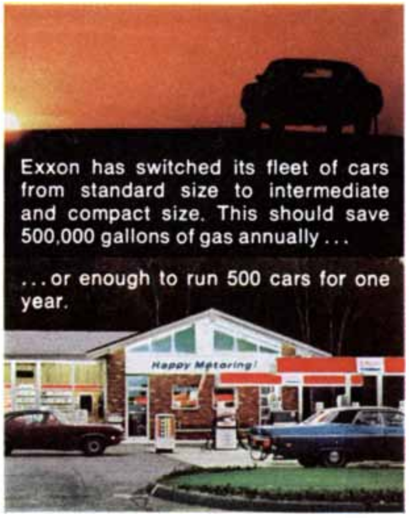
Last year, our 54-story headquarters in New York cut energy requirements by nearly 35 percent. Our Houston office reduced consumption of electricity by 7.3 million kilowatt-hours. That is enough electricity to power 575 average-sized homes for one year.


Exxon's Houston office has reduced its annual electric consumption by 7.3 million kilowatt-hours...

...or enough to provide electricity for 575 average-sized homes for one year.

Recently Exxon switched all company cars from standard size to intermediate or compact size. We expect that this will save 500,000 gallons of gasoline annually—or enough to run 500 cars for one year.

There is evidence of progress.
As a nation, there is evidence that we are making progress on curbing energy use. Figures from the U.S. Bureau of Mines and the American Petroleum Institute show that demand for energy *dropped* 3.3 percent in 1974 as compared to 1973. Gasoline consumption alone dropped 2 percent.


Exxon has switched its fleet of cars from standard size to intermediate and compact size. This should save 500,000 gallons of gas annually...

...or enough to run 500 cars for one year.

Take a good look at how you run your operation, whether it's a corporation, a small business, or a home in the suburbs. We think you'll be surprised at the ways you can use energy efficiently to conserve our nation's energy supplies. And you'll save money too.

EXXON

and are less often left untended. Perhaps also more communal hunting by their mothers makes for more successful hunting and hence a larger food supply.

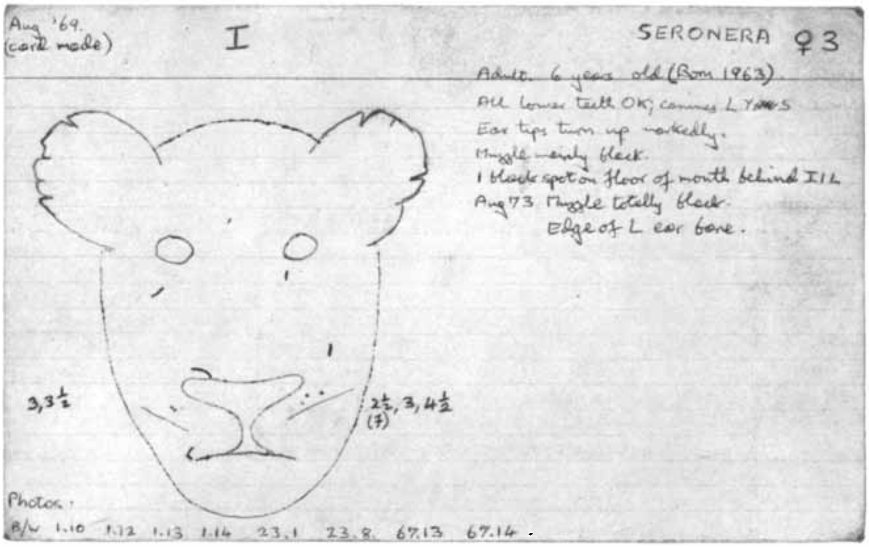
With survival as with births the takeover of a pride by new males appears to have an effect. On the basis of somewhat scanty data it seems that the mortality rate among cubs goes up for about three months after the new males arrive. An indirect cause might be stress effects on the females, making them produce less milk or hunt less efficiently. A direct cause is the killing of cubs by the new adult males. It is difficult to determine how prevalent this behavior is, but the available evidence suggests that males on taking over a pride are quite likely to kill cubs already there. The apparent avoidance of new males by females with cubs reinforces that evidence.

One is led to ask what evolutionary reason might explain the killing of cubs by new adult males. A plausible answer can be found in the short reproductive period of the males. It produces a strong selective pressure for the evolution of behavior that increases the reproductive output of the males during that period. A female whose cubs have died comes into estrus sooner and gives birth sooner to cubs fathered by a new male. Moreover, without the presence of older cubs to compete with new cubs for food the offspring of the new males will have a better chance of survival.

It is also interesting from an evolutionary point of view to ask why the matings of lions are so inefficient, that is, why the number of births is so low in relation to the number of matings. One reason appears to be that lions are under much less selective pressure than other animals to breed efficiently. They do not have a breeding season that must not be missed, and they run no risk of predation that would increase the importance of mating quickly and efficiently.

The question might be modified to ask why the females have evolved in such a way that they mate at times when it is highly likely that they will not produce cubs as a result. An explanation could be that an effect of the system is to reduce competition among the males by reducing the genetic value to them of each copulation. Assuming that lions mate every 15 minutes for three days, that only one in five three-day mating periods results in cubs, that the mean size of litters is 2.5 cubs and that the mortality among cubs is 80 percent, then a male must mate on the average some 3,000 times for each of his offspring reared to the next generation. When a single copulation is of such small potential consequence, the pressure on a male to fight another male for an opportunity to copulate is small. It is not to the evolutionary advantage of the females to have the males of a pride competing, because injury to the males causes them to be replaced more quickly by other males, which in turn results in increased mortality among the cubs of the females.

The same selective pressure may have a bearing on the synchrony of estrus among the females in a pride. It too tends to reduce competition among the males by ensuring that each male more often has an estrous female to be occupied with.


Aug '69.
(card made)
I
SERONERA ♀3
Adult. 6 years old (Born 1963).
All lower teeth OK; canines L Y[illegible]S
Ear tips turn up markedly.
Muzzle mainly black.
1 black spot on floor of mouth behind I1L
Aug '73 Muzzle totally black.
Edge of L ear bone.

Photos:
B/W 1.10 1.12 1.13 1.14 23.1 23.8 67.13 67.14.

IDENTITY CARD was prepared by the author for each lion in the prides he observed. It included photographs of the animal and a drawing that showed the animal's distinctive physical features, such as scars, nicks in the ears and the positioning of its whisker spots.

RURAL MARKET NETWORKS

They form geometrical patterns that provide a basis for close analysis. Two recent investigations bring out the critical role of such systems in contrasting countries: China and Guatemala

by Stuart Plattner

Periodic markets are essential to the articulation of agricultural societies. The markets usually meet once every few days; crowds come from the countryside to sell their farm produce and buy manufactured goods and foodstuffs from other areas. Government officials frequently visit on market days and local places of worship hold services, so that a farmer can combine economic, political and religious activities at one time and place. The market is also an arena of social interaction. It has much of the excitement of a fair, with friendships made, love affairs begun and marriages arranged. In many societies the end of a market day is often marked by drinking, dancing and fighting.

Anthropologists interested in the study of rural markets originally concentrated their attention on the interactions within the marketplace. An example is Sidney Mintz's analysis of the conflicts between peasants and changing coalitions of townsmen, tradesmen and officials in a typical Haitian market [see "Peasant Markets," by Sidney W. Mintz; SCIENTIFIC AMERICAN, August, 1960]. Recently interest has shifted from this kind of study to an examination of regional market networks as integrated (and integrating) systems. The shift is in large part due to the stimulus of an exciting conceptual development in economic geography: central-place theory. Here I shall report on one such study that demonstrated, with respect to China, the remarkable vitality of central-place systems of periodic markets and on another study showing, with respect to Guatemala, how normal economic evolution can be stifled by the imposition of a central-place system from above. By way of introduction to the two studies I shall first define central-place theory.

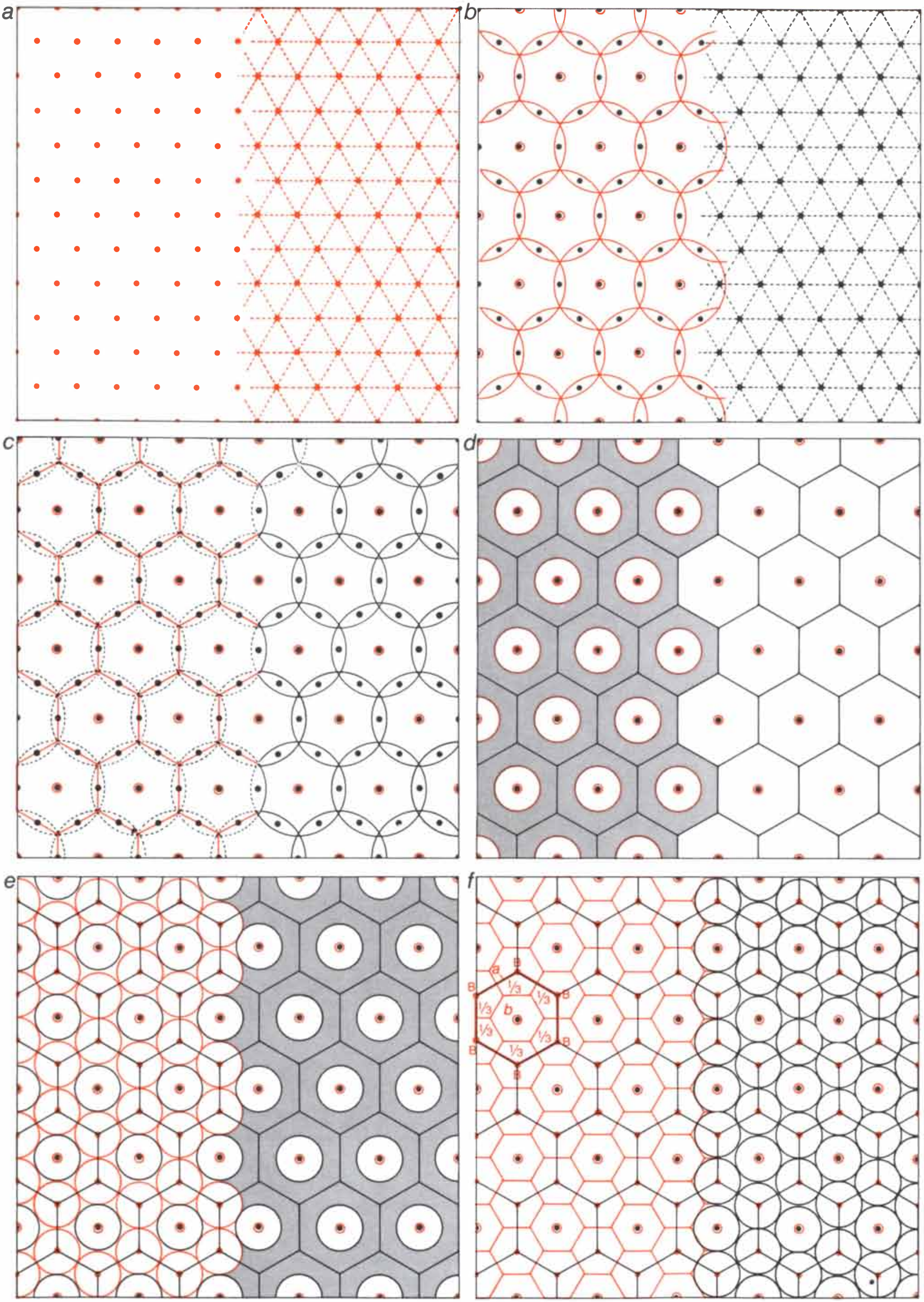
The basic elements of central-place theory were first proposed in the 1930's by the German economic geographer Walter Christaller and were elaborated by another German economic geographer, August Losch. The theory is built on several formal simplifying assumptions that are so abstract as to seem entirely unrelated to the real world. Yet the final model has been shown to be strikingly appropriate to several real-world agrarian societies; more important, it has brought a fresh understanding of how such systems work.

Christaller's fundamental model is predicated on the existence of a featureless landscape. It has neither such natural obstacles as mountains nor such natural highways as rivers. Transport is equally easy in all directions and, at least initially, the population is evenly distributed among "lowest level" settlements without markets [*see illustration on opposite page*].

It is further assumed that all the population's commercial actions are economically motivated, that the region is without social or political barriers to trade and that traders are freely competitive, so that markets can arise anywhere on the landscape in response to purely economic factors. The effect of these assumptions is that a constant demand for goods and services is evenly distributed across the featureless landscape. The assumptions also ensure that every member of the population will shop at the nearest market; as a result the markets that do arise in response to demand will be evenly spaced.

One additional assumption is that at any particular market all consumers will be charged the same price for the same article. This assumption has the following important implications. Given a stable market price for each kind of commodity, if the consumer is to calculate the real cost of an article, he must add to its price the cost of his visit to the market and his return home. This cost, of course, varies directly in proportion to the distance traveled. Therefore at some calculable distance from the market the real cost of any particular article will be so high that the consumer's demand for the article will drop to zero. The zero-demand distance will obviously be different for different kinds of goods and

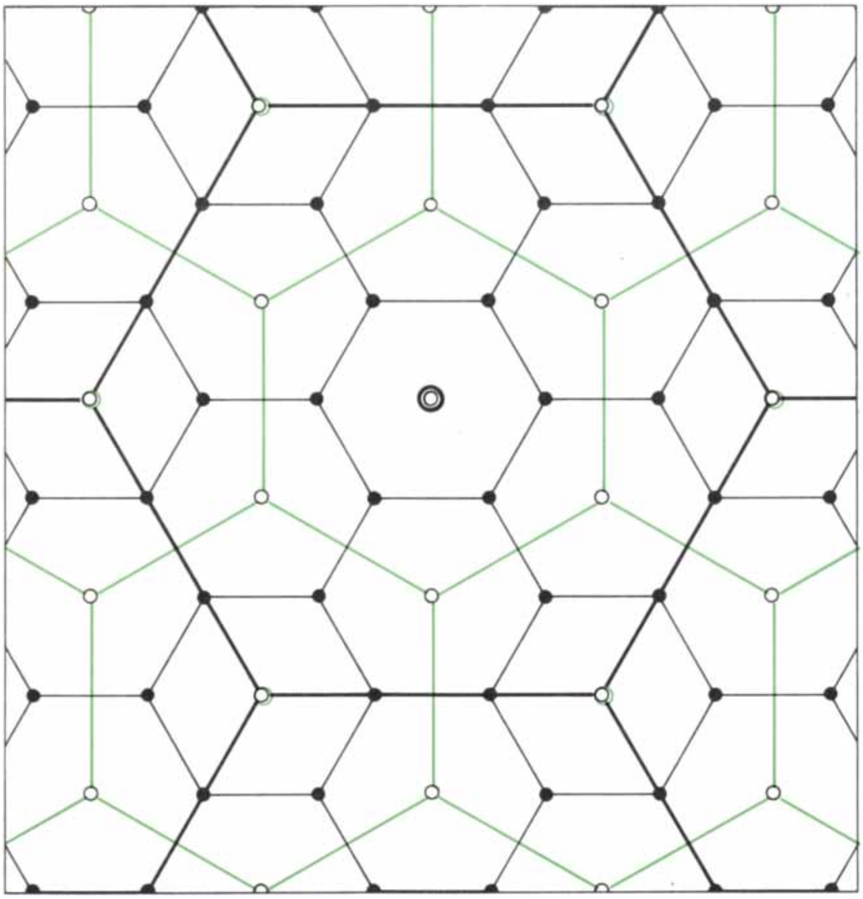
THEORETICAL GROWTH of a network of markets is traced through six successive stages in the illustration on the opposite page. The process begins, according to the central-place theory proposed by the economic geographer Walter Christaller, with a featureless landscape (*a*) occupied by equidistantly spaced marketless settlements, interconnected (*right*) by a grid of pathways that can all be traveled with equal ease. Given a demand within each settlement for high-value goods, evenly spaced markets for the goods will appear (*b, left*), creating a pattern of overlapping circles with radii equal to the maximum range of the goods. These are *A* centers (*small colored circles*); the overlap is soon eliminated (*c, left*) by the establishment of hexagonal trade boundaries between adjacent centers. The population will also express a demand for low-value goods with smaller maximum ranges. At first (*d, left*) the *B* centers that meet this demand (*larger circles*) will arise within existing *A*-center trade boundaries. Areas of unsatisfied demand (*gray*) remain between the *B* centers, however, and additional *B* centers arise to satisfy the demand (*e, left*). Packing converts boundaries between tangent *B* centers into smaller hexagons (*f, left*) and the network is complete. Each *A*-center area contains one *B* center and its entire hinterland (*upper left*) and one-third of the hinterland of each *B*-center hexagon tangent to the central hexagon.


a
b
c
d
e
f
a
B
B
1/3
1/3
B
1/3
b
1/3
1/3
B
B
1/3
B

services, but for each article it represents the radius of a circle that contains within its circumference all the consumers with a positive demand for the article. That radius is called the article's maximum range.

Every article offered for sale in a central-place market also has a minimum range. For example, whereas the number of consumers within the circle of positive demand for an article represents the total potential volume of sales of the article, that demand is not necessarily sufficient to ensure the economic success of the supplier. In order to cover expenses and receive a normal income each supplier must make a minimum number of sales within a specific interval of time. A circular territory large enough to generate this minimum number of sales is the supplier's trade area, and the radius of this circle is the article's minimum range.

Maximum and minimum ranges are related. The relation has been intensively explored by James H. Stine of Oklahoma State University. He notes that two real-world practices, itinerant marketing by peddlers and the periodic convening of markets, are related to a situation wherein the minimum range of an article exceeds its maximum range. For example, if a supplier needs more consumers than are willing to patronize a particular central-place market, he must periodically shift location in order to tap a larger area of positive demand. In effect he is adding together a series of maximum ranges until their sum is equal to the minimum range of the article he offers. That is why many real-world rural markets convene only periodically. For example, by meeting every seventh day the venders increase the maximum ranges of the articles being offered, since by then the consumer demands of each of the six previous days have been added together. Conversely, if the minimum range of an article is less than its maximum range, the supplier will not need to move from one market to another and may even find it worth while to increase the frequency of market meetings.



ELABORATED MARKET HIERARCHY is of the same kind outlined on the preceding page. Each *B* center (*black circle inside small black hexagon*), the lowest category, lies inside a triangle (*not shown*) connecting three centers of a higher category. The trade area of each *A* center (*colored circle inside larger colored hexagon*), the first higher category, encloses a *B* center and its hinterland and also one-third of the hinterland of each of the *B* centers tangent to the first. Each center in the next higher category (*black and colored circles inside the largest black hexagon*) encloses an *A* center and its hinterland and also one-third of the hinterland of each *A* center tangent to the first. No higher categories are shown in the diagram. The upward progression, however, can be continued indefinitely.

Obviously the ranges of some goods and services differ from the ranges of others by entire orders of magnitude. For example, an individual household will want commodities such as fresh vegetables quite often and yet the consumer is not willing to travel very far to buy them. The vegetable supplier in turn can be satisfied by the demands of a relatively small number of consumers, because in the course of a year each household will buy a great deal of his produce even though the individual transactions will be small. Commodities of this kind are said to be low-order goods. The area of positive demand for them, and thus their maximum range, is small; at the same time the frequency of the demand means that their minimum range is also small.

A contrary example is provided by what are called high-order goods and services, which are characterized by far greater maximum and minimum ranges. Goods and services of this kind, for instance articles of furniture or the advice of an attorney, have a great maximum range because their cost is high and the consumer's cost of travel is trivial by comparison. Furthermore, the consumer's positive demand for such goods and services arises infrequently, and so he is willing to travel a considerable distance to satisfy it. The minimum range of such goods and services is equally great for the same reason. The purchases made by any one consumer are infrequent; if the supplier is to carry out the number of transactions per unit of time necessary for economic success, his trade area must be an extended one.

It is a basic prediction of Christaller's central-place models that any center where high-order articles are available will also offer low-order ones but that a center where low-order articles are available will not necessarily provide high-order ones. It follows that different kinds of central places will come into existence, dealing in goods with different ranges. The models must therefore account for systematically integrated hierarchies of markets. How do such regional networks arise?

Beginning at the beginning, with a population living in marketless lowest-level settlements evenly distributed across a featureless plain, let there be a

How did Dodge Colt put so much in such a little car?

Introducing the '75 Dodge Colt Carousel hardtop. It comes with all the good things listed here, as do all the other Colt models: 2-door coupe, 4-door sedan, 4-door wagon and the Colt GT.
Prices start at $2,945.**

*Based on EPA test results for 1975 Dodge Colt, 1600 cc engine, 30 mpg highway cycle and 20 mpg city cycle.

**Manufacturer's suggested retail price for a '75 Dodge Colt coupe. Not included are state and local taxes, destination charge, license and title fees, and Dealer preparation charge, if any.

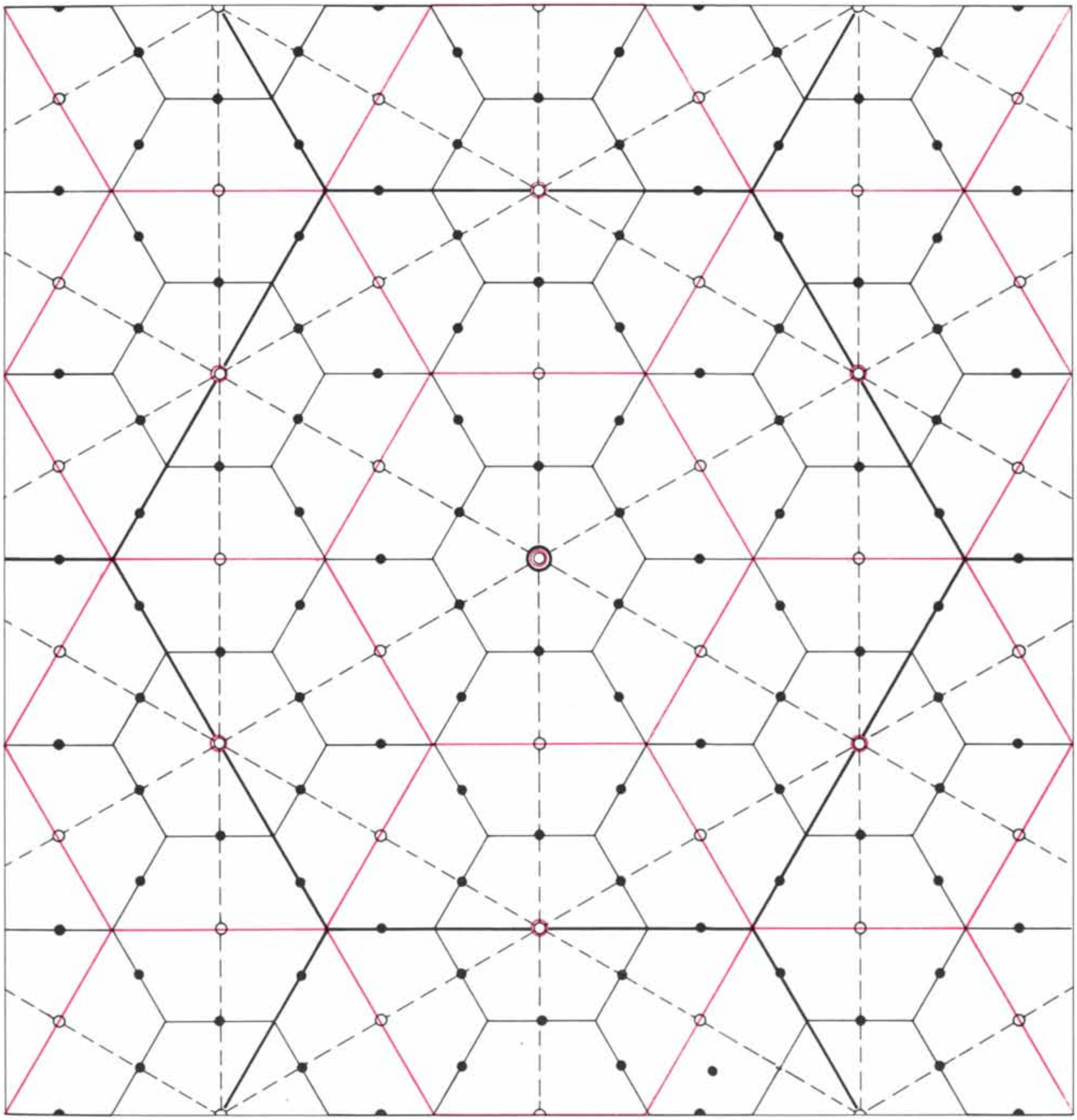
NO.1 FOR THE MONEY IS DODGE

CHRYSLER
MOTORS CORPORATION

Dodge
Dodge Trucks

positive demand for a spectrum of higher-order goods with an average maximum range M. Given the stipulation that no part of the plain will be left unserved, markets for these goods will arise at a series of lowest-level settlements so spaced that the circular trade area of each market slightly overlaps the trade area of the next. Packing efficiency demands that overlapping be eliminated by converting the circles into a grid of tangent hexagons, polygons that exactly fill the space and minimize the disadvantage to the consumers in the hinterland farthest removed from each central place. The market for higher-order goods at each of these central places we shall call an *A* center.

Given a similar positive demand for lower-order goods with a smaller maximum range, M_1, where will the markets for these goods, which we shall call *B* centers, arise? The first such markets will certainly be at the places where the *A*-center markets are already established. Economy of scale dictates this decision. Consumers are already visit-



SECOND MARKET HIERARCHY devised by Christaller makes allowance for the ease or difficulty of transport. *B*-center hexagons (*black*) are still found inside *A*-center hexagons (*color*), but now the additional *B* centers in the grid (*small black circles*) are located at halfway points on the communication lines (*broken black lines*) that connect the *A* centers (*colored circles*). Similarly, the next higher center (*black and colored circles*) contains within its hexagonal trade area one *A* center complete with hinterland and half of the hinterland of each of the *A* centers tangent to the first. In this hierarchy the market area (K) of each higher category includes the entire hinterland of one next lower center and half of the hinterland of the six centers tangent to it; thus K equals four.

There are times when only the very best will do.

A birthday.
Happy homecoming.
Welcome to fatherhood.
The Holiday season.
Bon voyage.
Welcome to the neighborhood.
Thank you.
Father's Day.
An anniversary.

...these are just some of the times you instinctively feel the gift should be equal in every way to the thought. You wish the best; you give the best.

What is best?

We'd like to give you an easy answer. We'd like to say that Maker's Mark is the finest whisky to come out of Kentucky.

But Bill Samuels, our president, won't let us say it.

He says there's no such thing as a universal "best" in whisky — that it's always a matter of personal taste.

We do know this, Bill set out to make the very best whisky he knew how — and as a fourth generation Kentucky distiller, he knew how.

And in his own eyes, he succeeded. In other words, he believes Maker's Mark is the best Kentucky whisky to be had at any price. And we can say that there's a large and steadily growing number of people who agree with him.

There is no way our little distillery can put up enough Maker's Mark to go around everywhere. So you may not be able to find it the first place you try — but that's another reason why those who receive a bottle will value it all the more.

Made from an original old style sour mash recipe by Bill Samuels, fourth generation Kentucky Distiller.

Maker's Mark Distillery, Loretto, Ky.,
Ninety Proof - Fully Matured.

Just for giving?

So far, we haven't touched on those times when you yourself might feel especially deserving. All we can say is this: Try a bottle of Maker's Mark. There's a good chance you will consider it the best you've ever tasted. And if so, you'll be all set for those times when only the very best will do.

Free booklet.

When you're in our neck of the woods, we cordially invite you to visit our little distillery on Star Hill Farm near Loretto. Meanwhile, if you have an interest in the history of whisky-making in the Bluegrass state, write for a free copy of our little booklet, "The Wonderful World of Kentucky Whisky."

ing the *A* centers. If while the consumers are visiting the higher-order market they also buy lower-order goods, the action splits the cost of transportation between the two classes of commodities, in effect reducing the real cost of the lower-order goods and increasing their maximum range.

The process will not, however, end there. Because the maximum range of the goods offered at *B*-center markets is smaller than the maximum range of the *A*-center goods the *A*-center hinterlands will still include large regions of unsatisfied demand for lower-order articles. If the unserved areas equal the minimum range of the M_1 suppliers, then additional *B* centers will arise in these vacant areas until no demand remains unsatisfied. When the process is complete, the kind of market hierarchy the model displays is known as the K-equals-three pattern. This is to say that the market area (K) lying inside a large *A*-center hexagon is made up of the hinterlands of three *B* centers as follows. One of the three is the *B*-center hinterland that lies entirely within the boundaries of the *A*-center hinterland. The other two are not entire *B*-center hinterlands; rather, they consist of a one-third share of each of the six surrounding *B*-center hinterlands that lie partly inside the *A*-center hexagon.

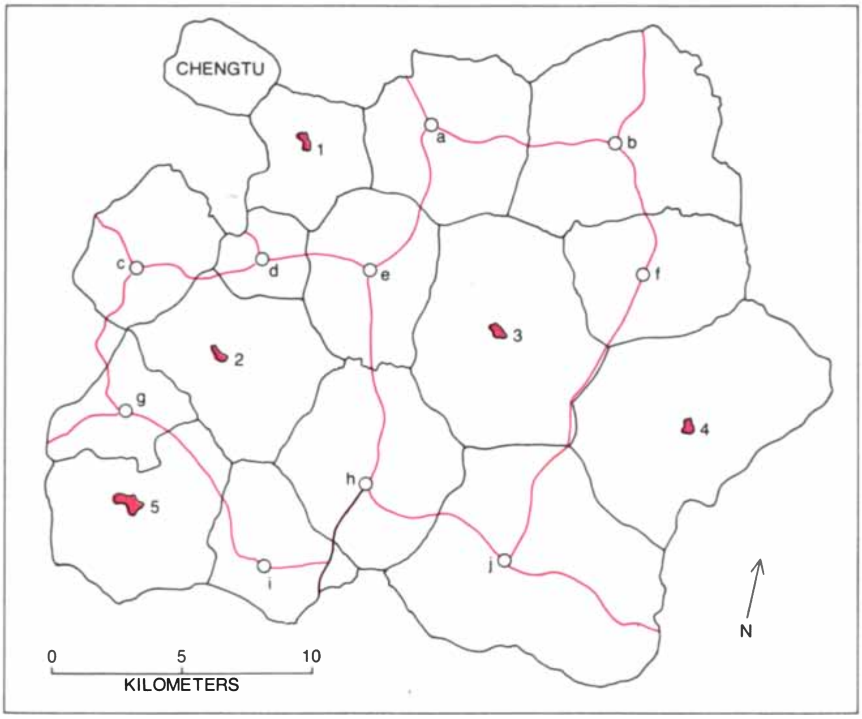
A K-equals-three hierarchy need not, as in this example, involve only two classes of markets. The same ordering process can be imagined for markets dealing in goods with any number of ranges. No matter how elaborate the hierarchy, the market specializing in goods with the greatest range would have a total hinterland equivalent to three market areas of the next lower order. Each of the three would embrace three of a still lower order, for a total of nine. The number of markets in the order below that would be 27, the number below that would be 81, and so forth.

A K-equals-three structure, as Christaller pointed out, is created by what he called the marketing principle. By this he meant that it is the structure most attuned to the needs of rural consumers because the ratio of high-level centers to low-level centers is at a maximum. Each successively lower-order market is equidistant from three higher-order markets, so that a *B*-center consumer or trader has a choice of three equally convenient places to shop for higher-order goods. At the same time an *A*-center consumer or trader in search of lower-order goods can tap not only the resources of his "home" *B*-center market but also those of six adjacent *B* centers.

By relaxing the constraint that travel within the model landscape be equally easy in all directions Christaller formulated a second kind of market hierarchy with a pattern different from the one I have been describing. Applying what he called the transport principle, he connected the higher-order *A* centers to one another by means of fully developed communication routes, such as roads or rivers. The existence of these developed routes meant that the consumer's travel costs were lower. Thus the true cost to the consumer of the goods and services obtained from such central places was also lower. A model of this kind approximates real-world situations such as when in mountainous terrain road construction is costly and it is economically more practical to travel greater distances on existing roads than to build more direct ones. A second real-world situation that is approximated by the transport-principle model is one where travel by land is not particularly difficult but where waterways, which allow freight to move at a lower cost per ton-mile than roads do, make the shipment of many categories of goods by barge economically attractive.

When the transport principle is applied, the *B* centers that arise to fill the unsatisfied demand for lower-order articles are no longer situated equidistant between three *A* centers, as in a K-equals-three hierarchy. Instead they arise at the halfway point along each of the communication routes that link the *A* centers. The market hierarchy that comes into being as a result is one of the K-equals-four type. That is, each *A* center controls a market area that embraces its own *B*-center hinterland and half (rather than a third) of the hinterland of each of the six surrounding *B* centers.

Christaller's two models, and they are not the only ones, are both intellectually and graphically elegant, but are they truly descriptive of real-world situations? Until a decade ago, when G. William Skinner, who was then at Cornell University, published the results of his 1949–1950 fieldwork in the Chinese province of Szechwan and his analysis of market reform during China's subsequent "Great Leap Forward," most central-place studies had been rationalizations of the nonexistence of Christaller's predicted patterns in the real world. Skinner's study reversed this trend. His diagrammatic abstractions of portions of



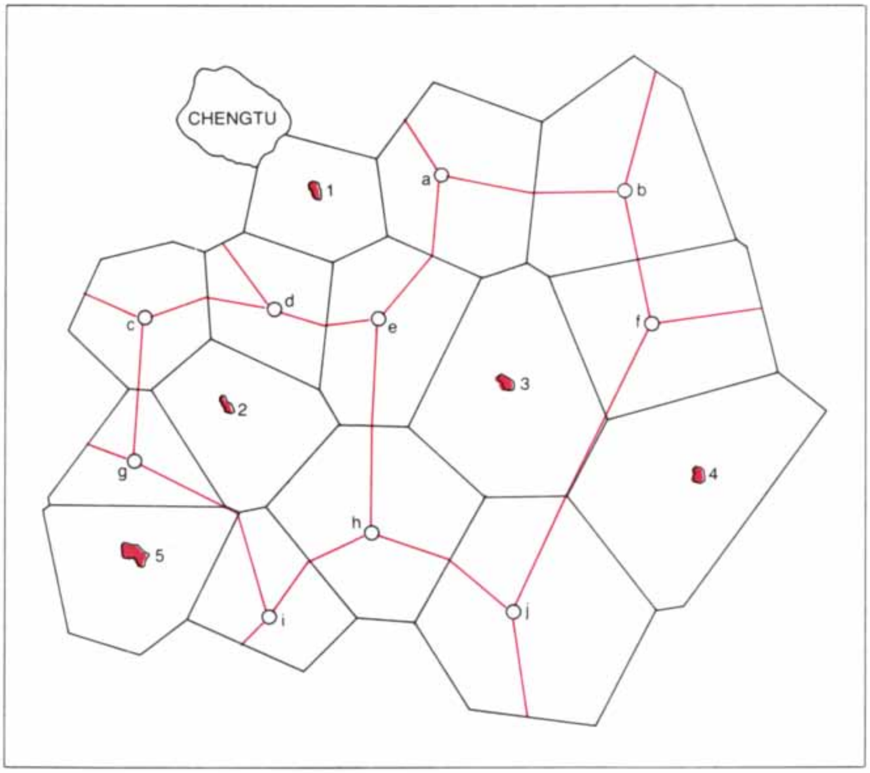
SZECHWAN COUNTRYSIDE southeast of the city of Chengtu was surveyed in 1949 and 1950 by G. William Skinner. Its five larger market towns (*colored areas*) are equivalent to Christaller *A* centers in the transformations shown on the opposite page; a number is assigned to each. The 10 smaller market towns (*black circles*) are equivalent to Christaller *B* centers; a letter is assigned to each. The lines define approximate market boundaries.

the Szechwan landscape excited many economic anthropologists. From this work emerged the possibility of using abstract models to analyze the complex ethnographic reality one encounters in fieldwork.

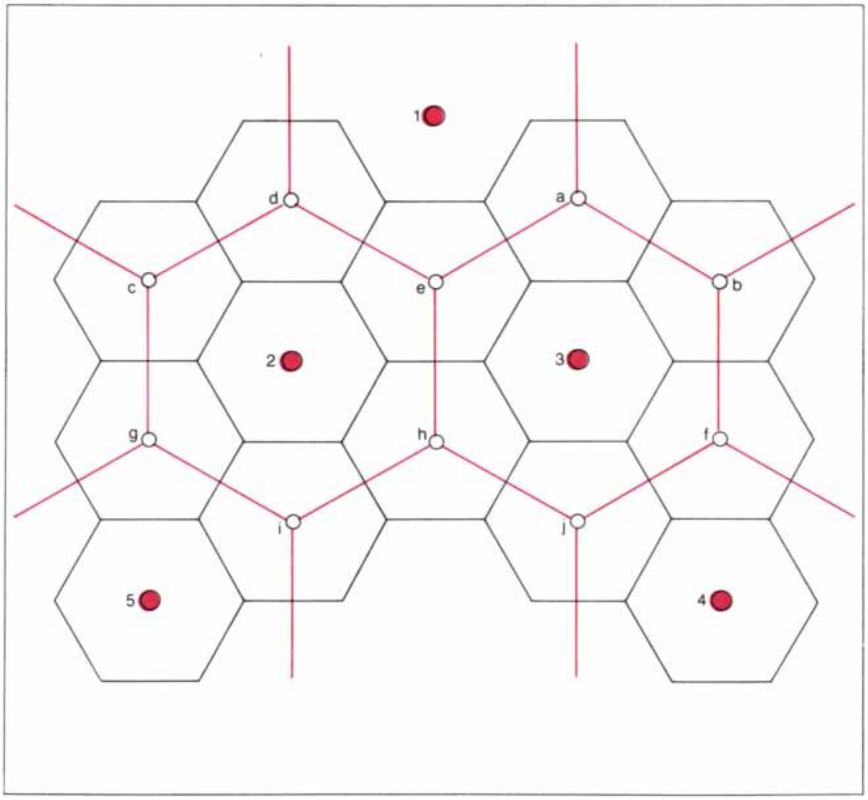
When Skinner observed the traditional rural marketing system in the vicinity of Chengtu, the second city of Szechwan, he found that both the geographical distribution of different kinds of markets and the hierarchical relations between them closely matched Christaller's models. Moreover, utilizing historical records for this part of China and other parts that covered a span of three centuries or more, he discovered that the rise and decline of rural markets was in accord with the models' predictions. Finally, by means of interviews and a close reading of contemporary accounts, Skinner was able to demonstrate that when the central government attempted to "leap forward" in the countryside by restructuring the traditional hierarchy of rural markets, the normal flow of trade was disastrously interrupted. Only by reinstituting the old marketing structure were the bureaucrats able to restore the orderly movement of agricultural products from the rural communes to the urban population.

The points of approximation between Skinner's findings and the predictions of Christaller's models include the following. First, Skinner found a broad network of what he calls standard markets that resemble the models' *B* centers. At these markets neighboring rural households could satisfy virtually all their normal commercial needs. The standard market was the downward terminus for manufactured goods and the upward entry point for the farm produce that flowed through the regional marketing system. (There was a lower category of market: a village-based meeting where some fresh vegetables were exchanged between farm households. Since the exchange was purely horizontal and did not involve the upward movement of produce, Skinner omitted these transactions from his analysis.)

Given an even distribution of the population, as in Christaller's model, such a *B* center should be surrounded by concentric rings of settlements: "marketless villages" numbering six in the first ring, 12 in the next, 18 in the third and so on. Skinner found that the usual Chinese pattern was a two-ring one: the average number of villages surrounding each standard market was 18 and the combined population of villages and market town averaged slightly less than



FIRST ABSTRACTION of the 15-center rural market hierarchy transforms meandering *B*-center (*black*) and *A*-center (*color*) boundary lines into two sets of intermeshed polygons. Unlike the *B*-center polygons, only two of the *A*-center polygons possess closed boundaries.



SECOND ABSTRACTION of the market hierarchy produces the Christaller-model pattern of larger and smaller hexagons. The five *A* centers (*color*) provide the foundation of the larger grid and the 10 *B* centers the smaller one. The K-equals-three pattern has emerged.

8,000. Such a "standard market community," the market town together with its satellite villages, was the basic unit of traditional Chinese rural society. As Skinner notes, an adult male member of such a unit has a "nodding acquaintance with almost every adult in all parts of the . . . system." That is because in the course of his life he has traveled to the *B* center literally thousands of times, as have all the farmers from the other villages in the market community.

Skinner found that the standard markets, or *B* centers, were geographically distributed in such a way with respect to higher-level markets, or *A* centers, that their trade boundaries could be rearranged into the pattern of tangent hexagons typical of the Christaller models. For example, 15 market towns southeast of Chengtu could quite readily be "abstracted" into a K-equals-three model [*see illustrations on preceding page*]. Another group of market towns to the northeast of Chengtu, where a portion of a mountain range restricts travel and a navigable stream provides an alternative to travel by road, could be abstracted into a K-equals-four model.

The spatial organization of the Szechwan markets was matched temporally by a complex interlocking system of alternate market days. If, for example, one *A*-center market met on the fourth, seventh and 10th days of a 10-day cycle, the adjacent *A* centers were likely to hold their markets on the third, sixth and ninth days or on the second, fifth and eighth days. The same nonconflicting periodicity characterized the schedule of *B*-center market days with respect to the *A*-center schedule. For instance, in the K-equals-three area southeast of Chengtu one *A* center, Chung-ho-chen, held its market on the first, fourth and seventh days of the 10-day cycle. The market days of the six surrounding *B* centers fell either on the second, fifth and eighth days of the cycle or on the third, sixth and ninth days.

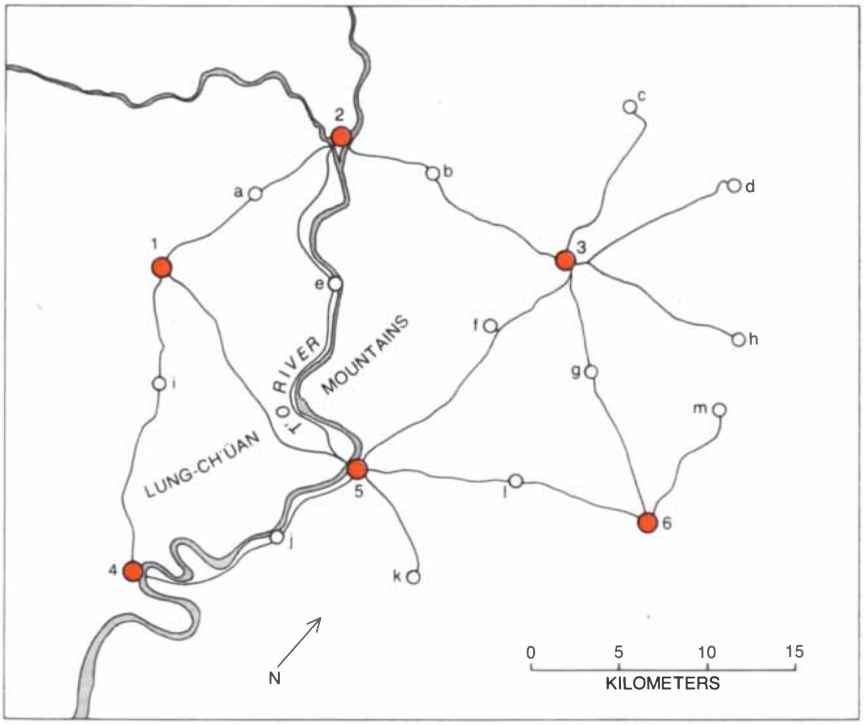
This, of course, served everyone's interests. If a farmer from a *B*-center satellite village happened to need some high-order article obtainable only at the *A* center, his trip to acquire the item need not have cost him a lost *B*-center market day. Similarly, one of the local elite who resided in the *B*-center hinterland and needed *A*-center goods or services, such as books or medical attention, could obtain them and still not leave his place of work on a *B*-center market day. Meanwhile the itinerant trader, who plays an important role in the vertical flow of goods, could make the rounds of the *B*-center markets on a schedule that enabled him to return to his *A*-center base at regular intervals, there to dispose of the lower-order commodities he had acquired during his travels and replenish his stock of higher-order merchandise.

One of Skinner's findings that was not predicted by Christaller's models is that the standard market community constitutes an endogamous unit. That is, whereas children who grow up in the same rural village may or may not marry one another, a marriage with someone from outside the community composed of the *B* center and its satellite villages is a rarity. Personal knowledge of the character of one's potential in-laws is, of course, an important factor in matchmaking, and information of this kind would be relatively easy to acquire within the confines of the marketing community. Skinner's finding provides an important corrective to a tendency among anthropologists to regard the village community as the only appropriate unit for investigation. It has made his colleagues more sensitive to the importance of supravillage levels of social integration.

Early in the 1950's the new regime in China introduced two institutions that were expected to "socialize" rural trade. The first was a network of state trading companies, each specializing in some category of commodities such as food grains or edible oils. The companies were headquartered in cities and major market towns and had branches in regional markets where sales and purchases were made. In effect the state trading companies constituted the national wholesale apparatus.

The second institution was a somewhat similar network of supply and marketing cooperatives. These were located in the countryside, collecting local products and distributing the goods provided by the state trading companies. By 1955 the cooperatives were handling at least 50 percent of the rural retail business. Skinner suggests that by that date most *B*-center shopkeepers and itinerant traders were "dependent on 'socialist commerce' for their supply of goods."

It was not long before public "self-criticism" began to reveal flaws in the new trade apparatus. The local and national press printed complaints that some rural farmers had to travel all the way to an urban center to buy such short-maximum-range articles as "a package of tobacco or [a few ounces] of



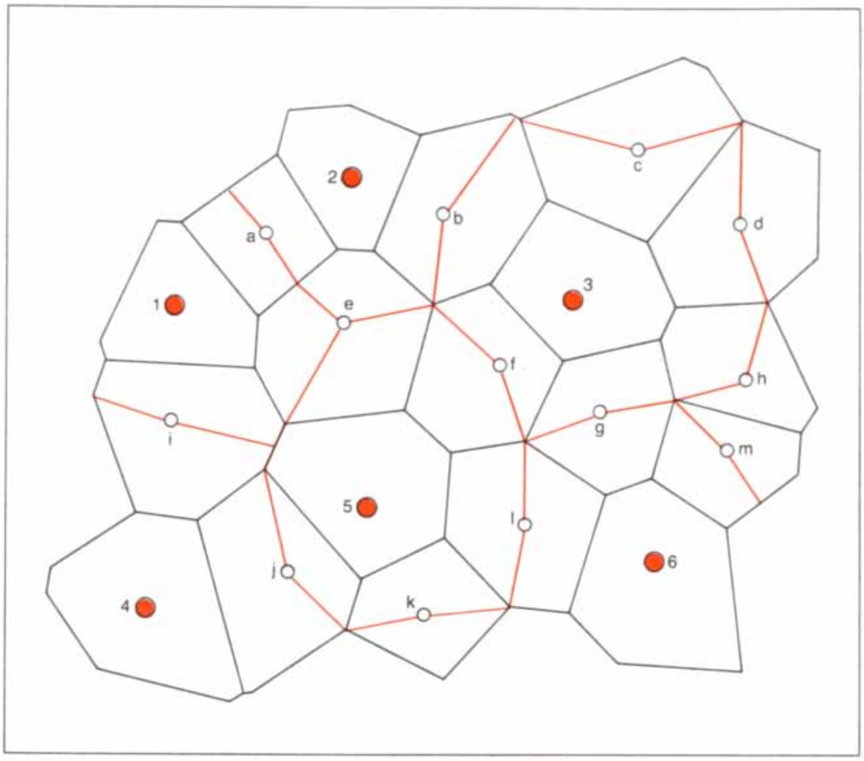
SECOND SZECHWAN AREA surveyed by Skinner lay northwest of Chengtu; the landscape included part of a mountain range and a navigable stretch of the T'o River. Skinner found six larger market towns (*color*), three of them on the river, and 13 smaller market towns (*black*), some of them in mountainous terrain. The two classes have been assigned numbers and letters respectively. Roads shown connect higher-level and lower-level towns.

wine," that perishables were "not purchased and sold in time" and that other agricultural produce, while "left unwanted" in the producing areas, was in short supply in "marketing areas." As the dislocations became more evident a government policy of relaxation was adopted. "Both rural production and rural-urban trade revived," Skinner notes. Traditional rural marketing practices continued thereafter until the government delivered a thunderbolt in 1958.

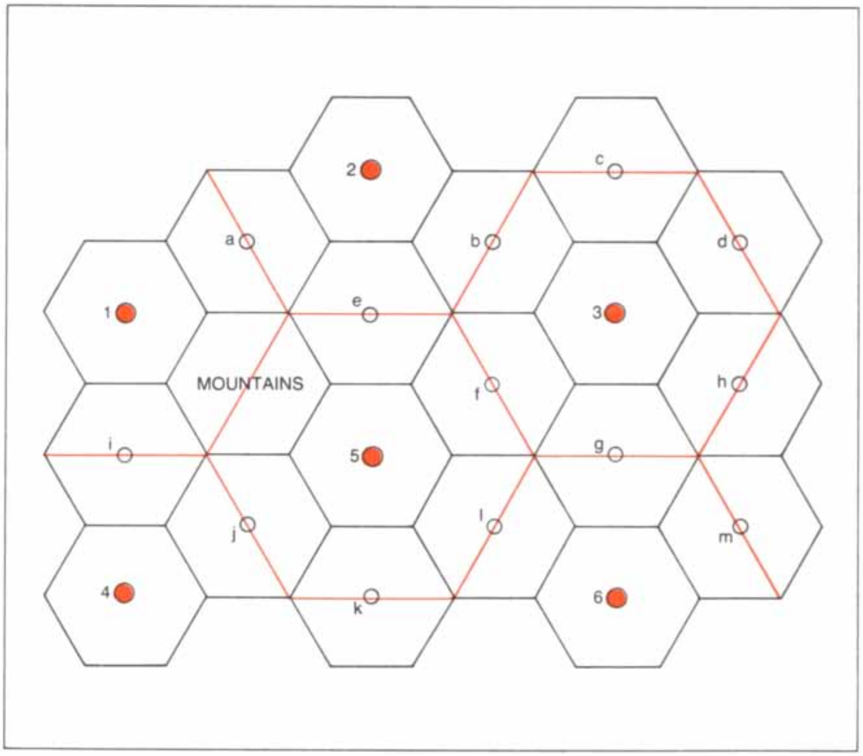
"With an agrarian economy . . . still largely unmodernized," Skinner writes, "with a rural marketing network over seven-eighths of whose nodes were traditional periodic markets, the regime attempted not merely to reform, not gradually to obviate, not eventually to bypass, but to *dispense altogether* with the traditional." In August and September, 1958, the supply and marketing cooperatives were merged to form a single department within each agricultural commune; the department was responsible for the sales of the commune's produce and the purchase of its necessities. Traditional marketing weeks were abruptly discontinued, Skinner notes, and the "periodic markets in most of agrarian China were closed."

Disruption followed almost immediately. For example, the planning commission in one county seat of Hupeh province reported that at the end of three months 400 kinds of products were in short supply and 32 categories of food were completely out of stock. Officials in one county in Shantung complained that such seasonal fruit crops as cherries and apricots were allowed to rot because of delays in organizing a market for the produce. In an effort to bolster the supply of young livestock for fattening, a group in Hopei province attempted a new procurement technique. A purchasing party was sent into two adjoining provinces to acquire 10,500 piglets. In August, 1959, the provincial newspaper noted angrily, the purchasing party had returned with more than 60 percent of its piglets dead or injured because of poor handling and the difficulties of long-distance transport. Incidents of the same kind were multiplied by the thousands; both the horizontal flow of produce between rural villages and the two-way vertical flow between city and countryside were slowed to a trickle.

In the countryside, Skinner reports, the farmers managed to tide themselves over the winter of 1958–1959 by using a quasi-religious marketing mechanism—the "temple fair"—that had not been specifically outlawed. By the summer of



FIRST ABSTRACTION of the 19-center hierarchy ignores river and mountainous terrain and divides the landscape into small (*black*) and large (*color*) polygons that represent the market areas of *B* centers and *A* centers respectively. Road network is also omitted.

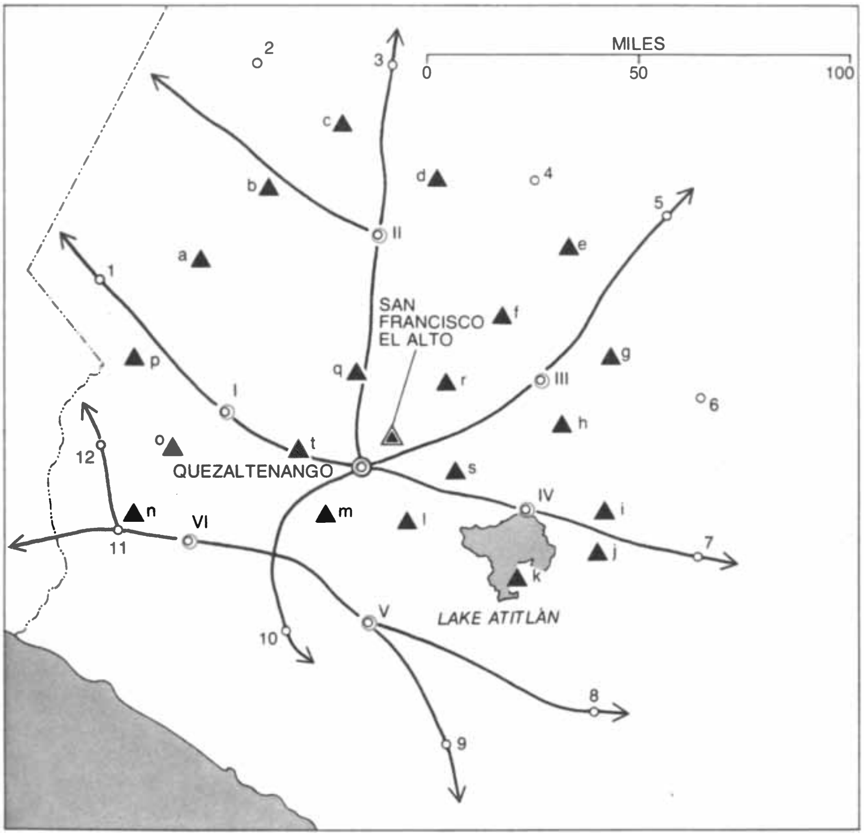


SECOND ABSTRACTION produces another Christaller-model pattern of larger and smaller hexagons. Note, however, that one of the smaller hexagons (*left*) does not include a *B*-center market town but is a space filler that makes allowance for the rigors of mountain transport. The pattern that has emerged is that of the Christaller K-equals-four model.

1959, however, as the piglets were dying on their way to Hopei, the regime realized the extent of its mistake. In August the unofficial voice of the government, the newspaper *Ta Kung Pao,* hailed a "new departure": in certain provinces "rural markets" were being introduced. A month later, on September 23, 1959, the State Council made the government's retreat official. "Rural markets," the council declared, "facilitate the exchange of commodities." Supply and marketing departments were therefore directed to organize rural markets "so as to smooth the flow of materials between town and countryside . . . and activate the rural economy." As Skinner points out, the fear of losing face prevented the State Council from speaking in the new directive of any reestablishment of a traditional system. The directive did note, however, that the schedules of the new rural markets were to be set "in accordance with old usage."

The retreat was total, although some years were to pass before the traditional rural marketing system recovered its full vigor and the government whittled down the size of the agricultural communes to the point where the size and population of an average rural "production brigade" were roughly equal to those of a village, and those of an average commune were roughly equal to the traditional town-plus-satellite-villages configuration of a *B*-center marketing system. As Skinner concludes, technological advances in the future are bound to transform the traditional Chinese system, but in this critical period of Chinese history it was not the government's design that imposed its form on rural socialization but the design of the traditional marketing community.

Skinner's account of the vitality of the traditional Chinese rural market system has greatly influenced other students of rural societies. One particularly interesting study is Carol A. Smith's detailed analysis of the market system in the western highlands of Guatemala. The rural markets of China and Guatemala have certain similarities: both are founded on periodic markets held at various locations, and the markets serve not only the townsfolk but also the rural population. As Smith's work demonstrates, however, the resemblance between the two systems is superficial. Although it is possible to abstract the geographical pattern of the Guatemalan central places so that it nearly duplicates a Christaller model, the difference between the Guatemalan and the Chinese models is profound. This should not be taken as demonstrating any serious deficiency in Christaller's theory, but it is a warning that his models cannot be expected to duplicate all the subtleties of real-world situations, which rarely meet the assumptions of central-place theory. The real-world situation in China is a culturally homogeneous one, although the culture is socially stratified. In Guatemala the culture not only is socially stratified but also, as a result of the region's colonial history, is divided into two quite different worlds: the Ladino and the Indian. In itself the dichotomy might not be significant. What gives it significance is the fact that the Ladino world controls the national economy.



WESTERN GUATEMALA within a 100-mile radius of Quezaltenango, the second-largest city in the nation, was surveyed in the 1960's by Carol A. Smith. She found a ring of six sizable Ladino market towns surrounding Quezaltenango (*Roman numerals*) and 12 lesser Ladino market towns (*Arabic numerals*) peripheral to them. An independent grid of markets dealing in Indian agricultural produce (*lettered black triangles*) was also present, but only a few of the 20 major bulking centers were located on the main highways. In the same way that Quezaltenango was central to the network of Ladino markets, a nearby major Indian market, San Francisco el Alto, was central to the network of rural bulking centers.

Indians, whose first language is Mayan rather than Spanish, make up about half the population of Guatemala. Most of them work as semisubsistence farmers and live according to Mayan cultural patterns. Ladinos, whose first language is Spanish, are heirs to a tradition that is culturally (and often genetically) a mixture of the Spanish and the Indian. The Ladinos usually live in the towns, and they are most often engaged in nonagricultural work. They are also socially, politically and economically the heirs of the conquistadors, and thus they are the dominant power in the nation.

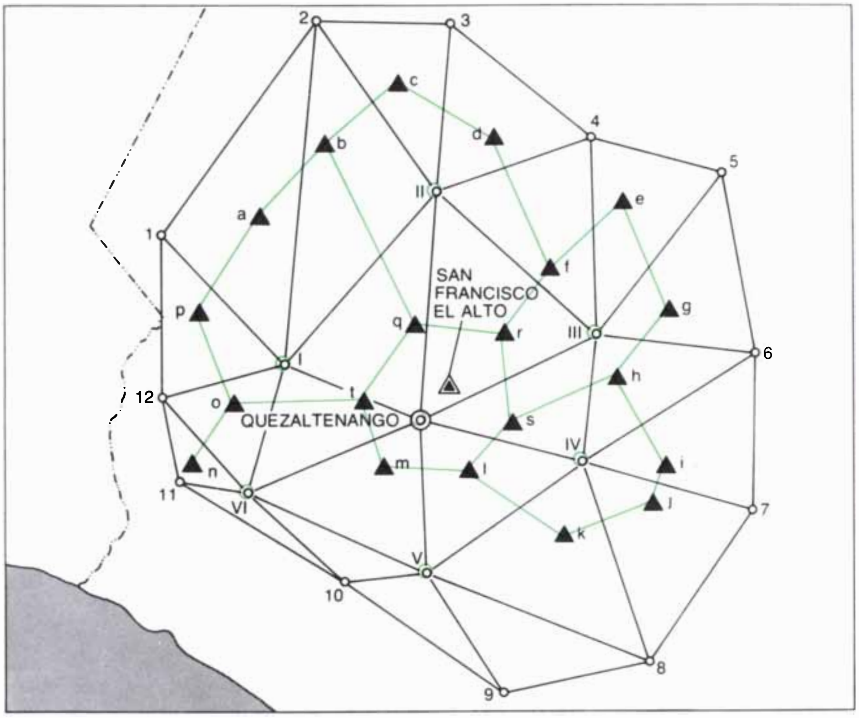
The market system studied by Smith surrounds the regional center of western Guatemala: Quezaltenango, a central place with a population of some 40,000 [*see illustration at left*]. Many commercial establishments are permanently located there, and manufactured goods of importance to the economy of the surrounding area are available in the city. Moreover, the size of the resident nonagricultural population means that Quezaltenango generates a strong demand for farm produce from the adjacent countryside. In the regional marketing hierarchy Smith classifies Quezaltenango as the pinnacle of the Ladino market-town system.

The city's major lines of communication lead radially outward to six intermediate-level Ladino market towns. For the sake of comparison the six can be likened to the *A* centers of the Christaller models. Beyond these *A* centers, lying either on the same major lines of communication or on branches or minor roads, are 12 lower-level Ladino market towns; they can be likened to the *B* centers of the models. These 19 central places, Quezaltenango included, constitute the framework of Ladino economic control: the vertical distribution of goods manufactured outside western Guatemala—fuel, machinery, iron tools and factory-made cloth and clothing—begins in the city and moves downward to the *A* centers and thence to the *B* centers. Quezaltenango and five of the six *A* centers are also the administrative capitals of their districts and hence control political disbursements as well.

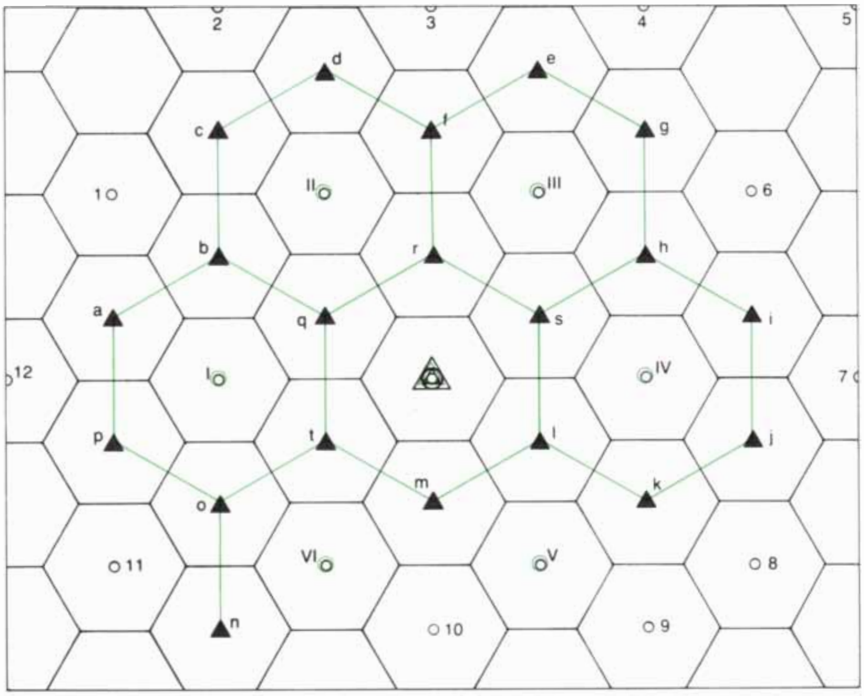
The geographical distribution of the Ladino central places differs from that predicted by the Christaller models. When the demand for goods is distributed evenly over the landscape, the models predict that the *A* centers will be widely spaced and that the lesser *B* centers will be distributed between the *A* centers. In western Guatemala, however, the *B* centers are peripheral to the *A* centers. According to Smith, the abnormal market distribution is due to an uneven distribution of demand: purchasing power declines significantly in proportion to distance from the city. Moreover, the number of Ladino households in the *A* centers is greater than the number in the *B* centers, and the Ladino population of Quezaltenango is the largest of all. The Indian population reflects the same economic gradient. It is most commercialized in the central zone. Fewer of the Indian households there engage exclusively in subsistence agriculture, and so the Indian family income is higher. In response to these factors the best roads are in the central zone and the worst ones are on the periphery.

The subsistence needs of the Ladino households in the region are largely met by purchases of Indian-raised produce. The trade network that delivers the produce is composed of a grid of lower-level central places that Smith calls rural bulking centers. The predominance of wholesale trade over retail trade distinguishes the bulking centers from their Ladino counterparts. They are situated in Indian communities where large commercial establishments are rarely if ever found. The Indian markets meet periodically, primarily to collect produce for export to the Ladino market towns but also to meet the Indian demand for such products as salt and cloth.

Smith identified 20 major rural bulk-



FIRST ABSTRACTION that allowed Smith to connect the six intermediate Ladino market towns to Quezaltenango and to one another with straight lines and similarly to connect the peripheral and the intermediate Ladino market towns produced a curious pattern. All 20 rural bulking centers (*black*) fell within one or another of the 24 triangles formed by lines connecting the Ladino markets. When the bulking centers were grouped into irregular but tangent hexagons (*colored lines*), each contained one of the higher-level Ladino markets.

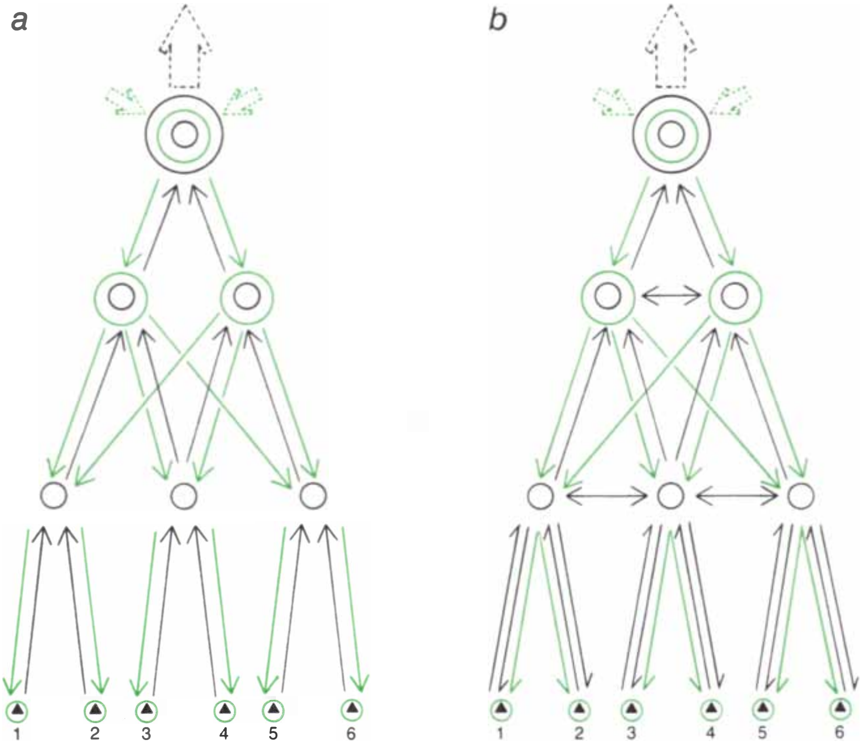


SECOND ABSTRACTION of the Guatemala market hierarchy is arranged to approximate a Christaller K-equals-three model. The divergence of the real-life situation from the model is evident. *B*-center Ladino markets are not interspersed among the Ladino *A* centers but are peripheral to them. Only in the highest central place is a rural bulking center found along with a Ladino market. Moreover, the shares of bulking-center hinterland "belonging" to Ladino *A* centers actually represent not a two-way flow between higher and lower markets in a single integrated hierarchy but a one-way Ladino levy of Indian foodstuffs.

ing centers in western Guatemala, not including the largest one. This was situated near Quezaltenango in the Indian town of San Francisco el Alto, and it played a predominant role in the wholesaling of Indian produce. In effect the San Francisco market occupied a position with respect to the rest of the rural bulking-center network very much like the position of Quezaltenango with respect to the network of Ladino market towns. Smith also found that when she connected the various Ladino market towns with abstract straight lines, most of the triangles so constructed contained a rural bulking center [*see top illustration on preceding page*]. Such a pattern, with lower-level markets enmeshed in a grid that connects higher-level markets, superficially duplicates Christaller's K-equals-three, or marketing, principle.

How does the Guatemalan marketing hierarchy differ from the traditional Chinese one, and what do the differences mean? Smith argues as follows. The regular distribution of central places in Guatemala, as in China, indicates that distance is an important influence on marketing behavior that will directly affect market location. The concentration of major central places near Quezaltenango, the economic center of western Guatemala, however, is not predicted by theory. It suggests that a spatial economy can be strongly influenced both by ethnic considerations and by the real-world distribution of political power. In Guatemala, Ladino political power translates into monopoly over certain aspects of the economy, and this monopoly control is realized in the centripetal pattern of market centers, where increasingly smaller centers feed into the Ladino political capitals.

How is the dominance of Ladino centers maintained? For one thing, goods are not free to flow in any direction in the market structure, as models of normal central-place systems assume. In Guatemala, Smith notes, the Indian rural bulking centers play a role that the Ladino market centers do not. The bulking of foodstuffs by the Indian market-places and their shipment to the Ladino towns is a one-way process. It is not reciprocated by a downward distribution of farm goods shipped to the Ladino market towns from other zones. When the foodstuffs reach the Ladino towns, they remain there, to be consumed by the local Ladino population. This process is facilitated by the road network (all of which feeds into the Ladino towns) and by the Ladino control of road building, trucking and storage facilities in the region.



TWO KINDS OF RURAL-URBAN FLOW are symbolized in this diagram. The movement of foodstuffs in both is indicated by black arrows; in both this flow culminates in agricultural exports (*broken black arrows, top*). The movement of manufactured goods in both is indicated by colored arrows; some imported goods (*broken colored arrows*) reaching import-export centers eventually flow to the lowest rural markets. In the first kind of flow (*a*) the one-way movement of produce means that foodstuffs from one rural market cannot reach any other rural market. Under these circumstances all farmers are under pressure to grow a little of everything. In the second kind of flow (*b*) produce grown in one part of the country reaches many markets, and local farmers can better their economic lot by raising specialized cash crops. The first system resembles the one imposed from above in Guatemala, analyzed by Smith; the second, the traditional one in China analyzed by Skinner.

Smith finds this to be a fact of the greatest importance because, far from facilitating the development of the Guatemalan hinterland, the marketing system actually blocks development. In the contrasting example of China the free flow of goods through the market system means that any local community can, if it chooses, specialize in the production of a single commodity, secure in the knowledge that the staples the community does not produce will be forthcoming from the market. In Guatemala, since the Ladino market towns do not redistribute foodstuffs, the only exchange of agricultural goods between villagers is at the level of the rural bulking centers. No bulking center is connected to any other bulking center through a higher-level market [*see illustration at left*].

The result, in Smith's interpretation, is that the market system represented by the Ladino market towns neither integrates the region nor assists in the development of regional economic specialization. Instead the Ladino market towns "divide and conquer" the local Indian communities, forcing the rural farmers to engage mainly in subsistence agriculture for the sake of self-sufficiency and thereby limiting the Indians' capacity to raise cash crops that might supplement their income.

Smith notes that the Guatemalan pattern is to be found in other parts of the world, and she argues that it is no coincidence that the areas where it exists are all former colonial possessions. The pattern, she suggests, arises from a process of "underdevelopment" that is impressed on the marketing structure of subject countries by the colonizing nations. In such marketing systems the flow of agricultural products begins in a network of low-level bulking centers and travels upward through a hierarchy of markets to a central place for consumption, for export or for both. Paralleling this upward flow is a downward flow of imported manufactures. There is, however, no horizontal flow of farm produce between the different agricultural districts of the region.

In such a system whatever underdevelopment may already exist in the region will be perpetuated. Farmers cannot specialize lest they gamble their survival on a market system that is not structured to deliver the food they need. Artisans cannot become manufacturers, not only because the goods imported from industrialized nations abroad are likely to be cheaper and better made than the local product but also because, as with foodstuffs, there is no provision for the horizontal flow of goods.

The dependency of the underdeveloped system is also perpetuated by the flow of profits as investment capital. In an independent system the profits from commerce and agriculture are available for investment in such support services as communication, transportation and education. Basic resources of this kind increase the productivity of labor over the long run. In a dependent system such as Guatemala the market facilitates the introduction of imported goods. These suppress the development of local manufacturing and allow the proceeds from commerce to flow up and out of the subject country to the already developed supplier. There the profits are available for investment in the basic resources of the developed nation, further increasing the gap in productivity between the developed and the underdeveloped countries. In Guatemala the gap widens not only between Indians and Ladinos but also between Guatemala and the developed nations.

I see no solution, easy or hard, for the dismal problem of the gap between the developed and the underdeveloped nations. The recent creation of cartels among the exporters of primary products such as petroleum is bringing a flow of capital from the developed nations into the exporting nations. Whether this capital will be used to create the basic understructure of education, communication and the other services that support the economies of the developed nations remains to be seen. Meanwhile nations without major natural resources to monopolize (for example India) will face hard times indeed. If studies such as Skinner's and Smith's hold any lesson for policymakers, it is that national development requires a free and vigorous system of internal exchange and that such exchange will be facilitated by a hierarchical network of central-place markets. As for external exchange, particularly exchange with more developed nations, the underdeveloped nations may find that judicious restriction is the better part of wisdom.

The Search for Extraterrestrial Intelligence

There can be little doubt that civilizations more advanced than the earth's exist elsewhere in the universe. The probabilities involved in locating one of them call for a substantial effort

by Carl Sagan and Frank Drake

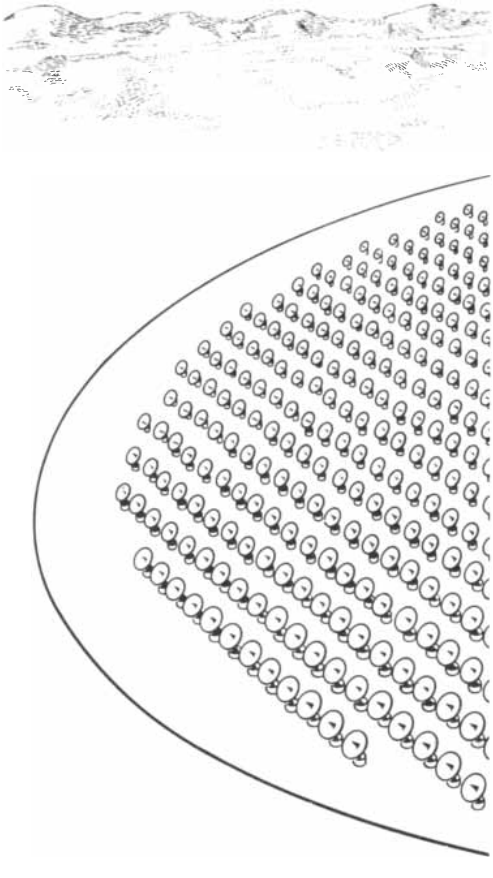
Is mankind alone in the universe? Or are there somewhere other intelligent beings looking up into their night sky from very different worlds and asking the same kind of question? Are there civilizations more advanced than ours, civilizations that have achieved interstellar communication and have established a network of linked societies throughout our galaxy? Such questions, bearing on the deepest problems of the nature and destiny of mankind, were long the exclusive province of theology and speculative fiction. Today for the first time in human history they have entered into the realm of experimental science.

From the movements of a number of nearby stars we have now detected unseen companion bodies in orbit around them that are about as massive as large planets. From our knowledge of the processes by which life arose here on the earth we know that similar processes must be fairly common throughout the universe. Since intelligence and technology have a high survival value it seems likely that primitive life forms on the planets of other stars, evolving over many billions of years, would occasionally develop intelligence, civilization and a high technology. Moreover, we on the earth now possess all the technology necessary for communicating with other civilizations in the depths of space. Indeed, we may now be standing on a threshold about to take the momentous step a planetary society takes but once: first contact with another civilization.

In our present ignorance of how common extraterrestrial life may actually be, any attempt to estimate the number of technical civilizations in our galaxy is necessarily unreliable. We do, however, have some relevant facts. There is reason to believe that solar systems are formed fairly easily and that they are abundant in the vicinity of the sun. In our own solar system, for example, there are three miniature "solar systems": the satellite systems of the planets Jupiter (with 13 moons), Saturn (with 10) and Uranus (with five). It is plain that however such systems are made, four of them formed in our immediate neighborhood.

The only technique we have at present for detecting the planetary systems of nearby stars is the study of the gravitational perturbations such planets induce in the motion of their parent star. Imagine a nearby star that over a period of decades moves measurably with respect to the background of more distant stars. Suppose it has a nonluminous companion that circles it in an orbit whose plane does not coincide with our line of sight to the star. Both the star and the companion revolve around a common center of mass. The center of mass will trace a straight line against the stellar background and thus the luminous star will trace a sinusoidal path. From the existence of the oscillation we can deduce the existence of the companion. Furthermore, from the period and amplitude of the oscillation we can calculate the period and mass of the companion. The technique is only sensitive enough, however, to detect the perturbations of a massive planet around the nearest stars.

The single star closest to the sun is Barnard's star, a rather dim red dwarf about six light-years away. (Although Alpha Centauri is closer, it is a member of a triple-star system.) Observations made by Peter van de Kamp of the Sproul Observatory at Swarthmore College over a period of 40 years suggest that Barnard's star is accompanied by at least two dark companions, each with about the mass of Jupiter. There is still some controversy over his conclusion, however, because the observations are very difficult to make. Perhaps even more interesting is the fact that of the dozen or so single stars nearest the sun nearly half appear to have dark companions with a mass between one and 10 times the mass of Jupiter. In



"CYCLOPS," an array of 1,500 radio antennas each 100 meters in diameter, is one system that has been proposed as a tool for detecting signals from extraterrestrial civ-

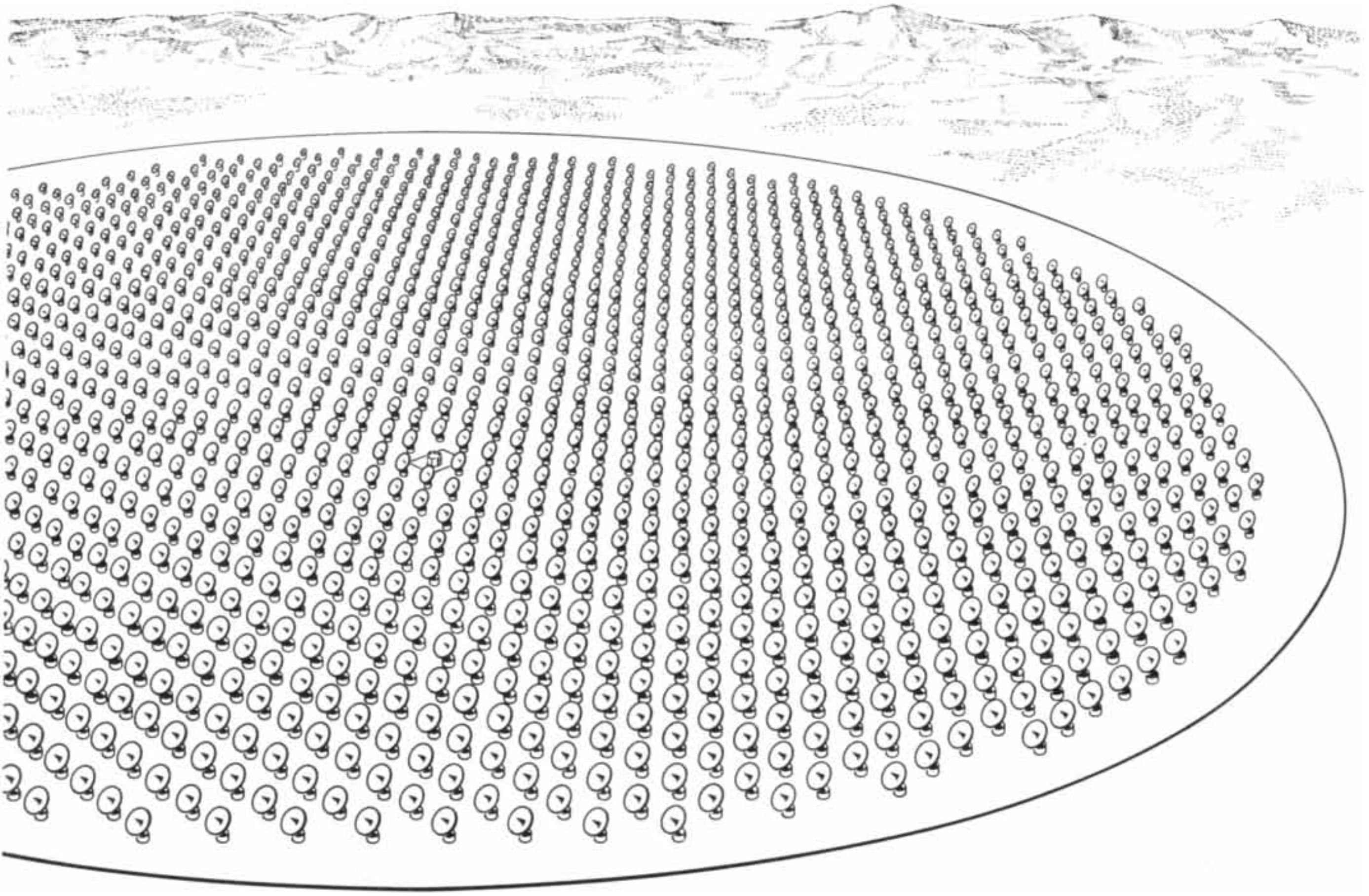
addition many theoretical studies of the formation of planetary systems out of contracting clouds of interstellar gas and dust imply that the birth of planets frequently if not inevitably accompanies the birth of stars.

We know that the master molecules of living organisms on the earth are the proteins and the nucleic acids. The proteins are built up of amino acids and the nucleic acids are built up of nucleotides. The earth's primordial atmosphere was, like the rest of the universe, rich in hydrogen and in hydrogen compounds. When molecular hydrogen (H_2), methane (CH_4), ammonia (NH_3) and water (H_2O) are mixed together in the presence of virtually any intermittent source of energy capable of breaking chemical bonds, the result is a remarkably high yield of amino acids and the sugars and nitrogenous bases that are the chemical constituents of the nucleotides. For example, from laboratory experiments we can determine the amount of amino acids produced per photon of ultraviolet radiation, and from our knowledge of stellar evolution we can calculate the amount of ultraviolet radiation emitted by the sun over the first billion years of the existence of the earth. Those two rates enable us to compute the total amount of amino acids that were formed on the primitive earth. Amino acids also break down spontaneously at a rate that is dependent on the ambient temperature. Hence we can calculate their steady-state abundance at the time of the origin of life. If amino acids in that abundance were mixed into the oceans of today, the result would be a 1 percent solution of amino acids. That is approximately the concentration of amino acids in the better brands of canned chicken bouillon, a solution that is alleged to be capable of sustaining life.

The origin of life is not the same as the origin of its constituent building blocks, but laboratory studies on the linking of amino acids into molecules resembling proteins and on the linking of nucleotides into molecules resembling nucleic acids are progressing well. Investigations of how short chains of nucleic acids replicate themselves in vitro have even provided clues to primitive genetic codes for translating nucleic acid information into protein information, systems that could have preceded the elaborate machinery of ribosomes and activating enzymes with which cells now manufacture protein.

The laboratory experiments also yield a large amount of a brownish polymer that seems to consist mainly of long hydrocarbon chains. The spectroscopic properties of the polymer are similar to those of the reddish clouds on Jupiter, Saturn and Titan, the largest satellite of Saturn. Since the atmospheres of these objects are rich in hydrogen and are similar to the atmosphere of the primitive earth, the coincidence is not surprising. It is nonetheless remarkable. Jupiter, Saturn and Titan may be vast planetary laboratories engaged in prebiological organic chemistry.

Other evidence on the origin of life comes from the geological record of the earth. Thin sections of sedimentary rocks

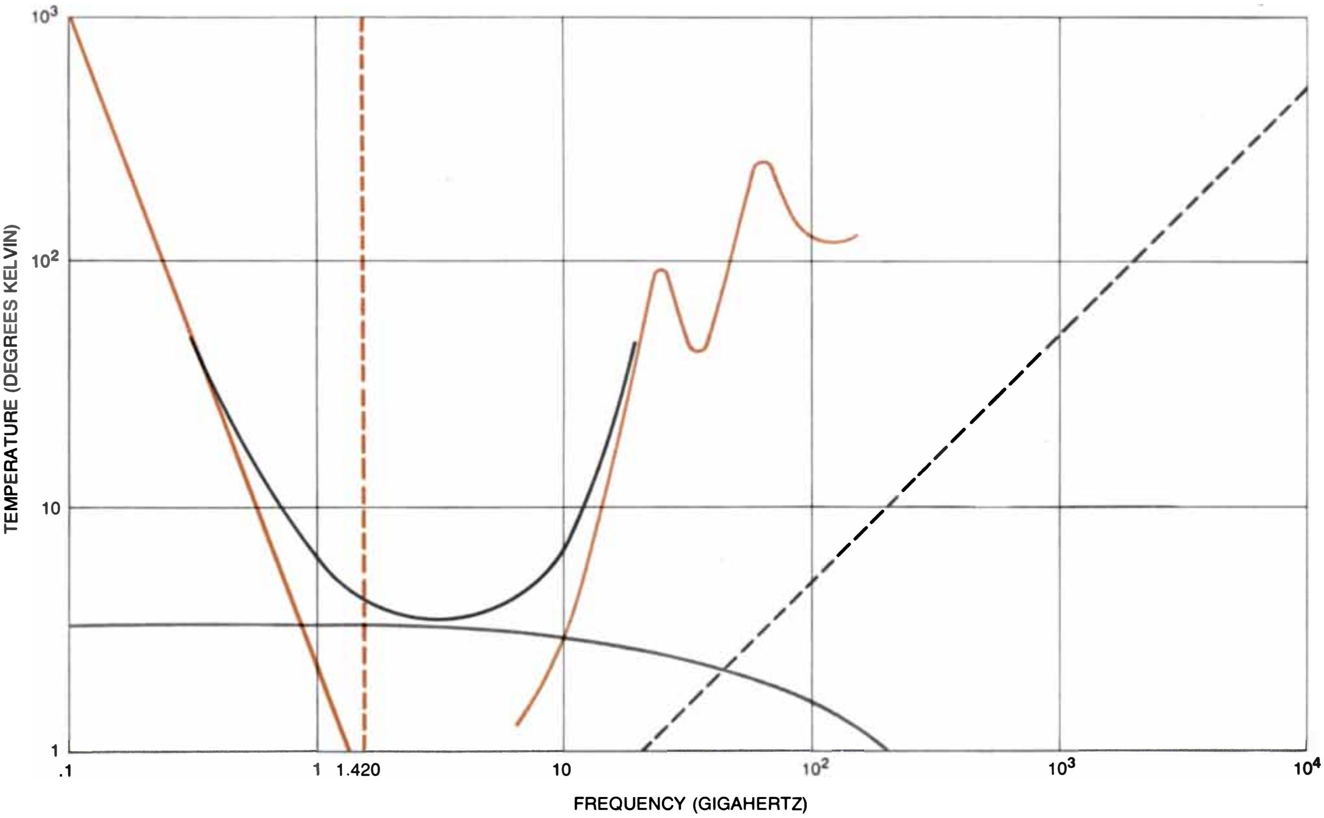


ilizations. The individual antennas would be connected to one another and to a large computer system. The effective signal-collecting area of the system would be hundreds of times greater than that of any existing radio telescope and would be capable of detecting even such relatively weak signals as the internal radio-frequency communications of a civilization as far away as several hundred light-years. Control building shown in center of array includes observatory with telescope operating at visible wavelengths.

between 2.7 and 3.5 billion years old reveal the presence of small inclusions a hundredth of a millimeter in diameter. These inclusions have been identified by Elso S. Barghoorn of Harvard University and J. William Schopf of the University of California at Los Angeles as bacteria and blue-green algae. Bacteria and blue-green algae are evolved organisms and must themselves be the beneficiaries of a long evolutionary history. There are no rocks on the earth or on the moon, however, that are more than four billion years old; before that time the surface of both bodies is believed to have melted in the final stages of their accretion. Thus the time available for the origin of life seems to have been short: a few hundred million years at the most. Since life originated on the earth in a span much shorter than the present age of the earth, we have additional evidence that the origin of life has a high probability, at least on planets with an abundant supply of hydrogen-rich gases, liquid water and sources of energy. Since those conditions are common throughout the universe, life may also be common.

Until we have discovered at least one example of extraterrestrial life, however, that conclusion cannot be considered secure. Such an investigation is one of the objectives of the Viking mission, which is scheduled to land a vehicle on the surface of Mars in the summer of 1976, a vehicle that will conduct the first rigorous search for life on another planet. The *Viking* lander carries three separate experiments on the metabolism of hypothetical Martian microorganisms, one experiment on the organic chemistry of the Martian surface material and a camera system that might just conceivably detect macroscopic organisms if they exist.

Intelligence and technology have developed on the earth about halfway through the stable period in the lifetime of the sun. There are obvious selective advantages to intelligence and technology, at least up to the present evolutionary stage when technology also brings the threats of ecological catastrophes, the exhaustion of natural resources and nuclear war. Barring such disasters, the physical environment of the earth will remain stable for many more billions of years. It is possible that the number of individual steps required for the evolution of intelligence and technology is so large and improbable that not all inhabited planets evolve technical civilizations. It is also possible—some would say likely—that civilizations tend to destroy themselves at about our level of technological development. On the other hand, if there are 100 billion suitable planets in our galaxy, if the origin of life is highly probable, if there are billions of years of evolution available on each such planet and if even a small fraction of technical civilizations pass safely through the early stages of technological adolescence, the number of technological civi-



RADIO SPECTRUM of the sky as it is seen from the earth is quite noisy. Any radio telescope picks up the three-degree-Kelvin background radiation (*gray line*), the remnant of the primordial fireball of the "big bang." The background radiation begins to fall off at about 60 gigahertz (billion cycles per second). At that frequency the quantum noise associated with all electromagnetic radiation (*broken black line*) begins to predominate, and the total noise level rises. Noise from within our galaxy (*dark colored line*) is due mainly to synchrotron radiation emitted by particles spiraling in around the lines of force in magnetic fields. Together these three sources of noise define a broad quiet region in the radio spectrum, between about one gigahertz and 100 gigahertz, that would be nearly the same for observers in the neighborhood of the sun and observers in similar regions of the galaxy. The earth's atmosphere is also a source of noise (*light colored line*) because molecules of water and oxygen absorb and reradiate energy at 22 gigahertz and 60 gigahertz. All sources of noise added together yield the curve in black, representing the total sky noise detected on the earth. The broken vertical line in color is frequency of spin-flip of the electron in un-ionized hydrogen atom at frequency of 1.420 gigahertz.

lizations in the galaxy today might be very large.

It is obviously a highly uncertain exercise to attempt to estimate the number of such civilizations. The opinions of those who have considered the problem differ significantly. Our best guess is that there are a million civilizations in our galaxy at or beyond the earth's present level of technological development. If they are distributed randomly through space, the distance between us and the nearest civilization should be about 300 light-years. Hence any information conveyed between the nearest civilization and our own will take a minimum of 300 years for a one-way trip and 600 years for a question and a response.

Electromagnetic radiation is the fastest and also by far the cheapest method of establishing such contact. In terms of the foreseeable technological developments on the earth, the cost per photon and the amount of absorption of radiation by interstellar gas and dust, radio waves seem to be the most efficient and economical method of interstellar communication. Interstellar space vehicles cannot be excluded a priori, but in all cases they would be a slower, more expensive and more difficult means of communication.

Since we have achieved the capability for interstellar radio communication only in the past few decades, there is virtually no chance that any civilization we come in contact with will be as backward as we are. There also seems to be no possibility of dialogue except between very long-lived and patient civilizations. In view of these circumstances, which should be common to and deducible by all the civilizations in our galaxy, it seems to us quite possible that one-way radio messages are being beamed at the earth at this moment by radio transmitters on planets in orbit around other stars.

To intercept such signals we must guess or deduce the frequency at which the signal is being sent, the width of the frequency band, the type of modulation and the star transmitting the message. Although the correct guesses are not easy to make, they are not as hard as they might seem.

Most of the astronomical radio spectrum is quite noisy [*see illustration on opposite page*]. There are contributions from interstellar matter, from the three-degree-Kelvin background radiation left over from the early history of the universe, from noise that is fundamentally associated with the operation of any detector and from the absorption of radiation by the earth's atmosphere. This last source of noise can be avoided by placing a radio telescope in space. The other sources we must live with and so must any other civilization.

INVESTIGATOR	OBSERVATORY	DATE	FREQUENCY OR WAVELENGTH	TARGETS
DRAKE	N.R.A.O.	1960	1,420 MEGAHERTZ	EPSILON ERIDANI TAU CETI
TROITSKY	GORKY	1968	21 AND 30 CENTIMETERS	12 NEARBY SUNLIKE STARS
VERSCHUUR	N.R.A.O.	1972	1,420 MEGAHERTZ	10 NEARBY STARS
TROITSKY	EURASIAN NETWORK, GORKY	1972 TO PRESENT	16, 30 AND 50 CENTIMETERS	PULSED SIGNALS FROM ENTIRE SKY
ZUCKERMAN PALMER	N.R.A.O.	1972 TO PRESENT	1,420 MEGAHERTZ	~600 NEARBY SUNLIKE STARS
KARDASHEV	EURASIAN NETWORK, I.C.R.	1972 TO PRESENT	SEVERAL	PULSED SIGNALS FROM ENTIRE SKY
BRIDLE FELDMAN	A.R.O.	1974 TO PRESENT	22.2 GIGAHERTZ	SEVERAL NEARBY STARS
DRAKE SAGAN	ARECIBO	1975 (IN PROGRESS)	1,420, 1,653 AND 2,380 MEGAHERTZ	SEVERAL NEARBY GALAXIES

ATTEMPTS TO DETECT SIGNALS beamed toward the earth by other civilizations have so far been unsuccessful, but the number of stars that have been examined is less than .1 percent of the number that would have to be investigated if there were to be a reasonable statistical chance of discovering one extraterrestrial civilization. "N.R.A.O." is the National Radio Astronomy Observatory in Green Bank, W.Va.; "Gorky" is the 45-foot antenna at Gorky University in the U.S.S.R.; "Eurasian network" is a network of omnidirectional antennas in the U.S.S.R. that is being operated jointly by V. S. Troitsky of Gorky University and N. S. Kardashev of the Institute for Cosmic Research (I.C.R.) of the Academy of Sciences of the U.S.S.R.; "A.R.O." is the Algonquin Radio Observatory at Algonquin Park in Canada; "Arecibo" is 1,000-foot radio-radar antenna at Arecibo Observatory in Puerto Rico.

There is, however, a pronounced minimum in the radio-noise spectrum. Lying at the minimum or near it are several natural frequencies that should be discernible by all scientifically advanced societies. They are the resonant frequencies emitted by the more abundant molecules and free radicals in interstellar space. Perhaps the most obvious of these resonances is the frequency of 1,420 megahertz (millions of cycles per second). That frequency is emitted when the spinning electron in an atom of hydrogen spontaneously flips over so that its direction of spin is opposite to that of the proton comprising the nucleus of the hydrogen atom. The frequency of the spin-flip transition of hydrogen at 1,420 megahertz was first suggested as a channel for interstellar communication in 1959 by Philip Morrison and Giuseppe Cocconi. Such a channel may be too noisy for communication precisely because hydrogen, the most abundant interstellar gas, absorbs and emits radiation at that frequency. The number of other plausible and available communication channels is not large, so that determining the right one should not be too difficult.

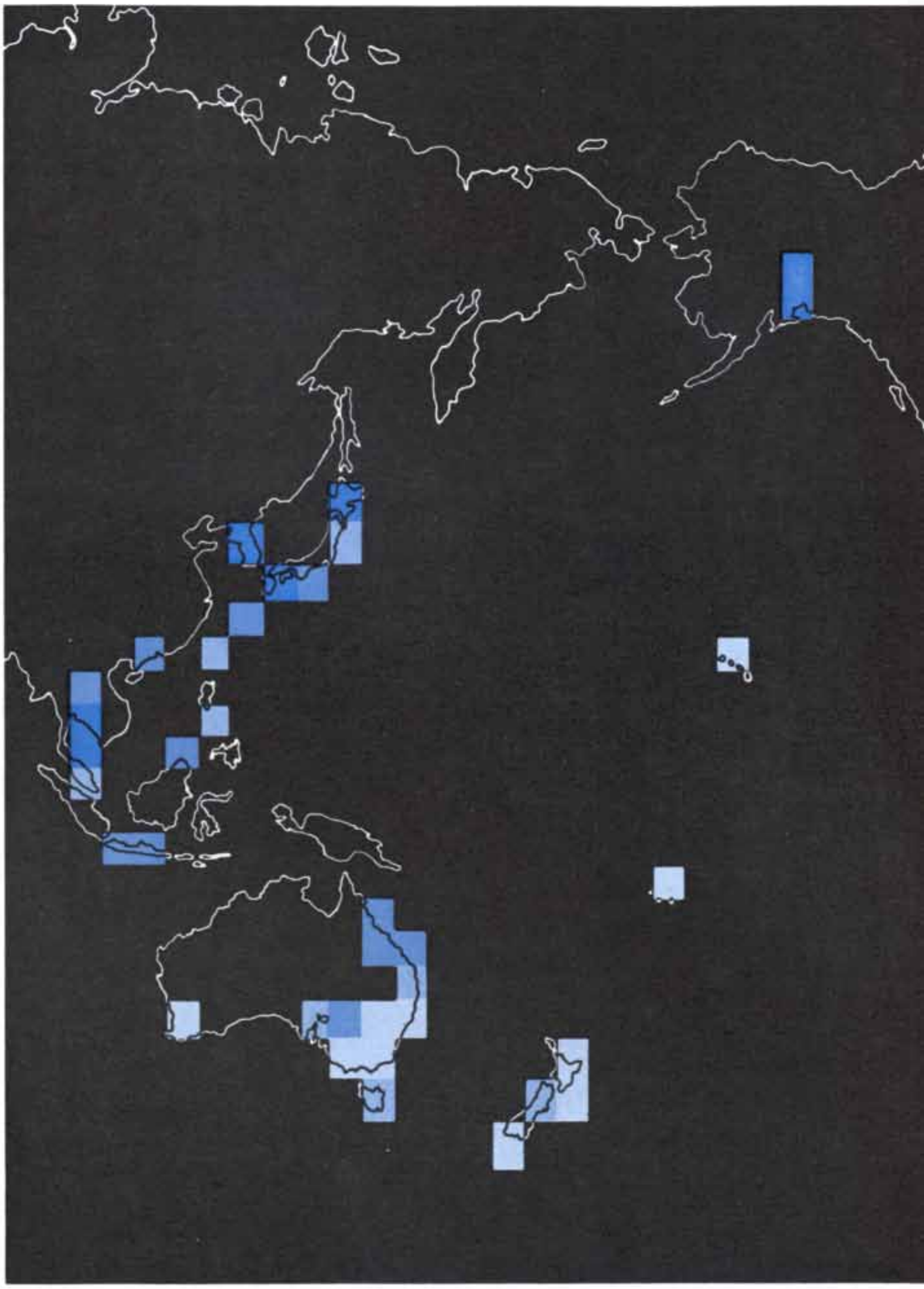
We cannot use a similar logic to guess the bandwidth that might be used in interstellar communication. The narrower the bandwidth is, the farther a signal can be transmitted before it becomes too weak for detection. On the other hand, the narrower the bandwidth is, the less information the signal can carry. A compromise is therefore required between the desire to send a signal the maximum distance and the desire to communicate the maximum amount of information. Perhaps simple signals with narrow bandwidths are sent to enhance the probability of the signals' being received. Perhaps information-rich signals with broad bandwidths are sent in order to achieve rapid and extensive communication. The broad-bandwidth signals would be intended for those enlightened civilizations that have invested major resources in large receiving systems.

When we actually search for signals, it is not necessary to guess the exact bandwidth, only to guess the minimum bandwidth. It is possible to communicate on many adjacent narrow bands at once. Each such channel can be studied individually, and the data from several adjacent channels can be combined to yield the equivalent of a wider channel without any loss of information or sensitivity. The procedure is relatively easy with the aid of a computer; it is in fact routinely employed in studies of pulsars. In any event we should observe the maximum number of channels because of the possibility that the transmitting civilization is not broadcasting on one of the "natural" frequencies such as 1,420 megahertz.

We do not, of course, know now which star we should listen to. The most conservative approach is to turn our receivers to stars that are rather similar to the sun, beginning with the nearest. Two nearby stars, Epsilon Eridani and Tau Ceti, both about 12 light-years away, were the candidates for Project Ozma, the first search with a radio telescope for extraterrestrial intelligence, conducted by one of us (Drake) in 1960. Project Ozma, named after the ruler of Oz in L. Frank Baum's children's stories, was "on the air" for four weeks at 1,420 megahertz. The results were negative. Since then there have been a number of other studies. In spite of some false alarms to the contrary, none has been successful. The lack of success is not unexpected. If there are a million technical civilizations in a galaxy of some 200 billion stars, we must turn our receivers to 200,000 stars before we have a fair statistical chance of detecting a single extraterrestrial message. So far we have listened to only a few more than 200 stars. In other words, we have mounted only .1 percent of the required effort.

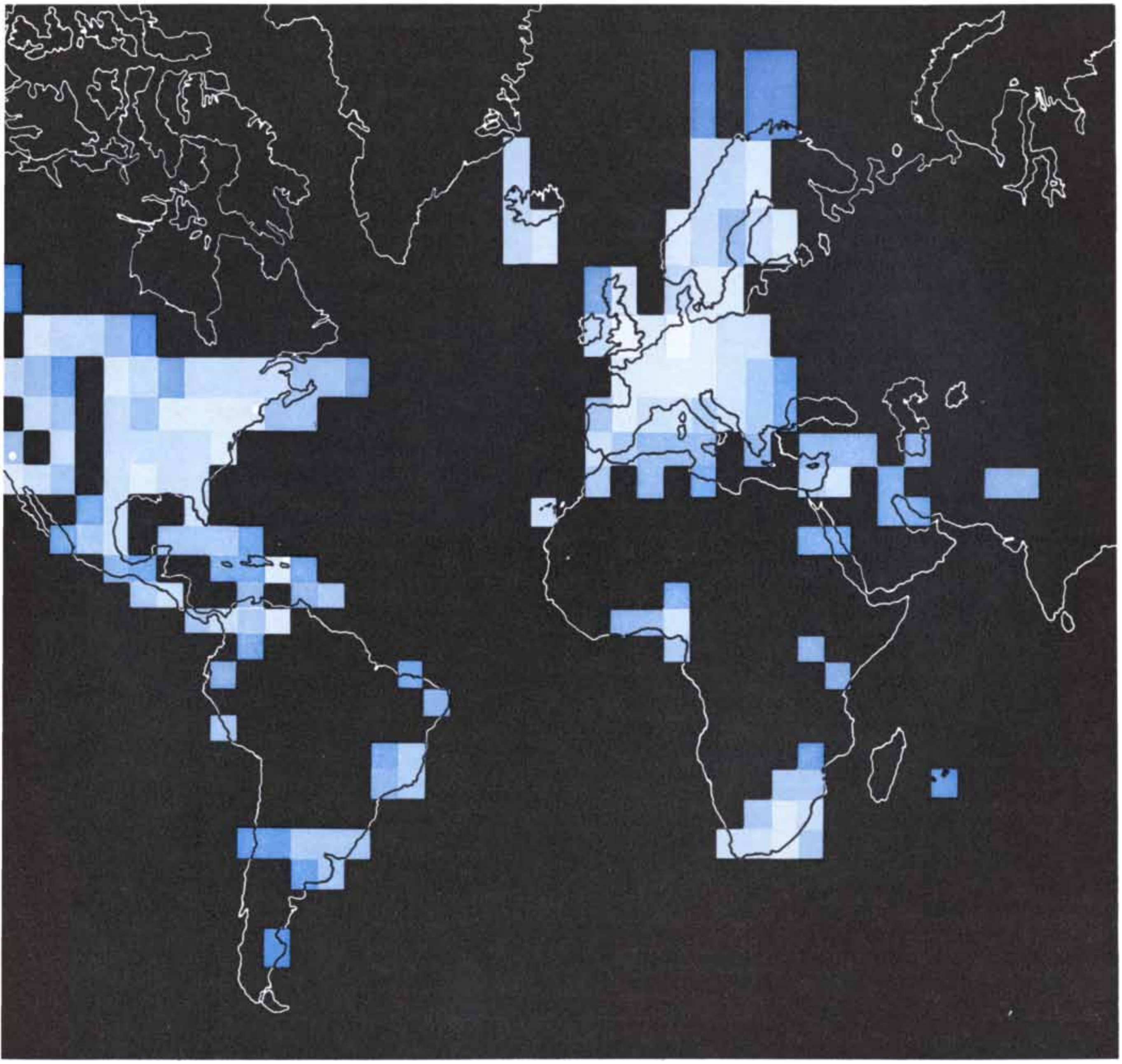
Our present technology is entirely adequate for both transmitting and receiving messages across immense interstellar distances. For example, if the 1,000-foot radio telescope at the Arecibo Observatory in Puerto Rico were to transmit information at the rate of one bit (binary digit) per second with a bandwidth of one hertz, the signal could be received by an identical radio telescope anywhere in the galaxy. By the same token, the Arecibo telescope could detect a similar signal transmitted from a distance hundreds of times greater than our estimate of 300 light-years to the nearest extraterrestrial civilization.

A search of hundreds of thousands of stars in the hope of detecting one message would require remarkable dedication and would probably take several decades. It seems unlikely that any existing major radio telescope would be given over to such an intensive program to the exclusion of its usual work. The construction of one radio telescope or more that would be devoted perhaps half-time to the search seems to be the



EARTH IS BRIGHT at the frequencies between 40 and 220 megahertz because of the radiation from FM radio and VHF television broadcasts. The power radiated by the stations is shown averaged over squares five degrees in longitude by five degrees in latitude. The radio brightness is equivalent to only practical method of seeking out extraterrestrial intelligence in a serious way. The cost would be some tens of millions of dollars.

So far we have been discussing the reception of messages that a civilization would intentionally transmit to the earth. An alternative possibility is that we might try to "eavesdrop" on the radio traffic an extraterrestrial civilization employs for its own purposes. Such radio traffic could be readily apparent. On the earth, for example, a new radar system employed with the telescope at the Arecibo Observatory for planetary studies emits a narrow-bandwidth signal that, if it were detected from another star, would be between a million and 10 billion times brighter than the sun at the same frequency. In addition, because of radio and television transmission, the earth is extremely bright at wavelengths of about a meter [*see illustration on these two pages*]. If the planets of other civilizations have a radio brightness comparable to the earth's



the temperature to which each area on the earth would have to be raised in order to produce the actual radio emission observed. The three brightest areas are the locations of three particularly powerful radar systems: the radio-radar antenna of the Haystack Observatory in Massachusetts, operating at a wavelength of 3.75 centimeters and giving a brightness temperature of 2.3×10^{20} degrees K., the 1,000-foot radio-radar antenna of the Arecibo Observatory, operating at a wavelength of 12.6 centimeters and giving a brightness temperature of 1.4×10^{21} degrees K., and the 210-foot antenna of the Jet Propulsion Laboratory at Goldstone, Calif., operating at a wavelength of 12.6 centimeters and giving a brightness temperature of 6.2×10^{19} degrees. Systems radiate so much power that at those wavelengths and in the direction of their beam they are brighter than the sun and should be detectable over interstellar distances.

from television transmission alone, they should be detectable. Because of the complexity of the signals and the fact that they are not beamed specifically at the earth, however, the receiver we would need in order to eavesdrop would have to be much more elaborate and sensitive than any radio-telescope system we now possess.

One such system has been devised in a preliminary way by Bernard M. Oliver of the Hewlett-Packard Company, who directed a study sponsored by the Ames Research Center of the National Aeronautics and Space Administration. The system, known as Cyclops, would consist of an enormous radio telescope connected to a complex computer system. The computer system would be designed particularly to search through the data from the telescope for signals bearing the mark of intelligence, to combine numerous adjacent channels in order to construct signals of various effective bandwidths and to present the results of the automatic analyses for all conceivable forms of interstellar radio communication in a way that would be intelligible to the project scientists.

To construct a radio telescope of enormous aperture as a single antenna would be prohibitively expensive. The Cyclops system would instead capitalize on our ability to connect many individual antennas to act in unison. This concept is already the basis of the Very Large Array now under construction in New Mexico. The Very Large Array consists of 27 antennas, each 82 feet in diameter, arranged in a Y-shaped pattern whose three arms are each 10 miles long. The Cyclops system would be much larger. Its current design calls for 1,500 antennas each 100 meters in diameter, all electronically connected to one another and to the computer system. The array would be as compact as possible but would cover perhaps 25 square miles.

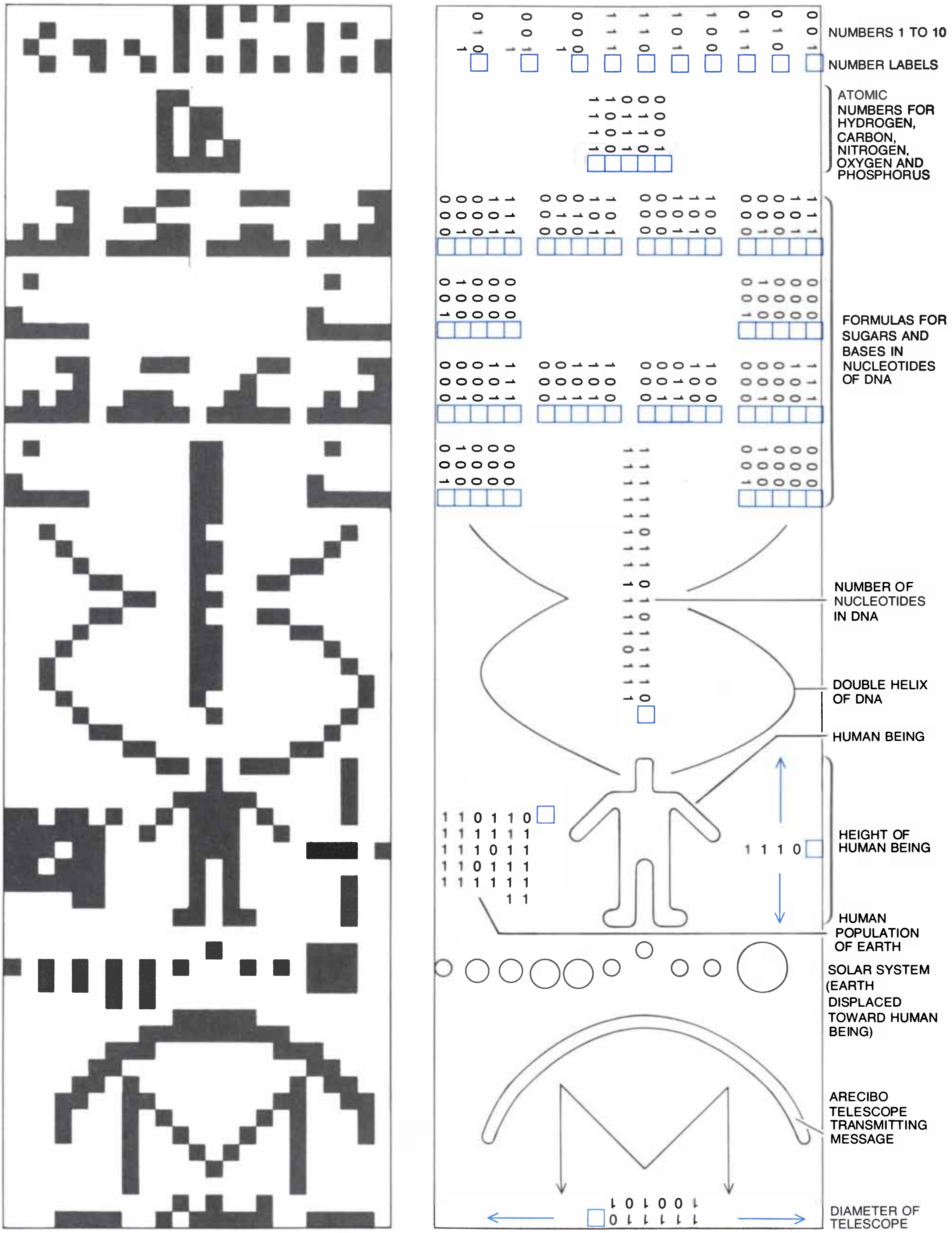
The effective signal-collecting area of the system would be hundreds of times the area of any existing radio telescope, and it would be capable of detecting even relatively weak signals such as television transmissions from civilizations several hundred light-years away. Moreover, it would be the instrument par excellence for receiving signals specifically directed at the earth. One of the greatest virtues of the Cyclops system is that no technological advances would be required in order to build it. The necessary electronic and computer techniques are already well developed. We would need only to build a vast number of items we already build well. The Cyclops system not only would have enormous power for searching for extraterrestrial intelligence but also would be an extraordinary tool for radar studies of the bodies in the solar system, for traditional radio astronomy outside the solar system and for the tracking of space vehicles to distances beyond the reach of present receivers.

The estimated cost of the Cyclops system, ranging up to $10 billion, may make it prohibitively expensive for the time being. Moreover, the argument in favor of eavesdropping is not completely persuasive. Half a century ago, before radio transmissions were commonplace, the earth was quiet at radio wavelengths. Half a century from now, because of the development of cable television and communication satellites that relay signals in a narrow beam, the earth may again be quiet. Thus perhaps for only a century out of billions of years do planets such as the earth appear remarkably bright at radio wavelengths. The odds of our discovering a civilization during that short period in its history may not be good enough to justify the construction of a system such as Cyclops. It may well be that throughout the universe beings usually detect evidence of extraterrestrial intelligence with more traditional radio telescopes. It nonetheless seems clear that our own chances of finding extraterrestrial intelligence will improve if we consciously attempt to find it.

How could we be sure that a particu-

```
0 0 0 0 0 0 1 0 1 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0 1 0 1 0 0 0 0 0 1 0 1 0
0 0 0 0 0 0 1 0 0 1 0 0 0 1 0 0 0 1 0 0 0 1 0 0 1 0 1 1 0 0 1 0 1 0 1 0 1
0 1 0 1 0 1 0 1 0 1 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
1 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 1 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 1 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0
1 1 1 0 0 0 1 1 0 0 0 0 1 1 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 1 0
0 0 0 1 1 0 1 0 0 0 1 1 0 0 0 1 1 0 0 0 0 1 1 0 1 0 1 1 1 1 1 0 1 1 1 1 1
0 1 1 1 1 1 0 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 0 0
0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 1 1 0 0 0 0 1 1 0 0 0 0 1 1 1 0 0 0 1 1 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0
0 0 0 0 0 0 1 0 0 0 0 1 1 0 1 0 0 0 0 1 1 0 0 0 1 1 1 0 0 1 1 0 1 0 1 1 1
1 1 0 1 1 1 1 1 0 1 1 1 1 1 0 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0
0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0
1 1 1 1 1 1 0 0 0 0 0 1 1 0 0 0 0 0 0 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 1 1 0
0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 1 0 0 0 0 0 1
0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1 1 0 0 0 0 1 1 0 0 0 0 0 0
1 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 1 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 1 1 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 1 0 0 0 0 1 1 0 0 0 0 0
0 0 0 0 1 1 0 0 0 0 1 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0
0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 1 0 0 0 1 0 0 0
0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 1 0 0 0 1 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0
1 0 0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0
0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0
0 1 0 0 0 1 1 1 0 1 0 1 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0
0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0
0 1 0 1 1 1 0 1 0 0 1 0 1 1 0 1 1 0 0 0 0 0 0 1 0 0 1 1 1 0 0 1 0 0 1 1 1
1 1 1 1 0 1 1 1 0 0 0 0 1 1 1 0 0 0 0 0 1 1 0 1 1 1 0 0 0 0 0 0 0 0 0 1 0
1 0 0 0 0 0 1 1 1 0 1 1 0 0 1 0 0 0 0 0 0 1 0 1 0 0 0 0 0 1 1 1 1 1 1 0 0
1 0 0 0 0 0 0 1 0 1 0 0 0 0 0 1 1 0 0 0 0 0 0 1 0 0 0 0 0 1 1 0 1 1 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0
0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 1 0 1 0 0 0 1 0 1 0 1 0 1 0 1
0 1 0 0 1 1 1 0 0 0 0 0 0 0 0 0 1 0 1 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 0 0 0 1 1 1
0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 1 1 0 1 0 0
0 0 0 0 0 0 0 1 0 1 1 0 0 0 0 0 1 1 0 0 1 1 0 0 0 0 0 0 0 1 1 0 0 1 1 0 0
0 0 1 0 0 0 1 0 1 0 0 0 0 0 1 0 1 0 0 0 1 0 0 0 0 1 0 0 0 1 0 0 1 0 0 0 1
0 0 1 0 0 0 1 0 0 0 0 0 0 0 0 1 0 0 0 1 0 1 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0
0 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 1 0 0
0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 0 0 1 1
1 1 1 0 1 0 0 1 1 1 1 0 0 0
```

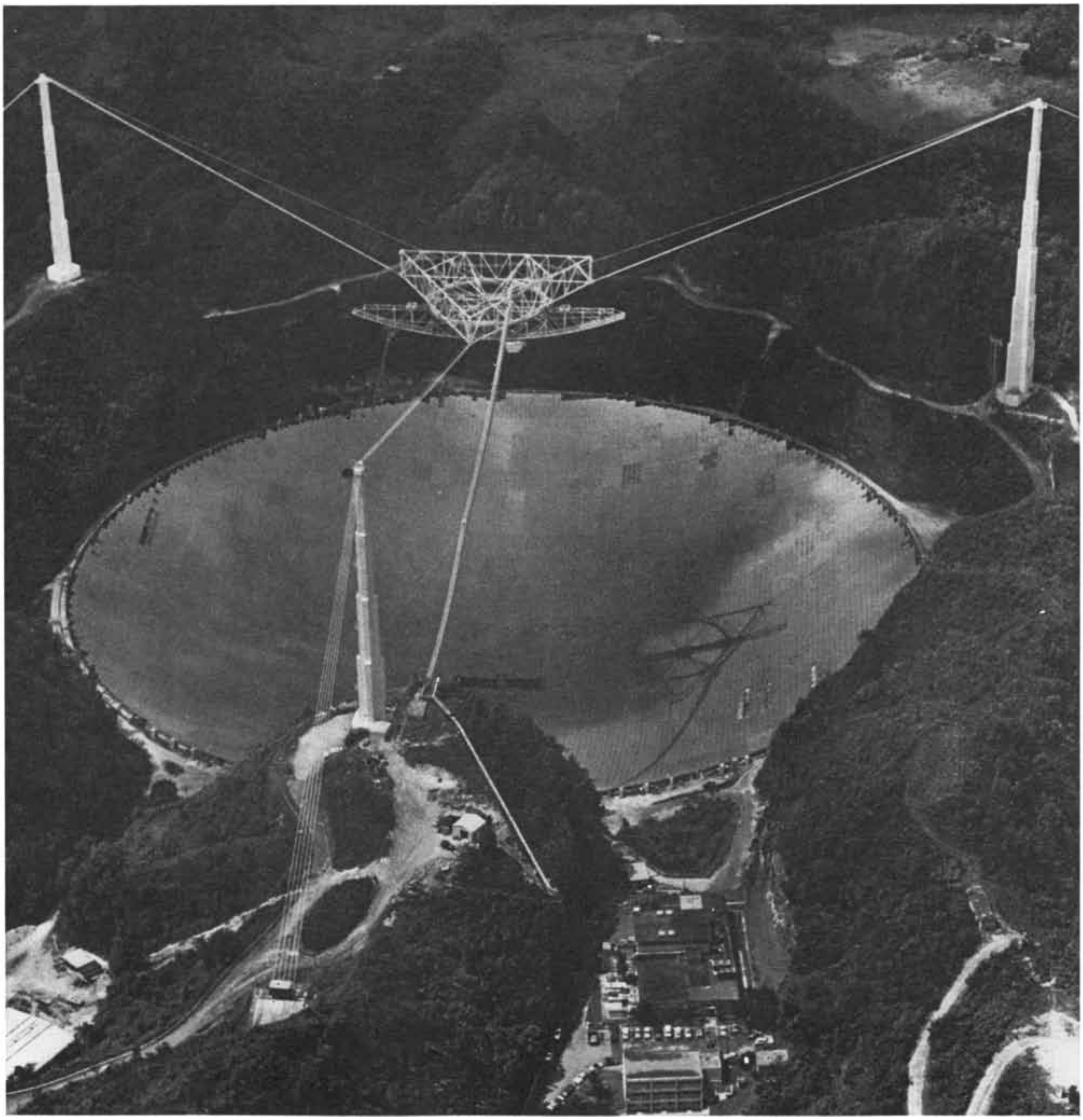
ARECIBO MESSAGE IN BINARY CODE was transmitted in 1974 toward the Great Cluster in Hercules from the 1,000-foot antenna at Arecibo. The message is decoded by breaking up the characters into 73 consecutive groups of 23 characters each and arranging the groups in sequence one under the other, reading right to left and then top to bottom. The result is a visual message (*see illustration on opposite page*) that can be more easily interpreted by making each 0 of binary code represent a white square and each 1 a black square.



ARECIBO MESSAGE IN PICTURES and accompanying translation shows the binary version of the message decoded. Each number that is used is marked with a label that indicates its start. When all the digits of a number cannot be fitted into one line, the digits for which there is no room are written under the least significant digit. (The message must be oriented in three different ways for all the numbers shown to be read.) The chemical formulas are those for the components of the DNA molecule: the phosphate group, the deoxyribose sugar and the organic bases thymine, adenine, guanine and cytosine. Both the height of the human being and the diameter of the telescope are given in units of the wavelength that is used to transmit the message: 12.6 centimeters.

lar radio signal was deliberately sent by an intelligent being? It is easy to design a message that is unambiguously artificial. The first 30 prime numbers, for example, would be difficult to ascribe to some natural astrophysical phenomenon. A simple message of this kind might be a beacon or announcement signal. A subsequent informative message could have many forms and could consist of an enormous number of bits. One method of transmitting information, beginning simply and progressing to more elaborate concepts, is pictures [*see illustration on preceding page*].

One final approach in the search for extraterrestrial intelligence deserves mention. If there are indeed civilizations thousands or millions of years more advanced than ours, it is entirely possible that they could beam radio communications over immense distances, perhaps even over the distances of intergalactic space. We do not know how many advanced civilizations there might be compared with the number of more primitive earthlike civilizations, but many of these older civilizations are bound to be in galaxies older than our own. For this reason the most readily detectable radio signals from another civilization may come from outside our galaxy. The relatively small number of such extragalactic transmitters might be more than compensated for by the greater strength



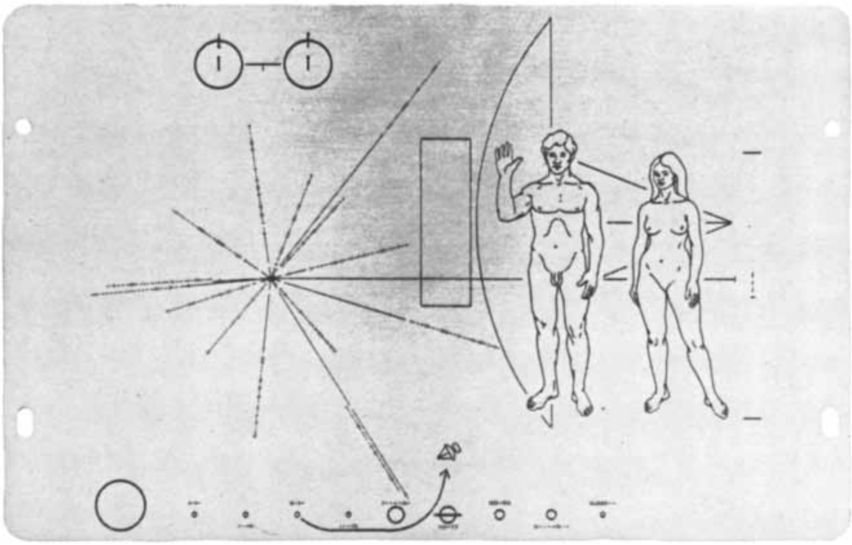
THOUSAND-FOOT ANTENNA of the radio-radar system at the Arecibo Observatory is made of perforated aluminum panels whose spherical shape is accurate to within 1/8 inch over the antenna's entire area of 20 acres. The triangular structure suspended above the antenna holds the receiver and the transmitter for the system. Control rooms and office buildings are to lower right of antenna.

of their signals. At the appropriate frequency they could even be the brightest radio signals in the sky. Therefore an alternative to examining the nearest stars of the same spectral type as the sun is to examine the nearest galaxies. Spiral galaxies such as the Great Nebula in Andromeda are obvious candidates, but the elliptical galaxies are much older and more highly evolved and could conceivably harbor a large number of extremely advanced civilizations.

There might be a kind of biological law decreeing that there are many paths to intelligence and high technology, and that every inhabited planet, if it is given enough time and it does not destroy itself, will arrive at a similar result. The biology on other planets is of course expected to be different from our own because of the statistical nature of the evolutionary process and the adaptability of life. The science and engineering, however, may be quite similar to ours, because any civilization engaged in interstellar radio communication, no matter where it exists, must contend with the same laws of physics, astronomy and radio technology that we do.

Should we be sending messages ourselves? It is obvious that we do not yet know where we might best direct them. One message has already been transmitted to the Great Cluster in Hercules by the Arecibo radio telescope, but only as a kind of symbol of the capabilities of our existing radio technology. Any radio signal we send would be detectable over interstellar distances if it is more than about 1 percent as bright as the sun at the same frequency. Actually something close to 1,000 such signals from our everyday internal communications have left the earth every second for the past two decades. This electromagnetic frontier of mankind is now some 20 light-years away, and it is moving outward at the speed of light. Its spherical wave front, expanding like a ripple from a disturbance in a pool of water and inadvertently carrying the news that human beings have achieved the capacity for interstellar discourse, envelops about 20 new stars each year.

We have also sent another kind of message: two engraved plaques that ride aboard *Pioneer 10* and *Pioneer 11*. These spacecraft, the first artifacts of mankind that will escape from the solar system, will voyage forever through our galaxy at a speed of some 10 miles per second. *Pioneer 10* was accelerated to the velocity of escape from the solar system by the gravitational field of Jupiter on December 3, 1973. *Pioneer 11* swung past Jupiter on December 4, 1974, and will travel on to Saturn before it is accelerated on a course to the far side of the galaxy.



ENGRAVED PLAQUE on the *Pioneer* spacecraft to Jupiter is another message that has been dispatched beyond the solar system. Meaning of symbols is given in text of article.

Identical plaques for each vehicle were designed by us and Linda Salzman Sagan. Each plaque measures six by nine inches and is made of gold-anodized aluminum. These engraved cosmic greeting cards bear the location of the earth and the time the spacecraft was built and launched. The sun is located with respect to 14 pulsars. The precise periods of the pulsars are specified in binary code to allow them to be identified. Since pulsars are cosmic clocks that are running down at a largely constant rate, the difference in the pulsar periods at the time one of the spacecraft is recovered and the periods indicated on the plaque will enable any technically sophisticated civilization to deduce the year the vehicle was sent on its epic journey. Units of time and distance are specified in terms of the frequency of the hydrogen spin-flip at 1,420 megahertz. In order to identify the exact location of the spacecraft's launch a diagram of the solar system is given. The trajectory of the spacecraft is shown as it leaves the third planet, the earth, and swings by the fifth planet, Jupiter. (The diversion of *Pioneer 11* past Saturn had not been planned when the plaques were prepared.) Last, the plaques show images of a man and a woman of the earth in 1973. An attempt was made to give the images panracial characteristics. Their heights are shown with respect to the spacecraft and are also given by a binary number stated in terms of the wavelength of the spectral line at 1,420 megahertz (21 centimeters).

These plaques are destined to be the longest-lived works of mankind. They will survive virtually unchanged for hundreds of millions, perhaps billions, of years in space. When plate tectonics has completely rearranged the continents, when all the present landforms on the earth have been ground down, when civilization has been profoundly transformed and when human beings may have evolved into some other kind of organism, these plaques will still exist. They will show that in the year we called 1973 there were organisms, portrayed on the plaques, that cared enough about their place in the hierarchy of all intelligent beings to share knowledge about themselves with others.

How much do we care? Enough to devote an appreciable effort with existing telescopes to search for life elsewhere in the universe? Enough to take a major step such as Project Cyclops that offers a greater chance of carrying us across the threshold, to finally communicate with a variety of extraterrestrial beings who, if they exist, would inevitably enrich mankind beyond imagination? The real question is not how, because we know how; the question is when. If enough of the beings of the earth cared, the threshold might be crossed within the lifetime of most of those alive today.

FOREST SUCCESSION

The study of a woods in New Jersey has yielded a predictive model for the succession of trees in a mixed forest. Two basic factors are the moisture in the soil and the geometry of leaf arrangement

by Henry S. Horn

When a forest has been devastated by man or by a natural catastrophe such as a hurricane, in time a semblance of the original forest returns. The regeneration of the forest goes through several stages called succession: trees that are dominant in one stage of regeneration succumb to trees of other species in the next. Forest succession raises a number of basic questions. Is there an orderly and predictable pattern to the changes? Does succession reach a stable end point if given enough time? Do different patterns of succession lead to the same final stage? Conversely, can one initial stage of succession lead to different final stages? Are some stages of succession more fragile than others? Answers to these basic questions are needed before we can hope to answer more practical questions. Can the natural regeneration of a forest be speeded up by enlightened human intervention? How can succession be controlled in order to keep a forest in its most productive state?

One of the more appealing aspects of studying forest ecology is that basic questions can be explored with a modest investment in technical equipment. Most of my own work is done with a pencil, a notebook and a measuring tape. The one specialized piece of equipment that is essential is an ingenious device for extracting a "core," or cylinder, of wood from a tree. The core provides a cross section of the tree's annual growth rings and hence a record of its age. Although I have access to a full-sized electronic computer, most of my calculations can easily be done with an inexpensive pocket calculator. My laboratory fits into a carpenter's apron.

Forest succession proceeds too slowly for it to be observed directly. Ideally one would like to find several plots of land with similar physical characteristics but different and well-documented histories. Unfortunately reliable historical records of land use are rare. Some investigators attempt to infer historical patterns of land use from the qualitative pattern of vegetation, arriving at successional patterns with a marvelous but spurious internal consistency. In order to avoid such circular deductions I make a comparative census of the ages of all the trees on various plots of land, all with unknown histories. I then infer successional patterns, and within these hypothetical patterns I measure changes in competition between different species of trees and differences in their growth rate. The empirical patterns I arrive at have led to some theoretical abstractions, which I have provisionally tested with further observations.

Before attacking the general questions about the succession of trees in a forest, I shall document the details of the particular succession I have examined. A 500-acre woods in Princeton, N.J., owned by the Institute for Advanced Study has provided a convenient natural laboratory for my work. A small portion of the woods apparently was never farmed, and the remainder is divided into plots that were last farmed at various times from 35 to 150 years ago. The only widespread disturbances were a blight that led to the loss of all the chestnut trees by 1920 and the uprooting of several trees in the hurricanes of 1938 and 1944. Throughout the woods' recent history each of the plots has been within seeding distance of trees that are characteristic of different successional stages.

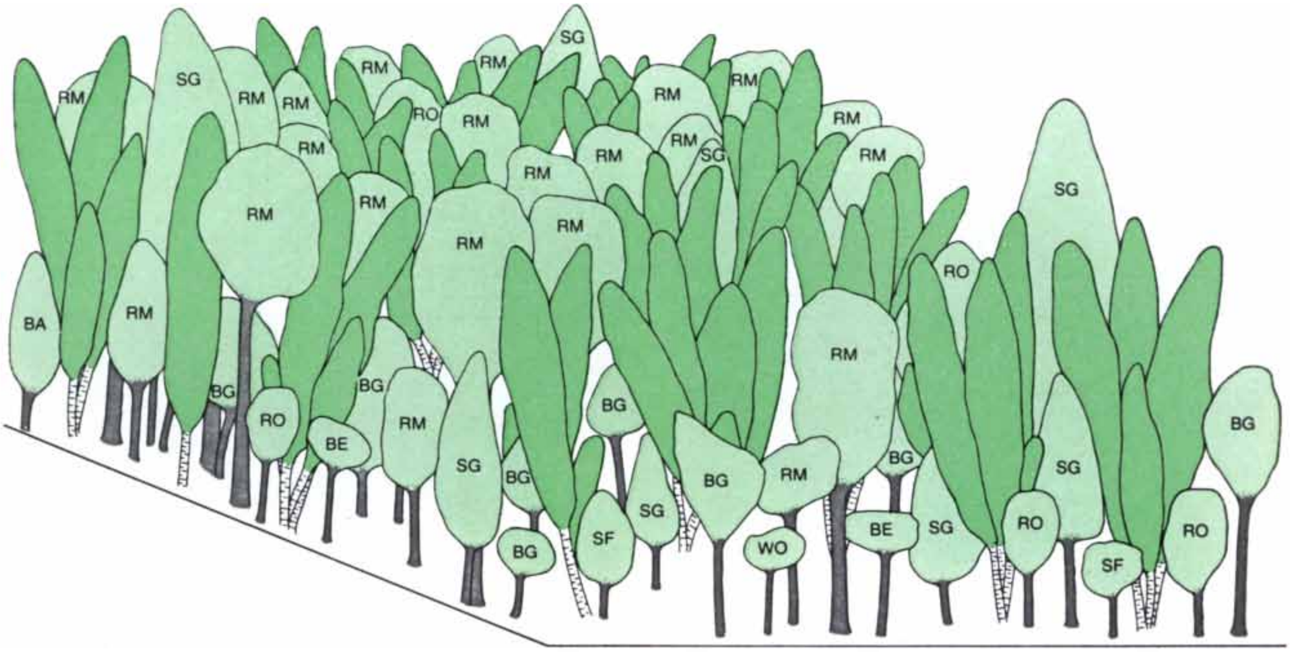
As a first approximation the stands are ordered by the age of the oldest trees of the dominant species: gray birch (33 years), bigtooth aspen (35 years), sweet gum (37 years), black gum (50 years), red maple (about 65 years), oak, hickory, tulip tree and dogwood (each about 120 years) and beech (about 250 years). For each stand I sought the answers to three questions: Is the stand under invasion by species that are characteristic of supposedly later stands? Is there a senile population among species that supposedly belong to earlier stands? Are the dominant species begetting their kind locally?

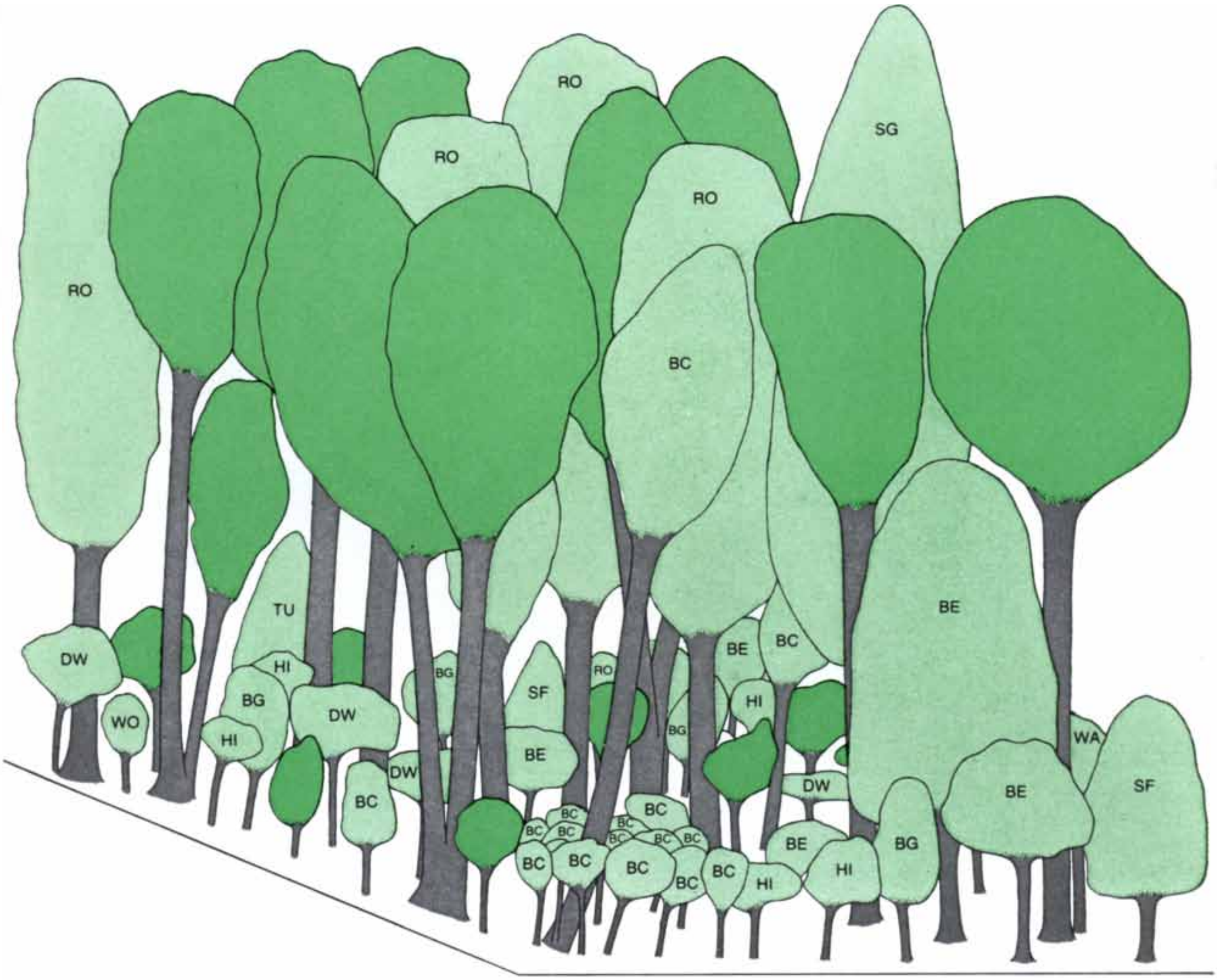
One cannot take the first approximation seriously or use the absolute ages of the trees to answer these questions because trees that have rotted away may be well represented by their offspring. On the other hand, the relative age profiles of the various species in each stand will clearly show how well each species will be represented in the near future. An invading species will be represented

FOREST STANDS in the 500-acre woods of the Institute for Advanced Study in Princeton, N.J., are identified by the dominant species of trees in each stand. The dominant species in the six stands on the opposite page, photographed with a "fish-eye" lens, are bigtooth aspen (*top left*), black gum (*middle left*), oak-hickory-tulip (*bottom left*), gray birch (*top right*), sweet gum (*middle right*) and beech (*bottom right*). The three stands at the left are on well-drained soils; the three at the right are on wetter soils. Note that the stands on the wetter soils have a greater number of canopy trees than those on the drier soils. Grapevines festoon the middle story of the bigtooth aspen stand. Gray birch stand has been invaded by the thorny vine greenbriar. Sweet gum has crowded out gray birch, which lies on forest floor. Oak, hickory and tulip are about equally abundant in their stand, which has understory of dogwood.







SG
RM
RM
SG
RM
RM
RM
RO
RM
RM
RM
RM
RM
RM
RM
RM
SG
RM
RM
RM
RM
RM
SG
RO
BA
RM
RM
BG
BG
RO
BE
RM
SG
BG
BG
SF
SG
BG
RM
WO
BE
SG
RO
SF
RO
BG
SG
BG


RO
RO
SG
RO
RO
BC
TU
DW
WO
HI
BG
HI
DW
BG
SF
RO
BE
BE
BC
HI
BG
DW
BC
DW
BE
WA
BC
BC
BC
BC
BC
BC
BC
BC
BC
BC
BC
BC
BC
BE
BG
BE
SF
BC
BC
HI
HI

by many seedlings, a respectable number of saplings and perhaps a few small trees. A locally reproducing species will have trees in several age classes, with a preponderance of young trees. If a stand consists of nothing but older trees, it is senile, and its fate depends on whether other species will invade it before the dominant population reproduces sufficiently for the trees that have died to be replaced.

Analysis of the age profiles of the trees in each stand shows that the gray birch and the bigtooth aspen are senile wherever they occur and that their stands are under heavy invasion by all the other local species. The black gum and the sweet gum are reproducing locally, but their understories are full of red maple, beech, oak and hickory. The red maple is also reproducing locally, but its stand is being invaded by beeches and a few oaks. The oak-hickory-tulip stand is senile, with a few beeches here and there. The beech stand is senile, but its understory is filled with root sprouts, so that the next generation undoubtedly will have an abundance of beech.

The soils in which the stands grow are also a factor in the pattern of succession. The oak-hickory-tulip stand and the bigtooth aspen stand are on coarse, well-drained soil, which is aerated to as great a depth as I can reach with a two-meter soil auger. (The aeration is indicated by the rusty color of the iron oxides in the soil.) In the red maple stand and the beech stand the average water table is within a meter of the surface of the fine, sticky soil. The stands of other species of trees are on intermediate soils.

I have developed a hypothetical successional pattern that is consistent with my observations of the actual stands [*see illustration on next page*]. Bigtooth aspen invades open spaces readily but gives way to oak, hickory and tulip on dry soil and to black gum on wetter soil. Gray birch is also an invader of open space, but it gives way to black gum or to sweet gum, depending partly on the history of the plot and partly on some subtleties of drainage. On dry soil black gum succumbs to oak, hickory and tulip, but on poorly drained soil black gum and sweet gum are displaced by red maple. Beech will replace red maple, although a series of dry years enables oak and hickory to invade red maple stands. Conversely, in wet seasons beech slowly invades the oak-hickory-tulip stand.

This hypothetical successional pattern is, of course, an abstraction. It represents what could happen, not what actually has happened or what necessarily will happen. In reality the oak-hickory-tulip stand has the characteristics of a former woodland pasture: compacted soil and a paucity of small trees. The red maple stand consists of sprouts from the stumps of an earlier stand that had been cut. The sweet gum stand has trees of almost uniform age and is just downwind from a huge and prolific sweet gum tree. The black gum stand is in the lee of the tall beech stand and is severely buffeted by turbulent eddies during windstorms. The black gum trees that dominate their stand have had their tops broken off several times and have regrown to their present height.

The only stands that are true to the hypothetical successional scheme are the gray birch, bigtooth aspen and beech stands. The beech trees may even belong to the original forest. Their trunks are free of branches to a great height, and the ground under them still has the mounds and pits created by the uprooting of trees in windstorms of the past, indicating that the land has never been plowed.

Several intriguing patterns arise from the hypothetical successional scheme. It shows that as long as there is enough soil moisture the succession converges on a beech forest. On dry soil, however, the succession either ends or stalls at an earlier stage, with the development of an oak-hickory-tulip forest. The successional sequence of gray birch to black gum to red maple to beech could actually take place on poorly drained soil. The sequence of bigtooth aspen to oak-hickory-tulip could occur on well-drained soil. I am justified therefore in assuming that real successions would show patterns like those I observe among these contemporary stands.

How do successional changes affect the productivity of a forest? Indexes of the annual production of new wood and the standing volume of wood can be calculated from the height, diameter and annual increase in radius of trees that are representative of each age class of each species in a stand. I plotted these "parabolic volume" indexes of the standing crop of wood and the annual production of new wood against the successional age of each stand [*see illustration on page 95*]. As an open field develops into a forest the amount of standing wood increases steadily, but the production of new wood first increases and then decreases as the forest's composition changes. The decline in the production of new wood is more obvious when production is expressed as the "interest rate," that is, the annual production of new wood divided by the volume of wood in the standing crop.

The geometrical arrangement of the leaves of a tree plays a role both in the tree's productivity of wood and in the successional pattern of a stand. The optimal leaf arrangement for intercepting and utilizing light most efficiently depends on the amount of light that is available to the tree. Leaves are able to conduct photosynthesis at about 90 percent of their maximal rate with as little as 25 percent of full sunlight. At lower levels of incident light, however, an optimal tree should intercept all the available light at its highest intensity. That is achieved by having a single layer of leaves in a shell around the tree. In a crowded stand a single-layer shell can provide no more than one unit of leaf area for each unit of ground area under the tree.

At high intensities of sunlight the tree should have leaves distributed throughout its total branch volume, subject only to the constraint that the interior leaves

BA	BIGTOOTH ASPEN	DW	FLOWERING DOGWOOD	SF	SASSAFRAS
BC	BLACK CHERRY	GB	GRAY BIRCH	SG	SWEET GUM
BE	BEECH	HI	HICKORY	TU	TULIP
BG	BLACK GUM	RM	RED MAPLE	WA	WHITE ASH
		RO	RED OAK	WO	WHITE OAK

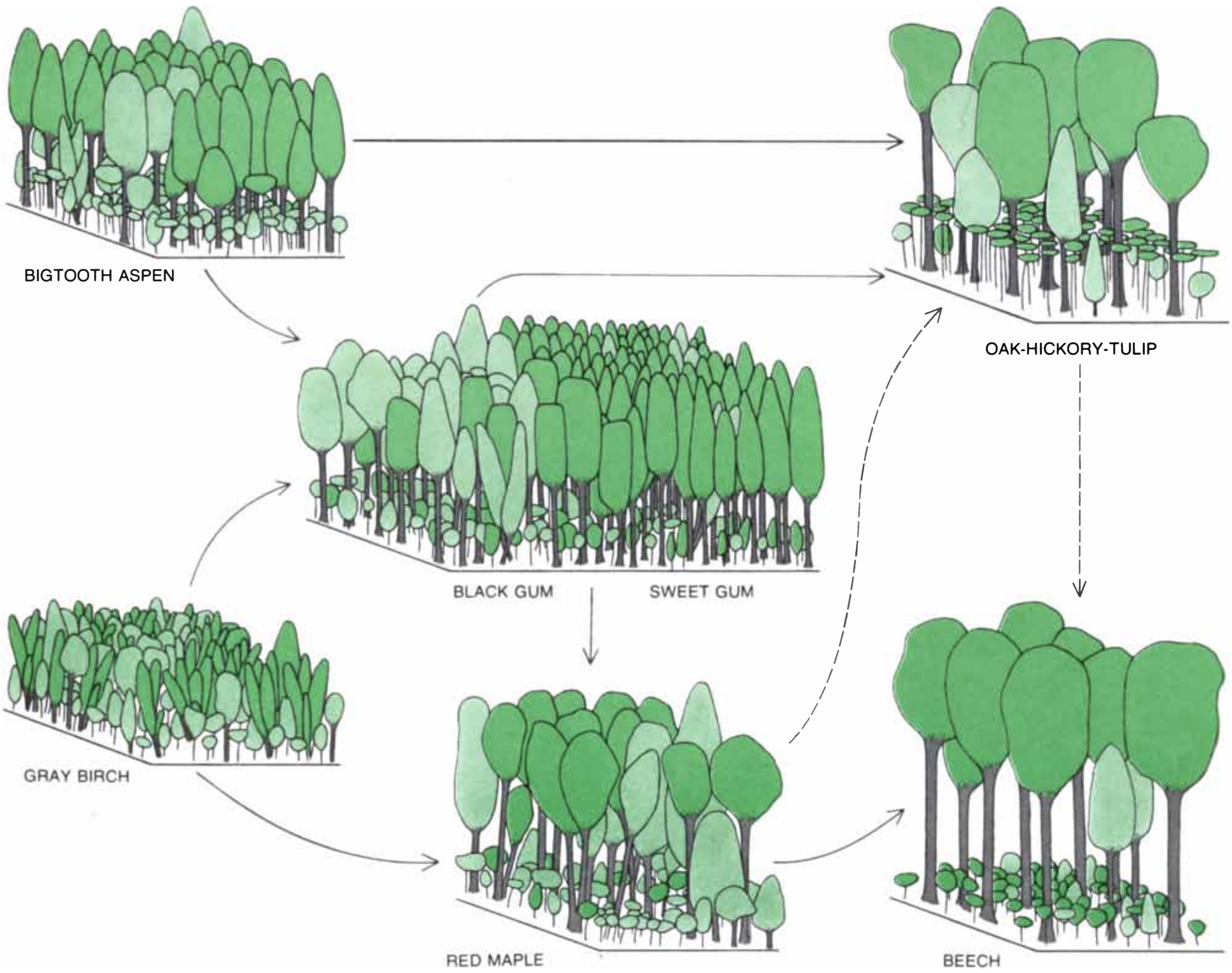
DIAGRAMMATIC STRUCTURES of a gray birch forest (*top*) and a red maple forest (*bottom*) are shown on the opposite page. The dominant species in each forest is dark-colored, and the relative heights and abundances of all the species on a 20-by-20-meter plot of ground are given. Although the gray birch dominates the canopy in its stand, it is being challenged by red maple. There are no gray birches in the understory of the red maple stand, but there are scattered young red maples. Beech appears to be invading the red maple stand. If the two stands are indicative of successional patterns, then gray birch gives way to red maple. Since the gray birch stand was originally colonized by both red maple and gray birch at about the same time some 35 years ago, the successional shift from gray birch to red maple is brought about by the late competitive dominance of red maple rather than by invasion.

must receive enough light to balance their own metabolic costs. Leaves need at least 2 percent of full sunlight to pay for themselves metabolically. If the distribution of leaves is optimal, the total leaf area can exceed the ground area under the tree severalfold even in a dense stand. In the open, therefore, the monolayered tree with its lesser leaf area is competitively inferior to the multilayered tree. Conversely, deep shade may prevent the growth of a multilayered tree because its heavily shaded interior leaves do not pay for themselves and must draw their sustenance from the peripheral leaves.

A monolayered tree produces dense foliage at the tips of well-lighted branches. The developmental program for a multilayered tree consists in producing lacy foliage at the tips of well-lighted branches and at the tips of any interior branches that are sufficiently lighted or that manufactured a local surplus of sugar in the previous season. An adjustment between dense and lacy foliage can be made by control of the shape and lobing of the leaves, the length of the leaf stems and the spacing of leaves along the twig. For leaves that are spirally arranged along the stem or are set in pairs at right angles, several of the parameters must be adjusted concurrently, but for leaves that lie in the same plane along the stem, either opposite another leaf or alternating with other leaves, only the distance between leaves need be altered.

Why cannot a tree have an adaptive leafing strategy that is flexible, depending on the existing light conditions? An optimal strategy would require either the ability to predict future lighting conditions or a system of communication between the inner leaves and the outer ones. I have no doubt that I shall eventually locate species of trees that make such an adjustment, but the few examples I have found so far are not convincing enough to warrant discussion. Hence I work currently with the assumption that a species of tree is characteristically either monolayered or multilayered in its arrangement of leaves [*see illustration on page 96*].

An appropriate and convenient expression of the distribution of leaves in a tree is the effective number of leaf layers within the tree with respect to the leaf area at the periphery of the tree. To obtain it I measure with a light meter the proportion of light that penetrates



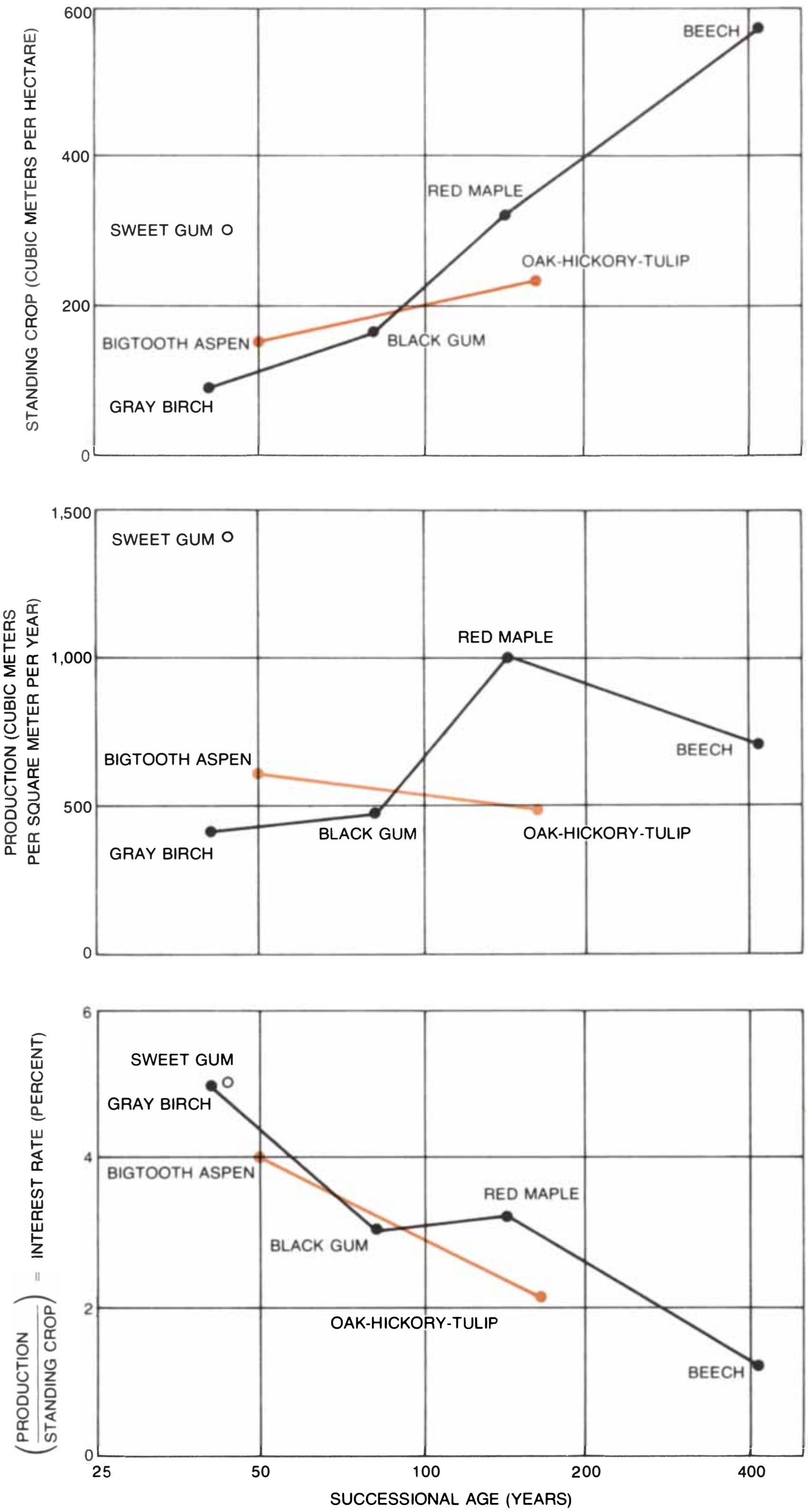
HYPOTHETICAL PATTERN OF SUCCESSION has been derived by the author from measurements of primary factors that determine succession in each stand. Dominant species in each stand are dark-colored; other species are light-colored. The capacity of the soil to hold water increases from the bigtooth aspen stand at upper left to the beech stand at lower right. The solid-line arrows represent successional changes that are consistent with the relative age distributions of the species involved. Any of the intermediate stages, except red maple, can be bypassed. There is some evidence that other changes, shown by broken-line arrows, can also occur.

a single densely foliated branch at the outer edge of the tree and then the proportion of light that penetrates through the densest part of the entire tree. From these measurements I calculate how many outside branches it would take to reduce the light to the amount that comes through the tree as a whole. That is done by determining how many times the proportion of light coming through the branch must be multiplied by itself in order to obtain the light intensity coming through the tree. The answer is given by the ratio of the logarithms of the two measured light intensities.

In the early successional stage the number of branch layers is 4.3 for gray birch, 3.8 for bigtooth aspen, 3.8 for white pine and 2.7 for sassafras. In the intermediate successional stage the number of branch layers is 2.7 for white ash, 2.6 for black gum, 2.7 for red maple, 2.2 for tulip, 2.7 for red oak and 2.7 for hickory. In the late successional stage the number of branch layers is 1.9 for sugar maple, 1.5 for beech and 1.6 for hemlock. From these numbers it is obvious that the early successional trees tend to have more branch layers than trees in later stages.

We now have a clue to understanding at least part of the successional pattern in the stands I have studied. Multilayered trees are able to grow faster than monolayered trees in the open environment of early succession, but the shaded understory of these multilayered trees is less suitable for their own offspring than it is for monolayered trees. When monolayered trees eventually reach the canopy, their understory is well shaded, and multilayered trees can live only in the gaps left by the death of large trees.

Several other successional patterns follow directly from differences in the geometrical distribution of leaves. A tree must dissipate the solar heat it absorbs. The heat to be dissipated (and hence the water lost through evaporation) is spread over a greater leaf area in a multilayered tree than it is in a monolayered one. This means that the water loss per unit of photosynthesis is lower for a multilayered tree than it is for a monolayered one. On well-drained soil where water is scarce multilayered trees should therefore be dominant. On moist soil, where monolayered trees eventually come to dominate the canopy by virtue of their competitive advantage in the shaded understory, wood productivity will fall because the monolayered trees are less productive than the multilayered trees they displace. Moreover, the species most

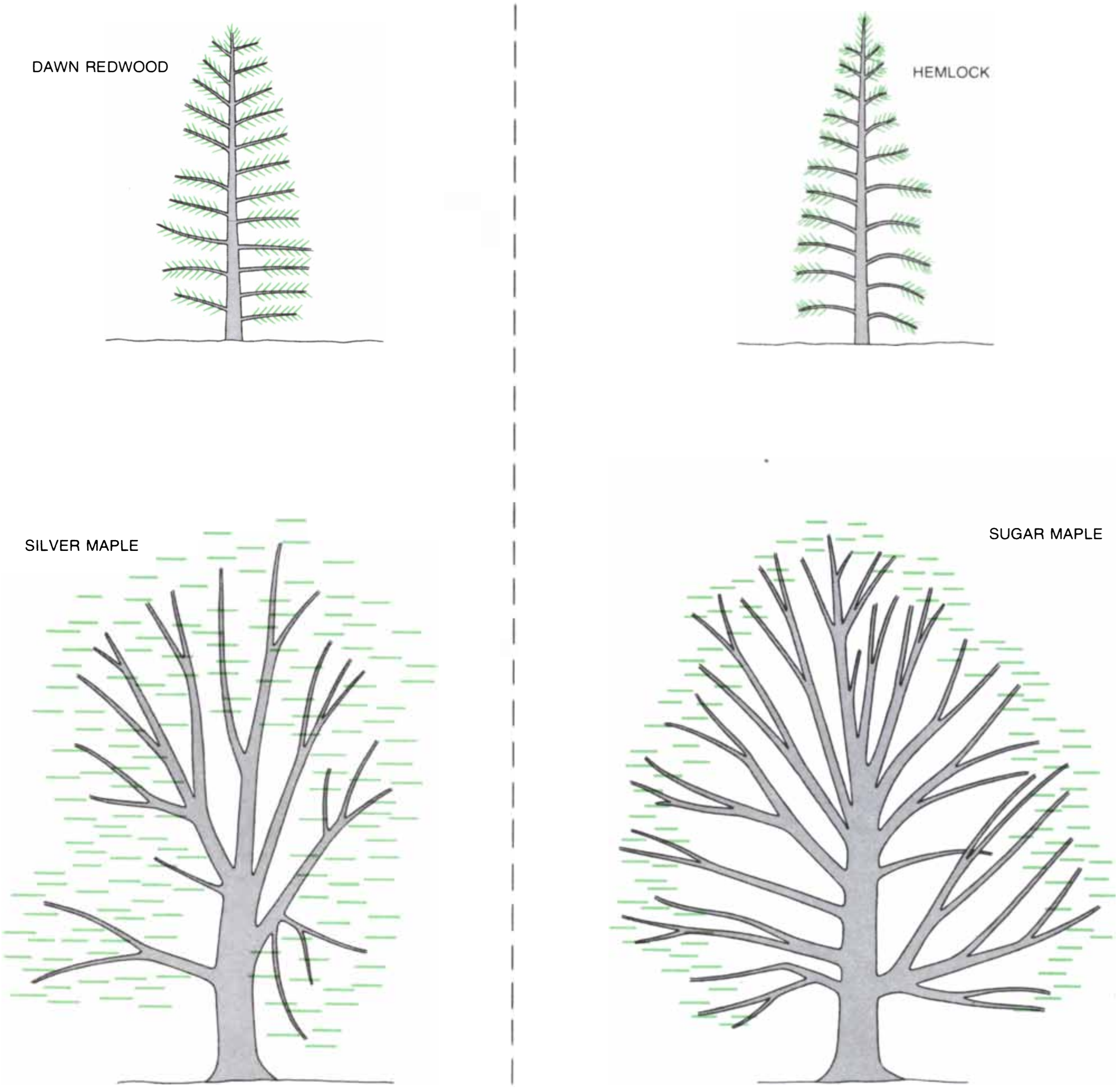


SUCCESSIONAL CHANGES IN PRODUCTIVITY are shown in these graphs, measured by indexes of the standing crop of wood (*top*), the production of new wood (*middle*) and ratio of the production of wood to standing crop (*bottom*). The successional age is given logarithmically. The black line in each graph connects plots of the trees on moist soils, from gray birch through beech. The colored line connects the two stands that grow on drier soils. The sweet gum stand, which grows on soil intermediate in moisture content, is plotted by itself. Note that with increasing age production first increases and then decreases, except for sweet gum, which maintains very high production in a young and densely stocked stand.

adapted to shade should have an unbroken peripheral layer of leaves and thus should cast a shade so deep that not even their own offspring can grow in it. We can expect, then, that undisturbed forests in which monolayered trees dominate the canopy will have little or no new growth and will be senile.

Another potentially important difference between monolayered trees and multilayered trees is their susceptibility to attack by insects. I have often seen spring inchworm caterpillars defoliate a multilayered tree by eating one layer, lowering themselves on silk threads to the next layer and so on throughout the tree. This behavior is less efficient on a monolayered tree. If the inchworms try to change their foraging position by lowering themselves, they find themselves on the ground and then must proceed to the base of the tree and up the trunk. I have the impression that inchworms defoliate the local monolayered trees less often than they do the multilayered ones, but the evidence is insufficient to demonstrate that the difference is ecologically significant.

To sum up, interspecific differences in the geometrical distribution of leaves are at least partly responsible for the pattern of succession, for the successional changes in the age structure and productivity of trees and perhaps even for differential predation by some insects. Furthermore, the number of branch layers is a measur-



DISTRIBUTION OF LEAVES in multilayered and monolayered trees is depicted schematically in a slice through the middle of the tree. Dawn redwood and silver maple are typical of multilayered trees, with leaves scattered throughout the volume of the tree. Hemlock and sugar maple are virtually monolayered, with leaves concentrated at the ends of branches. Although the two conifers have similar branching patterns and a similar pattern of leaf attachment. their total leaf distribution is quite different. The same difference holds for the two superficially similar maples. In open spaces, where the intensity of sunlight is high, the multilayered tree is competitively superior to monolayered tree, but in shaded understory of a forest monolayered tree has the competitive advantage.

able criterion for predicting a tree's rate of growth and its requirements for light and water. I believe predictions based on the number of branch layers could be useful in evaluating new varieties of trees for forests and gardens.

The hypothetical successional scheme predicts that the most aggressive species of monolayered tree will completely eliminate its competitors if given enough time and a favorable environment. In reality many factors other than tolerance of shade will determine the outcome of competition for places in the canopy of a forest. The effect of these additional factors can be taken into account by empirically determining the probability of whether a tree will be replaced by another of its kind or replaced by a tree of another species. I have estimated the replacement probability for each tree in a stand, and from the matrix of these probabilities I can calculate how many trees of each species can be expected to remain in the stand after any given interval [*see illustration at right*]. The probabilities are estimated by counting the number of saplings of each species under a canopy tree. I then gratuitously assume that every sapling under the canopy tree has an equal chance of replacing that tree in the next generation of the canopy.

SAPLING SPECIES

CANOPY SPECIES	BIGTOOTH ASPEN	GRAY BIRCH	SASSAFRAS	BLACK GUM	SWEET GUM	WHITE OAK	RED OAK	HICKORY	TULIP	RED MAPLE	BEECH
BIGTOOTH ASPEN	3	5	9	6	6	0	2	4	2	60	3
GRAY BIRCH	0	0	47	12	8	2	8	0	3	17	3
SASSAFRAS	3	1	10	3	6	3	10	12	0	37	15
BLACK GUM	1	1	3	20	9	1	7	6	10	25	17
SWEET GUM	0	0	16	0	31	0	7	7	5	27	7
WHITE OAK	0	0	6	7	4	10	7	3	14	32	17
RED OAK	0	0	2	11	7	6	8	8	8	33	17
HICKORY	0	0	1	3	1	3	13	4	9	49	17
TULIP	0	0	2	4	4	0	11	7	9	29	34
RED MAPLE	0	0	13	10	9	2	8	19	3	13	23
BEECH	0	0	0	2	1	1	1	1	8	6	80

NEXT GENERATION OF WHITE OAK = .02GB + .03SF + .01BG + .1WO + .06RO + .03HI + .02RM + .01BE

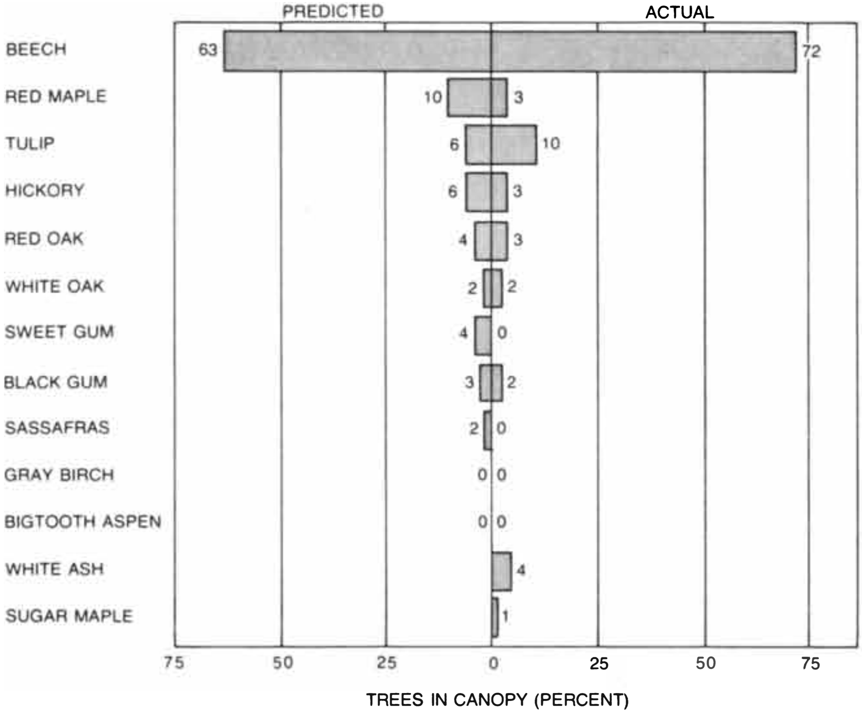
MATRIX OF PROBABILITIES for the replacement of existing canopy trees in a forest by trees of the same or other species is based on the assumption that every sapling under a canopy tree has an equal chance of replacing the tree. Each entry in a row is the percent of the total number of saplings found under a canopy tree of the species listed. For example, under the bigtooth aspen canopy tree 3 percent of the saplings are bigtooth aspens, 5 percent are gray birch and so on. The figures in each row add up to 100 percent. A sample calculation is given for the proportion of the forest that will be white oak in the next generation. The calculation shows that white oak will replace 2 percent of the gray birch, 3 percent of the sassafras, 1 percent of the black gum and so on. The new composition of the forest can then be substituted in the equation, and the composition of the second generation of canopy trees can be calculated. Repetition of the calculations leads to a stationary distribution in which the composition of the canopy remains unchanged in subsequent generations.

When the probabilities for successive generations are calculated, eventually a stationary distribution is reached where the proportion of different species in the canopy does not change. The probability matrix can be weighted to take into account the fact that species differ in longevity and that the generation span is not the same even for trees of the same species. When the actual distribution of trees that form the canopy of the beech forest—the oldest and least disturbed stand in the woods I studied—is compared with the predicted distribution of trees, the prediction is surprisingly accurate [*see illustration on next page*].

A word of caution is in order. I was fortunate to have found forest stands that began with the invasion of open fields. Later species grew up in the shade of the pioneering species, and the numerical abundance of saplings in the understory proved to be a reasonable predictor of a species' success in reaching the canopy. Other successions might be determined by a race among different species to invade a new gap after a canopy tree has fallen, by biochemical interaction among trees, by changes in the climate or in the soil or by a multitude of other interactions among plants, animals and the physical setting. Such situations would call for more complex and perhaps technically impossible measurement of replacement probabilities. Nonetheless, as long as the fate of a given spot in a forest is dependent only on its current occupant and a limited number of neighbors, a theoretical representation of succession as a tree-by-tree replacement process is at least reasonable in principle.

The replacement process has several intriguing properties. First, it converges on the same final distribution of trees no matter at which successional stage it begins. This suggests that the convergence is a statistical property rather than a biological one, although there are interesting biological reasons behind both the form of the replacement matrix and the final distribution to which the process converges. When the longevities of different species of trees are taken into account, the relative proportion of each species in the successive stages changes, but no new qualitative properties are seen to emerge.

When a forest that has reached a stationary state is disturbed, it takes longer for it to return to a stationary state if the first species to regenerate perpetuates itself rather than giving way to other species. Local self-replacement therefore slows down the achievement of a stable stationary state. Moreover, if each species' contribution to the next generation is made to be proportional to its



PREDICTED DISTRIBUTION AND ACTUAL DISTRIBUTION of canopy trees in a forest are surprisingly close to each other. The prediction is the stationary distribution that was calculated from the matrix of probabilities in the illustration on the preceding page and the weighted longevities of each species. There are no hidden circularities because the predicted composition and the actual one are based on completely different sets of data.

current abundance, then even small disturbances from the stationary stage may be self-augmenting. There might even be several stationary stages, depending on which species colonize first.

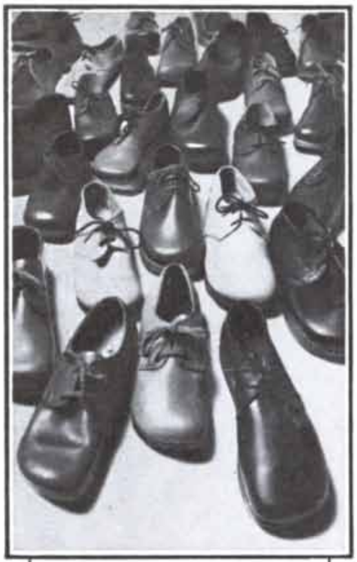
The model I have developed shows that it is dangerous to assign a unique successional status to a species; and that it is ridiculous to determine the successional status of a plot solely by the species of trees that are on it. The final stage of succession is really a mosaic of patches of varying successional age. Some species that are characteristic of early stages of succession, such as the black gum, persist in the late stages, both in the theoretical model and in reality. The age structure of each species within a stand and in several stands must also be taken into account.

The effects of eliminating one tree species or more from a forest can be calculated from the matrix of transition probabilities. Although such calculations are no substitute for real experiments, they may be useful as an aid in deciding which experiments would be worth conducting. One could attempt to find out how to speed or to slow certain successions. For example, the model indicates that patch-cutting, the harvesting of local groups of trees while leaving a large area of uncut forest around them, should result in a more rapid return to the stationary stage than the cutting down of all the trees or even "high-grading": the removal of all commercially valuable trees.

The dynamic stability of succession must decrease with time, almost by the definition of succession. If a late successional stage is disturbed, the regeneration must pass through a long series of changes before the late stage is regained. That is the exact opposite of the more traditional view that the stationary state of a forest persists as a result of an intrinsic stability of the entire community and that forests tend to resist the perturbations imposed on them by man or by natural catastrophes.

The traditional argument of stability has been invoked to justify the conservation and management of natural areas. A more careful analysis shows that a forest in a stationary state is in fact more unstable than a forest in an early stage of succession. Where natural forests in a stationary state are desirable, perhaps they should be protected not only from disturbance but also from well-intentioned attempts at management.

Why everybody's pretending they're us.



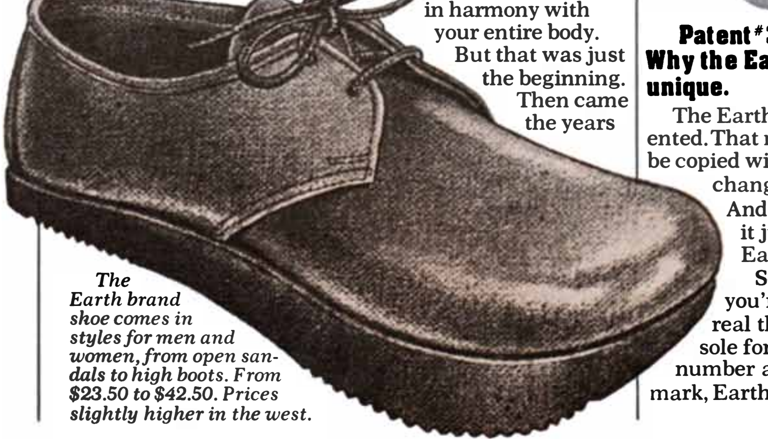
These are not Earth shoes. Just because they look like Earth shoes doesn't mean they are Earth brand shoes.

There was a time when the EARTH* negative heel shoe was the only shoe in the world with the heel lower than the toe.

In those days the other people who made shoes just laughed at us.

But things have changed.

And now that you love our Earth brand shoes, now that you're standing in line to get them, the shoe companies have stopped laughing and started copying.



The Earth brand shoe comes in styles for men and women, from open sandals to high boots. From $23.50 to $42.50. Prices slightly higher in the west.

The shoes that look like, seem like, but don't work like the Earth® shoe.

Today, a lot of people are trying to imitate our shoe. Some even use names that sound like ours, and have ads that look like ours!

It seems like everybody's trying to be us.

But what they don't understand is this. Merely lowering the heel of a shoe isn't enough. And imitating the outside of our shoe isn't enough. Just because a shoe looks like the Earth shoe doesn't mean it works like the Earth shoe.

It took many years to perfect the Earth brand shoe. And those years are crucial. They make our shoe different from all its imitators.

How the Earth® Shoe was invented.

It started years ago when Anne Kalsø had the original idea for the negative heel shoe.

She saw footprints in the sand, and realized that with every footprint the body was designing a shoe. A natural shoe. A shoe with the heel lower than the toe. A shoe that would work in harmony with your entire body.

But that was just the beginning. Then came the years of research and hard work to get every detail just right. To perfect the arch. To make the toes wide, comfortable and functional. To balance the shoe. To mold the sole in a special way so that it would allow you to walk in a natural rolling motion. Gently and easily even on the hard jarring cement of our cities.

To get an idea of how the Earth® shoe works, stand barefoot with your toes up on a book. Feel what begins to happen.

Patent #3305947. Why the Earth® shoe is unique.

The Earth shoe is patented. That means it can't be copied without being changed.

And if it's changed it just isn't the Earth shoe.

So to be sure you're getting the real thing, look on the sole for our patent number and our trademark, Earth. If they're not there, it's not the Earth brand shoe.

Sold only at Earth® shoe stores.

And there's one more thing that makes our shoes so special. Our stores.

Earth shoes are sold only at Earth shoe stores. Stores that sell no other shoe but ours, and are devoted entirely to the Earth shoe concept.

How our shoes fit you is very important to us. There's a special technique to fitting them. Our people are trained to fit you properly and we wouldn't trust anyone else to do it.

Find out for yourself.

To really appreciate Earth shoes you must try them.

When you do you'll see, perhaps for the first time in your life, what it's like to walk more gracefully, naturally and comfortably.

KALSØ ® earth ® shoe

**EARTH is the registered trademark of Kalsø Systemet, Inc. for its negative heel shoes and other products.*

©1974. Kalsø Systemet, Inc.

Anne Kalsø.

Inventor of the EARTH negative heel shoe.

You can only buy Earth® shoes at Earth® Shoe Stores in the cities listed on the facing page.

Introduction to Contemporary Psychology

Announcing a New Text

by EDMUND J. FANTINO and GEORGE S. REYNOLDS,
University of California, San Diego

An instructor's guide and study guide will be published along with the text. A free test item file will be sent to teachers upon class adoption of the text.

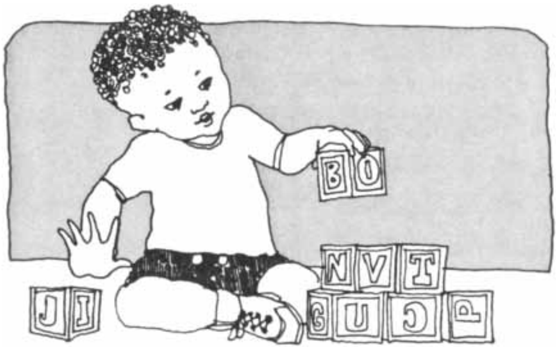
"Although we feel that there is not necessarily any incompatibility among prevailing psychological points of view, we also feel that the introductory student deserves a fair look at, and a sure feel for, all of them."—The Authors

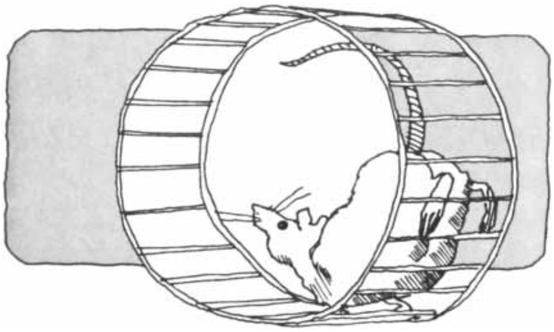
Many students take an introductory psychology course to satisfy an area requirement, others take it as a prerequisite to more advanced courses in psychology, still others take it because they find the subject fascinating. *Introduction to Contemporary Psychology* is a text that will serve all these students well.

This book takes a new approach to introductory psychology—it is an intentionally eclectic text, treating the subdisciplines and many schools of thought that constitute modern psychology from the point of view of the psychologists actually involved in them. Such an approach gives students an appreciation of the great scope of the science and prepares them for subsequent specialization.

Introduction to Contemporary Psychology is divided into two major parts. The first half of the book deals with such basic psychological processes as learning, remembering, thinking, perception, motivation, instinct, and heredity. Whenever appropriate, the authors have included a discussion of the physiological correlates of behavior. The second half of the book discusses the major subdisciplines of psychology and the important theoretical controversies of the science.

Publication date: February 1975


BO
NVT
JI
GUCP



MATHEMATICAL GAMES

On the remarkable Császár polyhedron and its applications in problem solving

by Martin Gardner

Some 20 years ago Donald W. Crowe, now a mathematician at the University of Wisconsin, discovered a surprising correspondence between the skeletons of n-dimensional cubes and the solution of a classic puzzle called the Tower of Hanoi. I described his discovery in this department in May, 1957, and the report is reprinted in *The Scientific American Book of Mathematical Puzzles & Diversions.* In "Mathematical Games" for August, 1972, the correspondence was extended to Gray binary codes and to the solution of another classic puzzle, the Chinese rings.

Professor Crowe has now done it again. In studying the skeletal structure of a bizarre solid known as the Császár polyhedron he has found some remarkable isomorphisms that involve the seven-color map on a torus, the smallest "finite projective plane," the solution of an old puzzle about triplets of seven girls, the solution of a bridge-tournament problem about eight teams and the construction of a new kind of magic square known as a Room square.

The polyhedron that leads into so many apparently unrelated recreations is enormously interesting in itself. It is the only known polyhedron, apart from the tetrahedron, that has no diagonals. (A diagonal is a line joining two vertexes that are not connected by an edge.) Consider, for example, the tetrahedron. It has four vertexes, six edges, four faces and no diagonals. An edge joins every pair of corners.

Skeletons of polyhedrons are isomorphic with graphs, that is, with sets of points (vertexes) joined by lines (edges). If an edge connects every pair of points in a set of n points, it is called a complete graph for n points. Any polyhedron without diagonals clearly must have a skeleton that is a complete graph. Since no polyhedron can have fewer than four corners, the complete graph for four points is the simplest such graph that corresponds to a polyhedral skeleton.

We must now be careful to make precise definitions. A simple polyhedron is one that is topologically equivalent to a sphere and whose faces are all simple polygons: polygons topologically equivalent to a disk. (If you think of a simple polyhedron as having elastic faces, the polyhedron can be inflated like a balloon to form a sphere.) This rules out such nonsimple structures as two polyhedrons joined by an edge or a corner, stellated polyhedrons with intersecting faces, solids with tunnels or interior holes and so on. Imagine a polyhedron with a surface like that of a cube with a smaller cube on the center of one face. The polyhedron can be inflated to a sphere, but is it simple? It is not, because one face is a ring.

It is not hard to show that the tetrahedron is the only simple polyhedron with no diagonals. Now, however, an interesting question arises. Let us define a toroid as a polyhedron whose faces are all simple polygons, but the solid itself is topologically equivalent to a sphere with one tunnel or more going all the way through it. Is it possible to construct a toroid with no diagonals?

The answer to the question was not known until the late 1940's, when a Hungarian topologist, Ákos Császár, succeeded in constructing such a polyhedron. Its skeleton is the complete graph for seven points, which is shown in symmetrical form at the left in the top illustration on page 104. The graph is isomorphic with the skeleton of a six-dimensional simplex, the 6-space analogue of the tetrahedron. Since an edge joins every pair of points, the polyhedron has 21 edges and 14 triangular faces.

It is not difficult to make a paper model of the Császár polyhedron. Copy the two patterns in the illustration on the opposite page on paper of good quality (or thin cardboard) and cut the copies out. Shade the seven colored triangles on both sides. Crease the paper to make "mountain folds" along each broken line and "valley folds" along each solid line.

1. With the pattern for the base, fold the two largest triangles to the center and tape the A edges to each other. Turn the paper over. Fold the two smaller triangles to the center and tape the B edges together to obtain a completed base.
2. The six-faced conical top is formed by taping the C edges together. Place it on the base as shown in the drawing of the completed model. It will fit in two ways. Choose the fit that joins triangles of opposite colors, then tape each of its six edges to the corresponding six edges of the base.

It is not yet known if there is another toroid without diagonals. By applying elementary Diophantine analysis (finding integral solutions of equations), however, Crowe has shown that if there is another, it will have at least 12 vertexes and six tunnels. His proof is as follows:

For simple polyhedrons there is a famous formula, discovered by Leonhard Euler, that relates the vertexes (v), edges (e) and faces (f). It is

$$v - e + f = 2 \, .$$

The formula is easily proved, and it takes only a bit of extra work to modify it for toroids. Letting h stand for holes, the formula is

$$v - e + f = 2 - 2h \, .$$

If a toroid is without diagonals, its skeleton is a complete graph, and for any complete graph, edges and vertexes are related by the formula

$$e = \tfrac{1}{2}v(v - 1) \, .$$

The faces and edges of a toroid without diagonals are related by the formula

$$f = 2e \,/\, 3.$$

Substituting the values for e and f in the formula $v - e + f = 2 - 2h$ and simplifying, we get

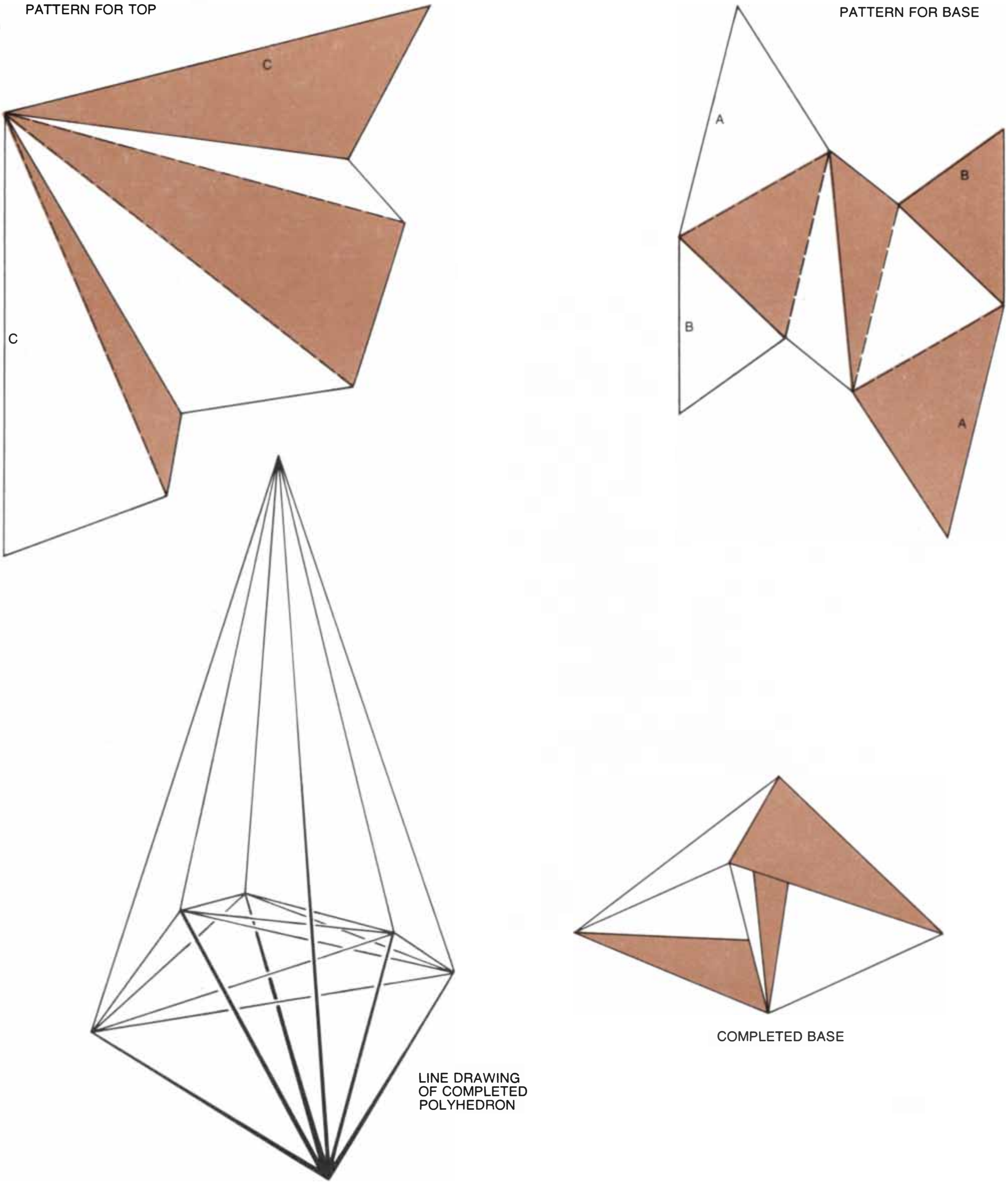
$$12h = v^2 - 7v + 12 \, .$$

Factoring the right side gives h a value of

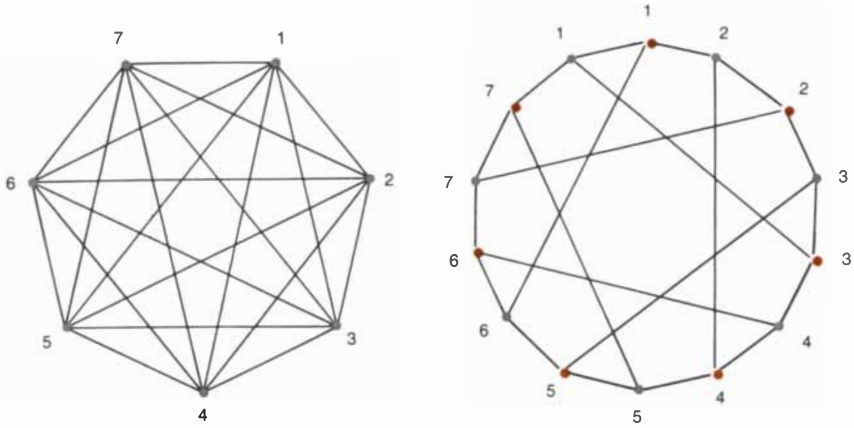
$$h = \frac{(v - 3)(v - 4)}{12} \, .$$

The values of v and h must be integers, and we also know that v must be greater than 3. If v is 4, h has a value of zero. These values fit the tetrahedron. If v is 5 or 6, h is not integral, proving that no toroid without diagonals can have five or six corners. When v is 7, h is 1. This corresponds to the Császár polyhedron. The next solution in integers is $v = 12$, $h = 6$. Whether a toroid can be constructed for these values remains unknown. Nor is it known if toroids exist for the next two solutions: $v = 15$, $h = 11$, and $v = 16$, $h = 13$. From here on, Crowe points out, the number of tunnels exceeds the number of vertexes, so that we can probably rule out all higher toroids.

The Császár polyhedron has one tunnel, which means that if our model's surface consisted of rubber rather than paper, we could blow it up to the shape of an inner tube. Its 21 edges would then



Patterns for making a model of the Császár polyhedron



Skeleton of the Császár polyhedron (left) *and its dual, the seven-color torus map* (right)

form a complete graph for seven points on the surface of the torus. None of these edges would intersect one another, proving that the complete graph of seven points has a toroidal crossing number of zero. (See the discussion on crossing numbers in this department for June, 1973.)

Suppose now that on this torus we change the complete graph to its dual. That is done by putting a point inside each of the 14 triangular faces, then drawing an edge from each point to the points inside its three neighboring faces. Since each new edge crosses one old edge, the number of new edges remains 21. The number of faces and the number of vertexes, however, are switched. The new graph is shown in a symmetrical planar form at the right in the illustration above. Crowe has made the 14 vertexes alternately colored and gray and has numbered them as shown for reasons that will shortly be clear. Its "faces" are hard to see, but you can trace out seven regions, each region surrounded by six edges.

This graph can also be drawn on a torus without any intersecting edges. When that is done, the result turns out to be the familiar seven-color toroidal map [*see illustration below*]. Note that every two of the seven hexagonal regions share a common edge. This means that if the map is colored so that no two contiguous regions have the same color, seven colors are required. On a plane no more than four regions can be mutually contiguous, but on a torus the maximum is seven.

As an exercise the reader may like to put a dot inside each region on the torus and see if he can convert the map to its dual by the same procedure described above. Simply draw an edge to connect each pair of dots. Each edge must cross just one boundary segment, and there must be no intersecting of the new edges with each other. Some lines must, of course, wrap around the torus. If you succeed, the new graph will be isomorphic with the complete graph of seven points and also with the skeleton of the Császár polyhedron.

Now, here is how Crowe used the dual graph of the skeleton of the Császár polyhedron to solve the following puzzle. Seven girls live in a house. Each day a triplet of girls is allowed to leave the house for a visit to town. How can the triplets be chosen so that at the end of seven days every pair of girls will have been in exactly one of the seven triplets?

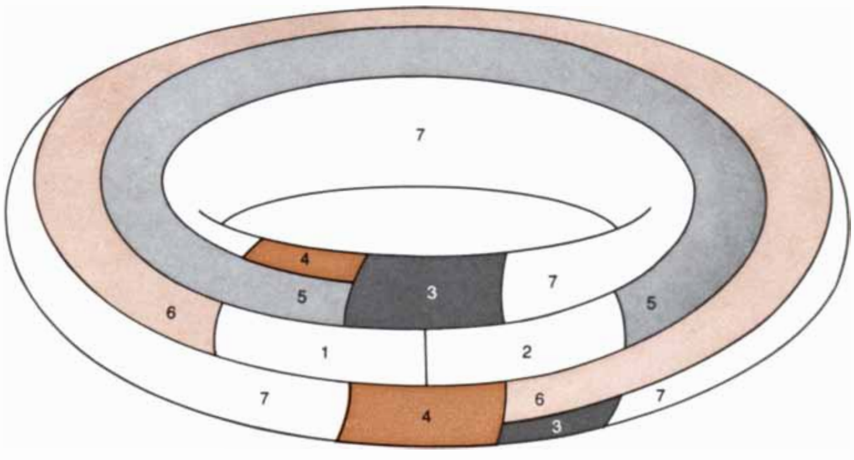
The graph at the right in the top illustration at the left provides two solutions. For one solution note each gray vertex and write down the (unordered) triplet of numbers on the three colored vertexes that are adjacent (connected by an edge) to it. For the other solution write down the numbers on the gray vertexes adjacent to each colored vertex. The two sets of triplets are

Gray vertex	*Colored vertex*
124	126
235	237
346	341
457	452
561	563
672	674
713	715

Each set is called a "Steiner triple system" of order 7 or a "finite projective plane" of order 2. Steiner systems and projective planes are topics of great importance in modern combinatorial theory, but we can (regrettably) mention them only in passing.

Because the seven-color graph is the dual of the Császár skeleton, with each of its 14 vertexes corresponding to one of the 14 faces of the model of the Császár polyhedron (gray spots to white faces, colored spots to colored faces), we can just as easily extract the two solutions from the model. On the model no face shares a border with a face of the same color. If the faces are numbered to correspond to the numbers on the vertexes of the graph, the triplet of numbers on the three white faces adjacent to each colored face gives one solution, and the triplet of numbers on the three colored faces adjacent to each white face gives the second solution.

Another way to find the same two sets of triplets on the model is to number the vertexes of the model, any way you like, from 1 through 7. The numbers at the three corners of each colored face give one set of triplets and those at the three corners of each white face give the other set. The two solutions will be equivalent



A seven-color map on the torus. (Regions 3, 4 and 7 wrap around)

to the two sets of triplets listed above, although the numbers may not match. The numbers are no more than arbitrary symbols on a symmetrical graph. To see the equivalence it may be necessary to permute them in some way, such as changing all 1's to 5's, all 5's to 3's and so on.

The two solutions obtained by any of these methods are also identical in the sense that one can be changed to the other by permuting the elements; in other words, there is only one basic solution. The next higher Steiner triple system is of order 9. It too has a unique basic solution. Nine girls go out in daily triplets for 12 days, each pair appearing in just one triplet. Can the reader find the solution before I give it next month?

The two variants we obtained for the order-7 solution are, as Crowe recognized, related in a curious way. No triplet appears in both sets, and if two pairs of girls in one set appear with the same third girl (for example, in the first set 1,2 and 3,6 each appear with 4), the same pair appears with different girls (5,6) in the second set. When both of these properties hold, the two solutions are called orthogonal.

A Steiner triple system of order *n* is possible only when *n* is equal to 1 or 3 (modulo 6). Every orthogonal pair of such systems of order *n*, Crowe goes on to explain, provides a solution to the following bridge problem for $n + 1$ teams. Suppose there are eight teams of card players and seven tables. Each team must play exactly once with each of the other teams and also exactly once at each table.

This is how Crowe tells us to construct the tournament. First draw a square matrix seven cells by seven cells. Consider the pair 1,2. In the first set of triplets it is associated with 4 and in the second set with 6, so that we put 1,2 in the cell at the intersection of the fourth column and the sixth row [*see illustration on this page*]. Consider another pair, 1,3. It is with 7 in the first set and with 4 in the other, so that 1,3 goes in the seventh column and the fourth row. Follow this procedure for all pairs of numbers. The final step is to combine 8 with 1, 2, 3, 4, 5, 6, 7 along a diagonal from the cell at the upper left to the cell at the lower right. Each column indicates a table, and each row indicates a round of simultaneous play at four of the seven tables. All conditions of the desired tournament are now met.

The matrix is called a Room square of order 8. Such a square is an arrangement of an even number of objects, $n + 1$, in a square array of side *n*. Each cell is either empty or holds exactly two different objects. In addition each object appears exactly once in every row and column, and each (unordered) pair of objects must occur in exactly one cell.

	TABLES						
ROUNDS	1	2	3	4	5	6	7
1	1,8			5,7		3,4	2,6
2	3,7	2,8			6,1		4,5
3	5,6	4,1	3,8			7,2	
4		6,7	5,2	4,8			1,3
5	2,4		7,1	6,3	5,8		
6		3,5		1,2	7,4	6,8	
7			4,6		2,3	1,5	7,8

The smallest nontrivial Room square

The smallest Room square is trivial. It is of order 2 and consists of one cell that contains 1,2. No Room squares are possible for four or six objects, so that the order-8 square is the smallest nontrivial Room square.

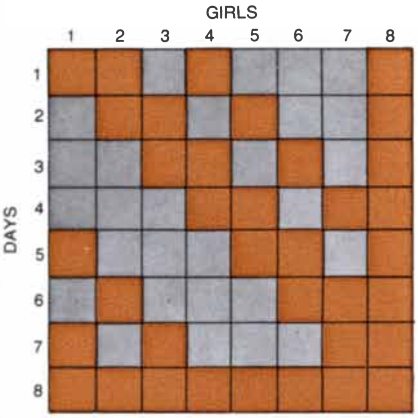
For years I assumed that such squares were called Room squares because they concern objects placed in "rooms," but in researching this column I found that they are named after their discoverer, Thomas G. Room, now at the Open University at Manchester in England. He first defined them in a paper, "A New Type of Magic Square," in *The Mathematical Gazette,* Volume 39, page 307, 1955. Combinatorial enthusiasts have been working on them ever since.

The interested reader can find out more about Room squares by checking the following papers in *Journal of Combinatorial Theory:* "On Furnishing Room Squares," by R. C. Mullin and E. Nemeth, Volume 7, November, 1969; "On Room Squares of Order $6m + 2$," by C. D. O'Shaughnessy, Volume 13, Series *A,* November, 1972, and "Solution of the Room Square Existence Problem," by W. D. Wallis, Volume 17, Series *A,* November, 1974. The last article proves that Room squares exist for all orders (which must be even) except 4 and 6.

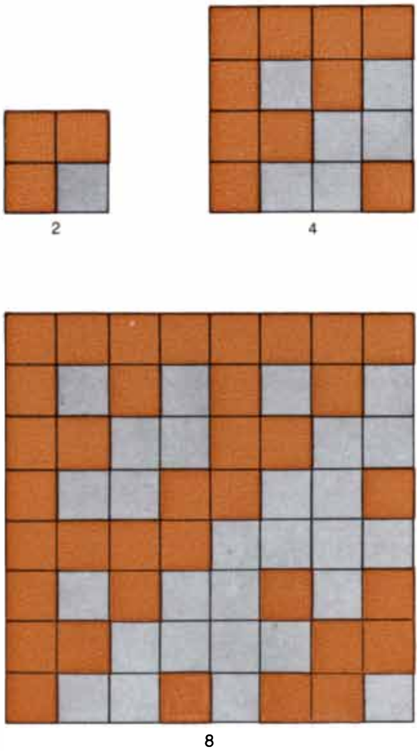
There is still more! H. S. M. Coxeter, in editing the 12th revised edition of W. W. Rouse Ball's classic *Mathematical Recreations and Essays,* published last year by the University of Toronto Press, explains how the Steiner triple system of order 7 can be used for constructing an "anallagmatic pavement" of order 8. Consider the order-8 chessboard. If we place any two rows alongside each other, either every cell in one row will match the color of its neighbor in the other or every cell will not match its neighbor's color. We want to color the 64 cells with two colors so that the following property holds: If any two rows are brought together, half of the paired cells will match and half will not, and the same will be true of any two columns.

Such squares are also known as Hadamard matrixes, after the French mathematician Jacques Hadamard, who studied them in the 1890's. Apart from the trivial case of order 2 no Hadamard matrix is possible unless the order is a multiple of 4. It is not yet known if matrixes for all such orders exist. The first doubtful case is order 188.

The upper illustration below shows how our first Steiner triple system provides a Hadamard matrix for the chessboard. The problem involves eight girls and eight days. Number the rows and columns as shown. The eighth girl is an older girl who chaperones the triplets daily, and on the eighth day all eight girls go into town. For each day (indicated by a row) color the cells that indi-



A Hadamard chessboard



Hadamard matrixes for powers of 2

cate the three girls (plus the chaperone) who walk to town. The result: a Hadamard matrix!

There is a simple technique for generating Hadamard matrixes for all orders that are powers of 2 [*see lower illustration on opposite page*]. The order-2 pattern is placed in three corners of the order-4 square, and its negative (colors reversed) goes into the lower right corner. The same procedure with the order-4 pattern generates the order 8, and so on for the higher powers.

More generally, given two Hadamard matrixes of orders m and n, a matrix of order mn can be created simply by replacing each colored cell of m by the entire pattern of n and each uncolored cell of m by the negative of n. The large matrix is called the tensor product of the two smaller ones. It does not matter whether m is larger, smaller or equal to n. If, for example, you have a Hadamard matrix of order 12, the tensor product of two such matrixes is a Hadamard matrix of order 144.

Hadamard matrixes are more than playthings. They are used for the construction of valuable error-correcting binary codes, and for descriptions of such applications the reader is referred to the marvelous new edition of Ball's book (so heavily revised as to be almost a new work) and to the references it cites. When the Mariner spacecraft of 1969 sent back pictures of Mars, Coxeter tells us, they were sent in an error-correcting code based on the order-8 Hadamard matrix.

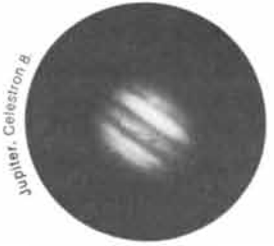
In closing, let us return to the Császár polyhedron and pose an intriguing problem. The Császár toroid cannot be constructed if all its faces are equilateral triangles. Suppose a one-hole toroid is made entirely of congruent triangular faces, all equilateral. What is the minimum number of faces it can have? Bonnie M. Stewart, a mathematician at Michigan State University, considers this problem on page 48 of his book *Adventures among the Toroids,* which he published in 1970. (This remarkable volume, which deals entirely with building models of regular-faced toroids, is available from the author, 4494 Wausau Road, Okemos, Mich. 48864, for $6.40 postpaid.)

Stewart gives the construction of such a torus with 54 faces, 27 vertexes and 81 edges. Recently one of his students, Kurt Schmucker, found a one-hole toroid with 48 equilateral triangular faces. Whether that is the minimum, however, is another tantalizing and unanswered toroidal question.

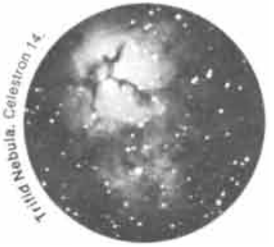
Celestron® Multipurpose Telescopes

On Display at Museums and Planetariums Throughout the Country

Squirrel at 70 ft. Celestron 5



Jupiter. Celestron 8



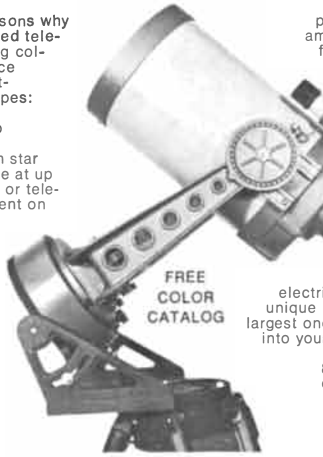
Trifid Nebula. Celestron 14.

The Choice. Three reasons why experienced telescope enthusiasts and leading colleges, universities and science centers the world over repeatedly select Celestron telescopes:

The Celestron 5. A tabletop observatory for exploring the Moon, planets, scores of open star clusters and gossamer nebulae at up to 300X. For casual observing or telephotography, rest the instrument on any flat surface, swing up the tube and focus. Close in on the whiskers of a squirrel at the near focus of 20 ft. or the face of a friend at half a mile. The 3½-lb. tube demounts for hand-held shots at 25X. (Size swung down: 7″ x 8″ x 16″, Wt.: 12 lbs., Base price, including electric clock drive and setting circles, $720)

The Celestron 8. Eight full inches of aperture make this portable observatory in a suitcase the amateur's favorite for studying the surface features of Mars, the subdivided rings of Saturn, the ever-changing belt structure of Jupiter, the intricate filamentary detail of deep-sky nebulae, the central regions of globular clusters at up to 500X. Also the ultimate terrestrial telescope or telephoto! (Size swung down: 9″ x 12″ x 22″, Wt.: 23 lbs., $845)

The Celestron 14. The appeal of this prestigious, fully electric dome instrument is enhanced by a unique design that also makes it the world's largest one-man-portable. Demount and load it into your compact car in five minutes! Within range of the Celestron 14, at up to 850X, are the delicate contrast levels of the diffuse and planetary nebulae, the spirals of remote galaxies, and the quasars. (Size swung down: 18″ x 22″ x 44″, Wt.: 108 lbs., $3,750)



Celestron Pacific 2835 Columbia • Box 3578-SA • Torrance, CA 90503

World's Leading Manufacturer of Compact-Portable Telescopes

Plant a THINK TANK anywhere and watch the minds grow!

home-office-school-park-club-churches-laboratory

Unique instructional games designed by university professors to make learning fun through brain-to-brain action. Beginning games can be mastered by young children— final games will challenge intelligent adults. These are the famous GAMES FOR THINKERS from WFF 'N PROOF Publishers.

WFF 'N PROOF (logic)	9.75*
QUERIES 'N THEORIES (science)	9.75*
EQUATIONS (mathematics)	6.50*
ON-SETS (set theory)	6.50*
PROPAGANDA (social studies)	7.50*
ON-WORDS (word structures)	6.50*
CONFIGURATIONS (geometry)	6.50*
TRI-NIM (problem solving)	5.50*
REAL NUMBERS (arithmetic)	2.25*
WFF (beginner's logic)	2.25*
QWIK-SANE (puzzle)	2.25*
TAC-TICKLE (pure strategy)	1.25*
Teachers Manual	1.25*
THINKERS Bookends	16.00*
12-Kit THINK TANK & Teachers Manual	
With THINKERS Bookends (save $9.25)	74.50*
Without THINKERS Bookends(save $3.25)	64.50*

postage & handling included in above prices

Order from: **WFF 'N PROOF**
1490-YD, South Boulevard, Ann Arbor, Mi. 48104

Gifts that are a COMPLIMENT to receive!

You can smell the pines in this catalog.

The full-color Browning Catalog is filled with some of the finest backpack gear going. Carefully built Down and Fiberfill II® sleeping bags, large body-hugging packs, lightweight backpack tents, country-covering boots, sharp knives, and really comfortable socks, shirts, trousers, and Fiberfill II® and down-filled parkas. If you love backpacking, you gotta' have Browning. Send this coupon today!

BROWNING
We're out to make you comfortable

Sirs: Please send me your full-color 112-page Catalog. Browning, Dept. C70, Morgan, Utah 84050. I am enclosing $1.00.

Name ____________
Street ____________
City ________ State ______ Zip ______

© 1975 SCIENTIFIC AMERICAN, INC

An Evel's eye view of the Snake River Canyon

It commanded the world's attention. Some 30,000 converged on Idaho's Snake River Canyon to see it. Evel Knievel, daredevil motorcyclist, would attempt to jump the canyon in a steam-driven "sky cycle." Uncannily, it seemed a greater challenge than a moon shot.

How does it feel to be shot 3000 feet into the sky and gaze down into a canyon 600 ft. deep?

To find out, photographer Stu Allen bought the rights to mount remote-controlled cameras aboard the sky cycle, bringing back an Evel's eye view of the flight.

"With $70,000 and weeks of preparation invested in the project," says Allen, "I had to be able to recover my film — even if the sky cycle was blown all over the desert. I checked with NASA and found that their Nikons, similar to my off-the-shelf models, withstood impact tests of 0 to 50 G's in 5 microseconds — the equivalent of hitting a camera with a sledge hammer!"

The cameras would have to function in desert air, where humidity hovered at zero percent. The aft camera, mounted atop a pressurized tank of 485°F water, became so hot the film was seared.

At liftoff, the parachute opened prematurely, and the sky cycle crashed into the canyon wall with a force of 12.5 G's, tumbling down the precipice.

But Allen's Nikons worked perfectly throughout the flight, descent and afterwards, despite a smashed lens shade, cracked battery pack and scratches.

Like most photographers, you'll never send your cameras on a canyon jump. But if you're serious about photography, you'll want to be ready for that once-in-a-lifetime picture opportunity. A day when you'll need the camera so versatile, so reliable, it's the choice of about 90 percent of photojournalists. Write for Folio 10A. Nikon Inc., Garden City, N.Y. 11530. (Canada: Anglophoto Ltd., P.Q.)

Someday, you're going to need a Nikon.

THE AMATEUR SCIENTIST

Making a refractometer for the identification of liquids

Conducted by C. L. Stong

A pure and transparent liquid of unknown identity can often be identified by comparing the speed of light in the liquid with the speed of light in air. This ratio, which is termed the index of refraction, has been accurately measured and tabulated for thousands of substances by institutions such as the National Bureau of Standards. The indexes are listed in handbooks of chemistry and physics.

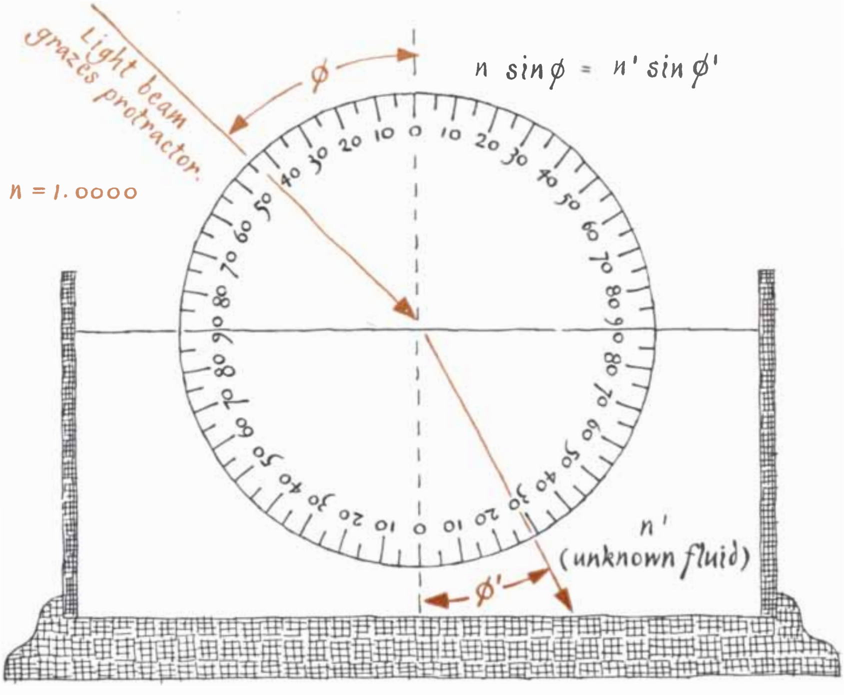
With this information and a modest collection of apparatus (an instrument for measuring the speed of light in a liquid, a thermometer and paper strips that have been treated to indicate acidity and alkalinity) the experimenter can identify many reagents and can also determine the concentration of some solutions. A homemade refractometer for measuring the speed of light in liquids has been built by Gene and Gary Frazier (2705 Gaither Street SE, Washington, D.C. 20031). The instrument includes as the source of light a helium-neon laser, a classroom version of which can now be bought for less than $100. A helium-neon laser can also be made at home, although the task is not an easy one [see "The Amateur Scientist," SCIENTIFIC AMERICAN, September, 1964, and December, 1965]. The Frazier brothers have written as follows about the principles and the operation of their instrument.

"Our refractometer measures the speed of light indirectly. It is based on the fact that a ray of light changes direction when it passes at an oblique angle from one medium, such as air, into another, such as water, having a different optical density. That is why a straight stick that has been pushed into clear water at an angle appears to bend at the interface of the water and the air.

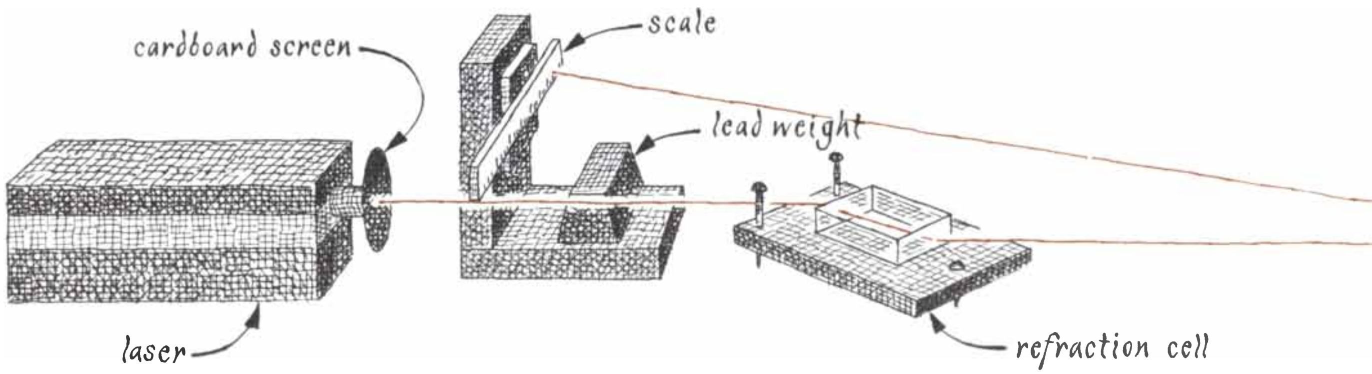
"The amount that a ray of light bends at the interface of adjoining mediums was first described quantitatively 354 years ago by the Dutch mathematician Willebrord Snell. Snell multiplied the index of refraction of each medium by the trigonometric sine of the angle made by the ray in each medium with respect to a line forming a right angle with the interface. He found that the two products are always equal. Expressed symbolically the relation is $n \sin \phi = n' \sin \phi'$, in which n and n' symbolize the indexes of refraction of the respective mediums and ϕ and ϕ' symbolize the respective angles made by the ray with a line that is perpendicular to the interface. The relation, which is named Snell's law, is basic to the design of lenses and prisms. It also figures in the production of many familiar compounds, including sugar (in solution) and alcohol.

"A simple apparatus that demonstrates the principle of the refractometer can be set up with parts available at a novelty store. The demonstration requires only a circular protractor, a rectangular aquarium (or some other kind of rectangular tank made of glass or clear plastic) and a narrow beam of light that can be directed downward at an oblique angle into a liquid contained in the tank. Immerse the protractor to its midpoint in the liquid. The light should barely graze the upper part of the protractor at an angle of precisely 45 degrees with respect to the vertical. The table of trigonometric functions lists the sine of 45 degrees as .707107.

"The beam of light enters the liquid from the air. The index of refraction of air can be assumed to be 1.0000, because the speed of light in air differs from its speed in a vacuum by only three parts in 10,000. For this instrument the numerical values of the first medium in Snell's law are $n = 1.0000$ and $\sin \phi = .707107$. To demonstrate the operation of the instrument fill the tank with water to the middle of the circular protractor. Switch on the light beam and adjust it to strike the water at exactly 45 degrees. Read the angle made by the submerged beam with the vertical. If a laser is the light source, the experimenter should be



A simple refractometer



The refractometer made by Gene and Gary Frazier

able to read the angle to within one degree.

"In water the observed angle ϕ' should be about 32 degrees. The table of trigonometric functions lists the sine of 32 degrees as .529919. According to Snell's law, the index of refraction of water, n', is equal to $n \sin \phi / \sin \phi'$, or $1.0000 \times .707107 / .529919 = 1.33437$. The speed of light in air is 186,228 miles per second. The speed of light in water is equal to the speed of light in air divided by the index of refraction of water: $186{,}228 / 1.33437 = 139{,}562.5$ miles per second. This speed is lower than the best recorded value of the speed of light in water at 20 degrees Celsius by only 112 miles per second, or about .08 of 1 percent. The error stems from the difficulty of measuring angles closely with a protractor.

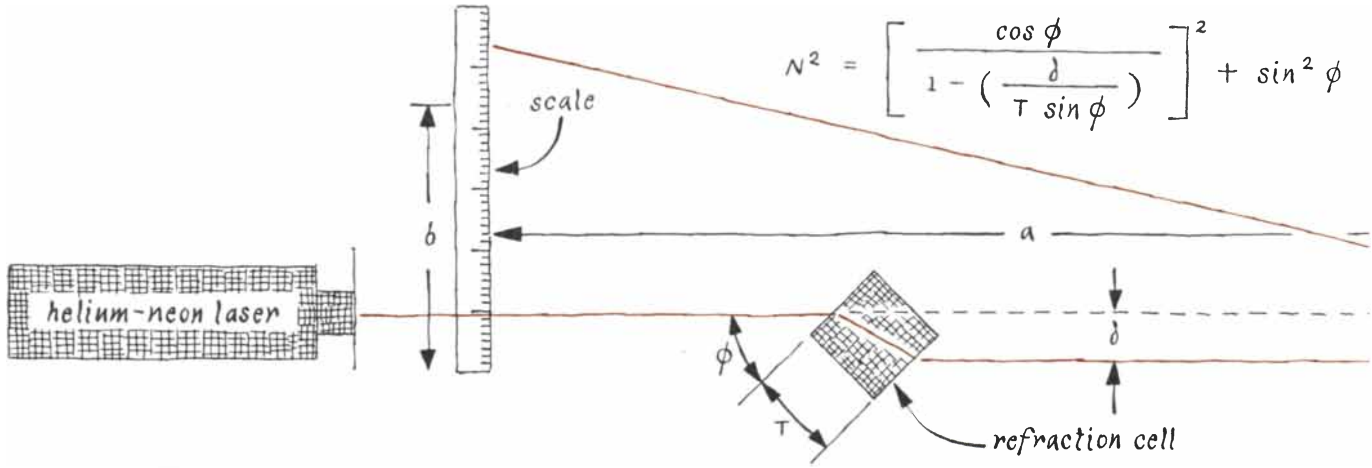
"Although an accuracy of .08 of 1 percent is adequate for rough measurements, the identification of unknown liquids and the study of the concentration of solutions require accuracies better than .004 of 1 percent. Finely wrought apparatus is needed to measure angles directly to this accuracy. Measurements of length to the same accuracy can be made with apparatus that is much easier to improvise at home.

"We cast about for a scheme that would enable us to measure angles indirectly to within at least a tenth of a degree. We finally hit on the idea of causing the specimen to displace a beam of light sideways and of employing a concave mirror to amplify the displacement. In this scheme the beam of light in air falls obliquely on one side of a glass cell with flat, parallel walls. The cell contains the specimen liquid.

"The beam is bent in the horizontal plane at the interface of the cell and the air. It then propagates through the cell and emerges into the air through the opposite wall. At this interface the beam bends exactly the same amount as it did at the first one but in the opposite direction, because it now proceeds from a medium of relatively high refraction into a medium of lower refraction. The path of the beam after its encounter with the cell exactly parallels its path prior to the encounter. The displacement increases with the index of refraction of the specimen. The apparatus is aligned when the cell is empty so that the beam proceeds to the center of the concave mirror, which can be of spherical form.

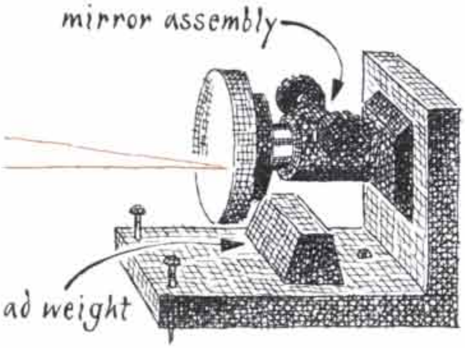
"When the cell contains a liquid, the displaced beam falls on the mirror at a point some distance from the optical axis. The beam is reflected by the mirror at an angle that increases with the displacement. The reflected beam falls on a horizontal scale that can be positioned at any convenient distance from the mirror.

"The displacement of the spot of light on the scale increases with the distance between the scale and the mirror and inversely with the focal length of the mirror. Mirrors of spherical figure deflect the beam at disproportionately larger angles as the beam departs from the optical axis. For this reason the excursion of the spot of light does not increase from its zero position on the scale in direct proportion to the displacement of



$$N^2 = \left[\frac{\cos \phi}{1 - \left(\frac{\delta}{T \sin \phi} \right)} \right]^2 + \sin^2 \phi$$

Optical train of the instrument



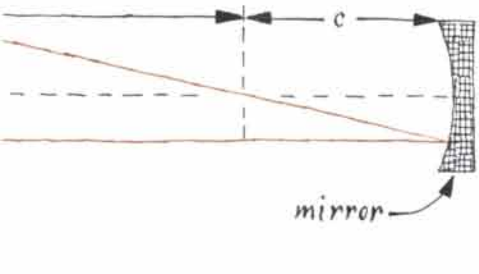
the beam by the specimen, as would be the case if the mirror were a paraboloid. The distortion is taken into account when the instrument is calibrated.

"The sensitivity of the instrument is determined by the thickness of the cell (*T*) and by the magnification of the displacement by the mirror. Our cell is made of standard microscope slides 25 millimeters wide, 75 millimeters long and one millimeter thick. The thickness of the cell (*T*) is 75 millimeters.

"The displacement (*d*) of the beam can be calculated with the formula that is included in the accompanying schematic illustration of the instrument [*bottom of these two pages*]. For example, in water, which has an index of refraction of 1.3333, the calculation for displacement is $75 \times .707107 \times 1 - \{.707107 / [(1.3333)^2 - (.707107)^2]^{\frac{1}{2}}\} = 19.86$ millimeters. If the mirror had a paraboloidal figure of 100 millimeters focal length (*c*) and the scale were located 1,500 millimeters from the focal plane (*a*), the displacement of the spot of light on the scale (*d*) would amount to $a \times d / c$, or

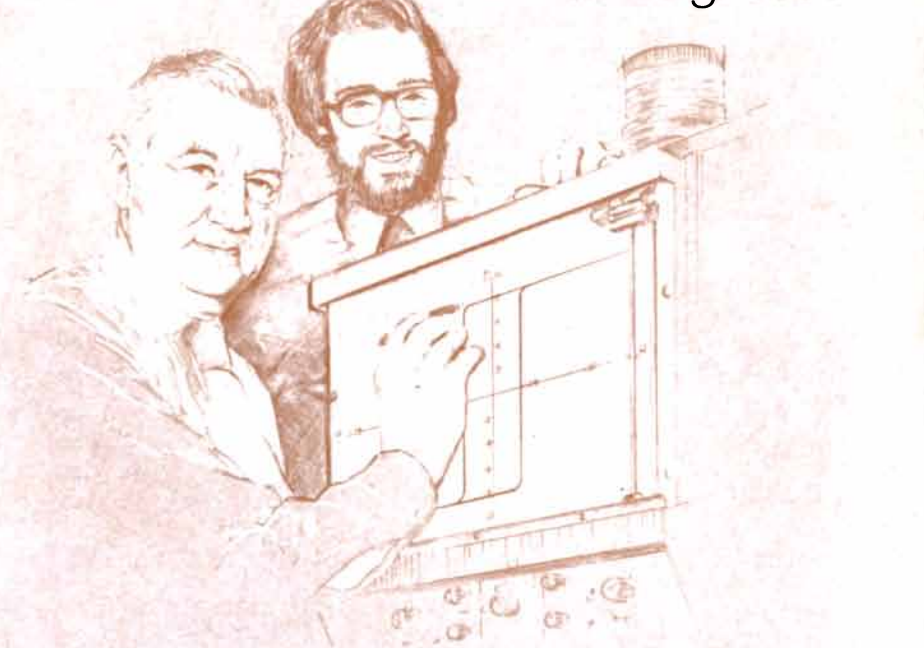
$$\delta = T \sin\phi \left(1 - \frac{\cos\phi}{\sqrt{N^2 - \sin^2\phi}}\right)$$

$$\delta \cong \frac{ad}{c} \;(\text{for } f/\infty \text{ to } f/5)$$



MATERIALS
RESEARCH
CENTER
REPORTS . . .

On the easy magnetization of metallic glasses.



Ribbons and wires made of ferrous metallic glasses can be fully magnetized by applied fields of less than one Oersted and demagnetized by fields of several milli-Oersteds. Indeed the earth's field can magnetize and demagnetize them. These METGLAS^{tm} materials are typically composed of 80% transition metals (Fe, Co, Ni, etc.) and 20% metalloids (P, B, Si, etc.). They are quenched very rapidly from the liquid state (rates of about 10^6 °K/sec) to retain the amorphous liquid structure.

Ribbons of one alloy show the following properties under moderate tensile stress: coercive force = 0.007 Oe; induction at 1 Oe = 7800G; maximum permeability = 10^6; loop squareness = 0.99. For other compositions, the saturation magnetization ranges from 8 to 15,000G. The electrical resistivities are about 3X larger than for crystalline Fe-Ni alloys.

These excellent properties are believed to result from the absence of grain boundaries which usually inhibit domain-wall motions and the absence of crystal anisotropy which usually inhibits domain rotations. Glasses in which magnetostriction is also absent should have nearly infinite permeabilities.

Unlike conventional soft magnetic materials, METGLAS alloys are extremely hard, strong and ductile. Therefore, their magnetic properties are much less sensitive to handling and are maintained after mechanical working.

Possible applications include: current and pulse transformers, magnetic amplifiers, switches and memories, recording heads, transducers, delay lines and magnetic shields.

Allied Chemical Corporation / Materials Research Center
P.O. Box 1021R ■ Morristown, New Jersey 07960.

tm Trademark of Allied Chemical Corporation

Allied Chemical

© 1975 SCIENTIFIC AMERICAN, INC

AN INVITATION TO JOIN THE . . .
AMERICAN LITTORAL SOCIETY

- To learn about what lives in coastal waters, marshes, and estuaries.
- To support coastal zone conservation.
- To receive Society publications — a quarterly journal, newsletters, conservation alerts.
- To take part in Society activities — fish tagging, field trips, seminars, dive/study expeditions.

Annual dues are $10 ($7.50 for students).

AMERICAN LITTORAL SOCIETY
HIGHLANDS, NEW JERSEY 07732

Please enroll me in the American Littoral Society. Enclosed is my check for $________ for first-year dues.

NAME ....
ADDRESS ....
CITY ....
STATE .... ZIP ....

Dues are tax deductible.

Martin Gardner's Sixth Book of Mathematical Games from SCIENTIFIC AMERICAN

Hours of instructive mathematical diversion in Gardner's biggest collection yet—25 columns originally published in *Scientific American*, now updated and expanded to include source references, letters from readers, and answers to all the problems. 183 diagrams in 2 colors. $5.95 paper

At your bookstore or order from:
CHARLES SCRIBNER'S SONS,
Dept. MN
597 Fifth Ave., New York, N.Y. 10017

Please send me______copy(ies) of **Martin Gardner's Sixth Book of Mathematical Games from *SCIENTIFIC AMERICAN*** @ $5.95. I enclose check/money order, which includes 50¢ for handling and postage, plus necessary sales tax.

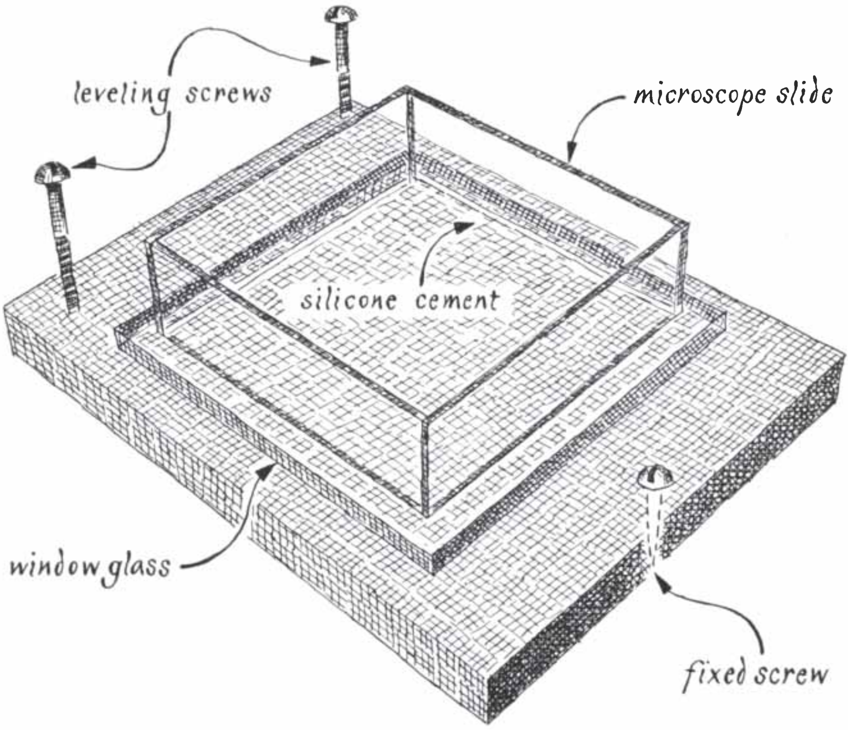
Name ____________________
Address ____________________
City ____________________
State ____________ Zip ________
(This offer expires 12/31/75)

1,500 × 19.86 / 100 = 297.9 millimeters.

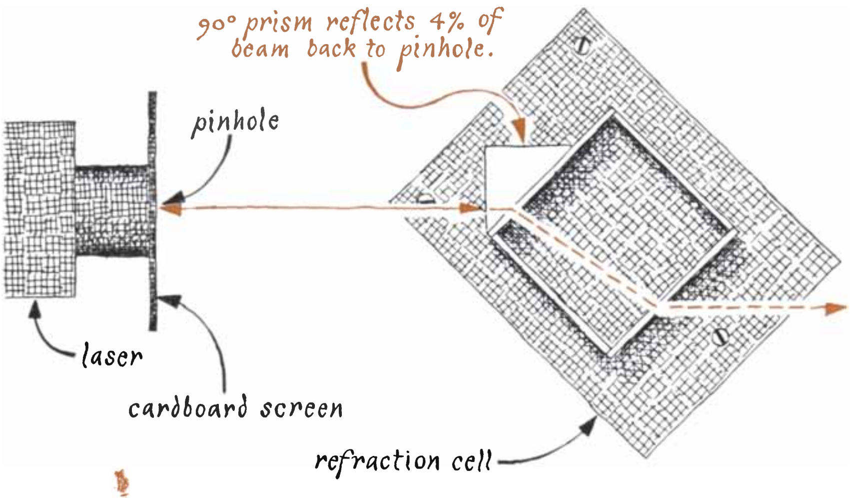
"A beam of light that falls at an angle of 45 degrees on water at a temperature of 20 degrees C. is actually bent about 32.027 degrees at the interface of the air and the water. Our instrument measures the angle in terms of the excursion of the light spot on the scale at the rate of 9.25 millimeters per degree of displacement. This sensitivity enables us to easily measure the angle of refraction to within .1 degree and to determine the refractive index of liquids to within one part in 1,000.

"A helium-neon laser is our light source for aligning the instrument and calibrating it and for measuring indexes of refraction. The output window of the laser is fitted with a short cylindrical cardboard tube that supports at its outer end a white cardboard disk. The disk is pierced with a centered pinhole about half a millimeter in diameter. A distant reflector that picks up the laser beam can be manipulated to reflect the beam back on itself. When the reflected spot of light is centered on the pinhole, the plane of the reflector is perpendicular to the beam.

"To make the displacement cell cut from window glass a 3¼-inch square.



Details of the specimen cell



Prism arrangement for setting the cell

© 1975 SCIENTIFIC AMERICAN, INC

Support the glass on a 3/4-inch wood base measuring five by eight inches. Insert a wood screw through one end of the base at the center to function as a leg. The leg should support the base about half an inch above the workbench.

"Insert somewhat longer wood screws through the corners of the opposite end of the board. These screws serve as adjustable legs for leveling the wood base. Attach the glass square lightly to the center of the base with a dab of silicone-rubber cement.

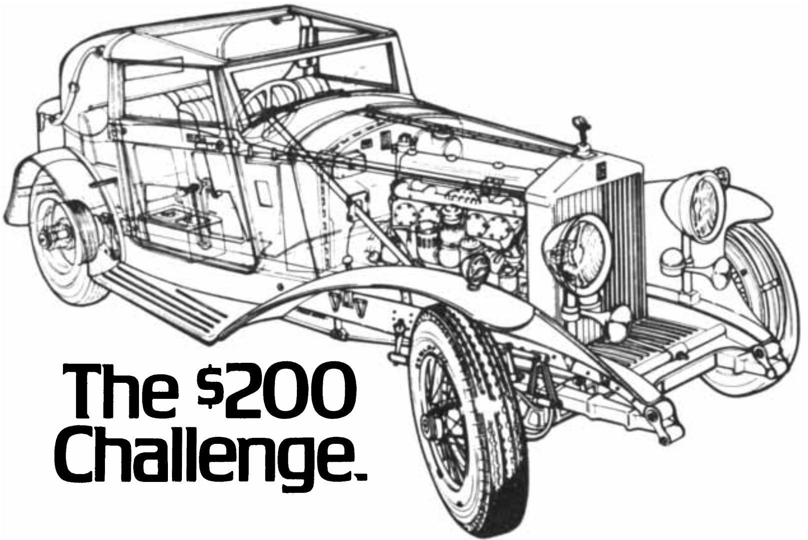
"Direct the beam of the helium-neon laser across the center of the glass square at a height of about 12 millimeters. Apply a few thin dabs of epoxy cement to one long edge of a clean microscope slide. Center the tacky edge of the slide above the glass square about six millimeters from one edge of the square and lower it into contact. Turn the wood base as required to reflect the laser beam from the microscope slide to the cardboard disk of the laser.

"Tilt the slide backward or forward to reflect the beam to the height of the pinhole. Center the beam exactly on the pinhole by manipulating both the base and the slide. Brace the slide in this position by placing one end of a light wood slat on the bench and resting the other end on the upper edge of the glass. Turn off the laser.

"When the epoxy has hardened, add microscope slides to form the sides of the cell. They need not be accurately aligned, but do not disturb the position of the wood base or the aligned slide. Check the alignment of the first side. If the alignment has been disturbed, adjust the base to recenter the beam on the pinhole. Complete the cell by adding and aligning the fourth side.

"Manipulate the slide to center the beam on the pinhole without disturbing the position of the base and the first slide. When both the front and the back slides center the spot on the pinhole, both must be perpendicular to the laser beam and optically parallel. Make the completed cell watertight by coating all joints lightly inside and out with silicone-rubber cement.

"One of the parallel sides must now be positioned at an angle of exactly 45 degrees with respect to the laser beam. To make the adjustment fasten the long side of a 45-90-45-degree prism to one of the parallel sides of the cell with two short pieces of sticky tape. Adjust the beam of the laser to fall on the prism at a height some eight millimeters above the bottom of the cell. Adjust the base of the cell assembly to the position at which one 45-degree facet of the prism



reflects the beam back on the pinhole. The side of the cell now makes a 45-degree angle with the beam [*see lower illustration on page 112*].

"Remove the prism without disturbing the position of the cell. A substantial amount of subsequent labor can be saved by making a metal plate containing three detents that fit the three adjustment screws of the cell assembly. The plate can be aligned, with the adjustment screws in the detents, and screwed to the bench. Thereafter the cell can be removed (for cleaning) and automatically replaced in alignment by returning the tapered ends of the alignment screws to the detents.

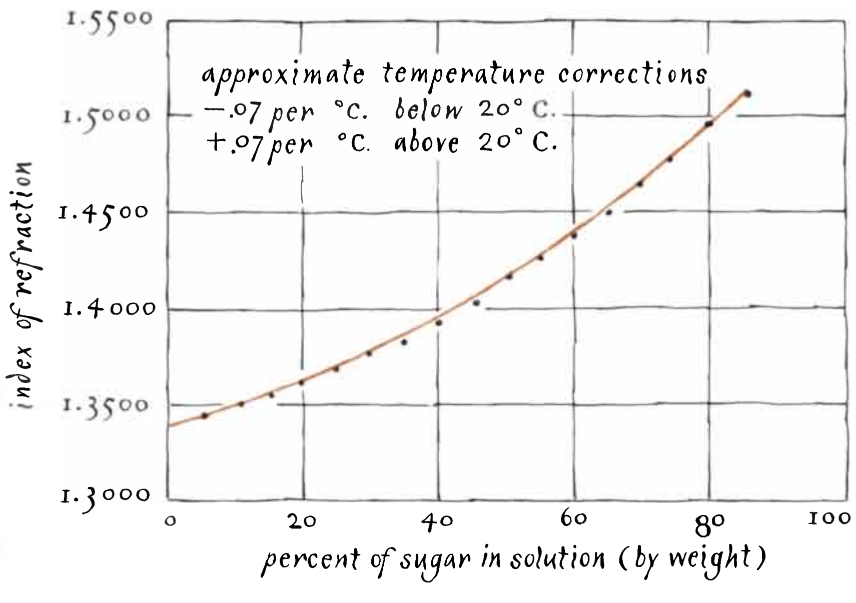
"As we have mentioned, the horizontal displacement of the laser beam by the specimen cell is magnified at the position of the scale by reflection from a concave mirror. Ideally the mirror should be a paraboloid. Parallel rays that fall on all parts of the paraboloid pass through the focal point. For maximum convenience the mirror should have a small ratio of focal length to aperture, say *f*/1, because for a given magnification the distance between the focal plane of the mirror and the scale varies inversely with the *f* ratio. The smaller the *f* ratio, the more compact the instrument.

"On the other hand, the difficulty of making parabolic mirrors varies inversely with the focal ratio. It is all but impossible for an amateur to make a good *f*/1 parabolic mirror. Mirrors of spherical figure not only are easier to make but also are available commercially at modest cost from suppliers such as the Edmund Scientific Co. Rays that fall on the edge of such mirrors, however, are reflected at angles corresponding to significantly shorter focal lengths than rays reflected from zones closer to the center. We accepted this spherical aberration and compensated for it by calibrating the scale of the instrument with a series of test liquids of known refractive index.

"The mirror is mounted on a wood *L* bracket fitted with adjustment screws that resembles the cell bracket in principle. We attached the back of the *f*/1 mirror to a wood disk with silicone-rubber cement. The disk fits tightly into a wood rod 1¼ inches in diameter. The rod is supported by an eyepiece holder of the rack-and-pinion type used in reflecting telescopes.

"The scale, which can be a 450-millimeter length cut from a meterstick, is also supported by a wood *L* bracket, but no screws are needed for making fine adjustments. The scale is fastened to a small sub-base of 1/4-inch plywood that is attached to the main bracket with wood screws. The screws pass through slots in the sub-base. The slots make it possible to adjust the height of the scale through a range of about an inch. The laser beam passes under the 'zero' end of the scale on the way to the mirror. The mirror deflects the beam laterally and also upward at an angle of about one degree.

"To set up the instrument turn on the laser and mark the path of its beam above the bench with two widely separated pushpins. Draw a straight line on the workbench between the pins. With a carpenter's square draw a perpendicular to this line a few inches from the laser. Place the zero end of the scale



Refraction indexes of sugar solutions

above the line and adjust the position of the scale to parallel the perpendicular. If the focal length of the mirror is about 100 millimeters, place the mirror about 1,600 millimeters from the scale.

"With the cell empty switch on the laser. Shift the mirror to the position where its center intercepts the beam. Adjust the angle of the mirror to center the reflected beam on the pinhole. Tilt the beam upward so that it falls on the scale. Adjust the scale sideways to center the beam on the zero graduation. The beam is not displaced significantly by the empty cell.

"Switch off the laser. Measure the interior thickness of the cell to determine *T* as accurately as possible, at least to within .04 millimeter. Fill the cell with clear water at room temperature, preferably at 20 degrees C.

"Switch on the laser. In our instrument water causes the beam to appear upscale about 300 millimeters from the zero mark. The beam should remain fixed. Any movement of the beam may indicate temperature differences in the specimen liquid that give rise to corresponding differences in the index of refraction and even convection currents. The effect can be minimized by working with specimen liquids that are at room temperature.

"The instrument can now be calibrated. We begin with a graduated scale such as a meterstick. To make the calibration we measure a series of increasingly refractive known solutions. An inexpensive series can be made up with cane-sugar syrup in increasing concentration.

"Tabulate the known index of refraction of each test solution and the corresponding excursion of the light beam on the scale. The spot of light that indicates the position of the beam varies in diameter with its distance from the laser. Center the spot exactly on the zero graduation of the scale and thereafter read all upscale excursions with respect to the center of the spot.

"In general, indexes of refraction range from 1.0010 for acetone at zero degrees C. through 1.8033 for a sugar solution of 85 percent at 30 degrees C. Within this range we established 15 calibration points at approximately equal intervals on the scale. On a strip of paper we plotted the calibration points to make a replacement scale divided into increments of refractive index. We cemented the strip to the meterstick to make the instrument read directly. Alternatively, the tabulated data could have been plotted as a calibration graph."

How to select a turntable like an expert, without having to become one.

Selecting a turntable calls for more know-how than any other component. After all, no other component physically handles your largest investment in music: your record.

Unless you know an audio expert, you'll probably depend on an audio salesman for advice. In that case, make sure he knows you want a quality turntable whose tonearm will preserve your records while getting the most out of them.

Next, consider the convenience and safety of an automatic turntable versus handling a manual tonearm. And whether you'll ever want to play two or more records in sequence. (Chances are you'll want a fully-automatic multi-play turntable.)

Be sure to note the workmanship of the turntables that you're considering. Operate the switches and tonearm settings. If they are not precise, record wear will accelerate and the sound will deteriorate.

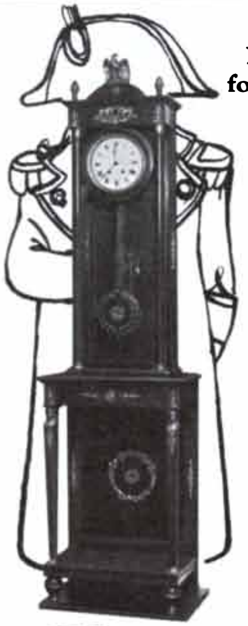
Finally, ask the salesman which turntable he owns. Most audio professionals—record reviewers, audio engineers, hi-fi editors and salespeople own a Dual. And that's all you really need to know to select a turntable like an expert.

The multi-play automatic Dual 1229Q, $269.95; Other multi-play automatics from $139.95. All less base and dust cover. Single-play automatics are the Dual 601, $295, and the electronic direct-drive Dual 701, $400. Both include base and dust cover.

Dual®

United Audio Products
Dept. SA, 120 So. Columbus Ave.,
Mt. Vernon, N.Y. 10553
Exclusive U.S. Distribution
Agency for Dual

The clock in our lobby was a gift from Napoleon to Comte Jean François de Jean in 1806.



But is that any reason for you to stay with us?

We think so. Because your stay in any hotel is made up of a thousand small details. And they all have to be right. That's why we pay such extravagant attention. With real antiques. Fresh flowers. Food and drink worried to perfection. And most important, a staff that knows how to wait on you hand and foot without invading an inch of your privacy. The clock in our lobby is a symbol of that kind of attention to detail.

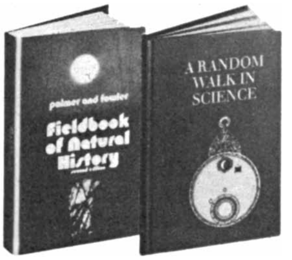
Any hotel can give you the time of day. We give you a perfectly beautiful time.

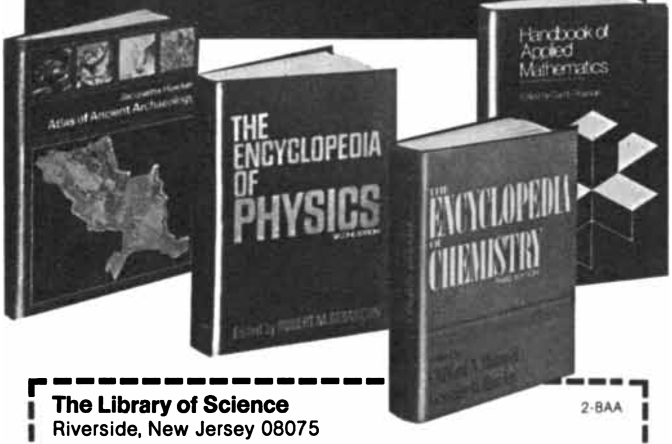
The STANFORD COURT
A Hôtel on San Francisco's Nob Hill (415) 989-3500

For people who understand the subtle differences.

For reservations dial toll free in the U.S. (800) 227-4248, in California (800) 622-0812, or call Quick Reservations or *Hotel Representatives Inc. (212) 838-3110. *H.R.I.—the leading hotels of Europe and World Wide Associates.

© 1975 SCIENTIFIC AMERICAN, INC


palmer and fowler
Fieldbook of Natural History
A RANDOM WALK IN SCIENCE


Atlas of Ancient Archaeology
THE ENCYCLOPEDIA OF PHYSICS
THE ENCYCLOPEDIA OF CHEMISTRY
Handbook of Applied Mathematics

BOOKS

Other conceptions of the search for extraterrestrial intelligence

by Philip Morrison

THE GALACTIC CLUB: INTELLIGENT LIFE IN OUTER SPACE, by Ronald L. Bracewell. W. H. Freeman and Company ($3.95). UFOs EXPLAINED, by Philip J. Klass. Random House ($8.95). The canon of space cartoons is full of wit, but no finer example can be found than the scene, imagined already in the year of Sputnik by Whitney Darrow, Jr., of the decisive moment in Eden mere seconds before the Fall. Eve is reaching for the apple, while the serpent lurks. But is it our Eden? Surely not, for the nude, still innocent pair, otherwise quite our own kind, sport little knobbed forehead antennae. Toward her rushes the astronaut, polite, tainted and knowing, his earthly rocket in the background: "Miss! Oh, Miss! For God's sake, stop!"

Ronald Bracewell is a distinguished radio astronomer and an artist with both brush and pen. Among the three dozen pictures in this delightful book, which is everywhere lighthearted but nowhere frivolous, is one of his own wash paintings, showing the swamp of another world. There are many other evocative outworldly prints and reproductions, including J. J. Grandville's juggler with planets, M. C. Escher's luminous surf where life lies implicit, Wassily Kandinsky and the California contemporary Jesse Allen. In this defining ambience he sets 15 graceful brief chapters (only one runs more than 10 text pages) that survey in an up-to-date, easily read way the issues of that most discussed conjecture of contemporary science: "Somewhere in this galaxy or another I think there is other intelligent life." He urges study, rather than action now, but he plainly looks ahead to a grasped opportunity for interstellar contact. On the way to this reserved conclusion he marshals the evidence without much technical demand on the reader, with a variety of graphs and tables. The usual arguments about planets, stars and life's origin are presented in a personal and engaging way. To them he adds much fresh material. Here is an account of Project Cyclops, the 1971 summer study of the Stanford/NASA/Ames Faculty Fellowship program that produced a valuable report describing an ambitious microwave system—its eventual 10,000 dishes covering an entire desert county—for searching out messages in the galactic noise. Then he explains his own conjecture of a set of automatic probes, each sent to search star systems until it finds a plausible candidate for membership in the Galactic Club. There it orbits, eavesdropping, until finally it can relay back home news of its success.

Two of Bracewell's chapters are explicit critiques, one of Immanuel Velikovsky's interplanetary "vermin," the other of Erich von Däniken's astronaut-gods. To Velikovsky he turns a forbearing cheek: he accepts that the learned man's literary sources admit the possibility of some cometary worldwide catastrophe about 1500 B.C. but is entirely skeptical in detail of the bizarre proposal that a comet brought vermin from Jupiter or Venus to the earth in those days. For the best-selling von Däniken, whose entire output is at a charitable estimate disingenuous, he shows less tolerance. That author denied the old Egyptians rope or wooden rollers to haul their pyramid stones, since their desert oases could not spare such goods! Von Däniken does not even mention the extensive Egyptian trade in cedar timbers from Lebanon or the museum samples of their heavy rope. A Russian author is cited to debunk tellingly the cosmonaut spacesuits seen by von Däniken on Ainu statuettes. Why are these figures mainly women, with protruding bare breasts? (Why should they "pose before the ancient Ainu partially depressurized...?") Those books, Bracewell says, "are a romanticist's fiction."

Bracewell himself has produced no fiction here; his hypothesis may be romantic, but he brings to it what it deserves: skeptical, evidential and measured test. His book is a model for popular science at its margins, tempted by the speculative a little past the stroke of Occam's razor but confident in its eventual ability to approximate the truth. If we are not alone, we should someday be aware of it.

The book by Philip Klass is in an entirely different mood. Here is an informed investigative reporter at work, testing the alleged reports of witnesses and pulling hard link by link at the chains of evidence for the flying saucers. Do you cite the radar-visual reports of pilots and radar men around Lakenheath in England in the mid-1950's? Then be prepared to understand the moving-target indicator of the CPS-5 radar at Lakenheath. You may be chasing spurious blips, and "if an operator decides that a succession of such blips all derive from a single target, then he can easily conclude" that he is watching a UFO that can stop and start instantly and move at impossible speeds. A single false visual report by a fighter pilot who demonstrably did not fully understand his own radar would remove this strange case from mystery into the commonplace puzzle of radar "angels." It is clear that even confirmed radar reports can be no better than our knowledge of radar illusion, just as eyewitness accounts can be no better than our knowledge of optical illusion. No UFO author has taken such a look at radar as Klass, whose decade of experience in the field as an engineer has been augmented by a later career as a knowing technical reporter covering the aerospace industry. (Indeed, his own earlier book seeking a unified physical cause behind many UFO's—plasma discharges on power lines—itself made a case too strongly based on a few erroneous reports and doctored photographs, some of which he mentions here.)

This is a good-sized, meaty, rather contentious work. It treats very successfully many of the classic "sightings" of the past, making a strong *prima facie* case of fraud in several of the best-known. The most recent flap (late in 1973) culminated in the tales of two shipyard workers from Pascagoula who reported that they had been taken aboard a flying saucer for examination by aliens. Within 48 hours they were famous, with coverage on the television networks and worldwide wire services and with their own booking agent. The only evidence was their verbal account of an encounter aboard a saucer with lobster-clawed, floating astronauts. It is hard to see why they were believed by professors from Evanston and Berkeley, who endorsed the unsupported tale with "something here . . . not terrestrial" and "no question . . . these two men have had a very terrifying experience." These two professors, whom it is fair to call friendly to the UFO, were after all not lawyers or even psychologists; their expertise lay in astronomy or in engineering. A new argument for credibility seemed to have emerged when one of the troubled men underwent a polygraph test, which at best is hardly fully probative. This particular test wilts under Klass's scrutiny. It was administered, he shows, not by the customary independent team of licensed and experienced polygraph operators but by "a young, inexperienced, 'uncertified' operator" on the staff of the brother of the agent's friend and former classmate.

It will all happen again: unsupported assertions of marvels seen, interviews by credential-bearing investigators who will accept the tale, a media delight and a subsequent set of talk shows. It ought to be self-evident that such "contact" stories can gain weight mainly from assay by those experienced in the study of testimony, not from astronomers and aerospace experts. It was Klass the investigative reporter, not Klass the engineer, who gave us relevant matter about Pascagoula, exactly in the way he might have done for a Watergate. One should not forget that truth in the next journalists' marvel we learn about.

During the 1973 scare the four-man crew of an Army helicopter in Ohio reported a night near-encounter with a very bright object—at first reddish, then green, then white—moving at high speed, which threatened a direct midair collision, "sucked" their craft upward and temporarily killed their radio. The pilot told his story on the inevitable talk show and the crew collected $5,000 as a prize for the "best UFO case of 1973" (which was not awarded to those Pascagoula passengers, even though two of the judges were the same men who had been so beguiled by their anecdote). The convincing Klass account is: A distant fireball (perhaps one of the Orionids, which would be right in date and direction), seen green through the green-tinted upper canopy of the aircraft, white when seen through the clear plastic below, dimmed to red at a distance. The pilot had instinctively pulled the craft out of his first startled evasive descent maneuver up to a safe altitude. After the light had vanished the crew, in a state of near-shock, found themselves climbing at an altitude of 2,500 feet. The pilot described his actions to Klass from memory; what he recalled doing would indeed have pulled his helicopter upward. The airports he sought to contact as last resorts were beyond his normal radio range, as later tests showed; the one nearby tower had by chance not responded to his hurried call—a common enough occurrence. On later inspection nothing was awry with the helicopter. It was not a fraud but a believable misunderstanding.

There is no more explicit and insightful account of UFO's than this one. Still, as the author reflects in the last lines of his text: "The myth of extraterrestrial visitors will persist . . . because so many people *want* to believe." The reader can profit a great deal, even though the tone is sometimes rather more indignant than seems wise. We owe to the spirit of scientific curiosity a skeptical willingness to listen with patience to every witness, but to set highest of all the close examination of every bit of evidence.

GEARS FROM THE GREEKS: THE ANTIKYTHERA MECHANISM—A CALENDAR COMPUTER FROM CA. 80 B.C., by Derek de Solla Price. Science History Publications ($8.50). In this thin, learned, well-illustrated monograph (first published in the 1974 *Transactions of the American Philosophical Society*) Professor Price recounts a strange tale with a most happy ending. Driven off course by a gale on their way home around Eastertide in 1900, two small oared cutters with sponge fishermen from Syme, an island near Rhodes, dropped anchor in a deep cove on the shore of the uninhabited rocky island of Antikythera, between Crete and the Peloponnesus. When the storm ended, they sent hard-hat divers over the side just where they lay, on the off chance of finding a few sponges. On the bottom, 40 meters down, lay a great wreck, a ship many times larger than their own, laden with amphoras and a treasure of statues of marble and bronze, sea-changed but still recognizable. By 1902 the ship, probably bound from Rhodes to Rome about 2,000 years earlier, had yielded its wealth of images, ever since on proud display in the Greek National Archaeological Museum in Athens.

A heavy formless lump of verdigris lay among the bronze fragments of a statue of Hermes as the Athens restorers tried to fit them together. There were many such pieces, and each was examined again and again to see where it would fit. After months—during which the lump must have dried and cracked open because of remnant wood inside it, this one piece was recognized as some kind of mechanism. In 1958 Professor Price, a student of the ancient history of clockwork and fine instruments, reexamined the relic, which by then was widely known but only meagerly described. The cover of *Scientific American* for June, 1959, bore a photograph of the object, an illustration for his article "An Ancient Greek Computer." That study was tantalizing but largely conjectural; too many parts were lost or still hidden in the corroded mass. Since 1971 Professor Price has revisited the enigmatic and wonderful machine. This time he can almost see through it all, not only with improved insight but also with a fine set of careful radiographs, thanks to his collaborator, Charles Karakalos of the Greek Atomic Energy Commission. The evidence is here, in analyses, old photographs and new, and many radiographs, some carefully marked to outline the gear teeth. (The all-important tooth count was much aided by the meticulous work of Emily Karakalos.) Overall the mechanism has been pretty reliably worked out. There is a capping schematic diagram giving the full logical reconstruction of this ancient analogue computer as far as we have it. It resembled a tall, narrow, rectangular mantel clock in size and form (a Roman foot high, roughly equal to an English foot) with a drive axle on the right side. One turn of the crank drive—which may have been done by hand or by water power—represented a year. It was either set ahead a little day by day or perhaps moved more quickly as a demonstration device. The output dials include a four-year dial and other dials for moon position and sun position among the stars, for the lunar year and for the synodic (moon phase) month. (The moon and sun dials are not quite firmly established.) The fragmentary inscriptions preserved include a text describing on the front face a rather familiar kind of Greek calendar, with pegs set daily by hand, and on the back face a scrap alluding to dials and pointers. The working parts include 40-odd gears—all with excellent 60-degree gear teeth—hand-sawed and filed out of a single sheet of bronze about three square feet in area. The metal is simple tin bronze, with a tiny lead content probably coming from an occasional solder joint, doubtless a repair. This suggests to the experts, Earle R. Caley of Ohio State University and Cyril Stanley Smith of the Massachusetts Institute of Technology, that the low-lead alloy was older than the time of construction—that a plain flat plate of an old-fashioned alloy was held in stock a century or more before it was shaped by the artisans. The best guess is that the mechanism was built under an "anonymous master," an original virtuoso mechanician in the shop of the stoic philosopher Posidonius of Rhodes, where Cicero reported seeing "the orrery recently constructed by our friend Posidonius, which at each revolution reproduces the same motions of the sun, the moon and the five planets that take place in the heavens every day and night." It might have been this very machine!

The most surprising attribute of the mechanism, established beyond doubt in this study, is the presence of an epicyclic differential turntable: a single big gear driven in opposing directions by two meshing smaller gears, each turning at its own rate. The numbers of teeth agree with what is expected for drives at the rate of the sidereal revolutions of the sun and the moon; the big gear then turns with the differential rate—the rate the moon displays when its motion is measured not by the stars but by moon phase, that is, against the shifting sky position of the sun. The next differential gear seen in Europe is found in a famous clock made in Kassel in 1575; the scheme was common for much the same purpose by the 17th century. (Quite recently the University of Groningen physicist André W. Sleeswyk has made a convincing case that the "south-pointing chariot," reinvented—after ancient Chinese accounts—at the Sung court by the imperial engineer Wu Tê-jen, used compound differential idler gears like your car's rear axle about the year 1107. The Chinese device, also a kind of analogue computer, was required to operate in real time, on the scale not of a clock but of a chariot. No connection between the two distinct inventions seems to be implied.)

The Antikythera apparatus displays "the cyclical sequence of . . . discrete phenomena rather than a continuum of events in a flowing time." It is more abstract than the automated sphere of the heavens described vaguely by classical writers and ascribed to Archimedes, which presented all the planetary motions. In his day, before the Alexandrians, those motions were approximated more roughly, and simple gear trains could do an adequate job. After Hipparchus the five planets were seen to have more complex orbits. This sophistication may have restrained the model-builders until the elaborations of the 14th century, which used a couple of hundred gears. The flat astrolabe, using the projection we owe to the Alexandrian school, then set the style of sun-moon-star models.

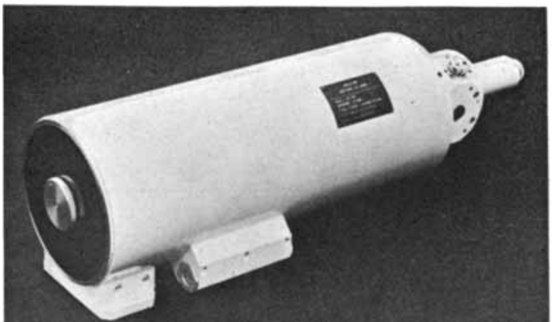
Professor Price accounts for this history of gaps, of brilliant peaks and long declines, by invoking the occasional presence of an innovative mechanical genius who, unlike Archimedes or Leonardo, did not write or paint and hence remains unnoticed by the scholars in their libraries. ("A technological tradition is . . . much more fragile than anything that was encoded into a written book.") The memory of technology lies even now in the artifacts themselves more than in the testimony of bookish fellows who cannot even understand what they see. Who can doubt that the *Pioneer 10* inscription will tell galactic archaeologists less about us than the substance of *Pioneer* itself will?

SCIENTIFIC ANALYSIS ON THE POCKET CALCULATOR, by Jon M. Smith. John Wiley & Sons ($12.95). Within certifiable memory one pushed a kind of motorman's lever attentively to and fro merely to generate an eight-digit product step by step on a whirring, clanking machine, heavy brass on its own wheeled stand. Nowadays a much cheaper wallet-sized integrated-circuit device sits soothingly in the hand and silently displays such a product in glowing red diodes as fast as your fingers fly on the small keys. In this interesting text—not aimed by any means at beginners but available in part to anyone—a software expert presents a practical guide for getting the remarkable most out of whatever modern pocket calculator you own. His exposition, clear if here and there a bit hasty (some errors in sign are to be found), is not tied to any specific make of calculator; he writes for three hypothetical machines, one of each of the three most popular types. The simplest hears only instructions to add, subtract, multiply and divide. Next is the engineering machine, with a four-register memory stack and a modest number of such wired-in functions as logarithms. Finally, the book treats the powerful programmable calculator, which possesses addressable storage. That device—it costs a few hundred dollars—provides "a quantum jump in pocket computing capability" with its software library. The four-function kind now costs an order of magnitude less than the programmables do; "what is amazing is that these small . . . machines . . . can provide tremendous computing power."

The first 50 pages of overview include a neat account of the logical architecture of possible machines. There is no description of any hardware behind the keys on the face of the machine. Most of the machines work with one of two logical designs: algebraic, with data and instructions strung along just as we write them with pencil on paper, or "reverse-Polish" logic, one of the newer schemes (developed by the school of modern logicians around Lukasiewicz) in which operations always follow the data. Smith makes clear the relation between the logical language used and the extent of the memory. Then follow the clever nesting-parenthesis schemes that enable one to quickly compute such quantities as logarithms, trigonometric functions and binomial expansions even on the simplest machines. Much more—Bessel functions and beyond—can be handled by one or another ingenious scheme, often recursive, for which processes, examples and the needed coefficients are given. Smith's aim is to free the analyst from extensive tables; a few file cards can hold the framework to span a wealth of tabulated truths. Extrapolation and interpolation, many statistical calculations and simple iterative means of finding the roots of a polynomial all remain practical on the simplest machines. Naturally the more complex machines transfer many keystroke operations to the chips within. A couple of keystrokes in addition to the number yield an eight-place logarithm on a machine with a chip that knows that function. On a simpler machine the same operation would require tens of strokes to get three- or four-place accuracy.

In the next category are Fourier analysis and integration methods; most of these are not, of course, new to pocket calculators but have been the tools of numerical analysts for some time. Here care has been taken to adapt the routines for direct use on the pocket calculator. Difference equations and powerful low-error polynomial approximations are included. After a systematic account of statistical practice the volume ends with matter for the user of the programmable device, with an entire chapter on its use for many-parameter optimization problems. There is an appendix on tricks of the trade (including a rather satisfying expletive to display for viewing upside down when things go awry). One lesson is well taught: the type of language or size of memory on your machine matters least of all; "what matters most is to begin to use some pocket calculator in advanced analysis." If your interest is in number games or in simple accounting, the book is not at all for you, but if you want numerical results for fairly complicated operations out of an inexpensive pocket device, it is admirably direct.

Raised Relief Topographic Maps, United States Quadrangles, prepared by the U.S. Defense Mapping Agency, Topographic Center. Scale 1 : 250,000. Hubbard Scientific Company, Northbrook, Ill. 60062 ($11.95 each). Plastic Relief Maps, World Quadrangles, Series 1301P, prepared by Army Map Service. Scale 1 : 1,000,000. Defense Mapping Agency, Topographic Center, Washington, D.C. 20315 ($8.75). At the lower right-hand corner the U.S. Naval Reservation at Kittery Point is marked against blue water; more than two feet away, at the other end of the map diagonal, there is the contrasting reservation of the Dartmouth Outing Club below Mount Moosilauke. The New Hampshire countryside is here rendered in the most dramatic way, in accurate relief. Kittery is sea flat, but Dartmouth's mountain lies among the jagged peaks of the southern Whites, full half an inch high. Roads and railroads, built-up sites, sheets of fresh water such as great Lake Winnipesaukee, towns and dams and lookout towers speckle the green forested areas among thin contour tracery on the map. The excellent paper-flat quadrangles of the U.S. Geological Survey have here been combined and generalized to a more modest scale, to present in three dimensions a piece of our nation, mapped in six colors on molded and printed plastic—a sheet only about twice as thick as map paper yet tough enough to hold durably the peaks and valleys of the intricate surface. The threefold vertical exaggeration pays proper tribute to the force of gravity. The scale is well chosen for prolonged attention to a region 100 miles across, say several visits on wheels to different spots and a saunter here and there on foot. The climber or hiker will need his more detailed and foldable maps on flat paper, but for people who are lucky enough to have access to a mountain vista or who want an amateur's overview of a region, this stiff, crinkled map at about four miles to the inch is a resource and a pleasure. (Study the fine structure of dawn and dusk with your desk lamp.) The commercial producers offer a clear order form with a fine index map; their prices are nowadays well above the old days of Government issue, but the service is good and the maps remain a unique possession for any who love some distant countryside or their own. Libraries require their region's map, and one map ought to lie near any window with a mountain view. Three



hundred and three three-dimensional maps are available; they cover nearly the entire country west of the Black Hills, plus the Ozarks, the Upper Peninsula of Michigan and a wide belt along the Appalachians from Birmingham to Bangor. Much of Alaska and the Hawaiian Islands is also available.

The Army Map Service is by no means out of business, although it has had its name changed to the wider, euphemistic form. Now that the U.S. plastic relief maps are on the commercial market the Defense Mapping Agency remains the source of beautiful three-dimensional maps prepared in the same way, but for regions over the entire world. These maps are similar in actual size to the U.S. maps, but they are at the less detailed scale of 1 : 1,000,000, just 10 kilometers to the centimeter. The sheet at hand—mostly blue sea—shows that wonder cone, Fuji-san, towering half an inch high, four inches from flat downtown Tokyo. (The map is too small for all 36 views of Fuji, but fine for half a dozen confirmations of Hokusai.) The map is well printed, with expected detail, but the color tints are related to height and not to vegetation; the effect on land is yellow-brown and autumnal. Only 150 sheets are available; they include Alaska, the Rockies and the Coast Range, central Mexico and the Antilles, Europe and the Mediterranean from Rabat to the North Cape, east to the Caucasus, with the Baltic and Russian plains omitted. Southeast Asia is ready, from Java on to Japan, Korea and Kamchatka, and there are sheets covering New Zealand, South India and a few other scattered places. The Alps would be a treasure, although they occupy several sheets. An order form will be sent free on request; if you enclose five cents and ask for it, an index map for these plastic relief maps of the world will come along.

Paper maps require a strong effort of imagination to approach the thrill of the three-dimensional marvels, but we dare not hold them in low esteem. They too are wonders; these same maps are available much more cheaply in two dimensions. The U.S. maps can be obtained from the Geological Survey in Denver or in Arlington, Va., $1 a sheet for the 1 : 250,000 scale; the 1 : 1,000,000 world maps in flat paper—there are more than three times as many areas published in paper as there are in plastic relief—come from the Defense Mapping Agency. The same order form will work for these; for another five cents you can obtain the index map for the flat maps too, called Series 1301.

INDEX OF ADVERTISERS

MAY 1975

BIBLIOGRAPHY

Readers interested in further reading on the subjects covered by articles in this issue may find the lists below helpful.

EARTHQUAKE PREDICTION

Earthquake Prediction: A Physical Basis. C. H. Scholz, L. R. Sykes and Y. P. Aggarwal in *Science,* Vol. 181, pages 803–810; 1973.

Tilt Precursors before Earthquakes on the San Andreas Fault, California. M. J. S. Johnston and G. E. Mortensen in *Science,* Vol. 186, pages 1031–1033; 1974.

THE MOLECULAR BIOLOGY OF POLIOVIRUS

Structure of the Poliovirus Replicative Intermediate RNA. David Baltimore in *Journal of Molecular Biology,* Vol. 32, No. 2, pages 359–368; March 14, 1968.

Polyadenylic Acid Sequences in the Virion RNA of Poliovirus and Eastern Equine Encephalitis Virus. John A. Armstrong, Mary Edmonds, Hiroshi Nakazato, Bruce A. Phillips and Maurice H. Vaughan in *Science,* Vol. 176, No. 4034, pages 526–528; May 5, 1972.

Polyadenylic Acid at the 3′-Terminus of Poliovirus RNA. Yoshiaki Yogo and Eckard Wimmer in *Proceedings of the National Academy of Sciences of the United States of America,* Vol. 69, No. 7, pages 1877–1882; July, 1972.

Requirement of 3′-Terminal Poly(adenylic acid) for the Infectivity of Poliovirus RNA. Deborah H. Spector and David Baltimore in *Proceedings of the National Academy of Sciences of the United States of America,* Vol. 71, No. 8, pages 2983–2987; August, 1974.

MICROCOMPUTERS

Minicomputers: Hardware, Software and Applications. Edited by J. D. Schoeffler and R. H. Temple. IEEE Press, 1972.

Minicomputers for Engineers and Scientists. Granino A. Korn. McGraw-Hill Book Company, 1973.

The Use of Microprocessors as Automobile on-Board Controllers. R. H. Temple and S. S. Devlin in *Computer,* Vol. 7, No. 8, pages 33–36; August, 1974.

Monolithic Processors. Hermann Schmid in *Computer Design,* Vol. 13, No. 10, pages 87–95; October, 1974.

I^2L Takes Bipolar Integration a Significant Step Forward. Richard L. Horton, Jesse Englade and Gerald McGee in *Electronics,* Vol. 48, No. 3, pages 83–90; February 6, 1975.

RANDOMNESS AND MATHEMATICAL PROOF

Randomness and Gödel's Incompleteness Theorem. Howard De Long. Addison-Wesley, 1970.

Theories of Probability: An Examination of Foundations. Terrence L. Fine. Academic Press, 1973.

Universal Gambling Schemes and the Complexity Measures of Kolmogorov and Chaitin. Thomas M. Cover. Technical Report No. 12, Statistics Department, Stanford University, 1974.

Information-Theoretic Limitations of Formal Systems. Gregory J. Chaitin in *Journal of the Association for Computing Machinery,* Vol. 21, pages 403–424; July, 1974.

THE SOCIAL SYSTEM OF LIONS

The Serengeti Lion: A Study of Predator-Prey Relations. George B. Schaller. University of Chicago Press, 1972.

Lion Population Regulation. B. C. R. Bertram in *East African Wildlife Journal,* Vol. 11, pages 215–225; 1973.

RURAL MARKET NETWORKS

Marketing and Social Structure in Rural China: Parts 1, 2, and 3. G. William Skinner in *The Journal of Asian Studies,* Vol. 24, No. 1, pages 3–43, November, 1964; Vol. 24, No. 2, pages 195–228, February, 1965; Vol. 24, No. 3, pages 363–399, May, 1965.

Central Places in Southern Germany: A Pioneer Work in Theoretical Economic Geography. Walter Christaller. Prentice-Hall, Inc., 1966.

Examining Stratification Systems through Peasant Marketing Arrangements: An Application of Some Models from Economic Geography. Carol A. Smith in *Man,* Vol. 10, in press.

THE SEARCH FOR EXTRATERRESTRIAL INTELLIGENCE

What Shall We Say to Mars? H. W. Nieman and C. Wells Nieman in *Scientific American,* Vol. 122, No. 12, page 298; March 20, 1920.

Project Cyclops. B. M. Oliver et al. National Aeronautics and Space Administration, Publication CR-114445; 1972.

Communication with Extraterrestrial Intelligence. Edited by Carl Sagan. The MIT Press, 1973.

FOREST SUCCESSION

Plant Communities. Rexford F. Daubenmire. Harper & Row, 1968.

The Adaptive Geometry of Trees. Henry S. Horn. Princeton University Press, 1971.

Some Ecological Consequences of a Computer Model of Forest Growth. Daniel B. Botkin, James F. Janak and James R. Wallis in *The Journal of Ecology,* Vol. 60, No. 3, pages 849–872; November, 1972.

The Ecology of Secondary Succession. Henry S. Horn in *Annual Review of Ecology and Systematics,* Vol. 5, pages 25–37; 1974.

Reconstructing Forest History from Live and Dead Plant Material. J. D. Henry and J. M. A. Swan in *Ecology,* Vol. 55, No. 4, pages 772–783; Summer, 1974.

Markovian Properties of Forest Succession. H. S. Horn in *Ecology and Evolution of Communities,* edited by J. M. Diamond and M. L. Cody. Belknap Press of Harvard University Press, 1975.

MATHEMATICAL GAMES

A Polyhedron without Diagonals. Ákos Császár in *Acta Scientiarum Mathematicarum,* Vol. 13, pages 140–142; 1949–1950.

Euler's Formula for Polyhedra and Related Topics. Donald W. Crowe in *Excursions into Mathematics,* edited by Anatole Beck et al. Worth Publishers, 1969.

Steiner Triple Systems, Heawood's Torus Coloring, Császár's Polyhedron, Room Designs, and Bridge Tournaments. Donald W. Crowe in *Delta,* Vol. 3, pages 27–32; Spring, 1972.

THE AMATEUR SCIENTIST

Fundamentals of Optics. Francis A. Jenkins and Harvey E. White. McGraw-Hill Book Company, 1950.

Applied Optics and Optical Design. A. E. Conrady. Dover Publications, Inc., 1960.

TIME
LIFE
BOOKS proudly presents

THE LIFE HISTORY OF THE UNITED STATES

THE DELUXE
BICENTENNIAL EDITION
with Padded Grained
Burgundy Covers

- 12 volumes, each 8½ by 11 inches.
- Over 2,000 pages.
- Some one million words; more than 3,000 pictures, many in full color.

For every American family—a colorful, lavishly illustrated library of two centuries of struggle and triumph

At last the whole American story has been told the way it richly deserves! Not just as a recital of names, dates, battles, and treaties. But as the intensely *human* story of pilgrims, trappers, traders, gunfighters, homesteaders, cowboys, prospectors, doughboys. Conquering the wilderness, building a democratic nation, fighting to preserve it.

The Life History of the United States uses the picture-story techniques for which TIME-LIFE BOOKS are famous. Stunning paintings and photographs, many in full color. Compelling writing. Little-known facts and anecdotes.

■ As a result, the story of America now springs to crackling life. You'll meet such people as:

■ **John Hancock** signing the Declaration of Independence in letters large enough, so legend has it, for the English to read without spectacles.

■ **Our first Chief Justice,** John Marshall, who once defined judicial distinction as "the ability to look a lawyer straight in the eyes for two hours and not hear a damned word he says."

■ **Belva Lockwood,** the first woman to appear on a presidential ballot, who called for "domestic insurrection" to win the vote for women.

■ **Rough Rider Teddy Roosevelt,** whose only fear while charging up San Juan Hill was that he might lose his glasses (so he carried 12 extra pairs).

To make possible this fresh look at the whole American story, the Editors of TIME-LIFE BOOKS called upon outstanding historians to write about the periods they know best. Then the Editors searched museums, libraries, archives, and private collections around the country for some 3,000 photographs, woodcuts and paintings, many in full color, which would best recreate the look and sound and smell of the past.

Examine Vol. 1 FREE for 10 days.

The result is a rich library of text and pictures which belongs in the home of every American family that cares about our heritage as our nation approaches its 200th birthday. And the volumes may be easily collected at a price anyone can afford. See coupon for details—then mail coupon to examine Volume 1 free for 10 days.

TIME
LIFE
BOOKS
TIME-LIFE BOOKS, Dept. BDASC2
Time & Life Building,
Chicago, Illinois 60611

Yes, I would like to examine Volume 1 of the Deluxe Bicentennial Edition of your series *The Life History of the United States.* Please send it to me for 10 days' free examination and enter my subscription to the series. If I decide to keep Volume 1, I will pay $7.95 plus shipping and handling. I then will receive future volumes in the series, shipped a volume at a time approximately every other month Each is $7.95 plus shipping and handling and comes on a 10-day free examination basis. There is no minimum number of books that I must buy. I may cancel my subscription at any time simply by notifying you.

If I do not choose to keep Volume 1, I will return the book within 10 days, my subscription for future volumes will be cancelled, and I will not be under any further obligation.

☐ CHECK HERE IF YOU PREFER REGULAR EDITION. Regular binding with full-color front cover. Only $5.95 per volume.
BDASD0

*Print Name*____________________
*Address*____________________
*City*____________________
*State*________________ *Zip*________

Schools and Libraries: Order Library Style Bindings from Silver Burdett Company, Morristown, N.J. 07960. Eligible for Titles I, II funds.

© 1975 SCIENTIFIC AMERICAN, INC


Now the American dream
doesn't have to become a nightmare.

There could be as many as 20 million of them out there.

They lie rusting on streets and countryside like a pox on the American landscape.

Although they're 85% reusable metal, they've piled up for years.

Why? Simple economics: It often cost more to transport and process a hulk than it was worth.

But now, junked cars are moving faster than ever. Higher demand for steel and growing concern for conservation have made steel scrap look better and better.

Helping tip the scales is something Union Carbide helped develop: the ultra high-power electric furnace.

It can make new steel using 100% scrap. It's cleaner than other furnaces and it uses less energy.

It's also relatively small, so it may be located economically near the source of scrap, reducing transportation costs.

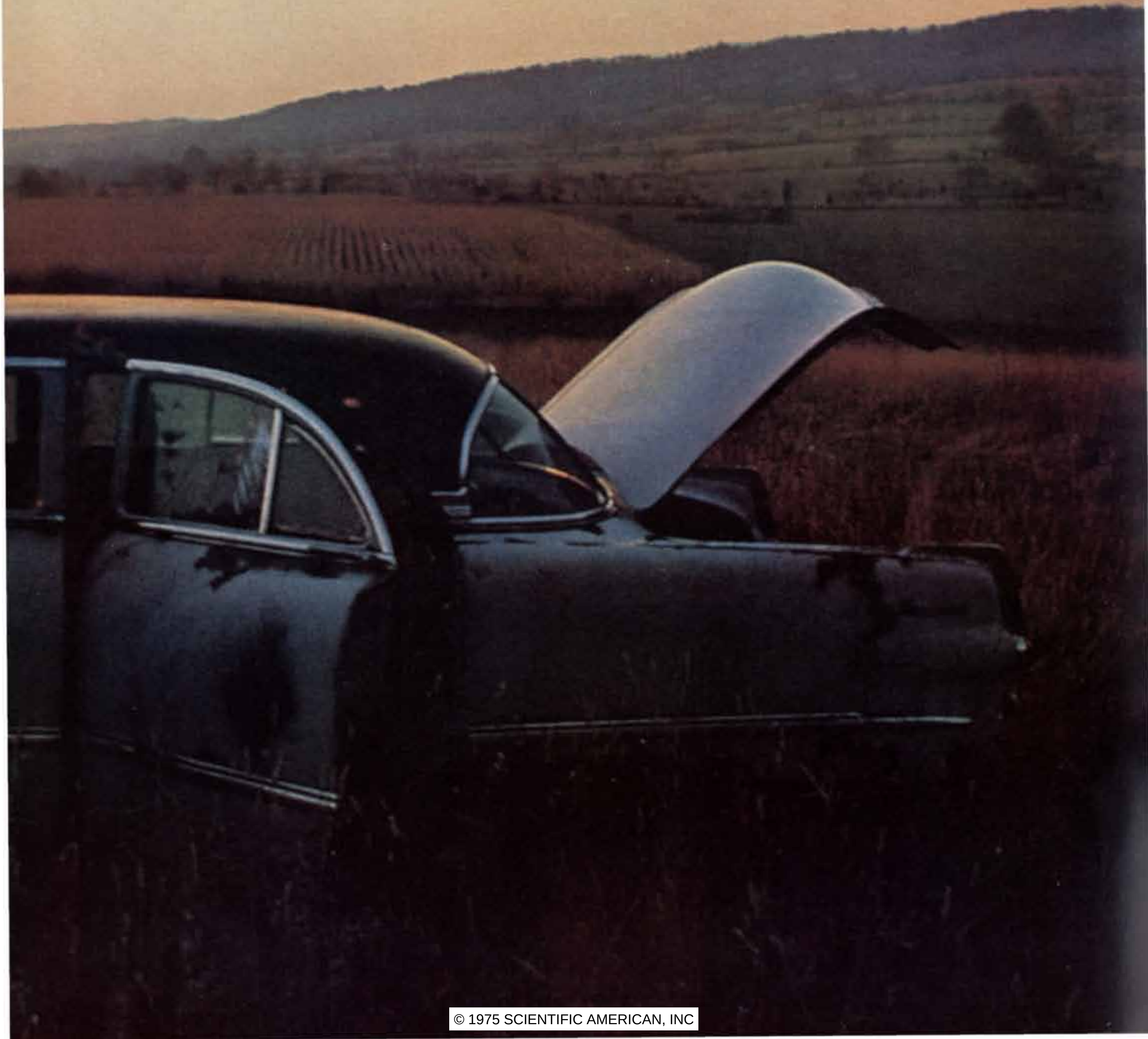
Union Carbide created the electrode that makes possible this versatile furnace.

It's only one of our many contributions to the steel industry.

But it's helping the country get rid of a nightmare.

UNION CARBIDE

Today, something we do will touch your life.




Brussels, 1893
Antwerp, 1894
London, 1888
Zurich, 1930
Paris, 1889
Edinburgh, 1890
"White Label"
DEWAR'S
BLENDED SCOTCH WHISKY
100% SCOTCH WHISKIES
PRODUCT OF SCOTLAND
John Dewar & Sons Ltd
PERTH SCOTLAND
4/5 QUART 86.8° PROOF
Schenley Import Company, New York, N.Y.