SCIENTIFIC AMERICAN

BICYCLE TECHNOLOGY

ONE DOLLAR

March 1973

Introducing the GMC MotorHome. It doesn't ride like a truck. It doesn't look like a box.

1. A NEW ERA IN MOTORHOMES IS BEGINNING.

Before we started building our MotorHome, we studied every other type of motorhome that exists.

We found that the simpler the basic construction, the fewer the problems. So we started with a strong, durable, steel perimeter frame and attached to this a cage of heavy aluminum ribs.

On top of this, we bonded both aluminum and fiberglass panels molded to a smooth finish. It's the same construction people are flying all over the world in.

Except now you'll be driving.

Then we sprayed the interior with a thick, rigid polyurethane foam for thermal insulation and noise suppression. This polyurethane foam has six times the insulation value of fiberglass. Which means the GMC MotorHome has better insulation than most homes. On or off wheels.

2. MORE POWER TO YOU.

To give you the excellent road performance, we installed a 455-cubic-inch V8 engine up front and coupled it to a 3-speed Turbo Hydra-matic transmission. We coupled that to a front wheel drive unit with a 3 to 1 ratio and put it all on top of torsion bar springs and stabilizer bar.

With our low overall body weight, it all means getting up to highway speeds quickly. Excellent traction. Excellent weight distribution.

3. A MOTORHOME IS NO FUN IF IT'S NO FUN TO MOTOR IN.

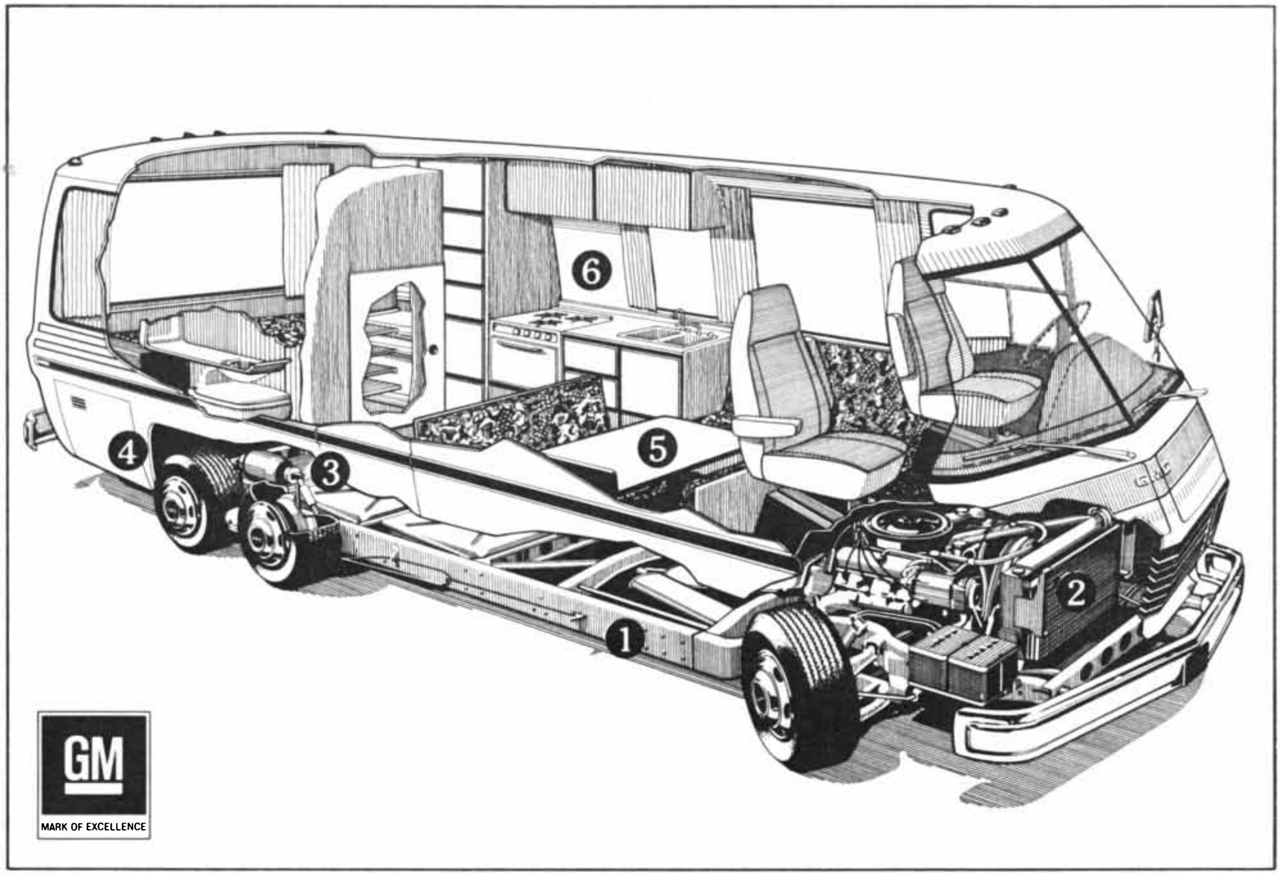
We took our basic construction and raised it only 15" from the ground. This puts the center of gravity only 37 inches above the ground. For easy handling.

See the rear wheels. We put one behind the other for four reasons: To give you a wider base. More room inside. Greater stability than you'd have with dual wheels. And so we could place a special air spring between the two wheels to pass the bumps from one to the other instead of to you!

These air springs are the only ones of their kind on motorhomes.

To keep the weight and balance of the interior within design limits, we fed all the data into a computer. It fed back what we needed to put things where they belong.

4. ABOUT OUR SIX-WHEEL BRAKING SYSTEM.



In addition to power steering, there's a six-wheel braking system with power disc brakes up front and four large finned-drum power brakes in the rear, plus an available leveling device operated from the driver's compartment for parking on uneven ground.

Incidentally, the parking brake grabs all four rear wheels.

5. CHOOSE FROM 15 DIFFERENT FLOOR PLANS. TWO LENGTHS.

The GMC MotorHome is available in 23- and 26-foot lengths. The standard 26' floor plan includes a dinette that converts to a double bed opposite a sofa that turns into double bunks. In the middle, a double sink, 6-cubic-foot refrigerator (it's electric so there's no pilot light that'll blow out), a range and oven with exhaust hood. There's also a bath with all the necessities plus ample cabinet space.

That's one floor plan. There are 14 more available.

6. WE INCLUDED TOP INTERIOR DESIGNERS IN OUR PLANS.

To put the finishing touches on the inside, we had *House and Garden* magazine's interior designers help us.

The driver and passenger seats are high, contoured seats with built-in arm rests. This high-level seating arrangement, combined with the big, wide-angle windshield, offers you panoramic visibility.

Every counter top has rounded corners. All cabinet knobs are eliminated. Every hinge is concealed.

There are thick, shag or cut pile carpets. And wood-grained vinyl on the walls and cabinets.

You also get a choice of four color-coordinated interior decors.

7. ONE-STOP, ROUND THE CLOCK SERVICE.

Your GMC MotorHome dealer services everything he sells. Inside and out. From the engine to the air conditioner and furnace. And there's a toll-free number you can call and immediately get the number of the nearest MotorHome dealer representative available for assistance around the clock.

For our 28-page, four-color catalog, write GMC MotorHome Headquarters, Drawer Y, Dept. 120, Lansing, Mich. 48909.

Better yet, see your GMC MotorHome dealer. He'll be glad to show you around the house. Have a good life.

GMC

The MotorHome from General Motors.

Meet the creator.

Today, the musical artist has a new instrument at his command—the recording studio. It's an instrument that can capture sound, manipulate and mold it, stack it and scramble it, equalize and echo it—a contemporary creative tool with possibilities confined only by the borders of imagination.

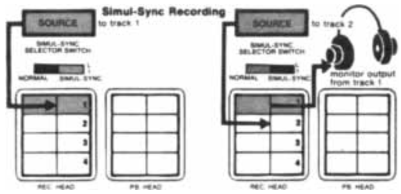
Some might call this musical soundfoolery, an adulteration of the pure musical art form. But throughout history, the truly creative artist has always used whatever instruments were available to reproduce the music he heard in his mind. The artist is no different today—but the instruments he uses are. And this has resulted in a dynamic new range of musical experiences for us all.

The creator—a 4-channel studio that fits on a shelf

With the needs of the contemporary artist in mind, TEAC tape technologists set out to design a precision musical instrument that would provide studio electronic flexibility and studio performance accuracy—yet be compact enough for home use and priced within the bounds of reason. The result: the creator, TEAC's amazing Model 3340 4-Channel Simul-Sync® Tape Deck—a recording studio that fits on a shelf.

The 3340, backed by TEAC's exclusive two-year Warranty of Confidence,* is carefully crafted in the TEAC tradition of professional quality. 10½" studio reels; a quick and gentle three-motor transport; four studio-calibrated VU meters; eight input controls for complete mic/line mixing; dual bias selection; 7½ and 15 ips studio-accurate speeds. And Simul-Sync.

Simul-Sync: what it does and how it works



Overdubbing has become a familiar term to every knowledgeable musician. Simply, it means a) recording a voice or instrument on one track of a multi-track tape machine, b) adding another voice or instrument to a different track at a different time, and c) matching the two tracks so it sounds as if they were recorded simultaneously when played back. To overdub properly, the artist recording on the second track has to listen to the material recorded on the first track while performing in perfect synchronization to it.

That's where the problem occurs with most tape recorders. Conventional record/playback monitoring systems only let you listen to the previously recorded material off the playback head. That means a time delay between the track being recorded and the track being monitored. A small delay, to be sure, but large enough to make perfect synchronization virtually impossible.

TEAC engineers solved the problem with Simul-Sync. They designed a studio-tolerance 4-channel record head, then added electronics that allow each track on that head to be switched independently to either record or playback modes. By doing so, they completely eliminated the time lag and permitted the artist to add track after track—all in absolute synchronization with each other.

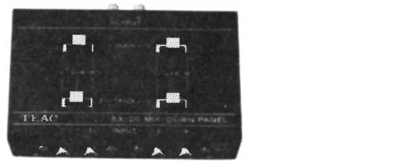
The TEAC Simul-Sync head, operating in conjunction with a 4-channel erase head and a hyperbolic playback head capable of reproducing either stereo or 4-channel material, served as the foundation for the TEAC 3340 concept. It also opened the door to a whole new realm of musical creativity and enjoyment.

Exploring the realm

Here are just a few of the sonic effects possible with the TEAC 3340:

1. Unlimited overdubbing. Up to nine individual instruments or voices can be recorded at different times without any track being used beyond second generation. Sensational signal-to-noise ratio is the result.

2. Professional quality mixdown. All four channels to a single track or ¼-track stereo masters. The optional TEAC AX-20 Mixdown Panel makes it a quick and easy process. Individual controls also allow for desired mixing level for each channel.



3. Special effects. Enter the world of psycho-acoustic phenomenon where the creative juices can really start flowing. Things like echo, cross echo, 4-channel rotating echo and pan pot effects (with AX-20 Mixdown Panels). You can put echo on some instruments and not on others. One-man group arrangements, with a single artist playing all instruments and singing all vocal parts. Backwards recording, an effect that gives any instrument a totally new sound. Dual speed recording, mixed down in perfect sync. With all these effects at his disposal, the professional musician can quickly save the cost of a 3340 in reduced studio experimental time alone.

4. Pseudo-quad recording through ambient delay to the rear channels. And, of course, full discrete 4-channel record and playback.

5. Are you creatively curious? If so drop TEAC a line, and ask for the "Meet the creator" booklet. It describes all of the 3340 effects in detail and explains how each is done. And it's free.

If creative involvement is what you're after, meet the creator—the TEAC 3340 4-Channel Simul-Sync Tape Deck. (or the 7" reel, 3¾ – 7½ ips version, the 2340).

When it comes to creative recording, they perform miracles.

*TEAC or one of its authorized service stations will make all necessary repairs to any TEAC tape deck resulting from defects in workmanship or material for two full years from the date of purchase, free of charge to the original purchaser. This warranty applies only to TEAC products sold in the United States.

The TEAC 2340 and 3340 are priced at $759.50 and $849.50, respectively. All prices subject to change without notice. For complete information, please write to TEAC, Dept. A-1, 7733 Telegraph Road, Montebello, California 90640. In Canada: White Electronic Development Corp., Ltd., Toronto. TEAC Corporation, 1-8-1 Nishi-shinjuku-ku, Shinjuku, Tokyo, Japan. TEAC EUROPE N.V., Kabelweg 45-47, Amsterdam—W.2, Holland. Hi-Fi, S.A. Alta Fidelidad Hidalgo 1679, Guadalajara, Jal., Mexico.

TEAC®

The sound of a new generation

Established 1845 **SCIENTIFIC AMERICAN** March 1973 Volume 228 Number 3

ARTICLES

DEPARTMENTS

THE COVER

The painting on the cover shows several of the key components of a modern racing bicycle (see "Bicycle Technology," page 81). This particular model employs a rim brake of the caliper type (*top right*) and a 10-speed gear system of the *dérailleur* type (*center*). The brake is operated by a cable in which an inner flexible wire in tension is contained within a flexible tube in compression; squeezing a lever mounted on the handlebar of the bicycle causes the brake to press on the flat outer sides of the narrow rim. This system, the first version of which was introduced by E. M. Bowden in 1896, provides a highly effective means for the remote operation of a mechanism such as a brake. The gear system consists of two large sprocket wheels at the driving, or pedal, end connected by a standard bush-roller chain to five small sprocket wheels at the driven, or hub, end. The gear ratio is changed by moving a hand-operated lever, which in turn operates (through another Bowden cable) a mechanism that transfers the chain from one sprocket to another. The necessary variation in the length of the chain on the tight side of the drive is accommodated by means of a spring-loaded jockey pulley on the slack side. The *dérailleur* gear system, which made its appearance in 1899, is light and efficient but needs careful adjustment and lubrication.

THE ILLUSTRATIONS

Cover painting by Ted Lodigensky

Page	Source
16–23	Ilil Arbel
25–33	Allen Beechel
35	O. L. Miller, Jr., Oak Ridge National Laboratory (*top*); Enid Kotschnig (*bottom*)
36	Enid Kotschnig
37	O. L. Miller, Jr., Oak Ridge National Laboratory
38–39	O. L. Miller, Jr., Oak Ridge National Laboratory (*top*); Enid Kotschnig (*bottom*)
40	Enid Kotschnig (*top*), O. L. Miller, Jr., Oak Ridge National Laboratory (*bottom*)
41	O. L. Miller, Jr., Oak Ridge National Laboratory
42	O. L. Miller, Jr., Oak Ridge National Laboratory (*left*); Enid Kotschnig (*right*)
50–52	Lick Observatory
53–68	Graphic Presentation Services, Inc.
71	Milford Small
72–76	Bunji Tagawa
77	Milford Small
78	Bunji Tagawa
81	Science Museum, London
82	Bettmann Archive, Inc. (*top*), Science Museum, London (*bottom*)
83–86	Science Museum, London
87	Bettmann Archive, Inc. (*top*), Mercedes-Benz of North America, Inc. (*bottom*)
88–90	Dan Todd
91	Oxford Mail & Times, Ltd.
92–96	Tom Prentiss
101–108	Eric Mose
110–113	Ilil Arbel
114–116	Roger Hayward
117	Roger Hayward (*left*), Harry L. Latterman (*right*)
118	Roger Hayward

Anatomy of a Nikon Camera (Part II)

Recently, we discussed the precision with which a Nikon is put together to make it the superb picture-taking instrument it is. But there is still another aspect of the way Nikon is put together that makes it the most versatile camera ever conceived.

We're now talking about the literally thousands of ways you can put a Nikon together yourself by way of its totally flexible, modular construction, on which the Nikon system is based.

Interchangeable Lenses

There are more than forty lenses in the Nikon system. Fourteen of them are wide angle lenses, for use at close quarters or to exaggerate perspective, including a 220° Fisheye Nikkor that actually "sees" objects behind the camera. To bring distant objects up close, you can choose any of fifteen telephotos, ranging up to 2000mm and giving up to 40X magnification compared with the "normal" 50mm lens. There are high-speed lenses for use in the dimmest available light and special lenses for extreme closeup work—even one that corrects linear distortion in architectural photos. You'll find all of these and more in the Nikon system—all sharing the exceptional quality, the surpassing sharpness and the magnificent color rendition that make Nikkor lenses the overwhelming choice of professionals.

Interchangeable Finders

Because a single lens reflex camera is built around the finder, Nikon pays particular attention to this aspect. It starts with the only 100% accurate viewing system in all of 35mm. That's right; the Nikon F and F2 are the only cameras whose finders show precisely the picture area appearing on the film—no more and no less. To implement this unique accuracy, there are five interchangeable finders for the Nikon F and six for the Nikon F2. Brilliant eyelevel prism finders. Photomic thru-the-lens meter/finder

systems, waist-level and magnifying finders — even one for viewing accurately while wearing goggles or with the camera in an underwater housing.

Interchangeable Finder Screens

The viewing screen is where the finder image is formed by the camera lens and the finder's optical system. In most cameras it's fixed; a few others accept a handful of interchangeable screens. There are no fewer than eighteen in the Nikon system, each offering practical advantages for some specific application. Even if you never use more than one, you can choose among seven universal types to suit your individual preference.

Interchangeable Backs

This feature is mainly for the professionals, who use more Nikons than any other 35mm camera. If you happen to be one, you may want to replace the Nikon back with one holding 250 (or even 800) exposure film loads for use with a motor drive.

Or, you might attach a Speed Magny back and shoot large-format Polaroid pictures (that's right, with a Nikon!).

Using just the parts discussed so far, you can come up with a staggering 21,600 different combinations for your Nikon camera. And that doesn't include closeup equipment, filters and dozens of other eminently practical accessories presently in the Nikon system, and new ones constantly being developed.

Of course, no one photographer will ever need all of it. The point is, no matter how exotic or far-fetched the situation may be, you're sure to find the equipment you need in the Nikon system.

That's versatility, Nikon style. To find out more about it and how it can work for you, visit your Nikon dealer. Ask him also about the Nikon School of Photography. Or write for detailed literature folio 19. Nikon Inc., Dept. SA, Garden City, N.Y. 11530. Subsidiary of Ehrenreich Photo-Optical Industries, Inc. EPOI (Canada: Anglophoto Ltd., Montreal 376, P.Q.)

It's the HP-80 Business Calculator.

It used to require up to 20 minutes to solve such time and money problems as bond calculations, figuring compound interest, loan repayments, depreciation amortization, investment analysis, sinking fund and statistics. To make matters worse they required constant reference to cumbersome tables. You'll notice this is all in the past tense.

With the advent of the new HP-80, the financial counterpart of the pocket-sized HP-35, these problems can be solved literally in seconds, and without reference to any tables.

The 9-ounce HP-80 has 40 specific capabilities built into its internal circuitry, which combine to perform more than 100 different financial calculations involving a relationship between time and money. All necessary programs, including a 200-year calendar, are built into the HP-80's solid-state memory.

You just enter the data, push a few keys, and read the answer — accurate to within the last penny on a million-dollar transaction.

Since it can perform most financial calculations in about 10 seconds, a user can save up to 20 minutes on a single problem. The price is $395 (*U.S. only, plus tax).

If you're in business, you can't afford to be without it. Ask for full information.

How to live happily in an increasingly digital world.

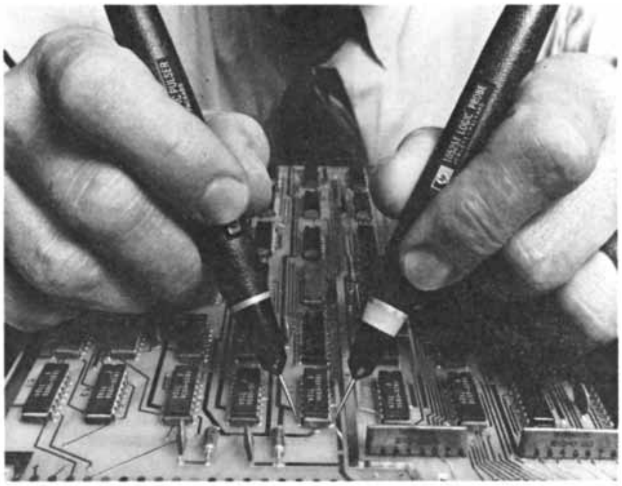
Most electronic products have changed drastically in the last few years. Everywhere you look, digital logic is displacing the familiar analog circuitry. And you'll be deceived if you believe that trouble-shooting these new products is the same as it was before the advent of the ubiquitous integrated circuit (IC). There are at least two important differences.

In the first place, you can easily check if any of the discrete components in an analog circuit is functioning properly by testing its input and output leads, usually two or, at worst, three. In a digital circuit, on the other hand, each IC has 14 or 16 leads and you may need to know what's going on at all the leads simultaneously.

Furthermore digital circuits are intentionally designed with low output impedances to make them insensitive to noise . . . perhaps their greatest single virtue. Unfortunately for circuit troubleshooters, this means that each IC input is clamped in either a high or a low logic state; hence the only way to perform a stimulus/response test on an IC has been to unsolder its input lead before injecting a test pulse and observing the output.

We're happy to report that things have changed for the better in the world of digital logic testing, with HP's IC Troubleshooters.

First there's the Logic Clip which instantaneously indicates the logic state of all 14 or 16 nodes on an IC without cables or power connections. Or the Logic



Comparator which compares an in-circuit test IC with a duplicate IC known to be good and identifies any dissimilarities.

For testing individual IC nodes, there is a series of Logic Probes covering the various IC families which immediately indicate logic states and pulse activity. New partner to the Logic Probes is the Logic Pulser which automatically injects a single pulse that drives a logic-low node high and a logic-high node low for 300 nsec. These are an effective stimulus/response test team — without unsoldering leads and despite low IC output impedances.

We've told the full story in an easy-to-read booklet that's yours for the asking.

For more information on the products described in these pages, fill out the coupon or write to: Hewlett-Packard, 1508 Page Mill Road, Palo Alto, Calif. 94304.

HEWLETT hp PACKARD

Sales, service and support in 172 centers in 65 countries.

00303A

Please send me information on the following:
( ) HP-80 Business Calculator
( ) IC Troubleshooters

Name ____________________

Title ____________________

Company ____________________

Address ____________________

City ____________ State ____________ Zip ________

LETTERS

Sirs:

Gunther Stent's thesis that Avery's transformation studies and Mendel's experiments were premature may be correct, but the evidence presented for this is questionable ["Prematurity and Uniqueness in Scientific Discovery," by Gunther S. Stent; SCIENTIFIC AMERICAN, December, 1972]. Muller, for example, in 1946 (*Proceedings of the Royal Society*, Series B, Volume 134, page 23), referring to Avery's work with the pneumococcus, claimed that the substances constituting the transforming principle "were, in effect, still viable bacterial 'chromosomes' or parts of chromosomes floating in the medium used. These might, in my opinion, have penetrated the capsuleless bacteria and in part at least taken root there, perhaps after having done a kind of crossing over with the chromosome of the host. In view of the transfer of only a part of the genetic material at a time . . . a method appears to be provided whereby the gene constitution of these forms can be analyzed, much as in the cross breeding tests on higher organisms. However, unlike what has so far been possible in higher organisms, viable chromosome threads could also be obtained from these lower forms for *in vitro* observation, chemical analysis, and determination of the genetic effects of treatment."

If Muller fully recognized the canonical implications of Avery's work when it was first presented, why then the reluctance of biologists until 1952 to accept DNA as a genetic material? In Muller's case the answer seems clear; he was not a biochemist and he believed Mirsky and other biochemists who claimed that "the extracted 'transforming agent' may really have had its genetic proteins still tightly bound to the polymerized nucleic acid." Hershey and Chase did away with this by showing that the *entire* virus could be constructed from the introduction of the viral nucleic acid.

At a lecture at the University of California at Los Angeles about 1966 I asked MacLeod, of Avery's team, if he had also considered Muller's interpretation at that time as the most likely mechanism for transformation. He pointed out that there were so many ways of interpreting the results (directed mutations, plasmagenes, inducers, repressors) that he, Avery and McCarty carefully avoided such speculation in these publications for fear of distracting attention from their key discovery—the DNA composition of transforming principle. If Avery's team did not see or push the implications of their own work, one can hardly criticize geneticists for missing out on a nonexistent bandwagon.

In Mendel's case Stent's canonical argument is also questionable. Mendel's laws of heredity were rejected not only by his colleagues such as Nägeli but by himself! When Nägeli correctly pointed out to Mendel that hawkweed (*Hieracium*) did not give Mendelian ratios, Mendel repeated Nägeli's work with seeds supplied by him and to his disappointment found that they did not Mendelize. Neither Mendel nor Nägeli knew that the pollen tube stimulates the hawkweed ovary to develop parthenogenetically. The pollen nuclei do not fertilize the ovule! Thus it was bad luck that Nägeli, the leading student of heredity, chose an unrepresentative form to launch his experiments in heredity.

Similarly, in 1900 the rediscovery and acceptance of Mendel's laws cannot be attributed, as Stent claims, to the chromosome theory. Morgan himself rejected attempts by Wilson and Sutton to force Mendelism into the chromosome theory and it was not until 1911 that he became converted. (His first paper on white eyes and sex-linked heredity carefully avoided any association of the phenomenon with the X chromosome!) Bateson, who gave genetics its name and who did so much to publicize Mendelism, was unconvinced by the chromosome theory. So too was Johannsen, who named the gene and who clarified the distinction between genetype and phenotype.

ELOF AXEL CARLSON

State University of New York
Stony Brook

Sirs:

I was delighted by Mr. Stent's article in your December issue, which demonstrates so convincingly that scientific research and artistic fabrication have much more in common than most people suppose. There are, however, certain differences between them to which I should like to draw the attention of your readers.

Every good work of art exhibits two qualities, Nowness and Permanence. Thanks to the former, it is possible for an art historian to give at least an approximate date for its making; thanks to the latter, it remains "on hand" in the world long after its maker and his society have disappeared. This means that, in the history of Art, there is Change but no Progress. Mozart does not supersede Monteverdi in the way that I suppose one must say that the Copernican picture of the universe superseded the Ptolemaic. It is only bad art which is superseded—by another kind of bad art. As we all know, much of the poetry, fiction, music, painting, etc., produced in any period is bad, though it may, temporarily, enjoy great popular success. Scientists are in the fortunate position of being judged by their peers, not by the general public. I should be curious to learn if it is possible for a scientist to say: "Much of the research being done at this time is worthless."

As for Prematurity, I can see that it is possible to call Mendel's work premature because, by the time it became known, genetics had already advanced beyond him. One cannot say that Cézanne's paintings were premature because, since they offended contemporary canonical taste, it was some time before their value was appreciated, since, as I said, in art there is no progress, whereas scientific discoveries are of permanent interest only to the historian of Science, not to scientists qua scientists.

W. H. AUDEN

Christ Church
Oxford

Sirs:

Evidently I failed to convey to Elof Axel Carlson the essence of my considerations on prematurity in discovery, or indeed the intent of my essay. First, far from meaning to "criticize geneticists for missing out on a nonexistent bandwagon," I tried to show just why scientists must not be blamed for not appreciating the premature. Second, Carlson's citation from Muller and his explanation of why Muller did not accept DNA as the genetic material, far from showing that the evidence I presented for my thesis is "questionable," supports my claim that Muller did not (and could not) appreciate the significance of Avery's discovery. For if in 1946 Muller really thought that the transforming principle represents "viable bacterial 'chromosomes,'" then it is all the more remarkable that in his 1950 essay on the nature of the gene he makes no reference to DNA. In other words, Muller's 1950 view of the gene would not have been very different after all had Avery's discovery not been made six years earlier.

Moreover, Carlson's explanation for the immediate impact of Hershey and Chase's findings in 1952 reveals his lack of familiarity with the details of that classic experiment. Since those findings allowed that as much as 20 percent of the protein of the virus might have entered the host cell, there was actually much more reason to suspect here that the "genetic proteins [are] still tightly bound to the polymerized nucleic acid" than there was in the case of Avery's purified transforming principle.

Finally, Carlson's comments on Mendel's case also show his failure to understand my argument. Thus, Nägeli's managing to talk Mendel out of his own conclusions merely shows that a discovery may turn out to be premature even for the discoverer himself. And for an account of the role of the chromosome theory in the rediscovery of Mendel's work I refer Carlson to Sturtevant's *A History of Genetics.* According to Sturtevant, it was the controversy stirred up by Weismann's chromosomal theory of heredity and development that led both De Vries and Correns to the design of Mendelian experiments and their eventual rediscovery of Mendel's paper. Furthermore, Sturtevant wrote that his teacher Morgan not only interpreted the discovery of sex-linked heredity from the very first in terms of the association of genes with the X chromosome but also as early as 1910 asked the question: "Is there recombination between genes that lie on the same pair of chromosomes?" In any case, in order for a theory to connect the implications of a discovery with canonical knowledge, it need only be *credible,* without necessarily being universally acceptable.

It is not without some trepidation that I dispute W. H. Auden's comments. First, contrary to Auden, I think that historical permanence and supersession are not necessarily mutually exclusive attributes of a creative work, and hence that they can hardly serve as criteria for distinguishing "good" art from "bad." One good work may live on through another that has superseded it. For instance, I do think that Monteverdi's *Orfeo* has been superseded by Mozart's *Don Giovanni* in the same sense that Mendel's paper has been superseded by the modern genetic literature; the earlier works, though they remain "on hand," are much more rarely heard (or read) than the successor works to which they eventually gave rise. Monteverdi lives on through Mozart, just as Mendel lives on through Watson and Crick. Naturally, by supersession I did *not* have in mind a case such as the abandonment of the Ptolemaic in favor of the Copernican theory. Here antithesis replaced thesis, a dialectical process that can occur only in science, which has resort to two-valued logic, and not in art, which is free of that encumbrance to the creative spirit. I would suggest that the criterion of badness in art, as in science, is not supersession but triviality. Thanks to that criterion nothing is easier for a scientist than to say, "Much of the research being done at this time is worthless." As Wolfgang Pauli once said of the work of another physicist, "It is not even wrong."

Second, I hold fast to the view that "progress" *is* a feature of the history of art, no more and no less than of the history of science. I can do so on the basis of the premise I stated in my article, namely that "a creative act on the part of either an artist or a scientist [means] his formulation of a new meaningful statement about the world." Hence there was progress in art and science as long as meaningful statements were being added to the accumulated capital of our cultural heritage. As I tried to show in my book *The Coming of the Golden Age* it is only in our time that in both domains this process of addition is coming to an end.

Third, Auden's remarks about Mendel and Cézanne suggest that he too did not quite fathom the prematurity concept I tried to develop. When Mendel's work finally became known, it was not because by that time "genetics had already advanced beyond him" but because meanwhile his experimental procedure and his inference of discrete hereditary units had become connectable with canonical biology. By the same token Cézanne can be said to have been premature, not because he offended contemporary taste but because his novel way of representing visual space was not connectable with contemporary artistic canon. And since the artist's audience depends on its knowledge of the artist's own canon for understanding the semantic content of his work, Cézanne's day came only when Western artistic sophistication had progressed to the point where geometrical formalization of natural shapes and departure from focal perspective could be appreciated. Finally, the reason even great scientific discoveries are of permanent interest only to the historian of science is, as I tried to explain in my article, that these discoveries can be readily paraphrased later by lesser men. Since in art it takes a greater man to paraphrase, it is mainly the lesser works of bygone times that become the exclusive interest of the historians of art.

I want to take this opportunity to thank the many other correspondents who have commented on my article. The sharp criticism of many of these letters has helped me to bring my ideas into better focus and I should have liked nothing better than to use up many more pages of this issue for a detailed response to all my critics.

GUNTHER S. STENT

Department of Molecular Biology
University of California
Berkeley

Scientific American, March, 1973; Vol. 228, No. 3. Published monthly by Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017; Gerard Piel, president; Dennis Flanagan, vice-president; Donald H. Miller, Jr., vice-president and secretary; George S. Conn, treasurer; 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 25 cents.

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

☐ Please enter change of address.

☐ Please enter subscription.
U.S. and Canada:
☐ 1 year, $10 ☐ 2 years, $18 ☐ 3 years, $25
All other countries:
☐ 1 year, $12.50 ☐ 2 years, $23 ☐ 3 years, $32

NAME

NEW ADDRESS

OLD ADDRESS

50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

MARCH, 1923: "Ernest F. Nichols and James D. Tear of the Nela Research Laboratory have developed an oscillator for the production of Hertzian or radio waves of such small length (one-fifth of a millimeter) that they are actually shorter than the longest heat waves radiated by a mercury-vapor lamp. The gap between the 'natural' waves of heat and light and the 'artificial' waves used in radiotelegraphy is thus completely closed, and it is now possible to produce and study at will waves of any length from many miles to a billionth of an inch."

"Every thoughtful observer is impressed with the fact that we are threatened in places with an absolute chaos in highway transportation. The rapid rate of increase in the number of motor cars and trucks on the highways of the country has brought about a congestion of traffic that calls for some form of regulation. Major Elihu Church would apply the methods that were used to control the enormous motor-truck traffic on the highways of France during the World War. His proposals include motor-truck trunk-line highways built wide enough to accommodate several lines of automotive vehicles, lighted for day and night service and equipped with telephone service for emergency calls from drivers in trouble. He also proposes that there be belt lines around heavy-traffic centers that are coordinated with the trunk lines, and that the trunk lines be controlled in principle precisely as a railway is controlled by the train dispatcher method."

"Several weeks ago the American Telephone & Telegraph Company conducted some interesting radio telephone tests between New York City and New Southgate, a London suburb. Although radio telephone words had been transmitted across the Atlantic before the recent tests, it remained for these tests to prove that entire conversations could be carried on with distinctness and full understanding. Mr. Thayer and Mr. Carty of the telephone company spoke from their offices in New York City, using an ordinary telephone instrument connected by a telephone circuit some 70 miles long, half of it being underground cable and half aboveground wire, leading to the powerful radio telephone transmitter at the Rocky Point station of the Radio Corporation. The distance from the Rocky Point station to New Southgate is 3,400 miles. For the transmission there were employed 20 of the new 10-kilowatt vacuum tubes arranged with modulators and filters to radiate about 100 kilowatts of radio energy."

"Only London can boast of 'solid smoke.' It is also known as soot, and some quite unpleasant statistics are made public regarding it in the annual report of the medical officer of the city. In order to examine the air in the city an apparatus consisting of a large rain gauge is mounted on the roof of a building, and the rain water from a known area of surface is collected monthly. The total solids varied from month to month, the highest being 18.33 metric tons per square kilometer in September and the lowest 7.79 metric tons per square kilometer in August. One metric ton per square kilometer is approximately equal to nine pounds per acre."

SCIENTIFIC AMERICAN

MARCH, 1873: "Professor Tyndall, in his eloquent plea for original research, has in mind the devoted band of workers who in the chill atmosphere of neglect have made the discoveries that are the boast of our present civilization. He cautions us to protect and foster such persons, and he has proved the sincerity of his motives by leaving in the hands of trustees all the earnings of his lecturing tour in the United States, for the support of needy students who wish to devote themselves to pure science. Let us start as many rills of discovery as we can, knowing full well that they will swell into torrents and ultimately flow through the valleys into the ocean of human wants. Keep the fountains of discovery pure and bright, and the streams of knowledge will not dry up and the world will not want for practical inventions."

"Mr. Willoughby Smith has been making a series of electrical experiments with selenium. Sticks of selenium were connected with platinum wire and hermetically sealed in glass tubes. The electrical resistance of some of the sticks was very great and the resistance of other sticks was much less. He was at a loss to account for this lack of constancy until after various trials he found that it was due to the action of light. When the sticks of selenium were shut up in a box so as to exclude light, the electrical resistance was highest and remained constant, but when the cover was withdrawn and light was allowed to fall on the sticks, the electrical resistance diminished according to the intensity of the light. These very singular observations may lead to new and useful discoveries."

"It appears singular that certain contagious diseases, especially small pox, spread more in the winter season (which in other respects is the healthiest time of the year) when the cold destroys the miasmata that flourish in tropical climes and hot summers sometimes visit portions of the temperate zones. But in order to explain this anomaly we have only to consider that in winter a large number of the lower classes of people huddle together in ill-ventilated rooms in order to shelter themselves against the cold. Of course this is favorable to the growth of miasmata, which only need suitable conditions to propagate themselves. Microscopists have succeeded in tracing the origin of many contagions to parasitic growth, either vegetable or animal, and it is probable that this will ultimately be the case with all, the denial of many medical authorities notwithstanding."

"Professor Agassiz evidently has no very high opinion of our educational institutions, and he publishes his adverse views with characteristic freedom. He says our colleges are nothing but high schools, and that even Harvard is far from being a university. The knowledge imparted is 'the traditionary learning of the middle ages' and only 'the dregs of scholarship.' In common with all progressive lovers of science Professor Agassiz strongly advocates a freer scope being given to the study of nature. It is probable that these ideas, from so eminent an authority, will give new impetus to the war between science and the classics that for some years past has been waged in our colleges. We adhere to the belief that, were the classics in our seminaries made subservient to thorough studies of the principles of nature, the graduates would leave their books much better prepared to encounter the world."

Fast-Poke.

One of our engineers lives in Schliersee, a picturesque village in the Bavarian Alps, about 60 kilometers south of our factory in Munich. The other day he wrote a memo to our marketing department conveying his impressions of our new BMW 2002 tii model. Here is an excerpt:

"Coming down the mountain I kept the engine constantly at red-line in first, second, and third gears, never falling below 5,000 rpm, and heeling and toeing accelerator and brake. I power-drifted the car through the tight turns and switchbacks and experienced no brake-fade. Once on the valley floor I shifted up into fourth, and as soon as I reached the Autobahn, went flat out, attaining 168 kmph (105 mph). I cruised at this speed all the way into Munich. Performance quite acceptable."

Our associate never mentioned that the world's best sports sedan is also highly utilitarian, with plenty of room for four long-legged adults. Or that it has a cavernous trunk, and a contempt for repair shops, and much more.

Some people, it seems, get one idea in mind, and then think of nothing else.

BMW's are available in four or six-cylinder models. U.S. Sole Importer: **Hoffman Motors Corporation.** 375 Park Avenue, New York, N.Y. 10022; 12541 Beatrice Street, Los Angeles, California 90066. Dealers Coast to Coast. Advantageous European Delivery Program. Write for free brochure.

Canada: East: BMW Distributors Eastern Canada, Ltd. 801 Progress Avenue, Scarborough, Ontario. West: BMW Distributors (Western) Co. 2040 Burrard Street, Vancouver 9, B.C.

Bavarian Motor Works

BMW

THE AUTHORS

TOMAS FREJKA ("The Prospects for a Stationary World Population") is a staff associate of the Population Council. Born in Czechoslovakia, he received the equivalent of a master's degree in 1959 at the Prague School of Economics and the equivalent of a doctorate at the Institute of Economics of the Czechoslovak Academy of Sciences in 1966. He held various industrial, government and academic posts in Czechoslovakia until 1966, when he became a visiting fellow in the Office of Population Research at Princeton University. In 1968 he was consultant to the Division of Social Affairs of the United Nations office in Geneva, and for a year before joining the Population Council in 1969 he worked at the UN Demographic Centre for Latin America in Chile.

BRUCE A. BOLT ("The Fine Structure of the Earth's Interior") is professor of seismology at the University of California at Berkeley and director of the 16 seismographic stations of the University of California. He was born and educated in Australia, obtaining his bachelor's degree at the University of New England in New South Wales in 1952 and his master's degree and Ph.D. (in mathematics) at the University of Sydney in 1954 and 1956 respectively. Before he joined the Berkeley faculty in 1963 he was lecturer in applied mathematics at Sydney. Bolt wishes to thank Howel Williams, professor emeritus of geology at Berkeley, "for kindly commenting on the manuscript" and the National Science Foundation "for essential financial support over the past decade."

O. L. MILLER, JR. ("The Visualization of Genes in Action"), is a member of the staff of the biology division of the Oak Ridge National Laboratory. On September 1 he will become chairman of the department of biology at the University of Virginia. Having received his bachelor's and master's degrees at North Carolina State College in 1948 and 1950 respectively, he spent six years as manager and co-owner of a 300-acre tobacco and livestock farm in South Carolina. He then resumed graduate study, receiving his Ph.D. from the University of Minnesota in 1960. He was a postdoctoral fellow there for a year before he went to Oak Ridge. Miller describes himself as "an avid football fan" and adds that he and his wife "are wild-flower enthusiasts, and we spend many of our weekends from early spring to late summer roaming the nearby Cumberland and Smoky mountains and areas in between, enjoying the flowers and searching for new specimens to transplant to our backyard wild-flower garden."

BARRY E. TURNER ("Interstellar Molecules") is associate scientist at the National Radio Astronomy Laboratory, where he has worked for the past five years. His initial education was in physics, in which he obtained his bachelor's and master's degrees at the University of British Columbia. "I then spent two years with the radio astronomy group of the National Research Council of Canada," he writes, "where I was involved in the development of electronic equipment for solar research. This was the time (1962 to 1964) when radio astronomy was just beginning the series of discoveries that established it as the exciting field it is today. I was irresistibly attracted to the field and decided to pursue the Ph.D. in radio astronomy at the University of California at Berkeley."

DOREEN KIMURA ("The Asymmetry of the Human Brain") is associate professor of psychology at the University of Western Ontario. Her degrees are all from McGill University: bachelor's in 1956, master's in 1957 and Ph.D. (in physiological psychology) in 1961. She did the research for her Ph.D. at the Montreal Neurological Institute and subsequently spent two years there as a fellow, also holding appointments at medical institutions in Los Angeles, Zurich and Hamilton, Ont., before receiving her present appointment in 1967. "My nonprofessional activities," she writes, "include caring for a small daughter, cooking and helping my husband to attract birds, squirrels and other wildlife to our home in the country."

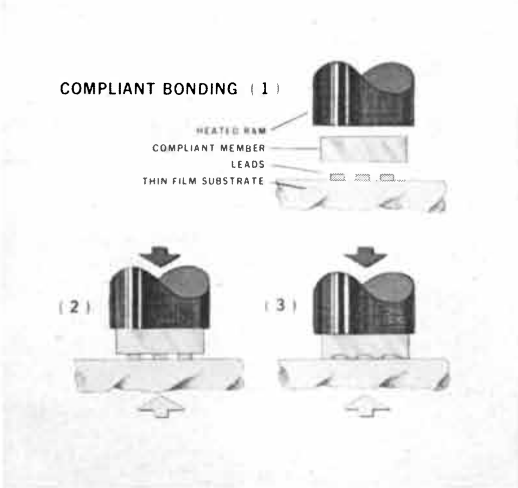
S. S. WILSON ("Bicycle Technology"), who describes himself as "a simple-minded engineer," is university lecturer in engineering at the University of Oxford and a fellow of St. Cross College. He was graduated from Oxford in 1944 and has taught there since 1946. "My work has been mainly in the field of heat engines and fluid mechanics," he writes. "My interest in bicycles dates back to school days. I have always owned and used a bicycle; during World War II, I several times cycled more than 100 miles in a day as a means of transport. Apart from the above I have done a lot of sailing. In 1958 I designed one of the first all-fiber-glass sailing dinghies, the 12-

foot Alpha, which was selected by the Design Centre in London and was the prototype of many later designs. I am a practitioner of 'do it yourself' in many forms, out of interest but also because I believe there is a fundamental need in *Homo faber* to design and make—a need quite insufficiently catered to in most education, particularly at universities. I am involved in the application of pedal power to tools and equipment for use in developing countries."

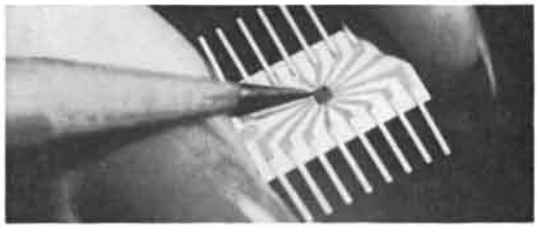
WILLIAM C. LEGGETT ("The Migrations of the Shad") is associate professor of biology at McGill University. "My choice of biology occurred more by accident than design," he writes. "My major interests as an undergraduate were English literature and history, and my B.A., received in 1962 from Waterloo University College (now Waterloo Lutheran University), records these as my major subjects. In my junior year I took a single biology course as an elective. Largely because of the quality of the instruction I registered for a second biology course taught by the same professor. By the end of that course I was, as they say, 'converted.' " Leggett went on to take bachelor's and master's degrees in science at Waterloo and to obtain his Ph.D. from McGill. He also serves as a project director with the Connecticut River Ecological Study. "My leisure activities tend to be more manual," he says. "High on the list are carpentry and cabinetmaking, at which I am proficient but slow. I am also a reasonably accomplished automobile mechanic, largely out of self-defense. Recreational flying and skiing are my main sporting interests."

CHARLES J. BRAINERD ("The Origins of Number Concepts") is assistant professor of psychology at the University of Alberta and an affiliated member of the Center for Advanced Study in Theoretical Psychology. He writes: "I grew up around Lansing, Mich., and received all my education and degrees from my hometown school: Michigan State University. I was granted a Ph.D. in developmental and experimental psychology in 1970. I believe I first became interested in the origins of the number concept at age 14, when I discovered to my horror that numbers other than integers are every bit as 'real' as integers. Even as I continued into higher mathematics, this fact still seemed curious to me. In addition to my number research, I have been studying the emergence in children's thinking of several other concepts that have been of historical importance in the evolution of mathematics and logic."

WESTERN ELECTRIC REPORTS



When heat and pressure are applied to the compliant medium, it begins to deform around the gold-plated leads. Deformation of leads is controlled by the flow stress properties of the medium. When the medium bottoms out, it stops the ram and the delicate metal parts are instantaneously and permanently bonded without damage.



Slightly magnified, the pencil points to a beam-leaded circuit chip which has been bonded to 16 gold conductors that converge on it. Within the silicon chip are dozens of microscopic transistors, diodes, resistors.

Greatly magnified, we can see gold beam leads projecting from a chip bonded to thin gold conductors on a thin ceramic substrate.

A new and better way to bond integrated circuits.

Engineers at Western Electric's Engineering Research Center (ERC) and Allentown Works have come up with a revolutionary but simple solution to some very complex circuit bonding problems. It's called compliant bonding.

As in other solid state bonding methods, heat and pressure are used to bond tiny integrated circuits to other components. (Some circuits have fifty or more delicate leads.)

However, our process differs in a very important way. ERC researchers added a compliant or yielding medium between the energy source and leads being joined. On contact, the medium compresses and transmits an equal amount of controlled, predictable bonding pressure to each lead.

There are many advantages to this new technique.

First, it is more reliable. Under heat and pressure, the compliant medium spreads the bonding pressure to all the leads uniformly. It automatically compensates for surface and lead irregularities. Strong, reliable electrical connections are assured for every lead.

It is also more versatile. We can now bond more than one circuit at a time, even with leads of different thicknesses or area widths. The compliant medium perfectly controls lead deformation in even the most complicated multiple bonding. It's no longer necessary to design and test complex bonding tools for each bonding job.

Engineers at Allentown are working to apply the process to large-scale manufacturing. They have developed the first production machines using the process. These machines are now in growing use at Allentown and many other Western Electric plants.

Conclusion: Compliant bonding is technically and economically superior to other solid state bonding techniques. Combined with automated production, compliant bonding promises reliable, high-speed production of circuit packages.

Western Electric

We make things that bring people closer.

Established 1845 SCIENTIFIC AMERICAN March 1973 Volume 228 Number 3

The Prospects for a Stationary World Population

The human population is now about 3.6 billion. An extrapolation of present world demographic trends that lies between two extreme projections shows it leveling off at some 8.4 billion by the year 2100

by Tomas Frejka

If a nation wants to achieve zero population growth, when can it reasonably expect to do so? If the policy-makers of another nation are concerned about a possible decline in population, what information is available to them about when the decline may begin? What alternatives of growth for the population of the world or any of its parts are realistic over the next several decades? What kind of demographic trends, mainly relating to fertility, will have to be generated if a certain desired rate of growth (and thus size of population) is to be achieved?

Questions of this kind have formed the basis for a research project in which I have been involved at the Population Council and earlier at the Office of Population Research of Princeton University. The main methodological tools were population projections, which are designed to facilitate understanding of the short-term and long-term implications of current basic demographic features and are meant to be instrumental in assessing realistic goals of population change. Among other findings, the research shows that a substantial growth of the world's population must be accepted and planned for and that the present ratio of 30 : 70 between the populations of the rich and the poor nations will swing inexorably to 20 : 80 and perhaps even to 10 : 90.

Until about 1700 the human population of the world grew very slowly: probably at an average rate of less than .002 percent per year. Now on a worldwide basis the rate is 1,000 times higher: 2 percent per year. The current state of affairs is the result of complex economic and social changes. They have been accompanied by a major change in the patterns of fertility and mortality that students of population call the demographic transition.

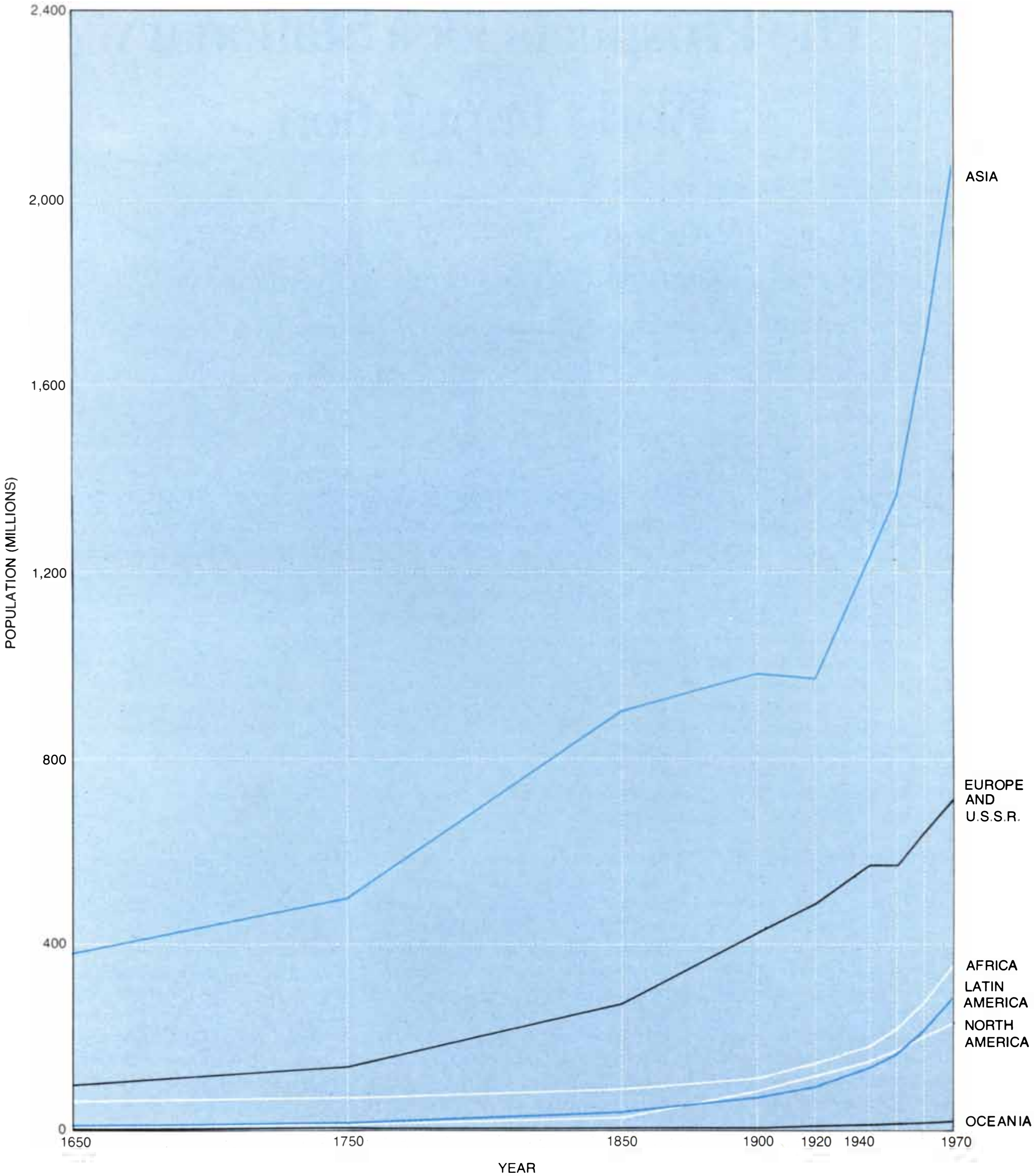
Three stages of the demographic transition can usually be detected in every population. The first stage is characterized by high and almost equal birth and death rates and by a low rate of growth. That pattern has existed in most populations throughout most of history. In the second stage mortality declines and is followed by a lagging decline in fertility, so that the rate of population growth is high. The third stage is characterized by low birth and death rates and therefore by a rate of growth that declines gradually.

The span of time during which the demographic transition takes place varies considerably among populations. So far the transition has run most of its course only in developed countries. The majority of the developing countries are in the second stage. Moreover, the decline in mortality has been more precipitous in these countries than it was in the developed countries, resulting in growth rates that are considerably higher than the ones experienced in the developed countries. That is the principal reason the growth rate of the total human population is at the high level of 2 percent per year.

The experience of history suggests that the population of the world may eventually reach a state close to nongrowth, that is, all countries will be in the third stage of the demographic transition. What demographic developments will have to take place for that stage to be attained? How long will it take, barring catastrophes of natural or human origin? How large will the world's population be when nongrowth is achieved?

One begins with the fact that a continuation of the present growth rate (not to mention an increase in the rate) would bring the population of the world close to seven billion around the year 2000, 14 billion in the 2040's and 28 billion in the 2070's. (It is now about 3.6 billion.) In the developed countries, however, the demographic behavior of populations has changed radically over the past few centuries. In those countries an increasing number of people have relatively small families and live to see their grandchildren grow up. Fertility and mortality patterns of this type yield rates of natural increase (a figure that ignores migration in or out) of less than 1 percent per year. Indeed, during the past decade the rate of natural increase has been decreasing in the developed countries.

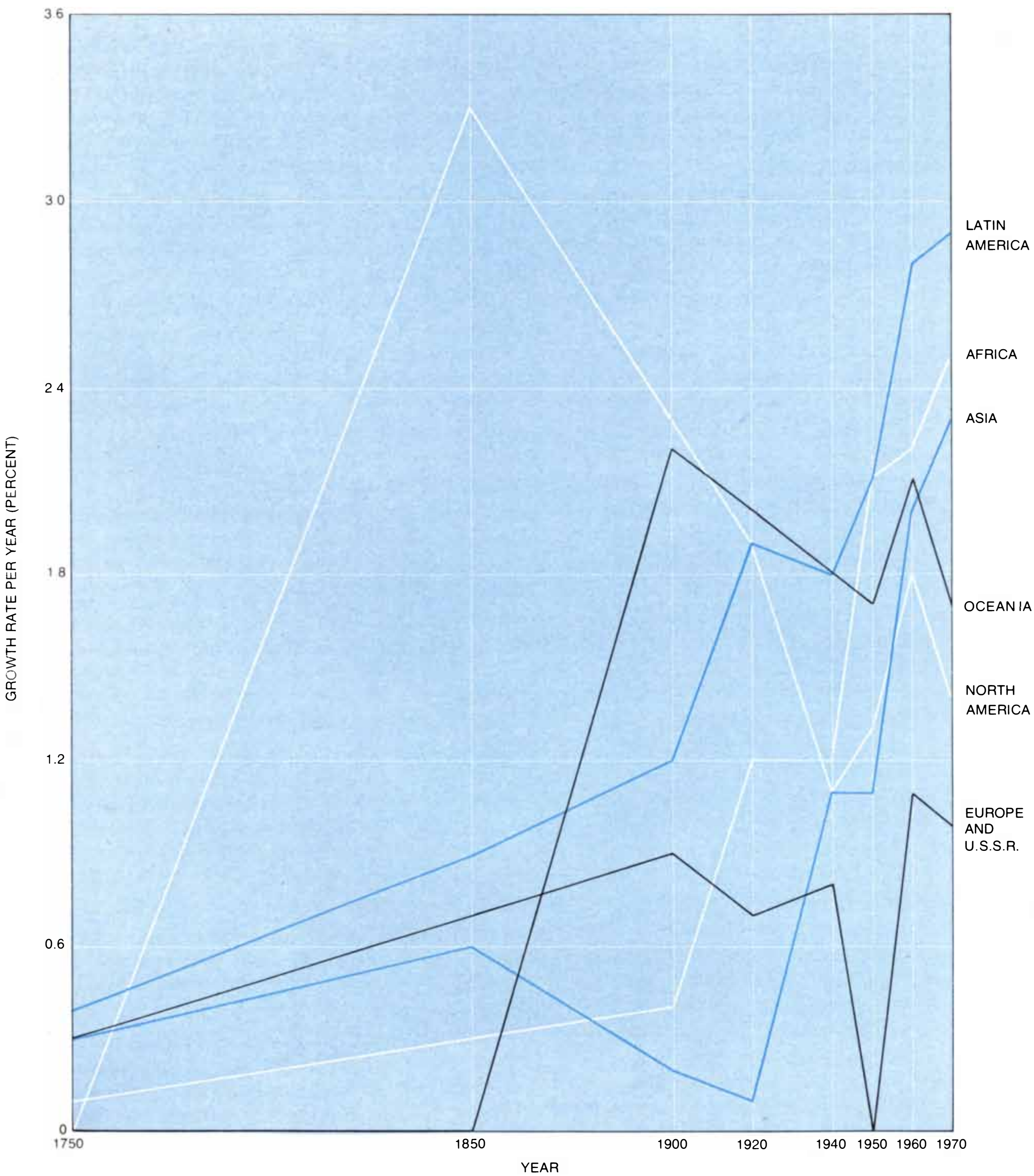
The demographic behavior of populations in the developing countries is distinctly different. The conditions that generated low mortality levels in the developed countries have been spreading to the developing countries rather rapidly, whereas fertility levels in those countries have changed slowly or not at all. The average rate of natural increase has been estimated at 2.6 percent per year in the late 1960's, with a crude birth rate of more than 40 per 1,000 inhabitants and a crude death rate of about 15 per 1,000. (All evaluations of the situation in de-



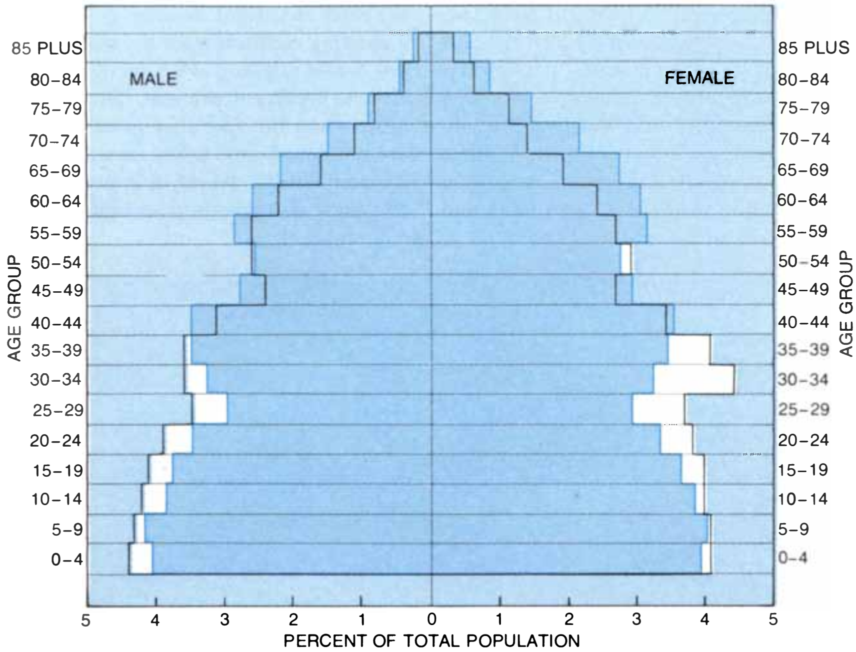
REGIONAL GROWTH of the world's population from 1650 to 1970 is portrayed. At the beginning of the period the total population of the world was about 553 million. By 1850 it was 1.3 billion. Since 1900 it has gone from 1.6 billion to the present 3.6 billion.

veloping countries as a whole must be qualified by the fact that there is increasing evidence of changes in demographic behavior in the People's Republic of China. The magnitude and rate of the changes are unknown, at least to the rest of the world.)

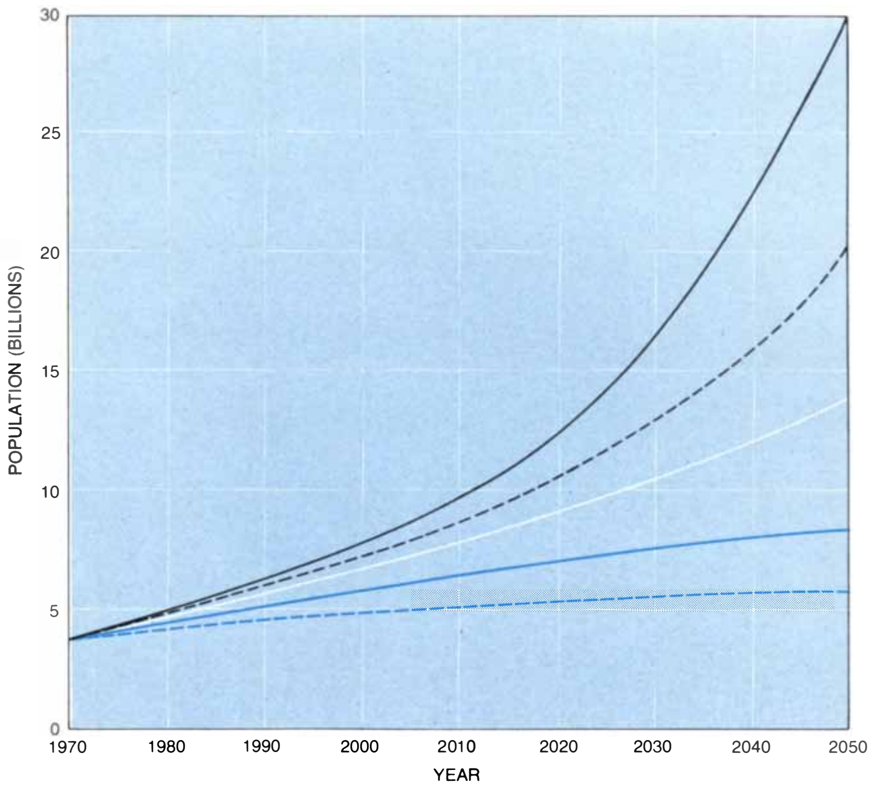
In the developed countries, even though demographic patterns have varied from country to country over the past 50 years, it is clear that the populations are reproducing themselves in a regime of low mortality and low fertility; indeed, the regime has intrinsic features close to nongrowth. Yet an examination of the crude death and birth rates and the rates of natural increase gives rise to a seeming contradiction: the populations are in fact growing at an average rate of close to 1 percent per year. The paradox arises from the fact that growth trends are not determined solely by the mortality and fertility levels of a population; one must also take into account the age



REGIONAL DIFFERENCES in growth appear when the population increases between 1650 and 1970 are expressed in terms of the growth rate, shown here as percent per year. For thousands of years the rate was about .002 percent. It is now some 2 percent.



AGE STRUCTURE of the population in Belgium (*pyramid in color*) and in Greece (*pyramid outlined in black*) differs and affects the number of births and deaths. In Belgium 13 percent of the people are older than 65, in Greece 9 percent. Hence Belgium has a larger proportion of people at ages where the risk of death is high. Belgium also has a smaller proportion of potential mothers than Greece, since only 18 percent of the females in Belgium are in the age group from 20 to 34 compared with about 23 percent in Greece.



DIFFERING PROJECTIONS of the world's population are made on the basis of different assumptions. Smallest increase (*bottom*) would occur if a net reproduction rate of 1, which implies replacement-level fertility, were achieved in the early 1970's. The population would still rise for a time because many people are of childbearing age. Other projections assume that a net reproduction rate of 1 is achieved between 2000 and 2005 (*solid color*) or between 2040 and 2045 (*white*), that present fertility and mortality rates continue (*broken black*) and that fertility remains constant while mortality declines (*solid black*).

structure, which is a product of past mortality and fertility patterns. The crude rates provide information on relative additions to and deductions from the present population, but they are not "true" measures of the level of mortality and fertility; they are products of the mortality and fertility levels for each age group and of the number of people at each age.

Greece and Belgium provide an illustration of the influence of age structure on growth trends. In 1967 the crude death rate was 12 per 1,000 in Belgium and 8.3 in Greece; the crude birth rate was 15.3 in Belgium and 18.7 in Greece. As a result the rate of natural increase was .3 percent in Belgium and 1 percent in Greece. (The difference may seem small, but one should bear in mind that a persisting growth rate of .7 percent per year causes a population to double in 100 years. Moreover, Greece was growing at a rate three times higher than Belgium.) The differences in the crude rates were primarily due to differences in the age structure of the two populations [*see top illustration at left*].

In both populations the patterns of mortality (the relative numbers of people dying at various ages) and the patterns of fertility (the average number of children born to women of various ages) were similar. The age structures, however, differed significantly because throughout the 20th century fertility has been higher in Greece than in Belgium. In 1967 Belgium had a much older population than Greece (13 percent of the people were older than 65, compared with 9 percent in Greece), meaning that significantly more of Belgium's people were in the age groups having a high frequency of death. At the same time Belgium had a smaller proportion of potential mothers than Greece: females in the age group from 20 to 34 accounted for 18 percent of the female population, compared with 23 percent in Greece. Therefore, although in both populations the same relative numbers of children were being born per 1,000 potential mothers, the absolute number of children born was higher in Greece than in Belgium, as was the crude birth rate.

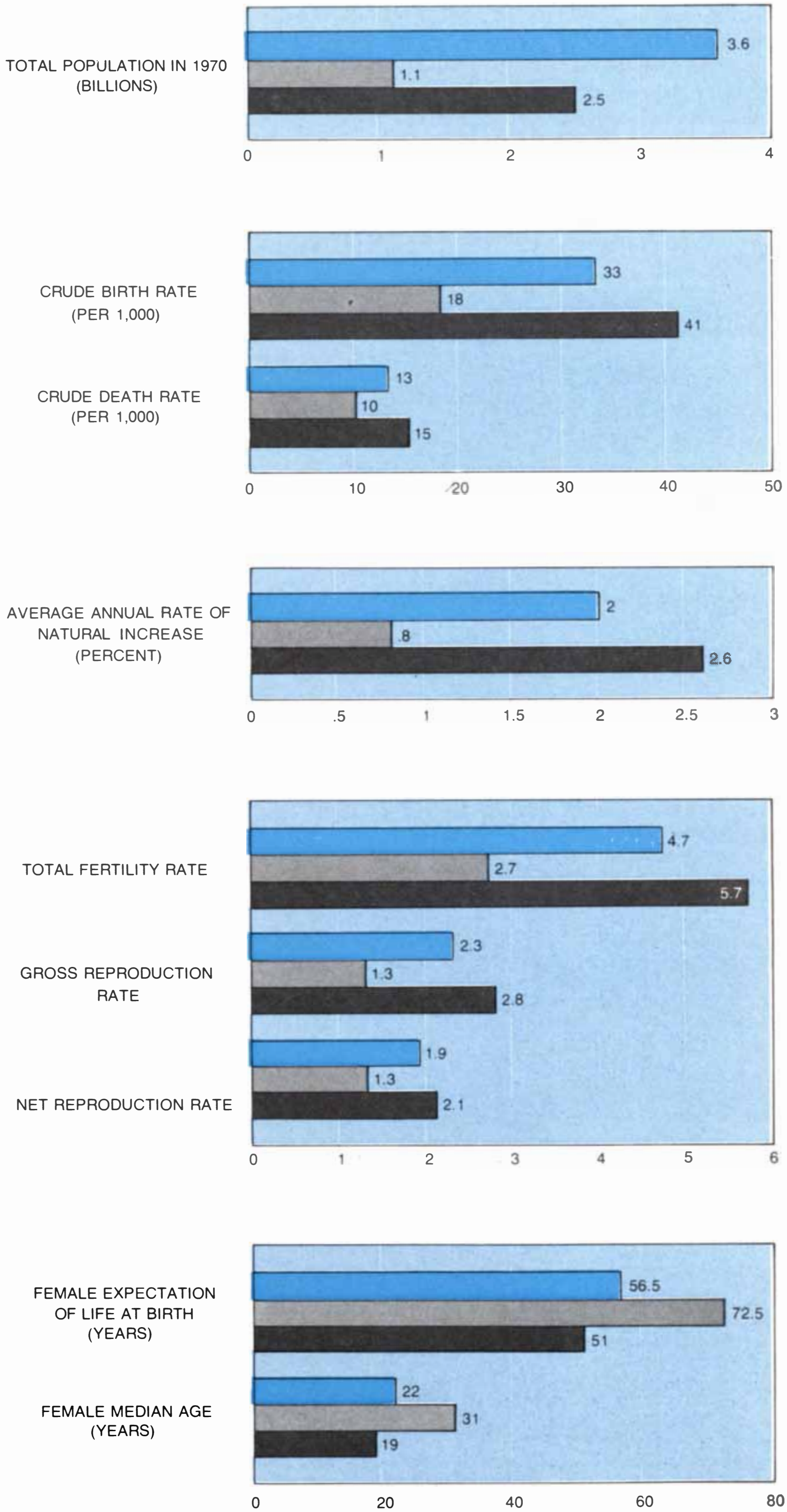
Demographic measures somewhat more complex than the crude birth and death rates provide information on fertility and mortality without regard to the age structure of the population. On the fertility side are the total fertility rate and the gross reproduction rate. On the mortality side is the expectation of life at birth. The combined effect of both fertility and mortality is expressed by the net reproduction rate.

The total fertility rate is defined as the average number of live children a woman would bear if she passed through her years of potential motherhood experiencing the current fertility of women at each age. In fact, the fertility of women at one age influences their later reproductive performance, so that in 1973 the fertility of 21-year-old women (who were the ones aged 20 in 1972) may differ from the 1972 rate for the women of 21. The total fertility rate is thus an artificial measure referring to a hypothetical woman, but it does provide a reasonable approximation of the average number of children borne by a woman in the respective populations. The gross reproduction rate is built on the same theoretical base as the total fertility rate but refers only to the average number of daughters who would be borne by a woman passing through her childbearing years.

The expectation of life at birth is about 70 years in the developed countries, implying low levels of mortality at all ages. Usually fewer than 30 infants in 1,000 die before reaching their first birthday, and about 95 percent of the children born survive at least to age 20. In many developing countries the expectation of life at birth is still no more than about 40 years, although the range of mortality differences is wide. Where the life expectancy is around 40 years, the usual picture is that more than 200 infants in 1,000 die before their first birthday and that only about 60 percent of all the children born survive at least to age 20.

The net reproduction rate provides information on how many daughters of the average woman would survive to childbearing age if they were subject to the current mortality patterns. Thus under conditions of high mortality the difference between the net reproduction rate and the gross reproduction rate is large, whereas under conditions of low mortality the net rate is not much smaller than the gross rate. These relations are illustrated by a comparison of Costa Rica and the Ivory Coast. The net reproduction rate is higher in Costa Rica than in the Ivory Coast, although fertility, as measured by the total fertility rate and the gross reproduction rate, is the other way around.

The net reproduction rate is an indicator of how a population is being reproduced. When the rate is around 1, it shows that the current fertility and mortality conditions are such that the cur-

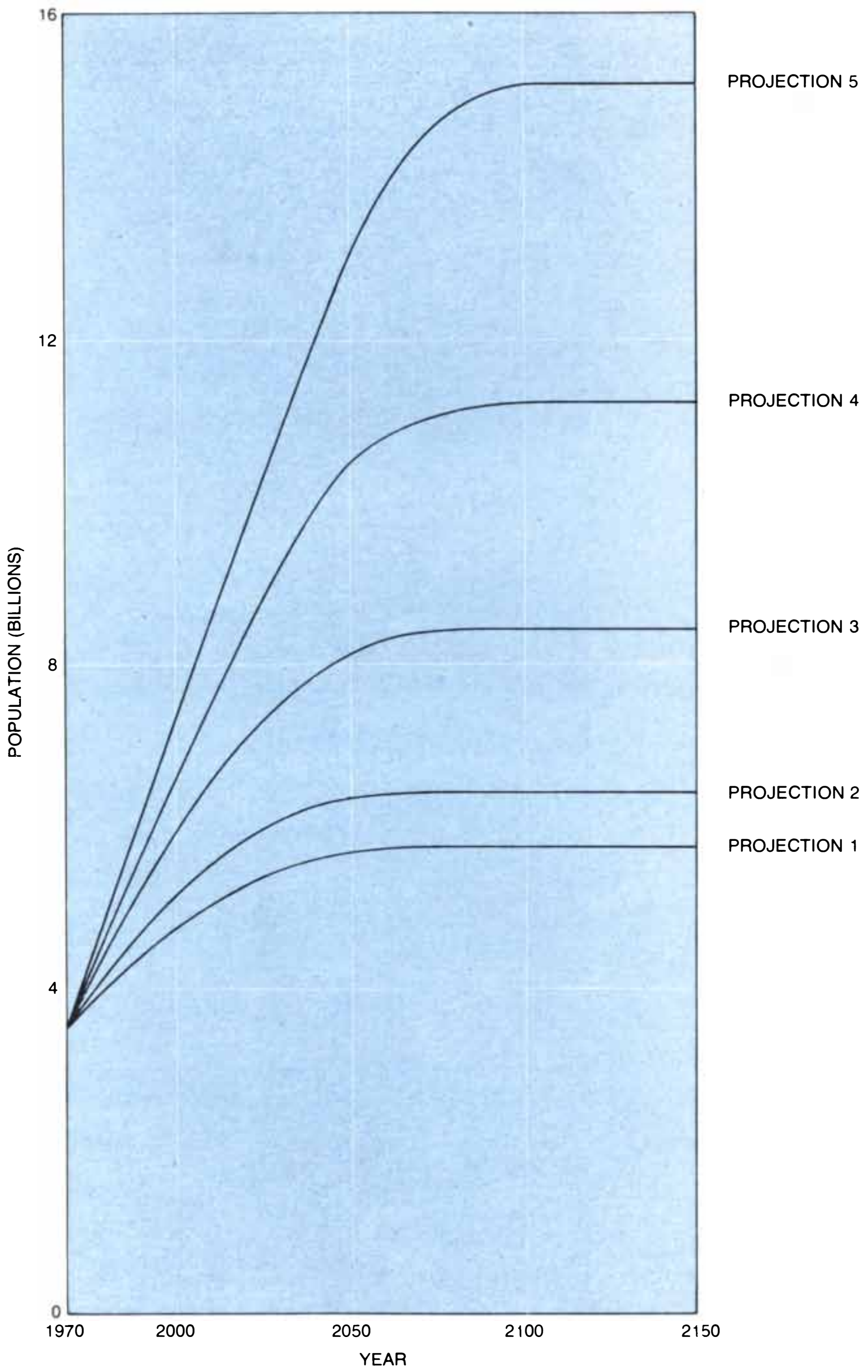


DEMOGRAPHIC FEATURES are depicted for the world (*color*), the developed nations (*gray*) and the developing nations (*black*). Total fertility rate is an average of live children born to a woman. Gross reproduction rate refers to the average number of daughters borne by a woman. The net reproduction rate indicates how one generation of childbearers is being replaced by the next one. A net reproduction rate of 1 implies bare replacement.

rent generation of childbearers would be replaced by a generation of the same size, provided that the conditions persisted. Fertility corresponding to this value is called replacement fertility. In the developed countries a total fertility rate of about 2.1 is the replacement fertility.

At the beginning of the 1970's a number of nations had fertility levels of this type. Among them were the U.S., Canada, Denmark, Finland, Sweden, Bulgaria, Czechoslovakia, Hungary, Poland, Japan and possibly other countries. Nonetheless, the populations of these countries were still growing, often at rates of close to 1 percent per year, because of their previous demographic history, which was present in and acting through the age structure.

In most of the population projections that I shall discuss it is assumed that a fertility and mortality regime with a net reproduction rate of 1 will be achieved and maintained. The value is justifiable for historic reasons (since the net reproduction rate must have been close to 1 for thousands of years) and because any other value persisting indefinitely would cause the population either to grow infinitely or to decline to zero. Moreover, in many developed countries new demographic patterns seem to be emerging that suggest a trend toward a net reproduction rate oscillating around a value close to 1.



POPULATION PROJECTIONS are based on five assumptions about when a net reproduction rate of 1 might be achieved and maintained. Reading from the bottom the dates are respectively 1970 to 1975, 1980 to 1985, 2000 to 2005, 2020 to 2025 and 2040 to 2045. With an index of 100 for the 1970 population, the indexes under the five respective projections would rise by 2050 to 153, 172, 224, 287 and 357 and by 2150 to 156, 176, 230, 306 and 416.

In the population projections that I have prepared (a set of five for each population) the current demographic features of the population under consideration serve as the starting point [*see illustration at left*]. The projections differ by the assumed point in time when the net reproduction rate of 1 is reached and thereafter maintained. Thus they also differ by the assumed rate of fertility decline. The projections assume that the present fertility level declines to a level corresponding to a net reproduction rate of 1 immediately [*projection 1*] or over a period of 10, 30, 50 and 70 years [*projections 2 through 5 respectively*]. The current level of mortality of every population is assumed to decline further, settling eventually at the lowest levels now experienced. For technical reasons the computations were done only with female populations, but the results are acceptable approximations of total population features and trends.

The range of possibilities for the future growth of the world's population is so large that it is well to set arbitrary limits that seem to represent the extremes. The minimum, which would be achieved if fertility declined so rapidly that a net reproduction rate of 1 was reached between 1970 and 1975, must be ruled out as highly unlikely. This minimum alternative, however, is informative because it illustrates developments that are not likely to materialize. For the other extreme one might assume constant fertility and declining mortality. This alternative seems unlikely but not impossible, since it portrays a situation that exists in a number of less developed countries.

At the minimum extreme the achievement immediately of a worldwide net reproduction rate of 1 would still result in a population increase of 30 percent in 30 years and more than 50 percent in 80 years (to a level of 4.7 billion in the year 2000 and 5.5 billion in 2050), since most populations currently have a preponder-

	YEAR	TOTAL POPULATION (BILLIONS)	CRUDE BIRTH RATE	AVERAGE ANNUAL GROWTH RATE	AVERAGE ANNUAL INCREMENT OF POPULATION (MILLIONS)	PERIOD	NET REPRODUCTION RATE	TOTAL FERTILITY RATE
	1970	3.6	33	2.0	68	1965-1970	1.9	4.7
	IF IN YEAR	THE FOLLOWING CHARACTERISTICS ARE TO BE ACHIEVED				THEN IN PERIOD	THE FOLLOWING RATES ARE NECESSARY	
PROJECTION 1	2000	4.7	18	0.8	37	1970–1975	1.0	2.5
	2050	5.6	14	0.2	9	1980–1985	1.0	2.4
	2100	5.7	13	0.0	0	1990–1995	1.0	2.3
						2000–2005	1.0	2.2
PROJECTION 2	2000	5.1	19	1.0	49	1970–1975	1.6	3.9
	2050	6.3	14	0.1	7	1980–1985	1.0	2.4
	2100	6.4	13	0.0	0	1990–1995	1.0	2.3
						2000–2005	1.0	2 2
PROJECTION 3	2000	5.9	21	1.2	70	1970–1975	1.8	4.4
	2050	8.2	14	0.3	21	1980–1985	1.6	3.7
	2100	8.4	13	0.0	0	1990–1995	1.3	2.9
						2000–2005	1.0	2.2
PROJECTION 4	2000	6.4	25	1.7	106	1970–1975	1.8	4.5
	2050	10.5	14	0.5	50	1980–1985	1.7	4.0
	2100	11.2	13	0.0	2	1990–1995	1.6	3.6
						2000–2005	1.4	3.1
PROJECTION 5	2000	6.7	28	2.0	124	1970–1975	1.9	4.6
	2050	13.0	16	0.8	97	1980–1985	1.8	4.2
	2100	15.1	13	0.0	5	1990–1995	1.7	3.9
						2000–2005	1.6	3.5

CHANGES REQUIRED in demographic features to achieve certain of the population levels projected in the illustration on the opposite page are set forth. Achievement of even the upper levels would require a significant decline in the present total fertility rate.

ance of people of childbearing age. With constant fertility and declining mortality the population of the world could be 7.4 billion in 2000 and close to 30 billion by 2050.

Since these limits are so far apart there are many alternatives between them. The actual path will depend on the trends of mortality and fertility. Assuming that mortality will continue to decline, one can foresee that with a rather rapid decline in fertility the world's population would be about six billion in 2000 [*projection 3*] and that with a moderate decline in fertility it would be 6.7 billion [*projection 5*]. In the first case the current average of 4.7 children born to each woman of childbearing age would have to decline during the next 30 years to about 2.2; in the second case it could be about 3.5. If a worldwide pattern of two children per family on the average—that is, a net reproduction rate of 1—were reached by 2000 and maintained thereafter, the world's population would be about eight billion by 2050 but would not grow much more. If a net reproduction rate of 1 were reached gradually by the middle of the next century, the population would be 13 billion in 2050 and some 15 billion in 2100.

Although the future size of the world's population is to a considerable extent predetermined by the current demographic features—levels and patterns of mortality and fertility, together with the age structure—much will depend on how such characteristics change. Mortality is already fairly low and is likely to decline further, barring major and persisting catastrophes. Fertility is still rather high, being more than twice the level that would ultimately result in a nongrowing population, and even if it declines, the decline is likely to be gradual. The age structure of the world's population is highly favorable to growth because a small proportion (fewer than 6 percent) of the population are more than 65 years old and a large proportion of the women are in the childbearing ages. It is of still more importance that relatively large numbers of women will enter the childbearing ages in the coming decades, since about 37 percent of the world's population are under 15 years of age. Moreover, at least during the 1970's generations of comparable size are likely to be born and to survive, judging from current fertility and mortality levels and patterns.

If the mortality trends of the future are roughly consistent with those of the past few decades and fertility declines rather rapidly, the world's population may develop along the path indicated by projection *3:* somewhat less than six billion people in 2000 and somewhat more than eight billion by 2050. A level as low as five billion by 2000 seems highly unlikely because of the built-in potential for growth and because it is unreasonable to expect average fertility behavior to change so rapidly throughout the world that by the 1980's only two or three children will be born to each woman on a worldwide basis.

The illustration on the preceding page presents data that give a general idea of the kind of population development that will result from various speeds of decline in the fertility rate. Projection *1*, which shows that a total fertility rate of 2.5 would have to be achieved in the early 1970's for the population of the world to be as low as 4.7 billion in 2000 and below six billion in 2050, illustrates developments that are most unlikely, if not impossible. Projection *3*, which assumes that the total fertility rate will drop to 4.4 in the early 1970's and 2.2 by 2000 for a world population of 5.9 billion in 2000 and about eight billion in 2050, can be considered as illustrating developments that would take place if the demographic transition (mainly the fertility decline in the developing countries) were relatively fast and the change in fertility behavior from traditional to modern were attained within about one generation. Projection *5*, wherein the total fertility rate is seen as 4.6 in the early 1970's and 3.5 by 2000, with a total population of 6.7 billion in 2000 and about 13 billion in 2050, illustrates developments that might occur if the demographic transition in the developing countries were to take as much time as it did in western Europe.

The data provide a basis for an answer to the question that is often raised about whether it might be possible to achieve zero population growth for the world by 2000. The answer is that such an achievement is most unlikely and probably impossible, unless there are unexpected changes in the world's economic, political and social relations, unanticipated evolutions in education and health or unforeseen high levels of mortality. To achieve zero population growth by 2000 the total fertility rate throughout the world would have to drop far below replacement level, reaching 1.1 by 2000.

No matter how rapidly fertility declines in the less developed countries, their population growth in the coming decades is likely to be different from the growth in the developed countries. Even if current fertility does not change, the developed countries would have only about twice their present population in 2050, whereas a drastic decline in fertility would be needed in the developing countries if their population is to be in 2050 no more than twice what it is now. Even if the less developed countries were to achieve a net reproduction rate of 1 by 2000, they would grow by about 2.5 times by 2050. With a moderate but nonetheless meaningful fertility decline that brought a net reproduction rate of 1 by about 2050, the population of the less developed countries would be about 4.5 times what it is now.

In the more developed regions a population growth of from 15 to 25 percent can be expected by 2000, and by 2050 they could have from 30 to 60 percent more people than they did in 1970. The population of the less developed countries can be expected to grow by 80 percent and perhaps by 100 percent between now and 2000. If current fertility remains constant and mortality declines, the population of the less developed countries could be more than 10 times its present size by 2050.

A logical consequence of the different growth rates of the two main segments of the world's population will be a change in their size ratio. The extent of the change will depend mainly on how fast fertility declines in the less developed countries. The shift from the present ratio of 30 : 70 to something like 20 : 80 by 2050 could also be affected by the fact that several of the nations now classified as less developed may by 2050 have become more developed.

A socioeconomic classification employed by the United Nations divides the world into eight major areas: East Asia, South Asia, Europe, the U.S.S.R., Africa, North America, Latin America and Oceania. Current demographic conditions differ from one region to another. Latin America probably had the highest growth rate in the late 1960's (2.9 percent per year), fairly high fertility (a total fertility rate of 5.5 from 1965 to 1970) and relatively low mortality (a female expectation of life at birth of 62.5 years). Africa had the highest total fertility rate (estimated at 6.4), and yet its population growth rate was 2.5 percent per year because mortality levels in Africa are the highest in the world. (Female expectation of life at birth was estimated at 44.5 years in the late 1960's.)

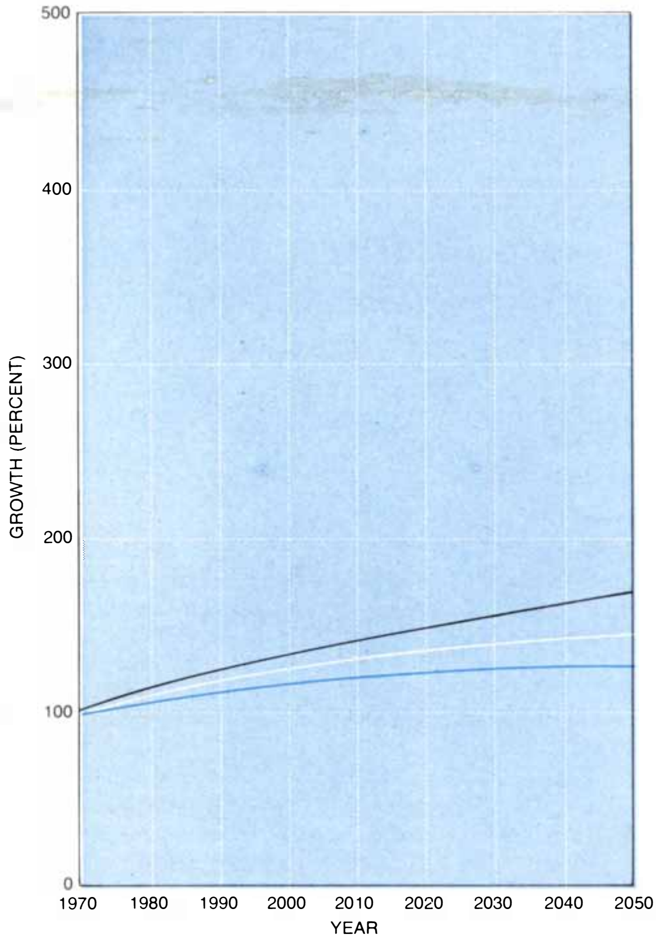
Clearly the major areas of South Asia, Africa and Latin America have a high built-in potential for further population growth. For example, with a fertility decline of the projection-5 type the population of South Asia could reach five times its present size by the middle of the next century.

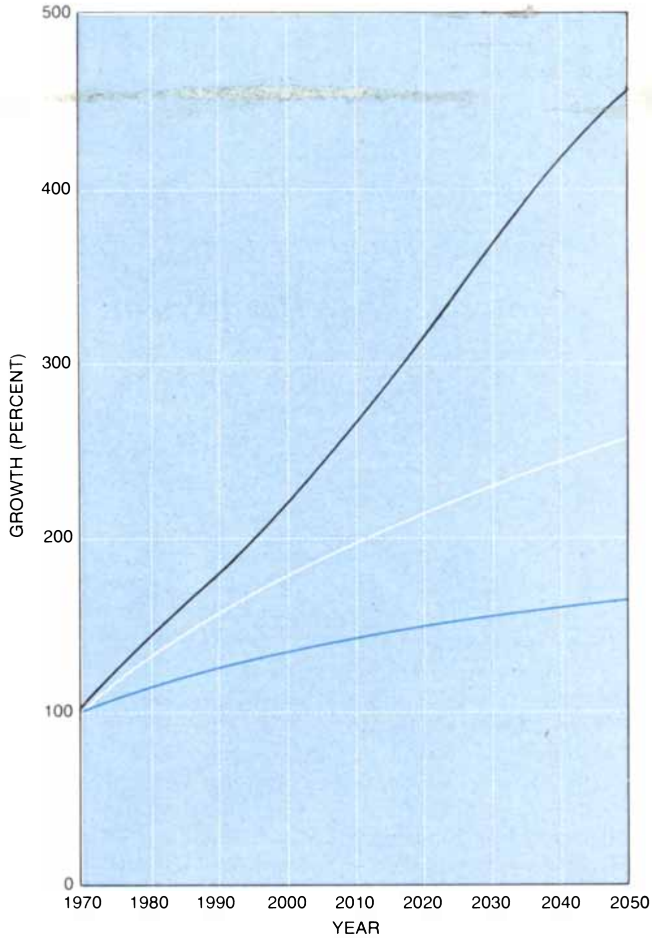
A considerable unevenness in the speed of population growth from region to region can be expected and will become evident in changes in the relative distribution of population. If growth proceeds along the lines of projection *4*, for example, the proportion of the world's population living in South Asia would increase from the current 30 percent to 40 percent, in Africa from 9 to 12 percent and in Latin America from 7 to 10 percent. By the same token the proportion of the population living in Europe would decline from 14 percent to 6 percent, in the U.S.S.R. from 8 to 4 percent and in North America from 7 to about 4.5 percent.

In the demographically less developed countries where there is a consensus that a slowing down of the population growth rate would be desirable, the starting date of the decline in fertility, the speed of the decline and thus the time when fertility levels approximating replacement are reached are crucial. For such countries it would be worthwhile to make moves toward this end as soon as possible. Because the potential of most less developed countries for population growth is so high, many of them could reach four times their present size within 50 years if they did not move toward the goal of a reduced growth rate and mortality rates continued to decline as they have in recent years.

Several demographically more developed countries, such as Poland and France, have expressed concern about a lagging growth of population or a lack of growth. In terms of total population the fear does not seem to be justified, since these nations will experience a growth of from 20 to 30 percent if they maintain replacement fertility from now on. Only a few nations—the Federal Republic of Germany, the German Democratic Republic, Austria and Sweden—can expect their population to remain at about the present level. A decrease of numbers seems unlikely but not impossible for most populations until at least the end of the 20th century.

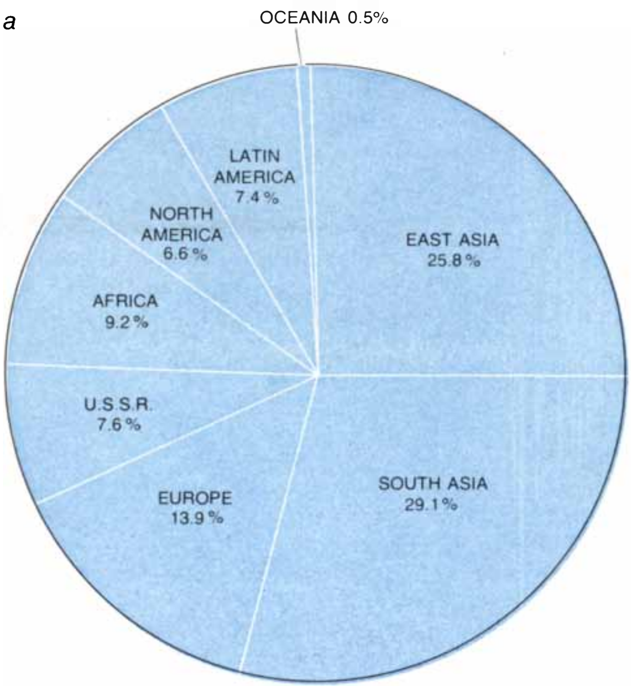
To achieve a nongrowing population in this century, even in most of the more developed countries, fertility would have to decline significantly below the replacement level. In the less developed countries such an event seems inconceivable during the coming 50 years, provided that mortality conditions do not deteriorate and there are no major disasters. The achievement of a nongrowing population is likely to take several decades in most of the more developed countries and many decades in the less developed countries.



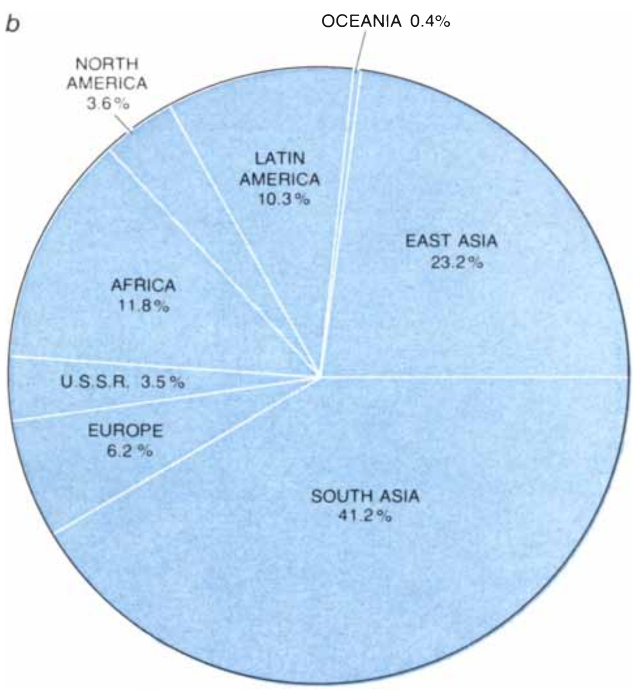


GROWTH PROSPECTS are charted for the developed countries (*left*) and the developing countries (*right*). The prospects are as envisioned in projections *1* (*color*), *3* (*white*) and *5* (*black*) of the five projections depicted in several preceding illustrations.

a



b



SHIFT IN PROPORTIONS of the population living in eight major areas of the world is portrayed as it could be expected under projection *4*, which assumes a net reproduction rate of 1 by the period 2020 to 2025. The present proportions (*a*) are compared with the projected proportions for the year 2100 (*b*). The eight areas are employed by the United Nations for socioeconomic classification.

The Fine Structure of the Earth's Interior

The waves sent out by earthquakes and nuclear explosions have been studied in detail with new seismometer arrays. They show, among other things, that the core of the earth has a solid kernel

by Bruce A. Bolt

The echoes of an earthquake ring deep into the earth. Depending on the medium through which they pass, the seismic waves are bent, speeded up, slowed down and in the case of some vibrations stopped altogether. When the faint echoes of the earthquake emerge at the earth's surface, they actuate the sensitive pendulum recorders known as seismographs. By correlating the records of seismographs at different locations it is usually a straightforward matter to establish the time and location of the original event, provided that it exceeded a certain magnitude.

During the past decade the worldwide interest in man-made seismic events (nuclear-test explosions) has led to a new generation of sensitive seismographs and the development of seismic arrays that have provided investigators with remarkably effective probes for studying the earth's interior. The delicate traces of the new instruments have enabled seismologists to confirm many of the older hypotheses of the earth's internal structure and have led to the discovery of fascinating new features. Just as radio telescopes have revealed celestial objects that were once invisible, the new generation of seismic instruments has detected fine details of our globe that were once unobservable.

As an example, geophysicists have speculated for years on the nature of the earth's center. Is it a solid or a liquid, and what is its density? The answers have been put on a much firmer basis within the past year. At the earth's center is a solid inner core with a density about 13.5 times the density of water. The radius of this core is some 1,216 kilometers, which makes it a little larger than the moon. The solid inner core is surrounded by a transitional region a little more than 500 kilometers thick. The transitional region in turn is surrounded by a liquid outer core about 1,700 kilometers thick.

Outside the liquid outer core is a mantle of solid rock some 2,900 kilometers thick, which approaches to within 40 kilometers of the earth's surface under the continents and to within 10 kilometers under the oceans. The thin rocky skin surrounding the mantle is the earth's crust. No drill has penetrated the earth's crust deeper than a few kilometers.

The basic technique in deep-earth "prospecting" is to measure the time of travel of a seismic wave from the instant of its generation by an earthquake or explosion to its arrival at a seismographic station. The distance between the seismic event and the station is expressed as the angle (designated delta) formed at the earth's center between radii drawn to these two points. Thus if the earthquake were anywhere on the Equator, it would be 90 degrees "distant" from a recording station at the South Pole. Similarly, if the earthquake were on the Equator in the Andes at 80 degrees west longitude, it would be 90 degrees distant from a recording station on the Equator in Gabon on the west coast of Africa at 10 degrees east longitude.

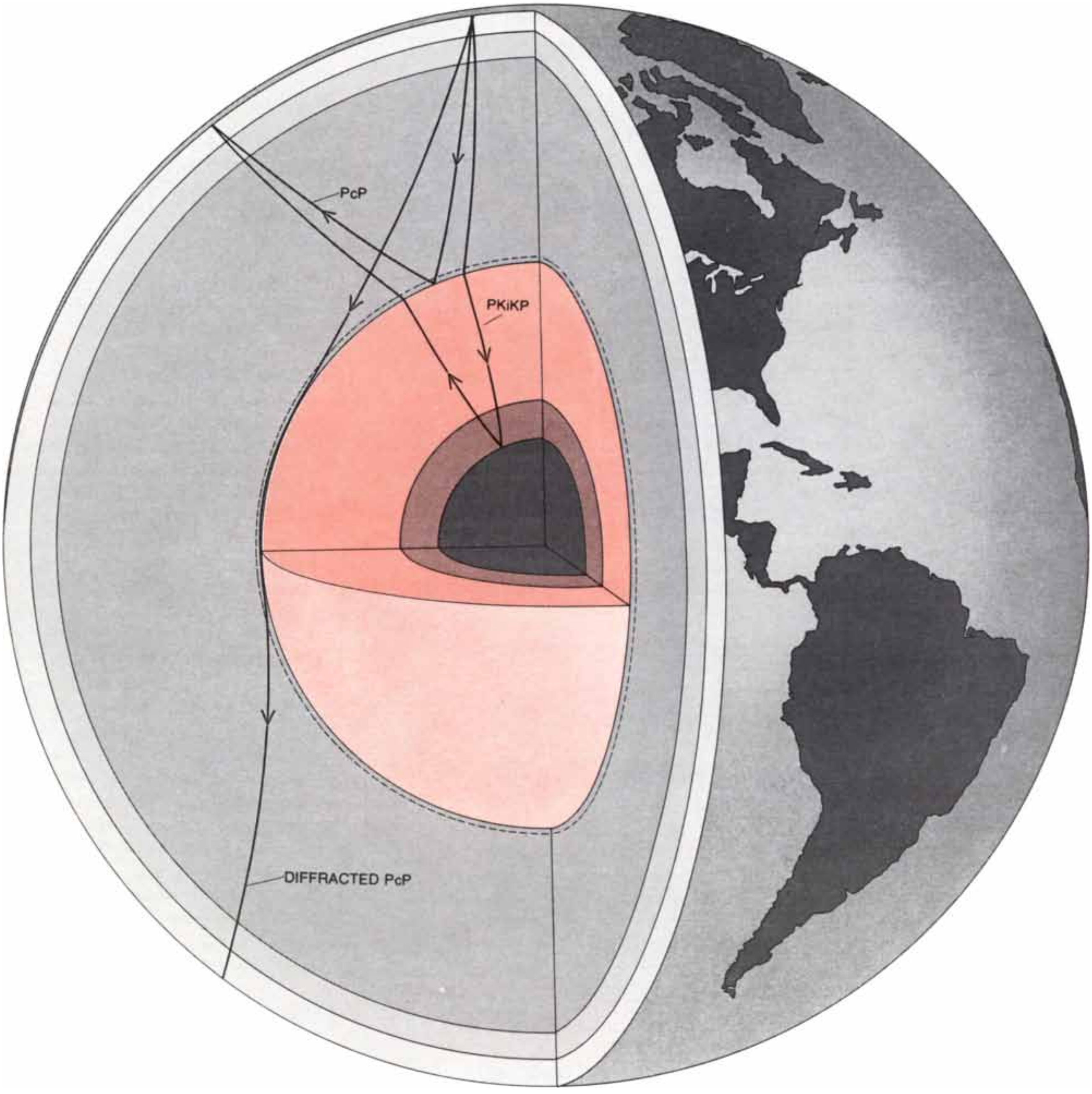
Earthquakes and explosions generate two types of wave that travel through the interior of the elastic earth. In seismology they are referred to as *P* (compressional) waves and *S* (shear) waves. The speed of the waves depends on the density and elastic properties of the rocks through which they pass. *S* waves are slower than *P* waves and do not pass through regions that are liquid. A wave that arrives at the seismograph without reflection is designated by a single *P* or *S*. If the wave is reflected once at the earth's surface, it is labeled *PP* or *SS*. In addition various letters and numbers can be inserted between the two (or more) *P*'s or *S*'s to indicate more specifically the region through which the wave has passed. Thus a compressional wave that passes through the earth's central core is called *PKP*, or *P′* for short. The wave that is reflected from the opposite side of the earth is labeled *P′P′*.

During the first half of this century seismologists painstakingly built up the curves of travel time against angular distance for *P* and *S* waves using hundreds of specially selected earthquakes from around the world. In 1906 the British seismologist R. D. Oldham first proposed that earthquake waves show that the interior of the earth is not featureless. He explained the pattern of arrival of the dominant *P* and *S* earthquake waves by postulating that the earth has a large central kernel. Although the details of his argument proved to be incorrect, Oldham's general conclusion was later verified in many ways. Somewhat stronger evidence of discontinuous features in the earth's structure was discovered in 1909 by Andrija Mohorovičić of Yugoslavia. He found that when he plotted the time-distance curve for *P* waves from Balkan earthquakes, there was a sharp bend near a distance of about 200 kilometers (an angular distance of about two degrees). Mohorovičić explained the bend by supposing that at a depth of about 50 kilometers there was an abrupt change in the properties of the earth's interior. That discontinuity, which separates the superficial crust from the mantle, was later found to be worldwide.

The detection of this discontinuity below the surface was soon followed by the firm location of the much greater discontinuity, nearly halfway to the earth's center, that had been proposed by Oldham. Seismologists had noted that at distances of up to 100 degrees *P* waves from earthquakes were recorded with more or less uniform amplitudes but that at distances beyond 100 degrees the amplitudes decayed dramatically. The short-period *P* waves appeared strongly and consistently again only at distances greater than 140 degrees, but at these larger distances they arrived some two minutes later than one would expect on the basis of simply extrapolating their velocity for distances less than 100 degrees.

An explanation that fitted the observed pattern of travel times was worked out numerically in 1913 by Beno Gutenberg of the University of Göttingen. He calculated that at a depth of about 2,900 kilometers the velocity of *P* waves falls precipitously by about 40 percent. This discontinuity marks the structural boundary between the mantle and the core. Seismograms show that *S* waves can travel anywhere above this major boundary, indicating that the mantle material must be rigid. For at least 2,000 kilometers below the core boundary, however, no *S* waves have been observed to propagate. For this reason and others the outer 2,000 kilometers of the core are regarded as being molten.

One might expect that a sharp boundary between the mantle and the core would reflect seismic waves. Indeed, such echoes are clearly observed on seismograms. A direct reflection of *P* waves from the boundary between the mantle and the core is designated *PcP* [*see illustration below*]. Echoes of the *PcP* type provide a direct means of determining the depth of discontinuities in the deep regions of the earth.



CROSS SECTION OF THE EARTH is based on the most recent seismological evidence. The outer shell consists of a rocky mantle that has structural discontinuities in its upper part and at its lower boundary that are capable of reflecting or modifying earthquake waves. Below the mantle an outer fluid core surrounds a solid kernel at the earth's center; between the two is a transition shell. The paths taken by the three major kinds of earthquake wave are depicted. The waves reflected from the outer liquid core are designated *PcP*; the waves reflected from inner solid core are *PKiKP*; the waves that creep around the liquid core are diffracted *PcP*.

In the 1920's, with the aid of somewhat improved seismographs and more refined earthquake surveillance, it became possible to detect the onset of delayed waves of the *P* type at distances between 110 degrees and 140 degrees. At such distances the waves evidently passed through the core and were of the *PKP* type. Inge Lehmann of Denmark suggested in 1936 that the pattern of travel times for these *PKP* waves could be explained if the core consisted of two regions, an outer one and an inner one. This notion was endorsed by Gutenberg, who now was working with Charles F. Richter at the California Institute of Technology, and independently by Harold Jeffreys of the University of Cambridge. In one of the most impressive sustained studies ever carried through in the physical sciences the two Cal Tech seismologists and Jeffreys (who worked in the early stages with K. E. Bullen) independently computed average velocity distributions for the whole of the earth's interior based on analyses of thousands of seismographic records of *P* and *S* waves. In general these independent solutions agreed as well as the measurements of that time allowed.

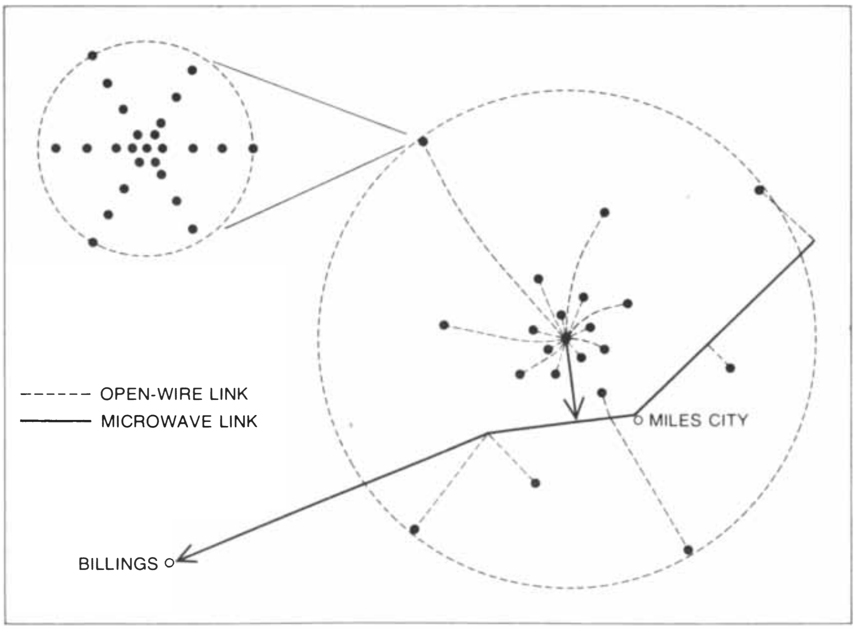
With the discovery of the inner core all the major boundaries within the earth had apparently been located. Other discontinuities were suggested from time to time to account for observed discrepancies in seismograms, but by and large measurements from the available instruments did not have the resolving power to clinch the case for further significant worldwide deep structures. In the period immediately following World War II, however, many seismologists became convinced (particularly from studies of seismic waves of long wavelength traveling around the earth's surface) that the structure of the upper few hundred kilometers of the earth was complex and also that it varied from place to place. The velocities of *P* and *S* waves in many geographical regions inexplicably decreased in a layer below the crust.

Around 1960 observational seismology took a major leap forward. Largely as a consequence of the attempt by several countries to find ways to discriminate between underground nuclear explosions and natural earthquakes, seismology was transformed from a neglected orphan of the physical sciences into a family favorite. A global network of more than 100 standardized sensitive seismographic units was established with U.S. support, and many other earthquake observatories were modernized throughout the world. Arrays of seismographs, comparable to giant radio antennas, were constructed by the U.S., Britain and other countries. In these seismic arrays the seismographs are connected in such a way that microseisms—the random small quivers of the earth—can be filtered out. One such giant antenna is the Large Aperture Seismic Array (LASA) located near Billings, Mont. It consists of 525 linked seismographs distributed over an area 200 kilometers in diameter [*see illustration on this page*].

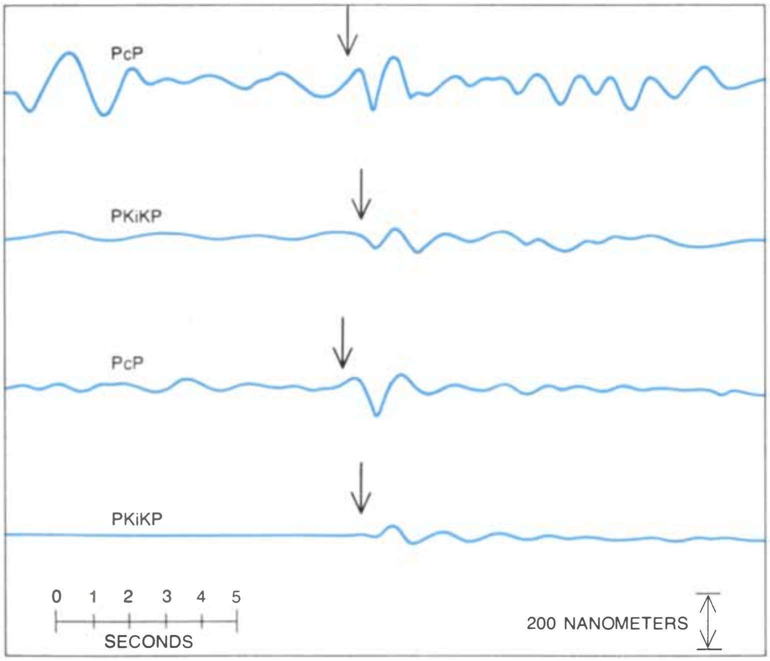
In the more than three decades following Lehmann's suggestion that the earth's core might itself have a core, no unequivocal evidence on the nature of the inner core's boundary had ever come to light. The evidence was finally supplied by LASA. In 1970 E. R. Engdahl of the National Oceanographic and Atmospheric Agency, Edward A. Flinn of Teledyne Incorporated and Carl F. Romney of the Air Force Technical Applications Center announced that the Montana array had detected the echoes, designated *PKiKP*, that had bounced steeply back from the boundary of the inner core. The source of the echoes was underground nuclear test explosions in Nevada as well as earthquakes [*see top illustration on opposite page*]. There were two immediate conclusions. First, the inner core has a sharp surface. Second, its radius is within a few kilometers of 1,216 kilometers (a value, incidentally, that I had predicted eight years earlier on the basis of other seismological evidence).

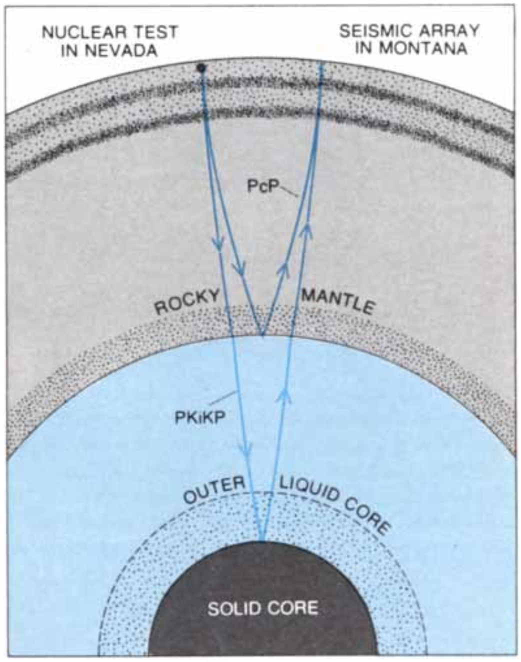
Even more can be done with these remarkable observations. By comparing the relative strengths of the *PcP* and *PKiKP* pulses one can calculate the density of the rocks at the top of the earth's inner core. There are a number of uncertainties in such calculations. The equations assume that the speeds of *P* waves are well determined throughout the earth and that the densities of the materials on each side of the boundary between the mantle and the core are known. Some recent calculations I have made in collaboration with Anthony I. Qamar indicate that the density at the center of the earth cannot be much greater than 14 times the density of water. Earlier estimates had ranged as high as 18 times the density of water. Our value agrees quite well with the density that iron is estimated to have when it is subjected to the pressure that exists at the earth's center.

Let us now turn our attention from the core to the fine structure of the upper part of the earth's mantle, the region just under our feet, so to speak. The presence of great mountain ranges, mid-ocean ridges and deep ocean troughs, together with a considerable amount of geological and geophysical evidence, indicates that large portions of the upper

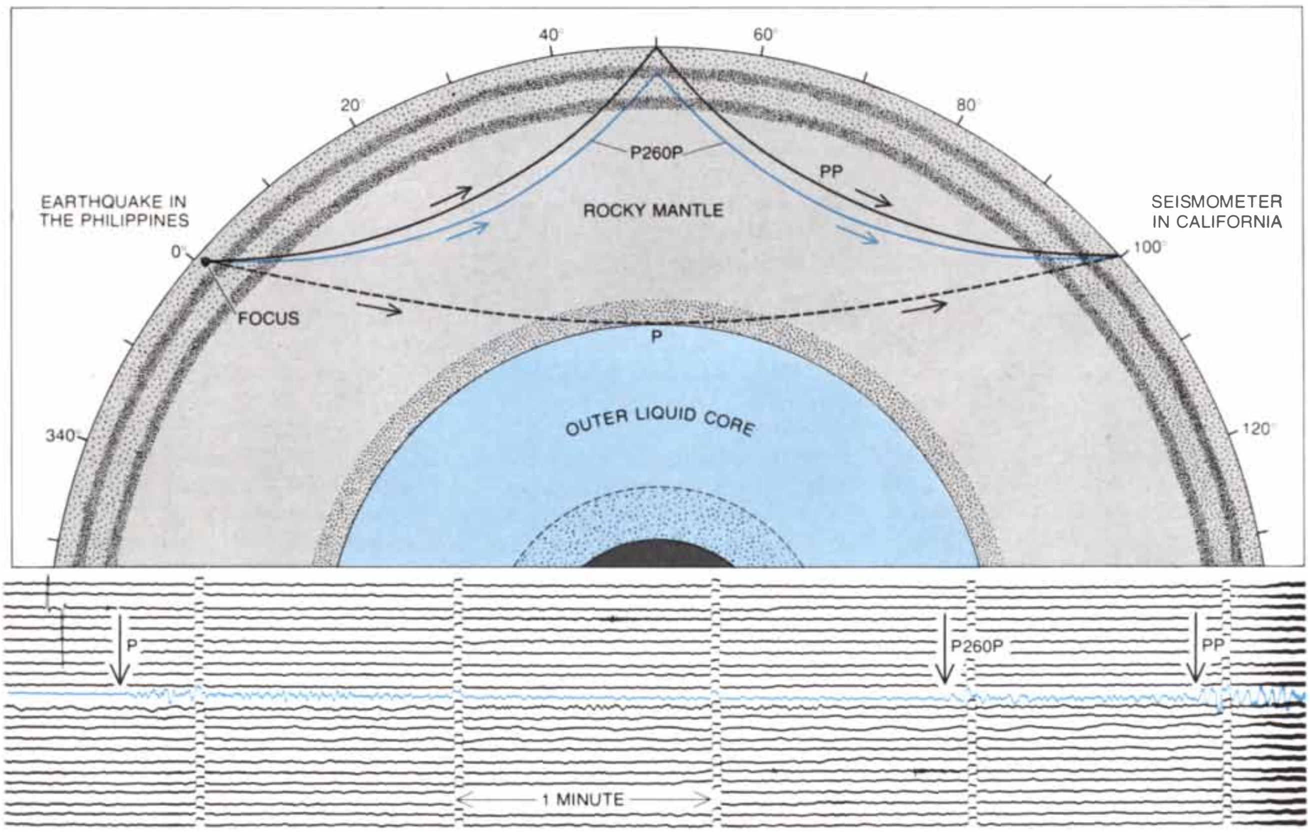


LARGE-APERTURE SEISMIC ARRAY (LASA) was installed near Billings, Mont., in the mid-1960's by the Department of Defense. One of the largest of more than 100 seismographic stations established throughout the world with U.S. support to detect underground nuclear explosions, it has provided much new knowledge of the earth's interior. LASA consists of 525 linked seismometers grouped in 21 clusters. In each cluster 25 seismometers are arranged as shown at the upper left. The array covers an area 200 kilometers in diameter.

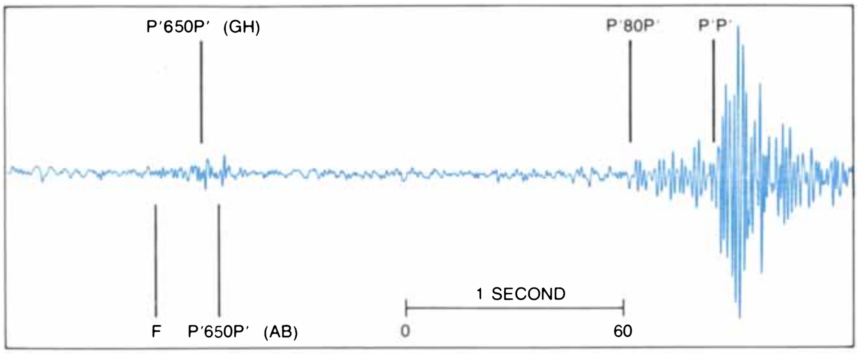




UNDERGROUND NUCLEAR EXPLOSION in Nevada on January 19, 1968 (code-named "Faultless"), produced the traces shown at the left on LASA seismometers. Time proceeds from left to right. The vertical scale shows the magnitude of ground movement involved; 200 nanometers is only half the wavelength of violet light. The *PcP* echoes from the outer core (*diagram at right*) are the closest to a straight-down-and-back path yet reported. Angular distance between explosion and recording instruments was only 11 degrees; this was the angle between lines from two points to the earth's center. *PKiKP* pulses represent echoes from solid inner core.



PHILIPPINE EARTHQUAKE generated waves that followed three distinct paths (*top*) before arriving at a high-gain seismograph located in an abandoned mine in the Sierra Nevada. The installation, the Jamestown Station, is operated by the University of California at Berkeley. In the seismograph a spot of lights is focused on a rotating drum, producing a series of parallel lines (*bottom*). The line in the Jamestown seismogram that contains three distinct echoes from the Philippine earthquake is shown in color. The *P* wave arrived first, followed some three minutes later by the echo *P*260*P* that probably bounced off a reflecting surface 260 kilometers deep in the earth under the Pacific Ocean. The reflection from the underside of the ocean floor, *PP*, arrived another minute later.



UNDERGROUND NUCLEAR TEST on October 14, 1970, on the Russian island of Novaya Zemlya produced this seismic trace at the Jamestown Station. The large phase *P′P′* was produced by a compressional wave reflected from the other side of the globe under Antarctica, as depicted in illustration on the opposite page. It was preceded by echo *P′*80*P′*, evidently reflected from a structure 80 kilometers below surface of Antarctica. Two minutes earlier still record shows doublet *P′*650*P′* (*GH*) and *P′*650*P′* (*AB*), evidently reflected from a layer 650 kilometers below surface of Antarctica. Origin of wave train starting at *F* is unknown.

part of the earth are in continuous slow motion. Indeed, it now appears that the surface of the earth is divided into six to eight large platelike regions that move with respect to one another. Africa is on one such plate; North America is on another. Dynamic processes operating particularly at the edges of plates would seem to account for much of the topographic relief on the surface of the globe. Most volcanoes are located and most earthquakes occur near the plate margins. One would expect large-scale movements of crustal masses to be reflected in the architecture of at least the upper few hundred kilometers of the earth. One would also expect this variable architecture to give rise to varied patterns on seismograms, and indeed different patterns are observed.

A few years ago, working with Qamar and Mary O'Neill at the University of California at Berkeley, I found a series of unexpected waves on seismograms. Measurements indicated that the waves might have been reflected from the underside of layers located perhaps hundreds of kilometers below the Mohorovičić discontinuity. The waves arrived at seismometers in the Berkeley network as much as 150 seconds before the corresponding waves reflected from the underside of the earth's surface.

We refer to these echoes generically as *PdP* waves. When we are able to calculate the depth below the surface of the layer from which the waves are reflected, we insert the figure in place of *d*. Thus a wave reflected from a layer 260 kilometers below the surface is designated *P*260*P*. A clear example of such an echo is one that was produced by an earthquake on May 22, 1972, near the Philippines and recorded at Jamestown, Calif., near the Nevada border [*see bottom illustration on preceding page*]. The distance between the focus of the earthquake and California is about 100 degrees.

Such reflections do not always show up on seismograms, and for this reason and others some seismologists have suggested that the wave paths are not as symmetrical as we have proposed. Notwithstanding the reservations of our colleagues, we are satisfied that many observations of *PdP* waves can be explained in terms of a roughness of the rocky material in the upper part of the earth's mantle that is capable of producing reflections. It is noteworthy that a significant discontinuity in the upper mantle was suggested as far back as 1926 by Perry Byerly of the University of California at Berkeley on the basis of travel times for waves produced by an earthquake in Montana. Byerly's plot of observed travel times against distance showed peculiarities that could be explained if the *P* waves had encountered some kind of surface at a depth of about 400 kilometers.

Our interpretation has been strengthened by slightly different types of analyses. In 1968 R. D. Adams of the Seismological Observatory in New Zealand and a year later Engdahl and Flinn in the U.S. independently observed small waves that arrived slightly earlier than the usual *P′P′* echoes. Ordinary waves of the *P′P′* type make the long journey from the focus of an earthquake to the other side of the earth and are reflected back to a station in the same hemisphere as the earthquake, having passed through the core twice. Adams and Engdahl and Flinn interpreted the precursor waves as *P′P′* waves that did not quite reach the opposite surface of the earth but were reflected back from a discontinuity in the upper mantle.

Waves of the *P′P′* type are particularly useful for probing the earth's structure. Their path is so long that they arrive some 39 minutes after they have begun to be generated by an earthquake. Therefore when they reach the seismograph most of the other waves sent out by the earthquake have already arrived at the observatory and the instrument is quiescent.

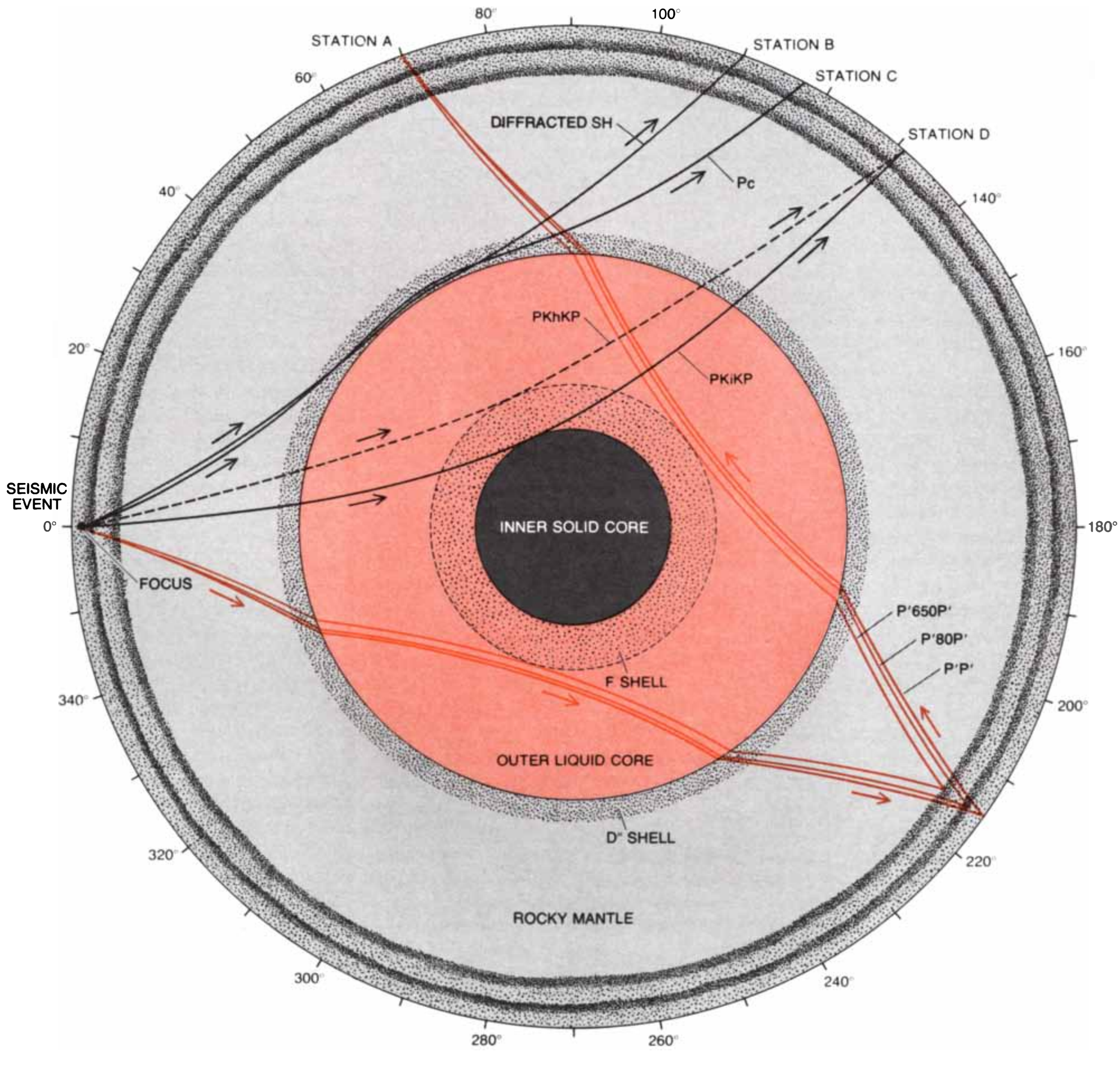
A particularly striking example of multiple long-distance reflections was provided by an underground nuclear explosion at the Russian test site in Novaya Zemlya on October 14, 1970. The *P′P′* waves passed through the earth's core, were reflected under Antarctica and returned to the Northern Hemisphere. In a recording made at Jamestown the main echo *P′P′* is the most prominent feature on the seismogram [*see illustration on this page*]. About 20 seconds before the onset of the large *P′P′* reflections a train of much smaller waves begins that can be explained as reflections from the underside of layers located in the 80 kilometers of rock below the surface of Antarctica. These forerunner waves are thus designated *P′*80*P′*.

As the eye scans the seismogram further from right to left only inconsequential waves are seen for more than a minute and a half; they are minor jiggles continuously produced by the background microseismic noise of the earth. All at once, almost precisely two minutes before the first *P′*80*P′* waves, one can see a beautiful doublet: two sharp peaks, separated by a few seconds, that stand out clearly above the background shaking. These sharp pulses agree nicely with the expected travel time of rays reflected by a layer located some 650 kilometers below the surface of Antarctica; hence they are designated *P′*650*P′*. The presence of a doublet means that there was only a slight variation in the paths of the two rays reflected from the 650-kilometer layer. Presumably one of the rays entered the transition layer, or *F* layer, that is thought to lie between the inner solid core and the outer liquid core and the other ray did not [*see illustration on opposite page*].

Many other examples have now been reported, particularly by James H. Whitcomb of the California Institute of Technology, of *P′P′* waves arriving earlier than one would expect if the upper mantle of the earth were uniformly smooth. Most seismologists now agree that these precursor waves at least indicate the existence of a rather sharp boundary at a depth of 650 kilometers and that the boundary is probably a worldwide feature.

It should be noted, however, that the particularly clear seismographic record of the Novaya Zemlya test explosion shows no spikes between the *P′650P′* waves and the *P′80P′* waves, as one would expect if there were a sharp reflecting surface at some intermediate depth, such as the 260-kilometer depth I have mentioned in connection with the record produced by the earthquake near the Philippines. The absence of intermediate spikes indicates either that such shallower discontinuities are not present under Antarctica or that they are not so easily detected by short-period *P* waves that pass through the earth's outer core because the shallower discontinuities are less sharp than the one at 650 kilometers.

Although the geophysical extent and the precise depth of some of the abrupt changes of structure in the upper mantle remain indefinite, the newly studied class of *PdP* and *P′dP′* reflections strongly indicates that sharp variations in rock properties do exist in the top few hundred kilometers of the earth. That is just the region where, according to the new plate-tectonic view of the earth's crust, there must be a driving mechanism to account for the movements of continents and crustal plates. The mechanism most frequently proposed is the slow flow of viscous hot rock in the form of convection cells immediately below the plates in the upper mantle. It is not obvious how such a flowing region could be reconciled with the observations of sharp structural discontinuities. Geophysicists



SEISMIC-WAVE PATHS are superposed on a diagram showing the shells that make up the earth's interior. The three nearly parallel rays arriving at Station *A* represent the paths taken by seismic waves from the Russian underground nuclear test whose seismographic record is shown on the opposite page. The two rays arriving at Station *B* and Station *C* represent the paths of diffracted *SH* waves and *Pc* waves depicted in the illustration on the next page and in the top illustration on page 31. Studies of such waves clarify the nature of *D″* shell. Deeper rays arriving at Station *D* represent the paths of waves shown in the bottom illustration on page 31.

are now busy examining whether or not convecting rocks can maintain sharp boundaries at certain depths.

Seismological evidence has long existed that suggests some kind of transition region or thin shell at the base of the rocky mantle just above the liquid outer core at a depth of about 2,900 kilometers. This inferred shell was given the label D'' even before much was known about it. Recently fresh light has been thrown on the velocities of *P* and *S* waves in the D'' shell, allowing some inferences about its properties.

In the first place modern instruments, unlike early seismographs, do detect clear waves of the short-period *P* type at distances well beyond 100 degrees. Such observations are surprising, because if the wave speeds remained constant or increased at the bottom of the mantle, the core would act as an opaque screen that would quickly cut off the direct, short-period seismic waves at about 105 degrees. Beyond this distance the seismographic stations would be in the shadow of the earth's core.

The actual situation can be compared to what happens when a shadow is cast on a wall by an opaque object illuminated by the sun. The shadow is not quite sharp because the light waves are diffracted, or bent, by the edge of the object. In the earth some seismic waves get diffracted into the shadow produced by the earth's core. Regardless of the actual velocity of *P* waves at the base of the mantle, therefore, the shadow region is always dimly illuminated with earthquake waves. It turns out, from work done by Lehmann and by Robert A. Phinney of Princeton University, as well as from the studies of others, that the strength of the short-period waves actually observed in the shadow of the core is greater than it should be if diffraction in a zone of constant velocity were the only mechanism operating.

Recently I have completed a study of waves designated *Pc* that have traveled out to more than 118 degrees and have arrived at seismographic observatories in the core shadow zone with substantial strength [*see illustration below*]. The time of travel and the amplitudes of such waves indicate that they have passed through a shell where the speed of propagation is perceptibly less than the speed that prevails only 100 kilometers or so above the core boundary. Putting the evidence together, the best explanation seems to be that the *P*-wave velocity drops by a few percent in the D'' shell above the earth's core.

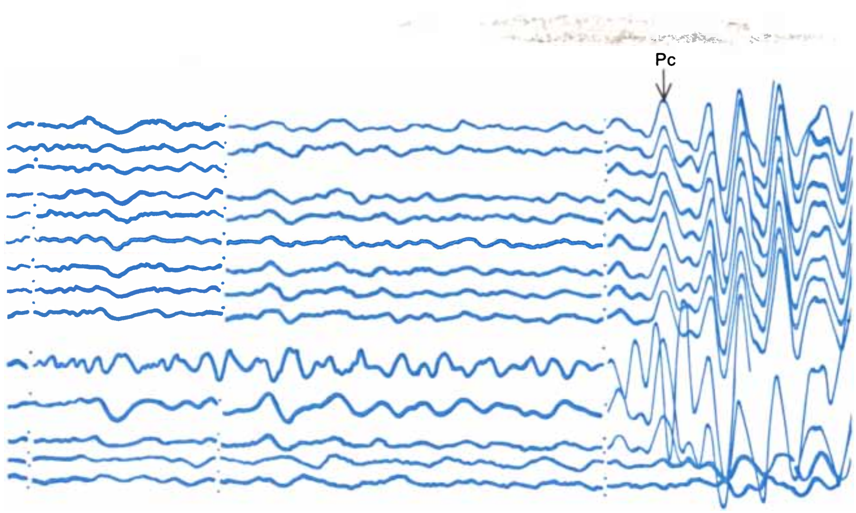
The evidence from the *Pc* waves does not stand alone. Because the outer core is liquid, certain kinds of shear (*S*) waves are inhibited from creeping around the core boundary; their energy leaks off into the core in the form of *P* waves. There is, however, one type of shear wave, called the *SH* wave, that cannot produce *P* waves in the liquid core. Its energy gets trapped at the base of the mantle, and thus it can travel great distances. Very large *SH* pulses have been recorded at distances ranging from 90 degrees to more than 115 degrees from an earthquake [*see top illustration on opposite page*]. From the measured travel times one can derive the velocity of the shear wave as it travels around the core boundary. Calculations carried out in recent years by John Cleary of the Australian National University, by Anton L. Hales and J. L. Roberts of the University of Texas at Dallas, by Mansour Niazi, then working at Berkeley, and by others indicate that the S-wave velocity decreases at the base of the mantle in the D'' shell in the same way that the *P*-wave velocity does.

What kind of earth model can be based on the fine-structure results I have been discussing? The model should also take other facts into account, notably the observed periods of the earth's free oscillations [see "Resonant Vibrations of the Earth," by Frank Press; SCIENTIFIC AMERICAN, November, 1965]. Such a model was computed at Berkeley in 1972. Named CAL 3, it shows the average variation in *P*-wave and *S*-wave velocities and in rock density from the surface of the earth to the center [*see illustration on page 33.*]

What does the low velocity in shell D'' mean? One suggestion is that in this narrow shell the mantle rocks become less rigid as they approach the liquid core. Another possibility is that the rock composition changes slightly in the D'' shell. Below the mantle-core boundary the measured physical properties indicate a material that consists mainly of iron. Perhaps there is about 10 percent more iron in the D'' shell than there is in the rocks above it. Such iron enrichment may represent iron that never settled into the core during the formation of the earth or it may represent liquid iron that has diffused outward to form an alloy with the solid mantle.

Currently there is controversy concerning the possible presence of bumps, perhaps 500 kilometers from one side to the other, on the boundary between the mantle and the core. The existence of such bumps has been suggested by Raymond Hide of the British Meteorological Office and by others to explain the variations in the strength of the earth's magnetic and gravity fields as they are measured at different points on the earth's surface. The only way to "X ray" the earth for such fine structure is to use short-period seismic waves that interact with the boundary of the core. It can be calculated, however, that undulations on the mantle-core interface of less than 10 kilometers cannot easily be resolved with waves of the *PcP* type that return to the surface after a single reflection.

Fortunately, with the advent of sensitive seismographs a method has come into being that has improved resolving power. Some of these seismographs are at sites so quiet that they regularly de-



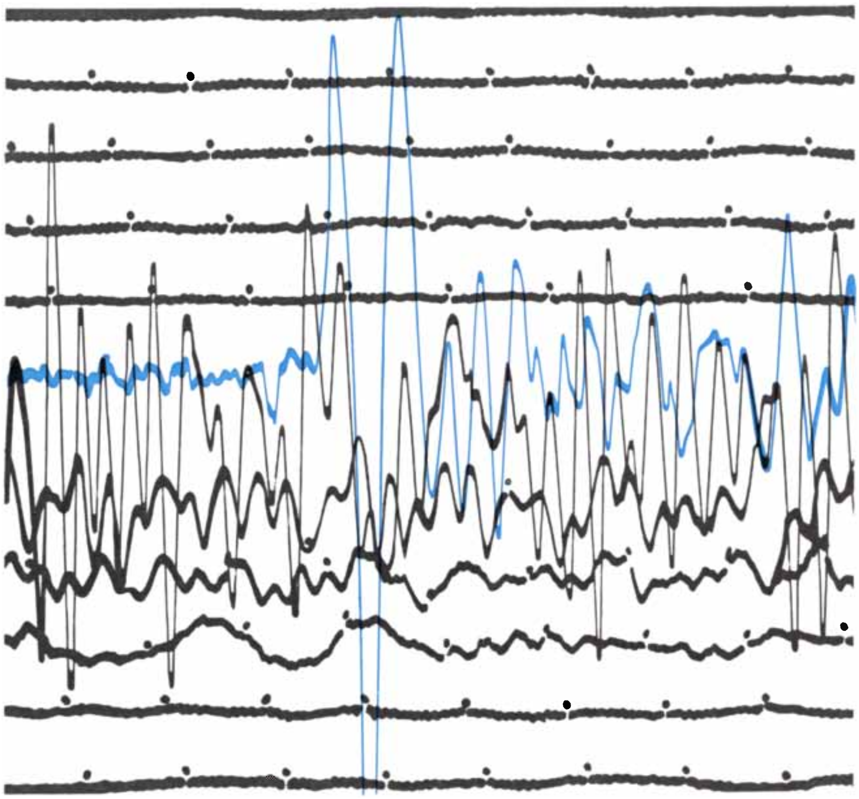
Pc WAVES FROM AN EARTHQUAKE in the South Sandwich Islands were guided with only small loss around the earth's core by the D'' shell. They were recorded at an angular distance of 118 degrees by an array in Uinta Basin of Utah. Here each wavy line is produced by a separate seismometer. Distance between breaks on the bottom line is 10 seconds.

tect waves that have been trapped inside the core and internally reflected four to seven times before emerging and returning to the surface. A wave that has been internally reflected seven times is designated *P7KP* [*see top illustration on next page*]. Such multiple reflections had been predicted but were never seen until the advent of the modern sensitive seismograph and of seismic arrays. It was a great thrill when we scanned along a seismogram made at Jamestown of an explosion on Novaya Zemlya and there —Eureka!—at the travel time predicted for a wave reflected seven times was an unmistakable tiny pulse nestling in the valley of microseismic background noise [*see bottom illustration on next page*].

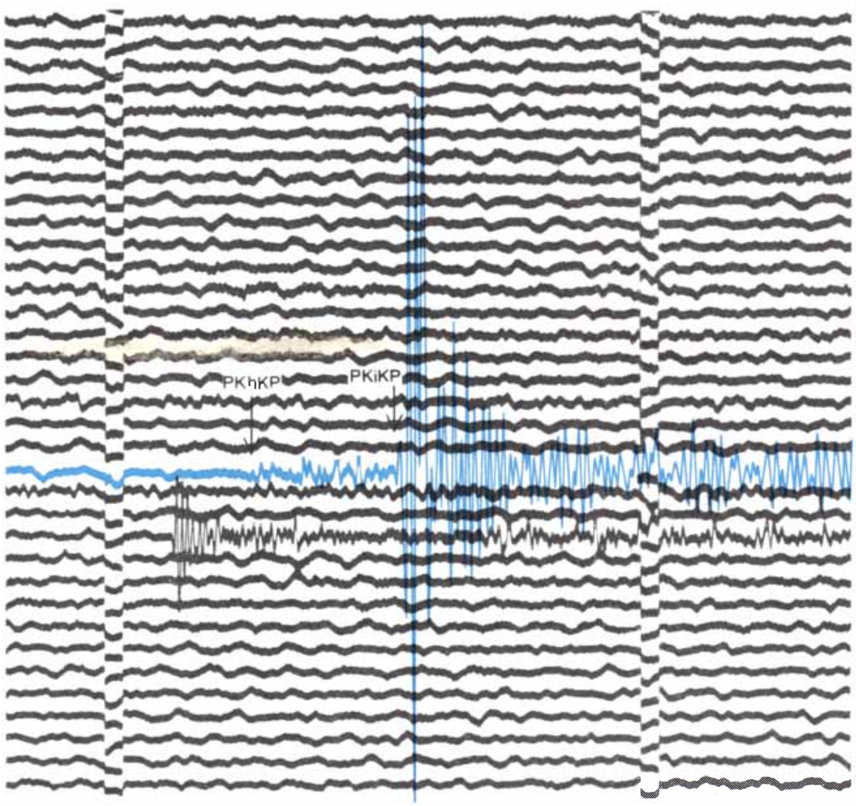
It is doubtful that we will ever see a natural earthquake generate such a clear example of a *P4KP* or *P7KP* wave. Sharp onsets can be seen when the source is a nuclear test because the energy is released explosively in a way that is simpler than the way energy is released in most earthquakes. In time planned experiments using nuclear explosions, set off in particularly favorable locations and designed to present no hazard to the human environment, should provide still more sensitive probes of the fine detail of the interior of the planet.

What conclusions can be drawn on the nature of the mantle-core boundary from the records of *P4KP* and *P7KP* waves? First, the onset of the waves is quite abrupt. This confirms that the mantle-core boundary is a sharp discontinuity, perhaps extending over no more than two kilometers. Second, the additional distance represented by the extra three legs in the *P7KP* wave reduces its amplitude to about a third of the amplitude of the *P4KP* wave. The small size of the decrease implies that the liquid outer core transmits short-period *P* waves very efficiently indeed.

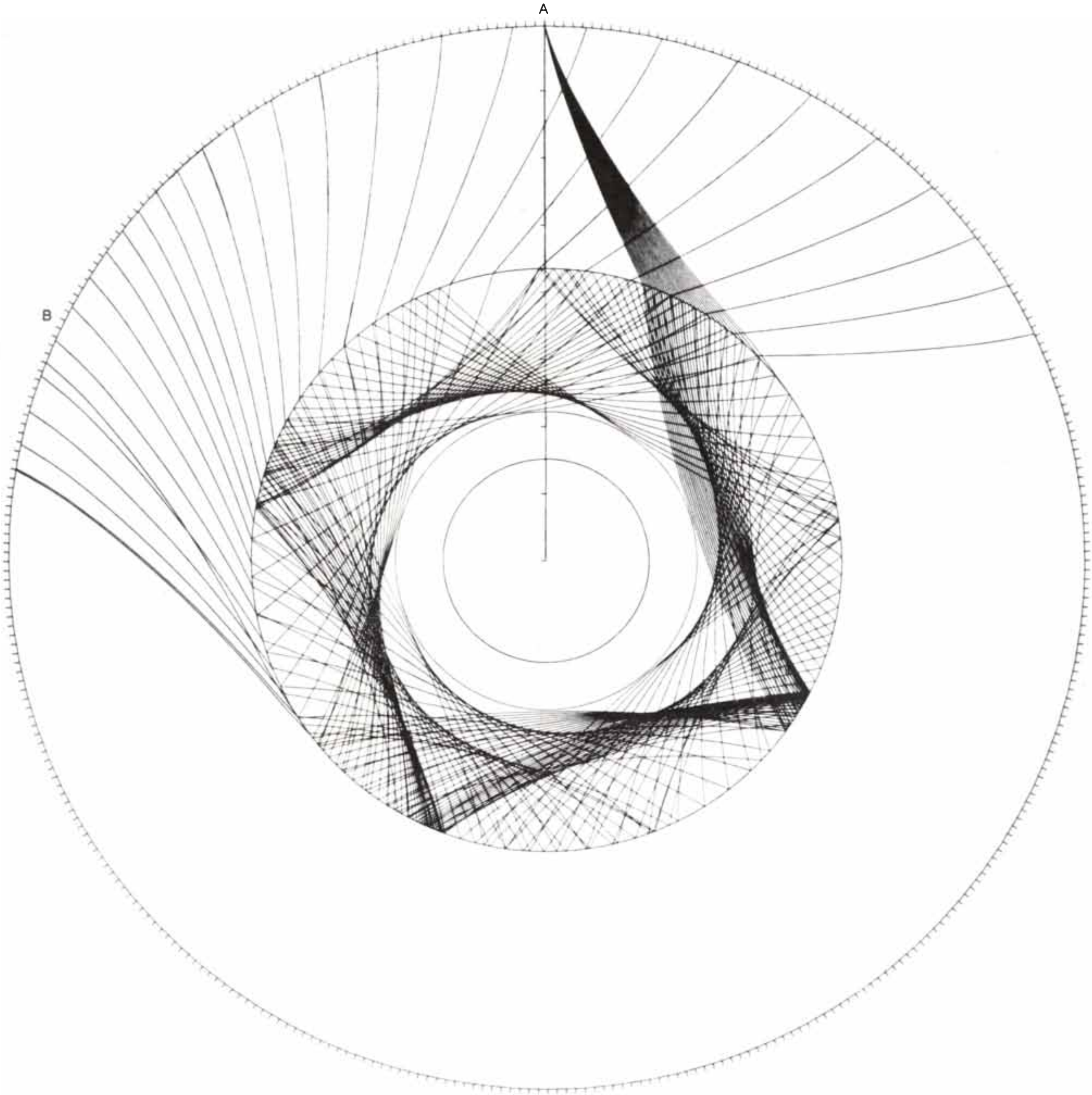
Finally, if the multiple reflections encountered topographic bumps on the mantle-core boundary more than two kilometers high, the travel times of the waves would be altered enough for the variations to be measurable. By comparing the travel times of many multiply reflected waves at the same seismographic station one should be able to derive the height of the bumps, if any, from the amount of variation in the times. Although studies of this kind are barely a year old, the present indications are that the variation is no greater than it is for waves that do not bounce from the core boundary. Therefore one can say tentatively that if there are topographic undulations, either their height is less than a few kilometers or, if their height is sig-



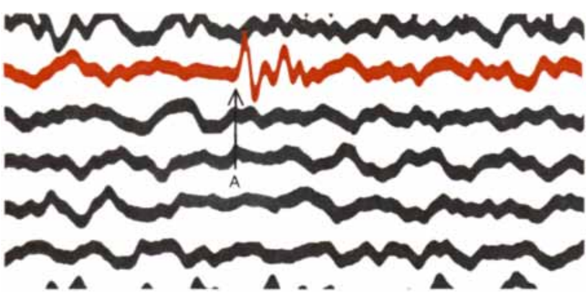
HORIZONTAL SHEAR WAVE of the type designated *SH* was produced by a major earthquake in Iran on August 31, 1968. It was recorded at a distance of nearly 100 degrees by a seismograph at Port Hardy in Canada that responds only to horizontal ground motion. The large *SH* pulse sent pen flying across recording drum, crossing traces made earlier and later.

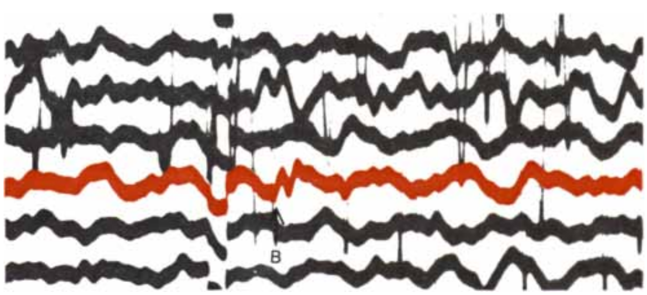


WAVES THAT PENETRATED EARTH'S CORE were created by a deep earthquake near Java and recorded at a distance of 132 degrees by a seismograph at Golden, Colo. The large-amplitude wave *PKiKP* was refracted through the earth's solid inner core. It was preceded 17 seconds earlier by smaller *PKhKP* waves, which were probably reflected from *F* layer.



MULTIPLE REFLECTIONS can result from *P* waves that get trapped inside the earth's liquid outer core. This computer plot depicts the paths of waves, generated by a seismic event at *A*, that have bounced inside the core seven times before reaching the surface, for example a station at *B*. Such waves are assigned a *K* for each bounce, hence the waves shown would be designated *PKKKKKKKP*, or *P7KP*. Computer program that produced ray paths was devised by C. Chapman for seismological average earth model called CAL 3.





FAINT PULSE OF *P7KP* ECHO can be seen at the right (*B*) in this seismogram made at Jamestown of an underground explosion on Novaya Zemlya in 1970. The stronger *P4KP* pulse, labeled *A* at the left, was recorded about 20 minutes earlier; it represents a wave that was reflected only four times inside the core of the earth. Such evidence indicates that the boundary of the core is rather smooth.

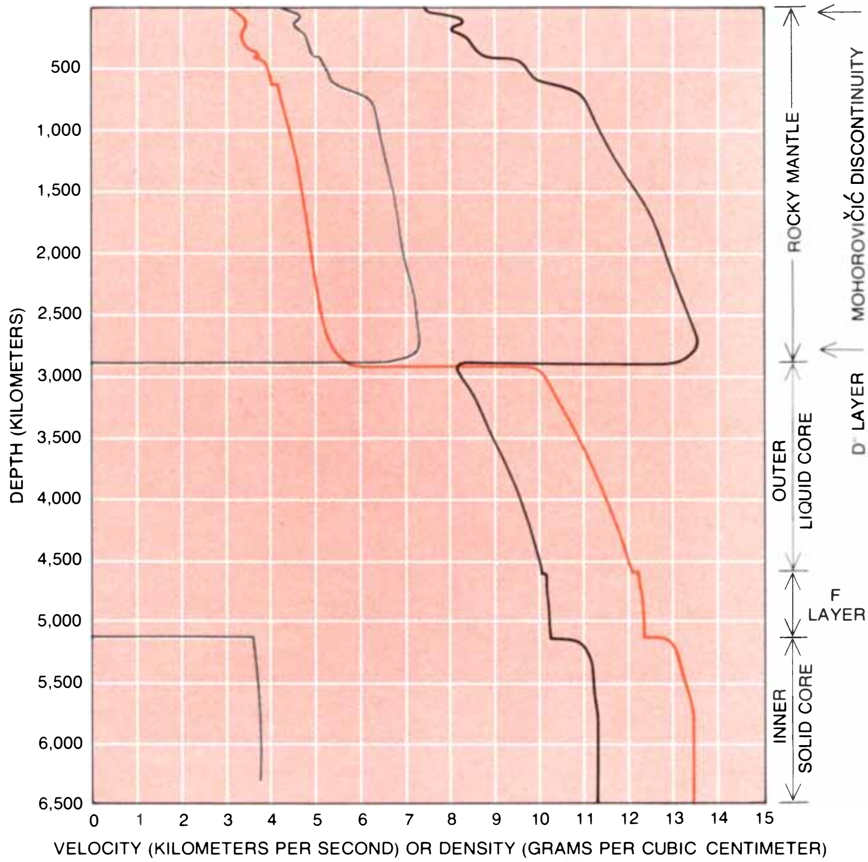
nificantly higher than that, they are few in number.

Let us now focus our attention still deeper in the earth. Even before 1940 the better-equipped seismological observatories reported tiny but clear *P*-wave precursors about 10 seconds before the main onset of the core waves (*PKP*) at distances of between 130 and 140 degrees. Quite unmistakable observations of such precursor waves were reported on seismograms from the nuclear explosions at Bikini Atoll in 1954. A recent typical precursor wave was generated by an earthquake near Java and recorded 132 degrees away at a seismic station in Golden, Colo. The precursor waves arrived 17 seconds before the onset of the much stronger *PKP* wave [*see bottom illustration on page 31*].

One straightforward explanation for these precursors occurred to me in 1962. Perhaps the wave velocities around the inner core were significantly different from the values that were accepted at the time. I suggested that between the inner and the outer core there might be a transition shell, designated *F*, that has a small jump in velocity at its upper boundary. After working out many mathematical models of the earth's core I found that, although other explanations are possible, the hypothesis of an *F* shell predicts travel times in reasonable agreement with all the *PKP* observations, including the precursor waves. Independent studies by Adams and M. J. Randall at the Seismological Observatory in New Zealand soon confirmed the general results I had obtained. The precursor waves reflected from the surface of the *F* shell were named *PKhKP*. Just in the past year an alternative explanation of some of the precursor waves has been put forward by R. A. Haddon of the University of Sydney. He suggests that they are *PKP* waves scattered by bumps on the mantle-core boundary.

There were several fruits of my proposal of 1962. Revised travel times for the core waves, based on the *F*-shell model, not only helped other seismologists to find more examples of the small *PKhKP* precursors but also helped in the search for the reflections (*PKiKP*) from the boundary of the inner core. It was these latter reflections that led to the conclusions that the inner core has a sharp surface and that its radius is about 1,216 kilometers.

An important by-product of the 1962 core-velocity estimates concerns the question of whether the inner core is solid or not. The notion that the inner core might be solid was put forward in 1940 by Francis Birch of Harvard University. Now, if the inner core is rigid, it should transmit not only compressional waves but also shear waves, giving rise to waves, designated *PKJKP*, observable under specially favorable conditions at the earth's surface.



SEISMOLOGICAL AVERAGE EARTH MODEL (CAL 3) was computed last summer in author's laboratory at the University of California at Berkeley. The model, which takes all available seismological information into account, is defined by the three curves. The colored curve shows the variation in the earth's density with depth. The solid black curve shows the average velocity of *P* waves. The gray curve shows the average velocity of *S* waves. Since *S* waves are not propagated through liquids, the gray curve is interrupted by the earth's outer liquid core. The energy in an *S* wave, however, can be transmitted as a *P* wave through the liquid core and then reconverted into an *S* wave for transmission through the solid inner core. Direct evidence for a solid inner core has only been obtained in the past year.

Working at the University of Sydney with the travel times of *P* waves and the wave velocities then assumed for the core, Bullen made estimates in 1950 of the travel times for *PKJKP* waves. There followed an enthusiastic but unsuccessful search for them. Seismologists at observatories around the world scrutinized seismograms to see if there were unexplained wiggles near the predicted times for *PKJKP* waves. In my own work on the core I estimated in 1964 an *S*-wave velocity in the inner core of about 3.7 kilometers per second. As a result the search for *PKJKP* waves was shifted to arrival times several minutes later than the ones that had first been tried.

Early last year Bruce Julian, David Davies and Robert Sheppard of the Massachusetts Institute of Technology provisionally announced that the large-aperture seismic array in Montana had recorded *PKJKP* waves close to the arrival time predicted by the 1962 model. If this announcement is confirmed, a long-sought key to the nature of the earth's kernel will have been found.

What further consequences follow from our new knowledge of the planet's fine structure? Many could be mentioned. For example, since wave speeds depend on rock densities, the revised *P* and *S* velocities in the earth make it possible to estimate the density variation with greater assurance, as incorporated in the CAL 3 model. Geochemists who try to describe the composition of the materials in the various shells will be aided by the more precise determinations of structural boundaries and density. Ultimately the entire body of geophysical and geochemical knowledge must be incorporated into a satisfying account of the evolution of the earth.

The Visualization of Genes in Action

The electron microscope reveals individual genes being transcribed into RNA and their RNA being translated into protein. The pictures look remarkably like diagrams based on genetic and biochemical data

by O. L. Miller, Jr.

Since the middle 1950's a major objective in biology has been to document and add detail to what is called the central dogma of genetics: that DNA is the hereditary material, that its information is encoded in the sequences of its subunits that constitute the genes and that this information is transcribed into RNA and then translated into protein. By a variety of remarkable biochemical and genetic techniques many of the steps in this process of information transfer and the substances and cellular elements involved in them were identified; the DNA-protein dictionary was worked out, and the effect of different conditions and foreign agents on transcription and translation was determined. The increasingly detailed picture of gene action that emerged was necessarily based largely on indirect and collective evidence, however; there was little direct evidence of how individual genes function.

At the Oak Ridge National Laboratory in 1967 my colleagues and I began attempting to make electron micrographs of individual genes in action. The success we have had indicates that electron microscopy is potentially a valuable tool for the study of cell genetics at the molecular level. Meanwhile one of the gratifying aspects of our work has been that many of our pictures bear an almost uncanny resemblance to diagrams of transcription and translation that have been published over the years in technical journals and in magazines such as *Scientific American*. In other words, the pictures tend to confirm what had been proposed through consideration of painstakingly accumulated quantitative data.

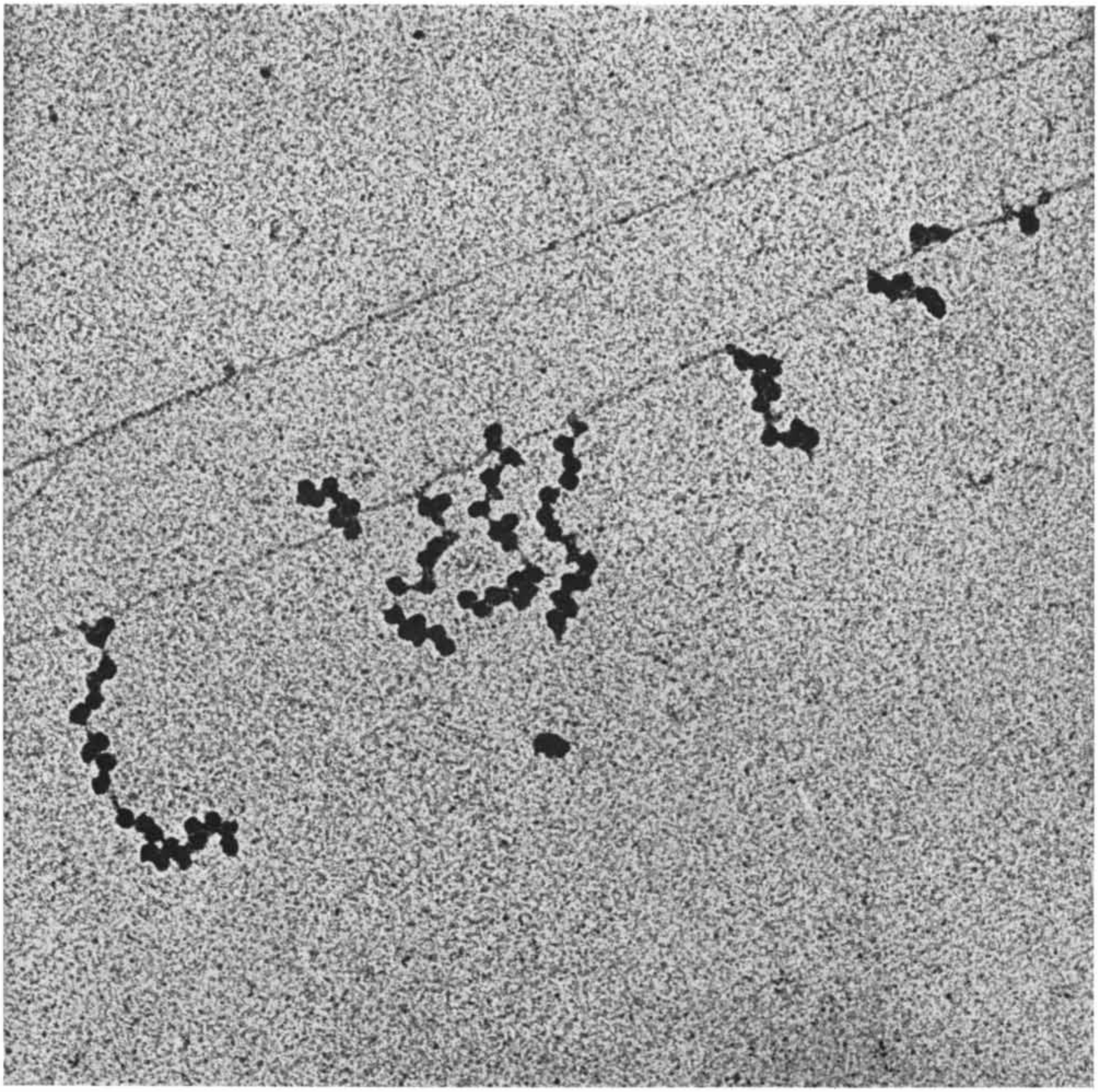
The transfer of information from the deoxyribonucleic acid (DNA) of the genes into protein is accomplished in a series of steps involving the transcription of DNA into three kinds of ribonucleic acid (RNA). One is messenger RNA, which serves as the template on which amino acid subunits are assembled into proteins in the translation step. Another is ribosomal RNA, a constituent of the structures called ribosomes on which translation is accomplished. The third is transfer RNA, which carries the amino acid subunits of protein to their proper site along the messenger-RNA template.

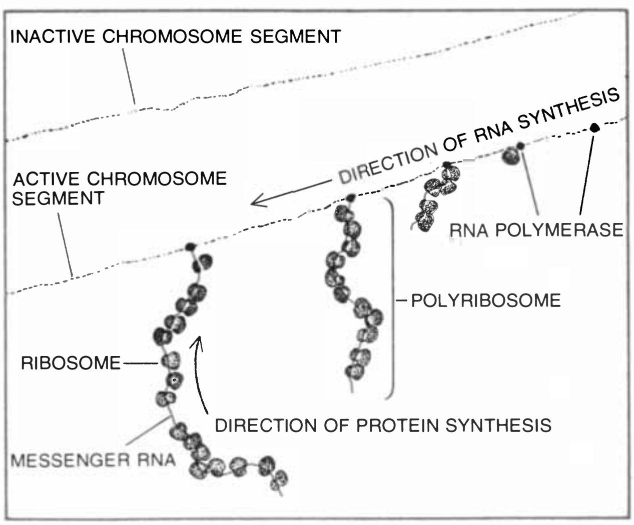
The steps involved in the transcription of DNA into RNA and the translation of RNA into protein are essentially the same in all living cells, but the temporal and spatial relations of the steps in eukaryotic cells, whose DNA is confined within a membrane-bounded nucleus during most of the cell cycle, are different from the steps in prokaryotic cells, which lack such a nucleus. In prokaryotes, which include the many species of bacteria, transcription and translation of messenger RNA occur at the same time and place. In eukaryotic cells, which range from single-celled organisms such as the paramecium to the cells of higher vertebrates, the two processes are separated in time and space: transcription of DNA into the various RNA's that accomplish protein synthesis occurs in the nucleus; then the RNA's migrate through the membranous nuclear envelope into the cytoplasm, the main body of the cell, where the machinery for protein synthesis is located [*see illustrations on page 36*].

We decided to begin by looking at the genetic machinery of a eukaryote: the oöcyte, or female reproductive cell, of amphibians such as frogs and salamanders. This type of oöcyte is a convenient cell with which to study genetic activity for a number of reasons; what was most important to us was that it would give us a good chance of seeing a great many identifiable genes, those that code for the two large molecules of RNA that are found in ribosomes. In most other eukaryotic cells these genes are found in clusters localized in nucleoli: dense bodies located at specific sites on certain chromosomes. In amphibian oöcytes large numbers of copies are made of the genes for ribosomal RNA, and these copies are in hundreds of nucleoli floating free and unattached to any chromosome. The result of this "amplification" of ribosomal-RNA genes is that a single oöcyte nucleus is equivalent to many hundreds of typical cell nuclei with respect to an identifiable genetic locus, and the amplified genes are not confusingly mingled with the other genes in the chromosomes.

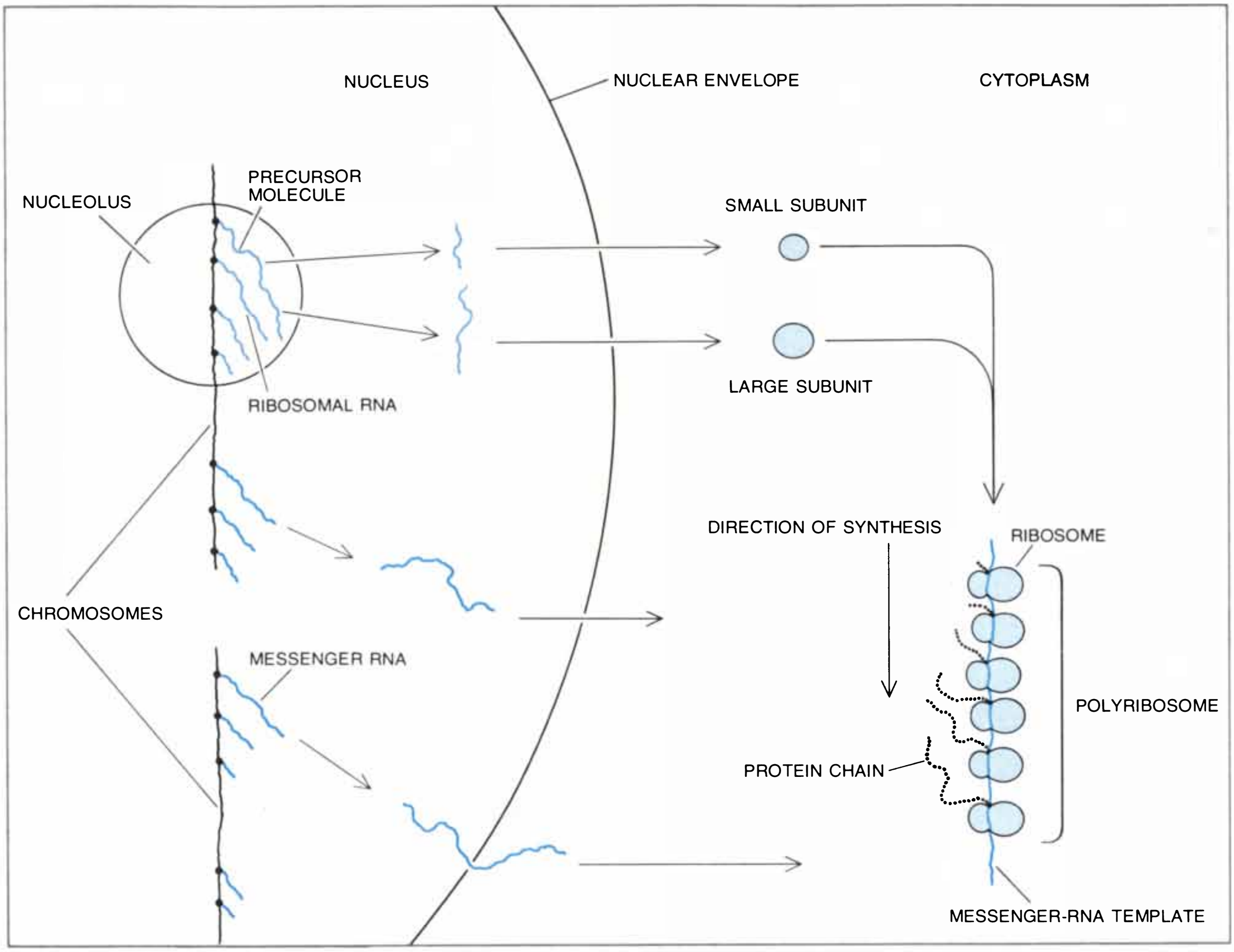
There are two other advantages in studying amphibian oöcytes. One is that while they are growing they are actively synthesizing messenger RNA, and their chromosomes are in a greatly enlarged and uncoiled "lampbrush" stage where we might be able to see structural details of DNA being transcribed into messenger RNA. The final advantage is that these oöcytes are remarkably large. In some species they are as much as two millimeters in diameter at maturity and have nuclei as much as a millimeter in diameter. With a low-power microscope and jeweler's forceps one can readily isolate individual nuclei and manipulate their contents to prepare them for electron microscopy.

Barbara Beatty and I found that if the contents of an oöcyte nucleus were isolated and then put quickly into distilled water, the granular outer layers of the extrachromosomal nucleoli rapidly dispersed, allowing the compact fibrous

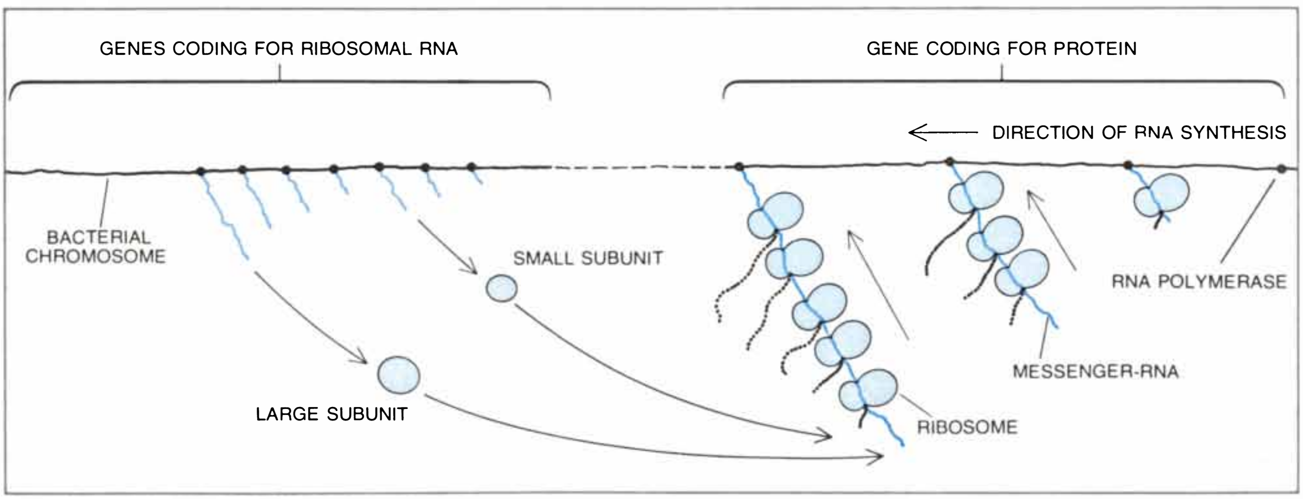




BACTERIAL GENE IN ACTION is enlarged 150,000 diameters in an electron micrograph made, like the others illustrating this article, by the author and his colleagues. The micrograph is interpreted in the somewhat simplified drawing at the right. One sees two segments of the chromosome of the bacterium *Escherichia coli*. The lower segment is active, that is, its DNA is being transcribed into messenger RNA and the RNA is being translated into protein. In the micrograph one can see molecules of RNA polymerase, the enzyme that catalyzes the transcription of DNA into RNA; one, at the far right, is at the approximate initiation site, where the transcription of each molecule of RNA begins. Successively transcribed messenger-RNA strands (*color in drawing*) in effect peel off toward the left; the longest one, now at the left of the micrograph, was the first to have been synthesized. As each RNA strand lengthens, ribosomes attach themselves and move along it, toward the chromosome, translating the RNA into protein (*not shown*).



TRANSCRIPTION AND TRANSLATION occur respectively in the nucleus and the cytoplasm of eukaryotic cells. Ribosomal RNA (*light color*) transcribed from genes in the nucleolus forms a precursor molecule that is cleaved within the nucleolus to form the two large RNA molecules found in the two units of a ribosome. (RNA's from the same precursor are actually unlikely to end up in the same ribosome.) Messenger RNA's (*dark color*) transcribed from other genes serve as the templates on which polyribosomes assemble amino acid subunits into protein chains. Both kinds of RNA are complexed with protein (*not shown*) as they are being synthesized. In amphibian oöcytes a large number of copies of ribosomal-RNA genes are also present in extrachromosomal nucleoli.



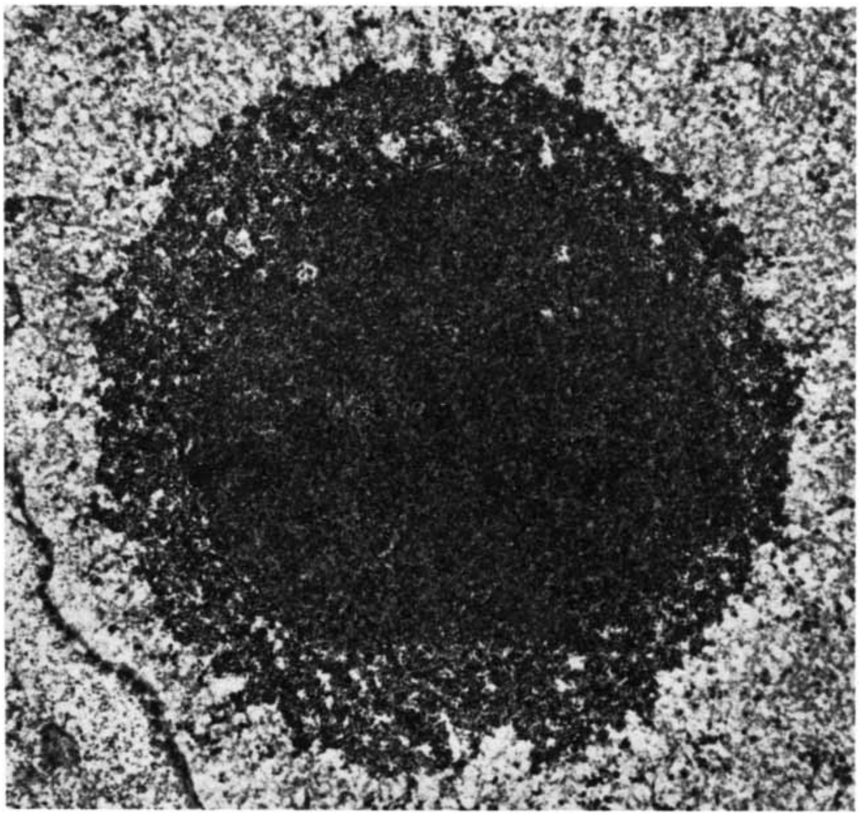
SIMULTANEOUS transcription and translation take place in prokaryotic cells such as bacteria, which have no true nucleus. Ribosomal RNA (*light color*) is transcribed from two genes, apparently contiguous, on certain segments of the single bacterial chromosome. Messenger RNA (*dark color*) is transcribed from genes at other sites and immediately translated into protein by ribosomes.

cores—which we hoped contained the genes we were looking for—to expand. We suspended such cores in a formalin solution to fix them and then spun them in a centrifuge tube where they were thrown against a thin carbon film on a wire-mesh specimen grid for an electron microscope. The problem was to spread the cores well without distorting them and destroying the pattern of gene activity we were looking for. We did that by treating the carbon films to make them hydrophilic, or water-attracting, and by rinsing the specimen grid in water containing an agent to reduce surface tension before drying it. Then we stained the specimen with a heavy metal to give it more contrast in the electron beam and made electron micrographs of the unwound cores.

Each unwound core appears as a long fiber in the form of a tangled, collapsed circle, along which there is a series of repeating shapes that we call matrix units [*see illustration on next page*]. Each of these units is about 2.4 microns long, and between each unit there is a segment of bare fiber that is usually somewhat shorter. Each matrix unit consists of a set of fibrils extending laterally from the long fiber. The length of the fibrils increases regularly along the fiber from one end of the matrix unit to the other, forming a structure with the outline of an arrowhead. All the matrix units on a single core fiber have the same polarity, that is, all the arrowheads point in the same direction. By treating the specimens with specific stains and enzymes we could identify the long axial fiber as a complex of DNA and protein and the lateral fibrils as a complex of RNA and protein. By means of autoradiography with radioactively labeled constituents of RNA we found that the matrix units were actively synthesizing RNA; the spacer segments between the matrix units were apparently not being transcribed. The nascent RNA molecules in the fibrils were complexed with protein that was made in the cell cytoplasm and that migrated into the nucleus and associated with the RNA molecules as they were being synthesized.

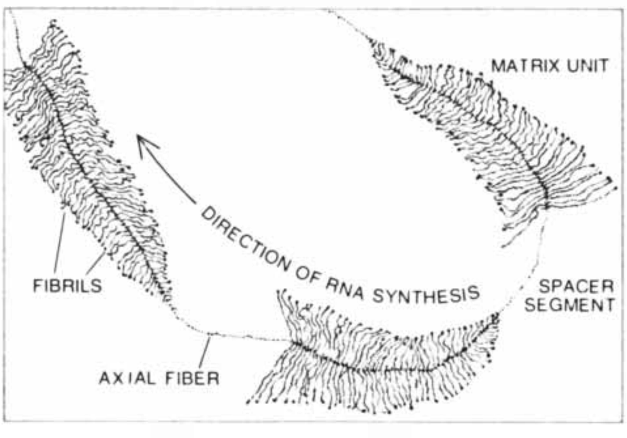
At about the same time we were making these observations some very pertinent biochemical information became available. Several investigators reported that the DNA in the extrachromosomal nucleoli of amphibian oöcytes contained subunit sequences that were complementary to the ribosomal RNA of ribosomes in oöcyte cytoplasm, thus confirming that the nucleolar DNA was the source of information for ribosomal RNA. The steps in the transfer of that information were also being elucidated in many laboratories. It turned out that in both mammals and amphibians the two large RNA molecules that are found in ribosomes are not transcribed from DNA separately but are snipped out of a single precursor molecule. (Subsequent studies have shown that this is probably true in all eukaryotes.) The precursor molecule has a molecular weight of about 2.7 million in amphibians, and about 2.7 microns of double-strand DNA would be necessary to code for such a molecule. The matrix units we were observing were about 2.4 microns long, a figure reasonably close to the 2.7 microns of DNA. Since extrachromosomal nucleoli had been shown to contain DNA coding for ribosomal RNA, and since our observations localized RNA synthesis in the matrix units of nucleolar cores, we concluded that the DNA fiber in each matrix unit was indeed a gene coding for a precursor molecule of ribosomal RNA, and that the lateral fibrils contained such molecules in the process of being synthesized.

On some fibers that were slightly stretched during preparation we could see something more: we could resolve the individual molecules of RNA polymerase, the enzyme that catalyzes the reading of DNA, that were active at each transcription site [*see illustration on page 39*]. Our success in visualizing nucleolar genes was due to the fact that when such genes are active, some 80 to 100 polymerase molecules are simultaneously transcribing each gene, forming fibrils of ribosomal-RNA precursor molecules (plus protein) in a graded set of lengths at successive stages of synthesis. The slight discrepancy between the observed 2.4-micron length of the matrix units and the 2.7-micron length expected for the genes can probably be explained by the fact that the DNA is unwound for transcription at each of the many transcription points; any bellying out of the DNA double helix would tend to shorten the overall length of the gene.



EXTRACHROMOSOMAL NUCLEOLUS in an oöcyte of the spotted newt, *Triturus viridescens*, is seen in thin section, enlarged 15,000 diameters. The nucleolus has a fibrous core and a granular cortex. Part of the nuclear envelope is visible (*lower left*), with the cytoplasm outside it. The core contains many copies of the genes for ribosomal RNA.

When we prepared amphibian oöcyte chromosomes in the same way that we had prepared the extrachromosomal nucleoli, we found that the fine detail of RNA synthesis could similarly be observed on the long lateral loops characteristic of the lampbrush stage [*see illustrations on page 40*]. Again we see the RNA polymerase molecules closely spaced along the active portions of the



ACTIVE GENES for ribosomal RNA are arrayed within an unwound nucleolar core that was stained with tungstic acid and enlarged 26,000 diameters in the electron micrograph, elements of which are interpreted in the simplified drawing (*left*). One sees a long axial fiber in the form of a collapsed circle, with a series of arrowhead-shaped "matrix units" along it. Each unit is composed of a set of fine lateral fibrils of graded lengths. By staining and enzyme testing the axial fiber is identified as a complex of DNA and protein, the fibrils as complexes of RNA and protein. The segment of DNA within each matrix unit is a gene for ribosomal RNA; each fibril is a protein-complexed ribosomal-RNA precursor molecule that was being transcribed from the gene in the living oöcyte. The stretches of the fiber between matrix units are inactive "spacer" segments the function of which is not yet known.

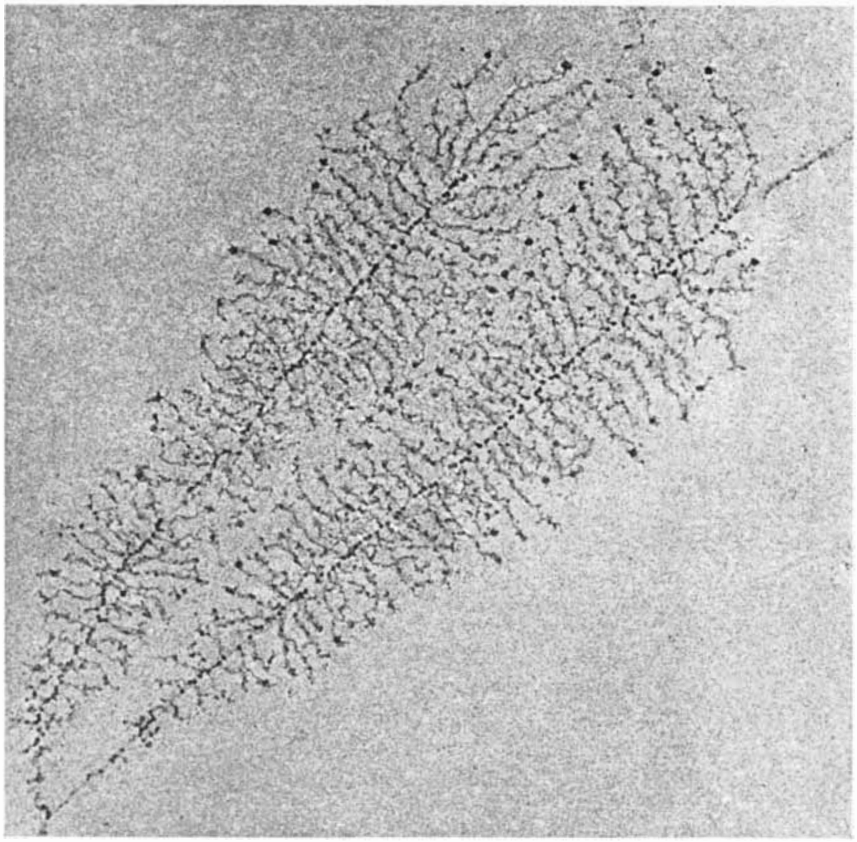
fiber, each at the base of one of the nascent RNA-protein fibrils that form a continuous gradient along the loop. The fibrils grow much longer than those on the nucleolar genes, however, and so they must contain very large RNA molecules. We do not know the subsequent function of the fibrils produced by the oöcyte chromosomes; presumably their RNA gives rise to messenger RNA's for the synthesis of specific proteins, but so far we cannot assign any specific messenger to any one loop.

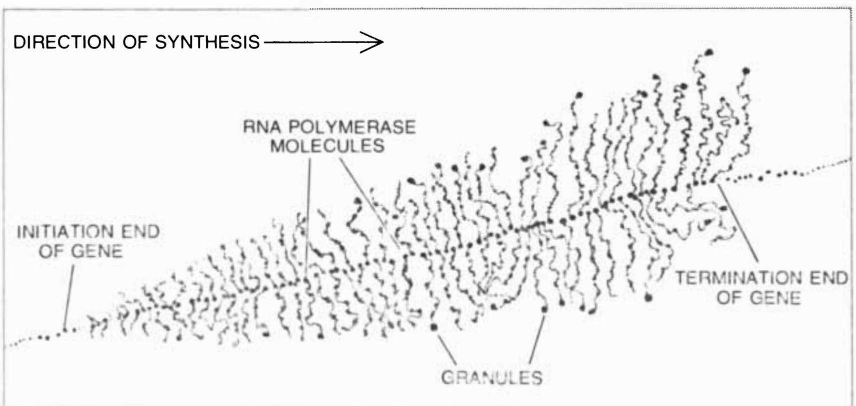
With my help, Aimée H. Bakken has recently extended our work with eukaryotes to a mammalian cell: the human tumor cell line called HeLa, which has been established in laboratory cultures for 20 years and in which the synthesis of RNA and proteins has been intensively studied by chemical techniques. The fact that these cells are much smaller than oöcytes (only some 30 microns or less in diameter) called for different methods of isolating and dispersing the contents of their nuclei. In an attempt to avoid tedious micromanipulation we sought to develop a chemical method for disrupting cells and nuclei so that we could spread out the nucleoli and chromosomes without degrading the structures we hoped to observe. After many trials with various agents we attained success by treating the cells with low concentrations of Joy, an ordinary household dishwashing detergent.

We have isolated from HeLa cells (presumably from the nucleoli) active chromosome segments that we have identified tentatively as genes for ribosomal RNA [*see top illustration on page 41*]. The identification is based primarily on their striking structural similarity to the ribosomal matrix units of oöcytes, whose identity is on firm ground. In both cases the active segments of DNA are fully loaded with RNA polymerase molecules, with as many as 150 such polymerases per gene in HeLa cells; small, dense granules appear at the free ends of the growing fibrils, and neighboring active segments are separated by inactive spacer segments, which in the HeLa cells are about the same length as the putative ribosomal-RNA genes. The active segments are longer in HeLa cells than they are in the oöcyte but shorter than would be expected on the basis of the molecular weight of the ribosomal-RNA precursor in mammals: the segments are about 3.5 microns long instead of an expected 4.5 microns. Again the discrepancy is probably due to some unwinding of the DNA at each of the numerous closely spaced transcription sites.

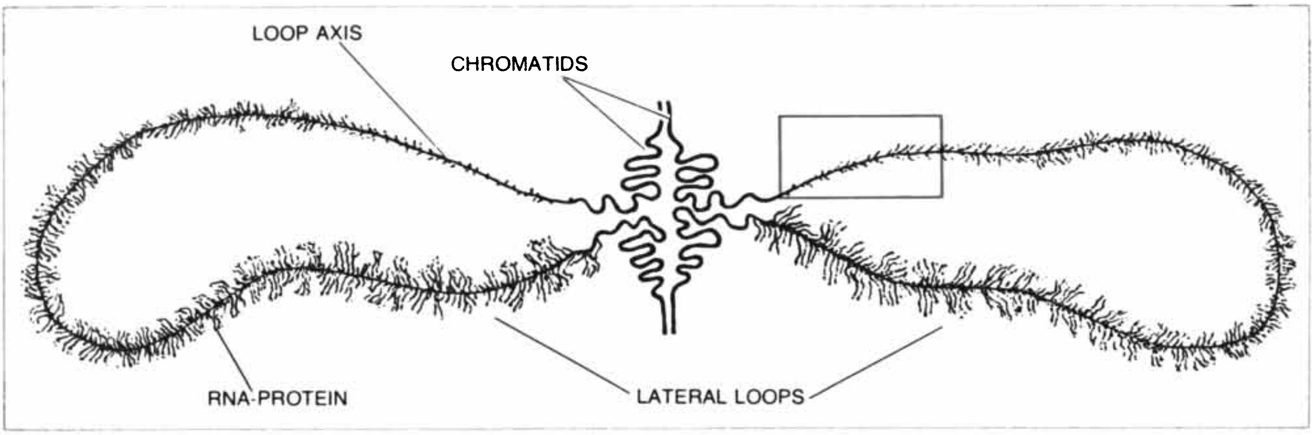
In considering non-ribosomal-RNA synthesis in HeLa cells, we knew from work done in other laboratories that 95 percent or so of the RNA synthesized in a HeLa cell at a given time does not survive to reach the cytoplasm but is degraded in the nucleus. This RNA is called heterogeneous nuclear RNA and its molecules range from about a tenth the size of the HeLa ribosomal-RNA-precursor molecule to many times that size. There is good evidence that small portions of the heterogeneous-nuclear-RNA molecules are not degraded and are precursors of the messenger RNA's found in the HeLa cytoplasm.

In addition to the distinctive ribosomal-RNA-precursor fibrils we do see RNA-protein fibrils of various sizes at scattered sites along dispersed HeLa chromosomes, but none of these fibrils are arrayed in a graded set of lengths. This fact indicates that RNA synthesis is initiated much less frequently, and that the level of polymerase activity is therefore much lower, on active sites of HeLa chromosomes than it is on the active





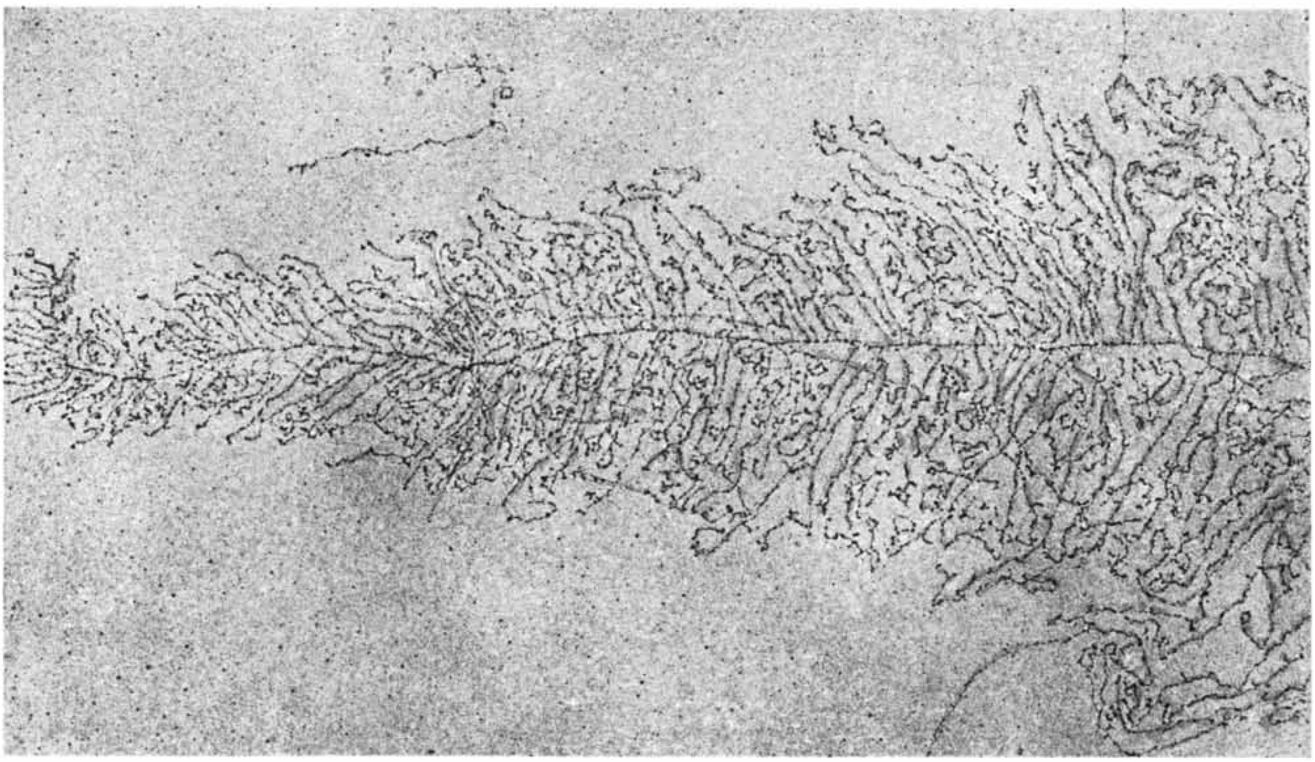
RNA POLYMERASE MOLECULES that are active at each transcription site are resolved on two nucleolar genes that were stretched in the process of preparation and are enlarged 48,000 diameters in the electron micrograph; one of the genes is reproduced in the drawing. The polymerase molecules are visible as dense granules at the base of each fibril; each polymerase was moving from left to right as it transcribed the gene. This micrograph also shows the granules, whose function is unknown, that are seen at the tip of maturing fibrils.



LAMPBRUSH CHROMOSOMES are a special form seen in oöcyte nuclei during such cells' long growth period. At the lampbrush stage the active portions of the sister chromatids that constitute each chromosome axis are greatly extended in lateral loops; the DNA axis of each loop is transcribed into RNA that is complexed with protein. The increasing length of the RNA molecules accounts for the increasing thickness of the loop from one point of juncture with the chromosome axis around to the next juncture. The colored rectangle shows approximately the segment of a lateral loop that is seen in the micrograph at the bottom of the page.

lateral loops of amphibian lampbrush chromosomes. Since heterogeneous nuclear RNA constitutes 95 percent or more of the RNA being synthesized in the HeLa nucleus, it seems likely that most of the dispersed RNA-protein fibrils we see attached to HeLa chromosomes contain nascent RNA of that kind. As in the case of the lampbrush chromosomes, however, no specific genetic role can yet be assigned to any of the observed RNA synthesis.

After our first success with amphibian oöcytes I collaborated with Charles A. Thomas, Jr., and Barbara A. Hamkalo at the Harvard Medical School in adapting the preparative techniques to a prokaryotic cell: the bacterium *Escherichia coli*. Later Hamkalo continued the work with me at Oak Ridge. Since there is a broad background of biochemical and genetic studies on the transcription and translation of various genes in *E. coli* and other bacteria, we could predict what active bacterial genes should look like if we could take their picture. The first problem, however, was to weaken the tough outer wall of the bacterial cells enough to open them up. We found that rapid cooling of growing *E. coli* followed by judicious application of the enzyme lysozyme removed enough of the wall to leave us with osmotically sensitive protoplasts: naked cells that could be burst open by osmotic shock when we diluted them in water. We then centrifuged the



LATERAL LOOP of a lampbrush chromosome from a newt oöcyte is enlarged 22,000 diameters. The thin end, near the juncture with the chromosome axis, is at the left. RNA synthesis begins there, as shown by the increasing length of the fibrils from left to right. RNA polymerase molecules are visible at the base of each fibril. No specific genetic function can yet be assigned to these RNA's.

burst cells much as we had the oöcyte cores and made electron micrographs of their extruded contents.

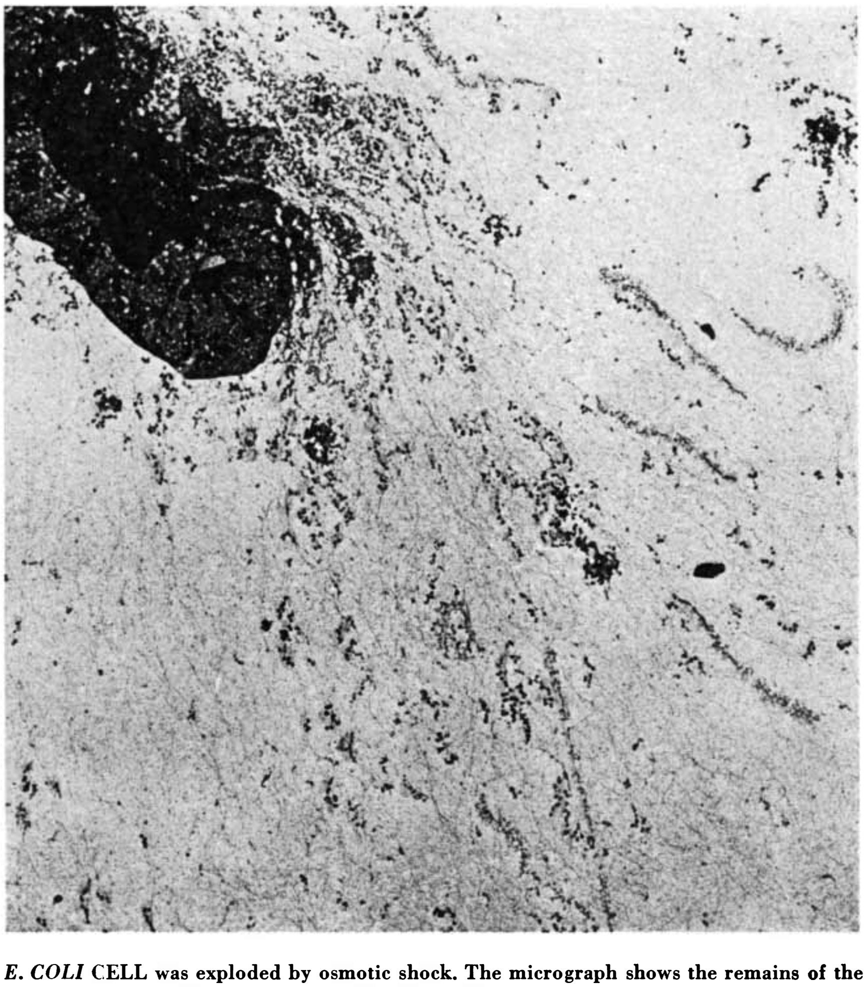
We knew that in bacteria, unlike in eukaryotes, nascent messenger RNA is normally complexed with translating ribosomes. As soon as a newly initiated strand of messenger RNA is long enough, a ribosome attaches itself to the strand and translation of the RNA into protein begins. As the synthesis of the messenger RNA proceeds, giving rise to a longer strand of RNA, more ribosomes attach themselves to the strand. They form a string called a polyribosome, which continues the translation of the elongating messenger RNA. Any active genes coding for protein synthesis in our specimens should therefore have a graded set of strings of polyribosomes attached along their length.

Micrographs of the extruded *E. coli* contents show masses of thin fibers, some of them with attached strings of granules [*see bottom illustration at right*]. If we treat a specimen with the enzyme deoxyribonuclease, the fibers are destroyed; if we treat one with ribonuclease, the granular strings are removed from the fibers. We conclude that the fibers are portions of the bacterial chromosome and that the granules are ribosomes that were translating messenger RNA. The fact that there are no attached polyribosomes at all along large stretches of the chromosomes is consistent with reports that a large part of the *E. coli* chromosome either codes for rare species of RNA or is never transcribed. Also, our bacteria are grown under optimal conditions, so that many of the biosynthetic pathways may be inactive and the genes coding for their enzymes may be repressed.

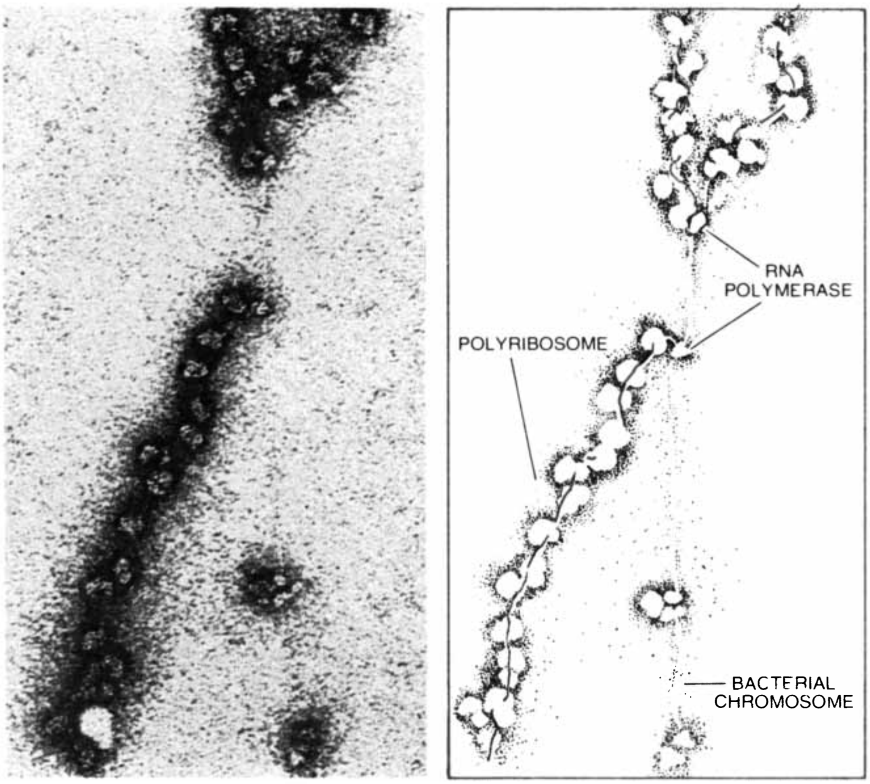
At higher magnification the detailed configuration of the ribosome-coated regions—the genes coding for the synthesis of protein—is revealed [*see illustration on page 35*]. The gene is visible as a thin axial fiber. Along it there is a graded set of strings of ribosomes. At the junction point of each string of ribosomes with the gene we can resolve the active RNA polymerase molecule that is catalyzing the transcription. With negative staining we can see that the enzymes are smaller and more irregular in shape than they are in eukaryotic cells [*see top illustration on next page*]. Unfortunately the pictures do not show the growing protein chains that are presumably associated with each of the ribosomes; the amino acids that make up the chain are simply too small for our procedures to resolve.

As for ribosomal-RNA genes in *E. coli,*

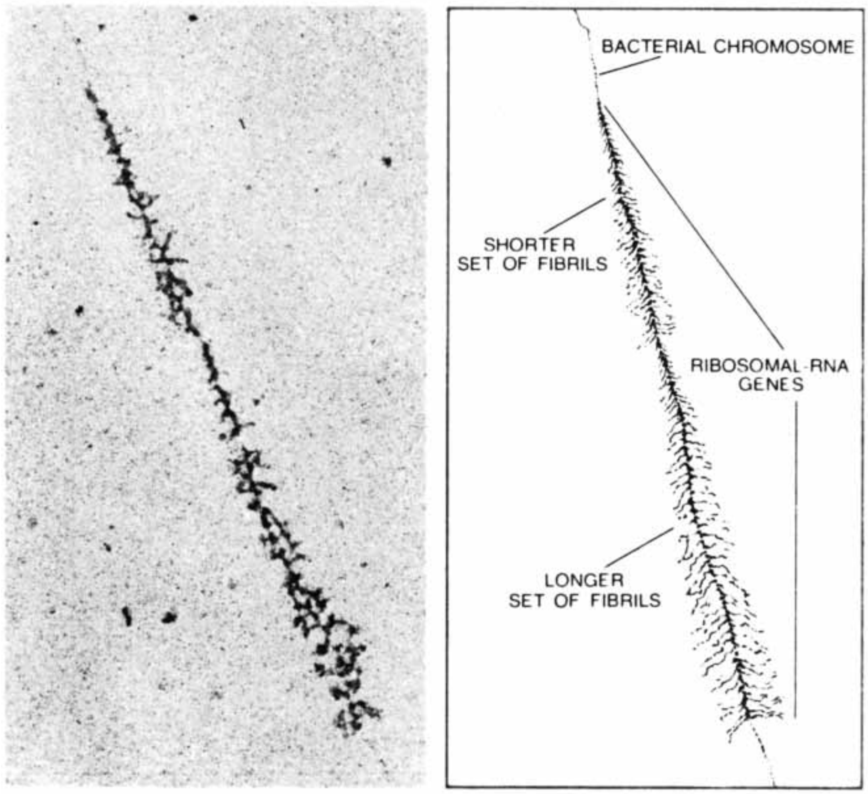
MATRIX UNITS isolated from the nucleus of a HeLa cell, a human tumor cell line, are identified as ribosomal-RNA genes on the basis of their similarity to the nucleolar genes of amphibian oöcytes: the graded set of closely spaced fibrils with a polymerase at the base of each, the dense granules at the tips of the fibrils and the occurrence of spacer segments (not resolved here) between the active genes. The enlargement is 16,000 diameters.



***E. COLI* CELL was exploded by osmotic shock. The micrograph shows the remains of the cell wall (*upper left*) and extruded cellular contents consisting of fine fibers with fibrillar segments and attached strings of granules. The fibers are portions of bacterial chromosome, the strings of granules are polyribosomes and the fibrillar segments are ribosomal genes. The enlargement is 22,000; the micrograph on page 35 further enlarges similar material.**



NEGATIVE STAINING (with uranyl acetate) outlines material from an *E. coli* cell, showing that the RNA polymerase molecules (*see drawing at right*) are smaller and more irregular than in eukaryotes. The staining reveals stretches of what is presumably messenger RNA but does not define the chromosome well. The enlargement is 150,000 diameters.



TWO CONTIGUOUS SETS of fibrils reflect transcription activity on genes for the two large RNA molecules of *E. coli* ribosomes. One matrix unit is about twice as long as the other and together they measure about 1.5 microns. The enlargement is 64,000 diameters.

we could predict that they should look very different from the genes for protein synthesis. Whereas the messenger RNA's of genes coding for proteins are contained within polyribosomes, the nascent ribosomal RNA should be complexed with some protein but not mature ribosomes. Quantitative studies of ribosomal RNA synthesis in rapidly growing *E. coli,* combined with estimates that within each *E. coli* chromosome there are no more than seven segments coding for ribosomal RNA synthesis, suggested that these segments should be relatively scarce in our preparations and should be densely packed with active RNA polymerase molecules. There was also evidence that the two large RNA molecules that are found in prokaryote ribosomes are synthesized separately rather than being cleaved from a single precursor as they are in eukaryotes; the genes for these two RNA's were thought to be contiguous, and so they should show up as two adjacent matrix units of fibrils. The molecular weights of the two bacterial ribosomal RNA's indicated that the matrix unit of one should be twice as long as the matrix unit of the other and that the total length of the two units should be close to 1.7 microns.

We have identified in our specimens segments of *E. coli* chromosome that meet these specifications rather precisely [*see bottom illustration at left*]. The segments have 60 to 70 attached fibrils arranged in two consecutive matrix units. As in the amphibian cells and the HeLa cells, these segments are somewhat shorter than had been predicted, again presumably because of the unwinding of DNA. The ribosomal RNA genes of *E. coli* appear not to be close to one another on the chromosome as they are in eukaryotes. In fact, they are separated by chromosome segments that show the polyribosomal activity of protein-synthesis genes. We have not yet determined the minimum length of these intervening segments.

The fact that rather simple techniques of isolation and preparation have enabled us to see such fine structural details of genetic activity in several types of cell gives us confidence that electron microscopy will become an increasingly important tool for the study of cell genetics. Refinement of the techniques may make it possible to observe directly active genes from almost any kind of cell, and eventually to identify those genes with their products and learn how they function at different times in the cell's life cycle, how they are affected by varying conditions and how they relate to specific cell activities.

We want to be useful
...and even interesting

A baking product of possibly cosmic significance

Here we go cuddling up to the astronomers again.

Inquiries about availability of strong red sensitivity in the "IIIa" type of emulsion will be welcomed by E. J. Hahn, Scientific Photography Markets, Kodak, Rochester, N.Y. 14650.

C. E. K. Mees, who started the Kodak Research Laboratories in 1912, was an astronomy buff. Instead of grinding telescope mirrors, he made photographic emulsions for astronomers. Some people refuse to distinguish between work and fun.

Mees numbered his emulsion types from I to V. From I to V, light sensitivity and granularity decreased and inherent contrast increased. Later he appended "a." This indicated a treatment against loss of effect of some of the photons when they are spaced out too much, as at low intensity. Capital letters designated sensitization to certain spectral regions. This nomenclature is now work lingo of the astronomer.

Last year astronomers from seven nations of Europe announced a project to map the sky of the southern hemisphere to "limiting magnitude." They were prepared to regard the capability of Type IIIa-J as "limiting." Why not Ia? Why IIIa?

Because of *detective quantum efficiency* (D.Q.E.). The photographic emulsion, like all detectors, responds to both signal and noise. It responds by grain blackening. From internal causes, some grains blacken that shouldn't; some grains that should, don't. Both cases add further noise. The noise concept comes to photography from radio. D.Q.E., the square of the ratio of output signal/noise to input signal/noise, varies directly with sensitivity and contrast and inversely with granularity. Which is hardly a surprise.

Eight years ago we started preaching to the astronomers to mind their D.Q.E., that there is an optimum choice of parameters of manufacture, exposure, and processing. On all sides of the optimum you got less information per precious hour of telescope time.

Then we presented them with KODAK "Spectroscopic" Plates, Type IIIa-J, the first photographic material designed above all else to maximize D.Q.E. if baked* before exposure. The "J" indicates sensitization only a short way into the blue-green beyond the sensitivity of unaided silver halide. To extend the sensitivity of the IIIa emulsion type beyond a practical photographic limit of 520 nm without losing response to baking has been difficult. Our modestly worded invitation above stems from tests where we almost hit 700 nm with a baked product. Man's window to the universe is opening wider.

*Typically for 18 hours at 55 C, as first worked out by W. C. Miller, Hale Observatories. Later, Miller tells us, he was able to bring the baking time down to 5 hours at 65 C.

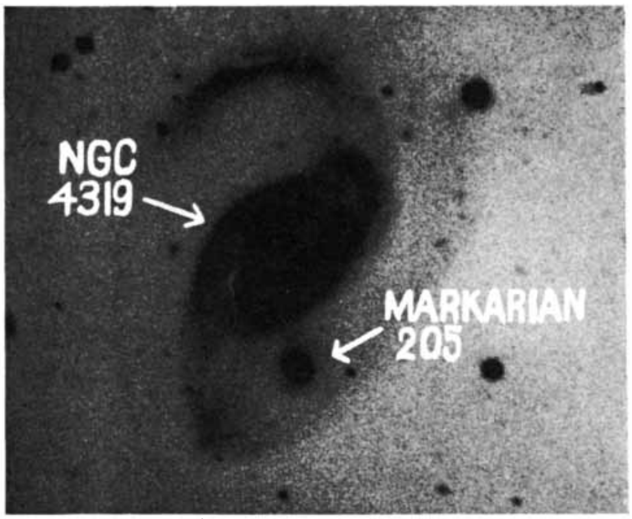
At the moment, though, we bow in the direction of Rosemary Hill. On that hill at Gainesville sits the University of Florida observatory. There close attention was paid to publications from the Kodak Research Laboratories about how moisture and O_2 depress sensitivity to dim light. There, as described in the July '71 issue of *Applied Optics* (*10*:1597), procedure and equipment were devised for *baking and exposing* IIIa-J plates under dry N_2.

Scarcely a month after the paper appeared, astronomers were telling each other of speed gains from $5\times$ to $8\times$ thus achieved at low intensity. This amounted to about the speed of IIa, but with about twice the contrast and half the granularity of IIa. Consequently D.Q.E. reaches 4%. Hitherto no actual photographic emulsion has gone much higher than 0.8%. Theoretical emulsions could go to 3.5% in theory. So our own theorists—conservative people—had theorized.

Let's hear it for Rosemary Hill!

Well beyond buffhood is the incumbent partisan of astronomy's cause at the Kodak Research Laboratories. During a Study Research Leave at Kitt Peak National Observatory he tried to confirm evidence from a Palomar IIIa-J plate that a filament luminous with $H\alpha$ light connects the galaxy NGC 4319 with the object known as Markarian 205. According to red shifts, these two are receding at 1700 km/sec and 20,250 km/sec respectively. Since no string can be stretching that fast, perhaps red shifts can occur from non-velocity causes. If so, then bang goes the big-bang theory of the expanding universe. Alas, with all the detective quantum efficiency at his command, our man has in vain searched for the string. In *The Astrophysical Journal* for August 15, 1972 (176:L5), a collaborator and he have nothing more startling to report than the possibility that one of the many faint extended objects in that region of the sky—presumably distant galaxies—happens to show up between the relatively nearby NGC 4319 and the apparently more distant Markarian 205.

Taken on N_2-baked KODAK "Spectroscopic" Plate, Type IIIa-J

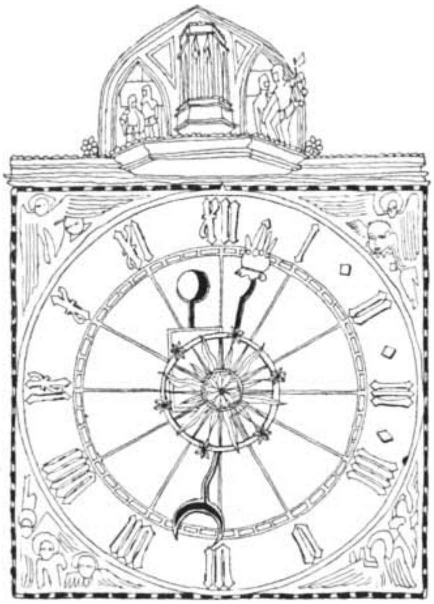


Are they connected? If a string of hydrogen joins them, availability of higher D.Q.E. for $H\alpha$ light offers additional opportunities to check a philosophically sensitive point by direct photography.

Kodak

May all enjoy their work

An equal-opportunity employer



SCIENCE AND THE CITIZEN

The Short View v. the Long

The diminished role of science and scientists in the high councils of the Federal Government, apparent throughout President Nixon's first term, was made perfectly clear immediately after his inauguration in January. In Reorganization Plan No. 1 of 1973 Nixon abolished the Office of Science and Technology and with it the post of the President's science adviser, a job last held by Edward E. David, Jr., who resigned early in January. The reorganization also marks the end of the President's Science Advisory Committee (PSAC), which came to prominence and influence in 1957, soon after the launching of the first Russian artificial satellite. Nixon has asked for and accepted the resignation of all 20 PSAC members, a group of scientists and engineers drawn from universities and industry. Under presidents Eisenhower, Kennedy and Johnson, PSAC helped to guide the U.S. into the age of intercontinental missiles and space exploration.

The downgrading of science by the Administration did not come as a surprise. Jerome B. Wiesner, president of the Massachusetts Institute of Technology and formerly science adviser to President Kennedy, commented that for the past four years the advisory group in the White House "has been more a facade than a reality. One cannot argue very hard against dismemberment of a facade." He said he regretted, however, "that this downgrading of science occurs when the role of science and technology is being questioned generally."

The functions of the Office of Science and Technology and of its director will be transferred to the director of the National Science Foundation, H. Guyford Stever, who will report to George P. Shultz, one of three men recently designated counselors to the President. Shultz is also Secretary of the Treasury and the chairman of the new Council on Economic Policy. The three counselors, all Cabinet secretaries, have been given "coordinating responsibilities in the broad areas of human resources, natural resources and community development."

President Nixon explained the reorganization as an effort to "streamline the executive branch of the Federal Government. By concentrating less responsibility in the President's immediate staff and more in the hands of the departments and agencies, [the reorganization] should significantly improve the services of the Government." Reorganization Plan No. 1 was only a small part of the President's attempt to "avoid inflation or higher taxes or both" by limiting the 1974 budget to $268.7 billion, a $19-billion increase over the estimated 1973 budget. The actual economy achieved by abolishing the Office of Science and Technology will be small: the elimination of 389 jobs with a payroll saving of $2 million.

In the proposed 1974 budget the total expenditure for research and development by all departments and agencies will be increased 1.7 percent to $17.4 billion. For the purposes of comparison the proposed Defense budget for 1974 is $81.1 billion, an increase of $4.7 billion, or 6.1 percent over 1973. In the "post-Sputnik" era the research and development budget had regularly increased about 15 percent per year. The research emphasis in 1974 will be on short-term results and on the technological application of scientific knowledge to social needs. In medical research, for example, more money will be spent on the problems of cancer and heart disease and less on basic studies. Thus the budget of the National Cancer Institute will rise from about $426 million to $500 million, and the budget of the National Heart and Lung Institute will be increased $18 million to $265 million. Overall the 1974 budget for the National Institutes of Health is to remain roughly unchanged at $1.57 billion. Taking inflation into account, this represents a decrease of about 5 percent.

The budget of the National Aeronautics and Space Administration is to be reduced about $400 million to $3.016 billion. The budget of the National Science Foundation, the chief supporter of basic research, shows a spurious increase from $615 million to $641.5 million; the 1974 figure includes nearly $60 million that will have been spent before the start of the new fiscal year. As evidence of concern for the energy problem, the 1974 budget will give the Department of the Interior $25 million for research on fossil fuels and other non-nuclear energy sources.

For all types of weapons research the new budget will provide the Atomic Energy Commission with $67 million. A one-sentence footnote in the budget indicates that half of the total will be allocated to development of a hydrogen bomb triggered by the heat of laser beams rather than by a nuclear-fission device.

The Abortion Decision

The U.S. now has what is in principle the world's most permissive legal policy on abortion. The procedure will henceforth be regulated on the basis of medical standards and practices, rather than according to socioreligious principles enforced by criminal law. Abortion will nevertheless not become a primary method of fertility control, although its availability as a backup procedure could have a beneficial effect on contraceptive practices. These are among the conclusions reached by legal and medical specialists in the wake of the Supreme Court's January decisions striking down almost all provisions of most of the state laws prohibiting and regulating abortion.

The decisions relied primarily on the right of privacy implied by the Constitution and on a finding that an unviable fetus is not a "person" within the meaning of the 14th Amendment. The Court held that abortion is so safe in the first three months of pregnancy that states may not interfere at all with a decision by a woman and her physician at that stage. In the second and third trimesters states may regulate the practice of abortion, but only in ways "reasonably related to maternal health." Only after the fetus is viable—has the capability of surviving outside the womb—may a state prohibit abortion. Significantly, the Court held that viability comes after 24 to 28 weeks of pregnancy, and it said that abortion must be allowed even after that "when it is necessary to preserve the life or health of the mother." ("Health" specifically includes mental health.) No state law complies completely with the Court's guidelines. Even New York's law permits abortion after 24 weeks only to preserve a woman's life, not her health; Alaska and Hawaii have residency requirements, which the Court disallowed. Moreover, health-code provisions (such as New York City's) that regulate abortions in the first trimester are now presumably unconstitutional. The extent to which the practice can be restricted later in pregnancy remains somewhat unclear. It will probably be tested by the promulgation and enforcement of a wide variety of state and local regulations that will be contested in the courts.

Some proponents of liberalized abortion indicated that they would now work for the complete elimination of all criminal legislation on the subject, holding that abortion should be treated like any other medical procedure and regulated by the medical profession, civil law and health codes. Others indicated they would be satisfied with laws such as New York's if they are modified to include the preservation of maternal health as a reason for abortion after viability.

Although abortion is now legal until late in pregnancy, the procedure is still far from generally available, and proponents of liberalized abortion plan to remedy that lack. Planned Parenthood–World Population announced that it would set up more abortion clinics and run a nationwide referral service. The National Association for the Repeal of Abortion Laws will hold regional workshops for physicians to promote safe abortion methods and advise on how to establish abortion clinics. Christopher Tietze, associate director of the biomedical division of the Population Council, estimated on the basis of New York's experience since 1970 that facilities should be available to provide abortions for some 1.6 million women a year in the U.S. It is now incumbent on the community to make abortions available to all at a fair price, he said.

Tietze pointed out that the Court's decision puts the U.S. well ahead of many northern European countries that have been considered relatively liberal but that still require approval of an abortion by either a medical board or more than one physician. In a legal sense, in authorizing abortion the U.S. now is more liberal than Japan, the U.S.S.R. and other countries of eastern Europe, which limit abortion on demand to the first trimester; on the other hand, it is easier to obtain an early abortion in those countries because more facilities are available.

In spite of the decision, Tietze said, he does not expect a large increase in the number of late-pregnancy abortions in the U.S. He cited the results of a study of abortion in New York. The number of late abortions decreased significantly in the period from mid-1971 to mid-1972 compared with the year before, the first in which legal abortion became generally available. He attributed the decline to the steady increase in facilities during the two years and to women's increasing awareness that early abortion is a great deal safer than late abortion, which, he said, is an indication of a "social failure," and is justifiable primarily by the development of heart or kidney disease in a woman or by an indication that the fetus is defective. In this connection he pointed to recent advances in the analysis of fetal chromosomes for evidence of certain congenital defects; such evidence can now presumably be discovered and adduced in support of abortion in the last trimester, since it certainly can be held to threaten a woman's mental health.

Along with most other students of the subject, Tietze does not think abortion should be a primary method of fertility control. The availability of abortion as a secondary method can be important, however. It means that "good but not perfect" methods of contraception—such as oral contraceptives that may be somewhat safer than the most effective ones—can be prescribed with confidence, with early abortion as a backup. A statistical study he made in 1969 showed that in-hospital abortion produces about the same low fatality rate as use of the most effective oral contraceptive method; the combination of "moderately effective" contraception and in-hospital abortion results in a much lower fatality rate.

Spare Parts

As a result of an effort mounted over the past few years by organizations interested in increasing the supply of kidneys, eyes and other organs for transplantation, it is now possible for people in all 50 states to be card-carrying donors. The card is crucial, since it serves the function of a will in a situation where time is usually too short for a will to be found and probated. Under the Uniform Anatomical Gift Act, which each of the 50 states has adopted, the gift becomes effective on the death of the donor.

The effort to gain adoption of the act arose both from a need for organs in the light of recent advances in transplantation and from the fact that before the act was adopted it was often legally difficult for a donation of an organ to be made, particularly by a prospective donor's relatives after his death. Now the wallet-size uniform donor card makes the donor's wishes plain and readily ascertained. The card must be signed by the prospective donor and two witnesses. It specifies that the donor gives "any needed organs or parts" or "only the following organs or parts," which are then enumerated by the donor, "for the purposes of transplantation, therapy, medical research or education." The act also enables a relative to donate "all or any part of the decedent's body," provided that there is no "actual notice of contrary indications by the decedent" and no opposition from closer relatives. A third option listed on the card enables the donor to give his entire body for anatomical study.

Under the act the gift can be made to "any hospital, surgeon or physician," any accredited medical or dental school, college or university, any organ bank or "any specified individual for therapy or transplantation needed by him." The number of people who have signed cards is unknown, although the National Kidney Foundation, which has sent out some five million cards, mostly on request, believes a significant number of the recipients have signed them.

Small Implosions

Discussions of the ultimate feasibility of schemes for generating useful electrical power from laser-induced fusion reactions have until recently tended to concentrate on the large disparity that exists between the amount of laser energy theoretically required to operate such a reactor system and the amount of laser energy practically available now or in the foreseeable future. It has been calculated, for example, that a typical controlled-laser-fusion scheme involving a small solid pellet of heavy hydrogen heated to thermonuclear temperatures by means of a single focused beam of pulsed laser light would require a laser energy on the order of a billion joules in order to produce more energy than is consumed. This amount of laser energy is so large as to be not only currently unavailable but also perhaps ultimately unattainable. Existing solid-state laser systems are capable of delivering peak energies of more than 1,000 joules only in extremely short pulses.

Two major developments that have taken place in the past few months have changed this situation. First, a new design concept, developed primarily by workers at the Lawrence Livermore Laboratory on the basis of computer simulations, promises to reduce by more than 10,000 times the amount of laser energy theoretically required to achieve controlled thermonuclear conditions. Thus it is calculated that a net output of energy can be obtained with laser input energies of from 10,000 to 100,000 joules. Second, laser technology has progressed to the point where both the U.S. and the U.S.S.R. have announced that they are planning to build 10,000-joule laser systems for the express purpose of testing the feasibility of the laser-fusion approach.

The new design concept for laser fusion, described by John L. Emmett and John H. Nuckolls of the Livermore group at a recent meeting of the American Physical Society in New York, involves the use of multiple laser beams to initiate a powerful implosion of a tiny hollow sphere of frozen hydrogen. The laser beams would be carefully controlled in order to heat the entire surface of the pellet uniformly to 100 million degrees Celsius in a few billionths of a second, causing the hydrogen to boil violently off the pellet surface at speeds of thousands of miles per second. As the hydrogen escapes, it would exert a large counterforce on the surface, much like the counterforce exerted on a rocket by the escaping exhaust gases. Applied evenly around the surface, these forces would suddenly implode the pellet to a density 100 times higher than that of lead. Under these conditions of density and temperature the hydrogen nuclei would fuse, releasing some of their nuclear energy.

According to Nuckolls, a billion-watt laser-fusion plant based on such a concept might emerge "in two or three decades." Such a plant would incorporate "a million-joule laser initiating up to 100 fusion-pellet irradiations per second, possibly 10 per second in 10 chambers, each chamber being several meters in diameter. Each fusion-energy release would result from the laser implosion of a cheap pellet of deuterium-tritium fabricated in a drop tower. Each pulse of thermonuclear energy would be equivalent... to the burning of about one gallon of gasoline. Virtually all of this energy would be in the form of neutron radiation. The energy generated would be deposited, mostly by the neutrons, in a liquid lithium blanket surrounding the walls of the chamber. The lithium would flow through a heat-exchange system, turning water into steam, which in turn would drive an electrical generator."

Although Emmett and Nuckolls expressed optimism about the ultimate achievement of power-generation by laser fusion, they cited three main obstacles that must first be overcome. First, even assuming that the planned 10,000-joule lasers prove to be technologically feasible and can be used to demonstrate significant "thermonuclear burn," laser systems with all the desired characteristics would involve considerable further development. Second, experiments needed to test the validity of the new design concept will be possible only after new facilities are available. Third, even if the Livermore approach is demonstrated to be correct and the scientific feasibility of laser fusion is proved in the next few years, formidable technological problems would remain. Nonetheless, the Livermore investigators conclude, "many of those in the field believe that these problems can be solved."

The Sane and the Insane

It has been known for some time that the traditional modes of psychiatric diagnosis of mental illness are not very reliable. Some members of the psychiatric profession (such as Thomas S. Szasz in his book *The Myth of Mental Illness* and R. D. Laing in *The Divided Self*) have challenged the entire concept of mental illness on philosophical, legal and therapeutic grounds, alleging that the term is useless, misleading and harmful to the patient. A new attack on the assumption that psychiatrists can distinguish the sane from the insane comes out of an experimental project conducted by David L. Rosenhan, professor of psychology and of law at Stanford University. Rosenhan and seven other volunteer subjects sought admission as patients to 12 mental hospitals by stating that they had been hearing "hollow voices." All the subjects easily gained admission (11 as being schizophrenic and one as being manic-depressive). None knew how long it would take to get discharged. The length of hospitalization ranged from seven to 52 days. Even though the pseudopatients stopped simulating any symptoms of abnormality immediately on admission, hospital staff members did not discover, or apparently even suspect,

Conversation Pieces

Technically intriguing items from TRW, guaranteed to add luster to your conversation and amaze your friends.

Oh Say Can You See Discussions about whether or not we should "go into space" often overlook a fundamental point—namely, that we are in space already. Each of us is an astronaut on a spacecraft called earth traveling around the sun at 18,000 miles per hour. The biological community that lives on the spacecraft has a fragile life support system–the thin film of soil, air and water in which we dwell. During the past century, the number of passengers aboard the spacecraft has increased tremendously; so also has their ability to consume its finite supplies. We see some of the results in the pollution of our environment and the decay of our resources.

As astronauts we need to take care of our spacecraft—to check on its status, to monitor its resources, and to see that we are not doing irreparable damage to its vital life support system. With this in mind, NASA has undertaken several programs which will bring the technology of space to bear on some very down-to-earth problems. One of these programs involves the development of advanced sensors. From the vantage point of a satellite orbiting the earth, these sensors will someday be able to monitor air quality, determine the condition of crops, or help locate mineral resources.

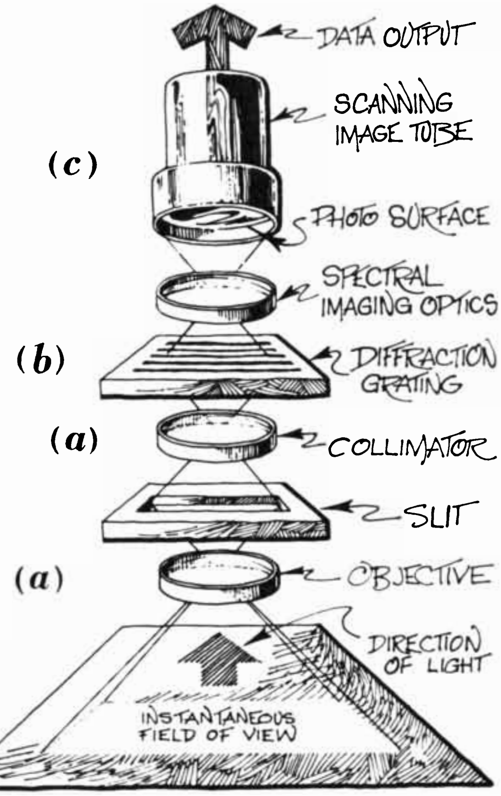
Currently, TRW is developing such a sensor for NASA. It is called MOCS, an acronym for Multichannel Ocean Color Sensor. MOCS is based on the principal that each object on earth reflects light in a unique manner. The light reflected from oil, for example, is quite different from that reflected from water. Thus the two objects can be distinguished on the basis of their spectral response or "signature." MOCS senses the radiation reflected from objects in its field of view and diffracts this radiation (see illustration) into its spectral components. It has better spectral resolution than any other available scanner.

MOCS weighs less than 20 lbs., has no moving parts, and uses only 7½ watts of power. Yet it has produced some fascinating information about the spacecraft on which we live. Recently, for example, we tested MOCS by flying it in an aircraft at 37,000 feet over Clear Lake in California. MOCS showed that the lake teemed with sediments, algae growth, and foreign matter. In fact, the only thing clear about the lake was its ironic name.

We're happy to be associated with the excellent work NASA is doing in remote sensing. We hope that MOCS and other NASA sensors will help make spacecraft earth a habitable home for us all.

The Multichannel Ocean Color Sensor

Lenses (a) focus light through a slit onto a diffraction grating (b) which separates the light into its spectral components. These are focused on a dissector tube (c) which shows spatial variations of light across the field of view in one direction and spectral variations in the other. The raster scan then gives an electronic signal proportional to the spectral radiance of each spectral band in each element across the field of view.



For further information, write on your company letterhead to:

TRW
SYSTEMS GROUP

Attention: Marketing Communications, E2/9043
One Space Park Redondo Beach, California 90278

that they had entered the hospital under false pretenses.

"It is clear," writes Rosenhan in *Science*, "that we cannot distinguish the sane from the insane in psychiatric hospitals. How many people, one wonders, are sane but not recognized as such in our psychiatric institutions?"

The pseudopatients did not go entirely undetected. In every case, Rosenhan says, a considerable fraction of the patients in the hospital detected that the pseudopatient was somehow different. Some voiced their suspicions vigorously: "You're not crazy. You're a journalist, or a professor. You're checking up on the hospital." For the first three pseudopatients 35 of 188 real patients in the admissions wards recognized them as being pseudopatients even though no staff member did.

The volunteers used pseudonyms, and those who were members of the mental health professions gave another occupation. Otherwise their life history was given factually. Three of the pseudopatients were psychologists; the others were a pediatrician, a graduate student in psychology, a psychiatrist, an artist and a housewife. Five were men and three were women.

Mariner 1

Who were the first men to make successful sea voyages? Until recently the oldest-known seafarers were Neolithic peddlers who during the middle of the fifth millennium B.C. managed to cross from the coast of Turkey to Khirokitia on the island of Cyprus carrying obsidian across a minimum of 70 kilometers of the open Mediterranean. Now the distinction belongs to a group of seminomadic fishermen and hunter-gatherers who flourished in Peloponnesian Greece at least 1,500 years earlier.

The evidence that establishes these men of the Mesolithic age (the transitional period between the Paleolithic and the Neolithic) as pioneer mariners comes from a stratified cave site, Franchthi Cave, located on a bay in the southern part of the Gulf of Argolis. Excavated by archaeologists from the University of Pennsylvania and Indiana University, the cave deposits proved to date from late Paleolithic levels (about 10,000 B.C.) at the bottom through Mesolithic strata to late Neolithic levels (about 4000 B.C.) at the top. Specimens of obsidian tools from the Mesolithic strata were sent by the excavators for analysis to the University of Birmingham. There a continuing study of Old World obsidian artifacts is in progress with the objective of determining by physical and chemical analyses the source of this distinctive volcanic glass.

One distinguishing mark of obsidian from a given site is its age, which can be determined from the number of uranium fission tracks in the material. Studying the fission tracks in the Franchthi artifacts, the Birmingham group found that the obsidian had been brought to the Greek mainland either from the island of Melos in the Cyclades, some 120 kilometers of open sea away, or from the island of Giali, some 280 kilometers farther away off the coast of Turkey. By island-hopping through the Greek archipelago the pioneer sailors could have shortened the distance from landfall to landfall to a maximum of 50 kilometers. Nonetheless, the fact that the Mesolithic inhabitants of the Mediterranean shore were capable of sea voyages of any kind, let alone economically motivated expeditions, comes as a surprise.

Helpful Hints

Where is the safest place to be during an earthquake? Probably an open meadow well inland, where there are no trees, utility poles or other objects to fall on you and where you are out of reach of any earthquake-generated tidal wave and safe from any fire fed by earthquake-ruptured fuel lines. The meadow, of course, had best not overlie an active geologic fault. In the absence of any early-warning system that would afford time to reach such a place the U.S. Geological Survey and the President's Office of Emergency Preparedness, assuming that earthquakes will surprise most people at home or at work, have issued a list of recommended actions.

Item: When you are at home or at the office, stay away from windows (broken glass) and chimneys (falling bricks). The best bet: at home get under a table or a bed; at the office get under a desk. Unless you are confronted with a fire, stay indoors.

Item: If you are in a public building such as a store, be wary of potential falling objects. Above all do not rush for a stairway or a door; many other people will probably have the same idea and such an area will be crowded.

Item: If you are outdoors, keep clear of high buildings (broken glass), walls (dislodged cornices) and utility poles (possible charged wires). Do not run.

Item: Whether you are indoors or outdoors, be prepared for small "aftershocks," which can be strong enough to shake loose debris.

The Pinto Runabout. It's the rugged Model A all over again, with a lot more carrying space.

Back in the 1930's, you could always count on the rugged Ford Model A to get you where you wanted to go.

But if you had to bring along much more than a toothbrush, it took a lot of imagination. And rope.

In today's Ford Pinto, carrying space is engineered in. The Pinto Runabout, for example, has a lift-up rear door, fold-down rear seat, and 41.3 cubic feet of cargo space. (See diagram.) With the seat down, the cargo area is more than four feet deep, and every inch of it is carpeted.

But, even more important, ruggedness and durability are also engineered into Pinto.

The Pinto engine (left) was improved and perfected in over 10 years of actual driving in small Ford-built cars all over the world. It's easy on gas, simple to maintain.

The Pinto body is welded into one solid piece of steel, with steel guard rails in the side doors and steel reinforcements in the roof. It's electrocoated to fight rust, and covered with five more coats of paint.

A four-speed floor-mounted transmission (left) is standard on Pinto. You can also get the optional automatic, of course. The transmission is fully synchronized, and designed to be "lubed for life." All it should need is inspection during routine dealer maintenance.

Everything about the Ford Pinto has been engineered and designed with durability in mind. From the ball-joint front suspension shown here to the special rust-resistant brake line coating.

We built Ford Pinto to be a basic, durable, economical car. With plenty of space in it to carry you and yours.

See the 1973 Pinto at your Ford Dealer's: two-door sedan, 3-door Runabout, and the popular Pinto Wagon.

Better idea for safety... buckle up!

Shown here is a 1973 Pinto Runabout with optional Ivy Glow paint, whitewall tires, luxury decor and deluxe bumper groups.

When you get back to basics, you get back to Ford.

FORD PINTO

FORD DIVISION Ford

INTERSTELLAR MOLECULES

Twenty-six kinds of molecule have now been discovered in the gas between the stars of our galaxy. Among them are carbon monoxide, water, ammonia, hydrogen sulfide, formaldehyde and methyl alcohol

by Barry E. Turner

Between the stars of our galaxy there are vast clouds of gas and dust. These interstellar clouds were discovered some 200 years ago by William Herschel, who described them as "holes in the sky" because they obscured the light of the stars behind them. The obscuring property was long attributed to the dust: tiny particles of unknown composition that redden and polarize the light from many stars. Over the past 35 years it has gradually been discovered that the interstellar medium also comprises many different species of free molecules, including some moderately complex ones. The molecules are densest where the dust is dense. They are found in regions where stars appear to be forming and in the outer atmosphere of cool stars. Observations of molecules provide information on the physical conditions of such regions, which until recently have been largely inaccessible to astronomers.

The quantitative study of the interstellar medium began after the spectrograph made it possible to analyze the light from stars in detail. In 1904 Johannes Franz Hartmann of Germany suggested that an absorption line of ionized calcium seen in the spectrum of certain bright stars originated in interstellar space, in other words, that there were calcium ions (atoms stripped of one electron or more) between the earth and the stars that absorbed the light of the stars at certain wavelengths. Later neutral, or un-ionized, sodium was also found to be a constituent of the interstellar medium. By 1937 it was recognized that hydrogen is the most abundant element in the cosmos and must therefore make up the bulk of the interstellar medium. Presumably the hydrogen was in the form of single atoms rather than the diatomic molecule (H_2). Today we know that only trace amounts of calcium and sodium are present in comparison with hydrogen.

In the visible region of the spectrum atomic hydrogen (H) can be observed only in the ionized state by its emission of what are known as the Balmer recombination lines. These spectral lines are produced when a free electron is captured by a hydrogen nucleus and cascades down the energy levels of the atom. Ionized hydrogen is found near very hot stars, where it is visible in the form of the glowing clouds known as emission nebulas or H II regions. The ionized hydrogen of these nebulas, which is associated with much smaller amounts of ionized atomic helium, oxygen, nitrogen, carbon and other trace elements, has been studied for many years. The ions are indicators of the temperature and density of the clouds of gas immediately surrounding the hottest stars—clouds out of which these stars are evidently born. The high temperatures (10,000 degrees Kelvin) and relatively high densities (100 atoms per cubic centimeter) that characterize the H II regions were long ago recognized as having no relation to the conditions that prevail in the cold, dark regions of space far from hot stars.

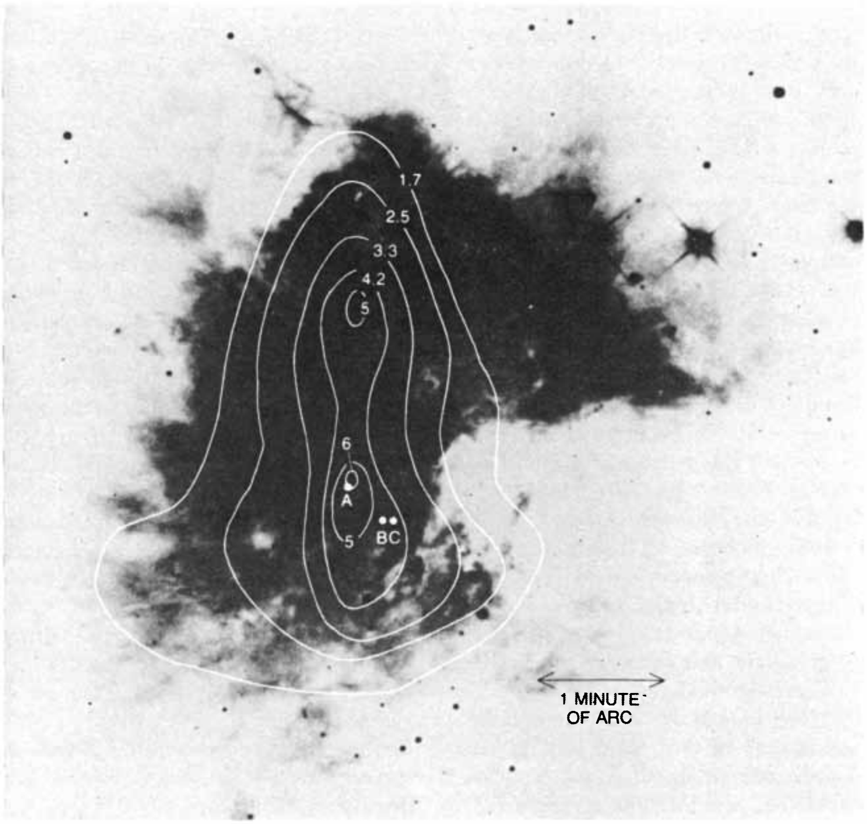
The first interstellar molecule was discovered in 1937. It was the free chemical radical of carbon and hydrogen (CH). The ionized radical (CH^+) and the cyanogen radical (CN) were identified during the next four years in the spectra of a few bright blue-white Type O and Type B stars, often the same stars against which the clouds of calcium and sodium had been observed. These trace constituents are not, however, useful for study of the general interstellar medium. One reason is that they can be observed only immediately in front of the brightest stars. Another is that the clouds containing them must be dense enough to produce an observable absorption line but not so dense that the light from the background star is too attenuated. Very specific conditions are thus required. Thirdly, observations at visible wavelengths can penetrate the interstellar medium only 2,000 or 3,000 light-years, because these wavelengths are effectively absorbed by interstellar dust. Fourthly, observations of trace constituents such as calcium, sodium, CH, CH^+ and CN give no information about the total amount of gas present between the stars. In addition, it cannot be proved that these constituents are not in some way physically associated with the very few hot stars in whose direction they are seen.

This deadlocked situation might never have changed except for the rapid development of radio astronomy over the past 20 years. In fact, one of the first major triumphs of radio astronomy, which came in 1951, was the detection of interstellar atomic hydrogen by means of its spectral line at the radio wavelength of 21 centimeters. It was quickly realized that observations at radio wavelengths could penetrate completely through the galaxy, a distance of some 60,000 light-years, because radio waves are not appreciably absorbed by the interstellar dust. The 21-centimeter line of hydrogen was therefore utilized immediately in the years after 1951 as a tool for studying the gas in our galaxy from one edge to the

GREAT NEBULA IN ORION is a cloud of dust and gas in which stars are being born. Embedded in the nebula are four bright, very hot young stars known as the Trapezium cluster. They are the energy source of the nebula itself, causing it to emit radiation at many wavelengths. The greenish color of the nebula is due to the fact that it was photographed in the light of the spectral line of atomic oxygen at the wavelength of 5,007 angstroms.

other. This gas accounts for between 5 and 7 percent of the total mass of the galaxy. In the course of the 21-centimeter explorations the entire spiral structure of the galaxy was laid out; before that only faint nearby glimmerings of it had been revealed by the distribution of stars.

What physical and chemical conditions of the interstellar gas could be deduced by examining the 21-centimeter hydrogen line? The first pictures were crude, but they indicated that on the average the gas had a temperature of 100 degrees K. and a density of between one atom and 10 atoms per cubic centimeter in the spiral arms of the galaxy and .1 atom per cubic centimeter between the spiral arms. Certain observational refinements and theoretical considerations have led to a "two-component" model of the interstellar medium. One component consists of cool clouds of gas with a density of between 10 and 100 atoms per cubic centimeter and a temperature of 100 degrees K. or less. These clouds are in a pressure equilibrium with the second component, a hot gas with a density of about .1 atom per cubic centimeter and a temperature perhaps as high as 10,000 degrees K. The gas within the spiral arms consists of both components; the gas between the spiral arms consists mainly of the hot component.



CONTOUR MAP superimposed on a negative print of the Orion nebula shows the distribution of the interstellar molecule formaldehyde (H_2CO) at the wavelength of two millimeters. Numbers on contour lines are brightness temperatures: a measure of the intensity of the radiation in degrees Kelvin above the cosmic background radiation of three degrees K. The two regions of strongest molecular emission do not correspond to any objects seen at visible wavelengths; they do, however, coincide with areas of strong infrared emission. One such object is an infrared nebula found by D. E. Kleinmann and F. J. Low of the University of Arizona (*A*). Also shown is a tiny source of powerful maser emission (*B*) from the hydroxyl radical (OH) and water vapor (H_2O), the spectra of which are shown on the opposite page. This source lies very near a bright starlike source of infrared radiation (*C*) discovered by E. E. Becklin and G. Neugebauer of the California Institute of Technology. It is believed that these molecular clouds actually lie behind the bright Orion nebula in a dense region of gas not yet ionized by the central Trapezium stars. Carbon monosulfide (CS) is distributed in much the same way as the formaldehyde within this region. A cloud of hydrogen cyanide (HCN) in the same region is about twice the size of the clouds of carbon monosulfide and formaldehyde, and is similar to a cloud of the unidentified molecule "X-ogen." Carbon monoxide (CO) is observed in a gigantic cloud about one degree in diameter (more than 20 times the size of the carbon monosulfide cloud). It may exist throughout the ionized region of the nebula as well as surrounding it.

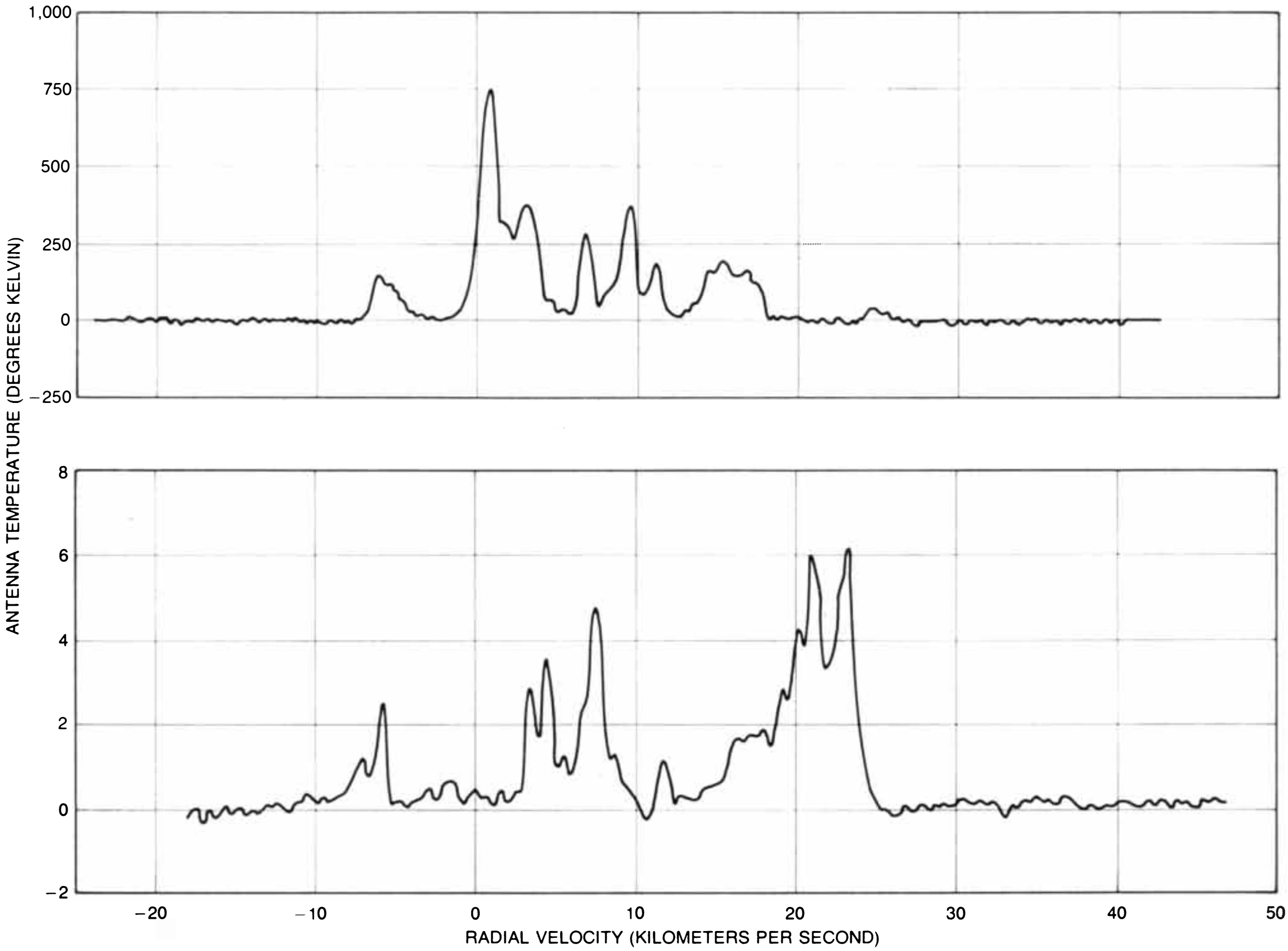
That was more or less the picture of the interstellar medium in 1968, the year marking the birth of molecular astronomy as we now know it. In 1963 radio astronomers had detected the hydroxyl radical (OH). By 1968 OH had been observed in a few dozen directions in the galaxy, nearly all of them being toward H II regions and young stars detectable primarily at infrared wavelengths. Because the H II regions radiate at a temperature higher than the temperature of the interstellar medium and the OH gas embedded in it, the observers expected that the OH gas would absorb some of the radiation from the H II regions. Therefore it was expected that the spectral lines of the OH gas would be seen as an absorption feature in the radio spectrum of the H II regions.

Emission from Hydroxyl

In many cases, however, the OH was observed not as an absorption feature but as intense, narrow emission features in the radio spectrum of the H II regions [*see illustration on opposite page*]. If the OH molecules had been absorbing the radiation as had been expected, it would have indicated that most of them were in a low energy state. The opposite behavior indicates that the energy states of the OH molecules are quite different from what one would expect if the OH gas were in equilibrium with its cool surroundings. The OH gas is evidently being "pumped" by some mechanism into an excited condition in which it is able to amplify the background radiation. Such behavior describes an interstellar hydroxyl maser.

The maser is the predecessor of the laser; it amplifies radio waves instead of visible light. In an interstellar maser the OH molecules are excited so that most of them are in a high energy state instead of a low energy state. When one of the molecules drops to a low energy state, it emits a quantum of energy at a certain narrow line in the radio spectrum that stimulates other molecules to do the same. Some source of energy within the interstellar cloud keeps pumping the molecules back up to the excited state after they have fallen out of it, so that the process continues.

Rather than contributing a fresh insight into the physical conditions of the interstellar medium, however, the observations of the peculiar OH emission presented many new questions about the excitation of the molecules of the medium. At the sites of strong OH emission the medium is very dense by interstellar standards: it consists of as many as 10^8



SPECTRA OF WATER (*top*) and of the hydroxyl radical (*bottom*) show that these interstellar molecules are emitting radiation rather than absorbing it. These emission spectra were taken in the direction of the infrared objects of the Orion nebula. They are calibrated in terms of the temperature of the antenna receiving the radiation (*vertical scale*); the brightness temperature is equal to the antenna temperature multiplied by the square of the ratio of the size of the telescope beam to the angular size of the source. Long-baseline interferometry has established that the angular size of these two emission sources is no more than .005 second of arc. This implies that the brightness temperature is in excess of 10^{13} degrees K. for both water and the hydroxyl radical. The emission from the water molecules is highly variable and changes on a time scale of a few days. Emission from the hydroxyl radical is strongly polarized. All these characteristics suggest that a maser is at work in interstellar space "pumping" the molecules into an excited state so that they amplify the weaker local or background radiation and produce the powerful emission signals that are observed.

particles per cubic centimeter. These sites seem to be confined to the atmosphere of cool young stars or to protostars: objects on their way to becoming stars.

The More Complex Molecules

Interstellar molecules consisting of more than two atoms were first discovered late in 1968. They have profoundly altered our concept of interstellar chemistry and modified our views of physical conditions in space. In 1968 most astronomers believed the density of the interstellar medium was so low that it would be difficult to get more than two atoms to combine. They expected to find only diatomic molecules in interstellar space, and even those molecules would be short-lived because of the destructive effects of ultraviolet radiation and cosmic rays. Such ideas needed drastic revision when a group at the University of California at Berkeley (Charles H. Townes, William J. Welch, A. C. Cheung, David M. Rank and D. D. Thornton) found ammonia (NH_3) in several interstellar clouds in the region of the galactic center. Soon thereafter the Berkeley group detected emission signals from water vapor in several regions of the galaxy.

Although more complex molecules were subsequently discovered, ammonia and water had already established two of the most prominent trends observed in interstellar molecular clouds. First, the clouds are quite dense compared with any other known interstellar region. Second, the physical conditions responsible for exciting the molecules to radiate or to absorb are quite different from terrestrial conditions. In some instances the conditions give rise to an interstellar maser. In others they create what might be called an interstellar refrigerator. At some wavelengths a molecule such as formaldehyde (H_2CO) absorbs more energy than the surrounding conditions would seem to allow if the conditions are such that the terrestrial laws of thermodynamics are applicable.

Where the Molecules Are Found

Since 1968 the rate at which new interstellar molecules have been discovered has increased exponentially. The list now stands at 26 molecules. Where are these molecules found? How are they formed and destroyed? How can the astronomer utilize their signals to understand the physical and chemical processes that are operating in interstellar space?

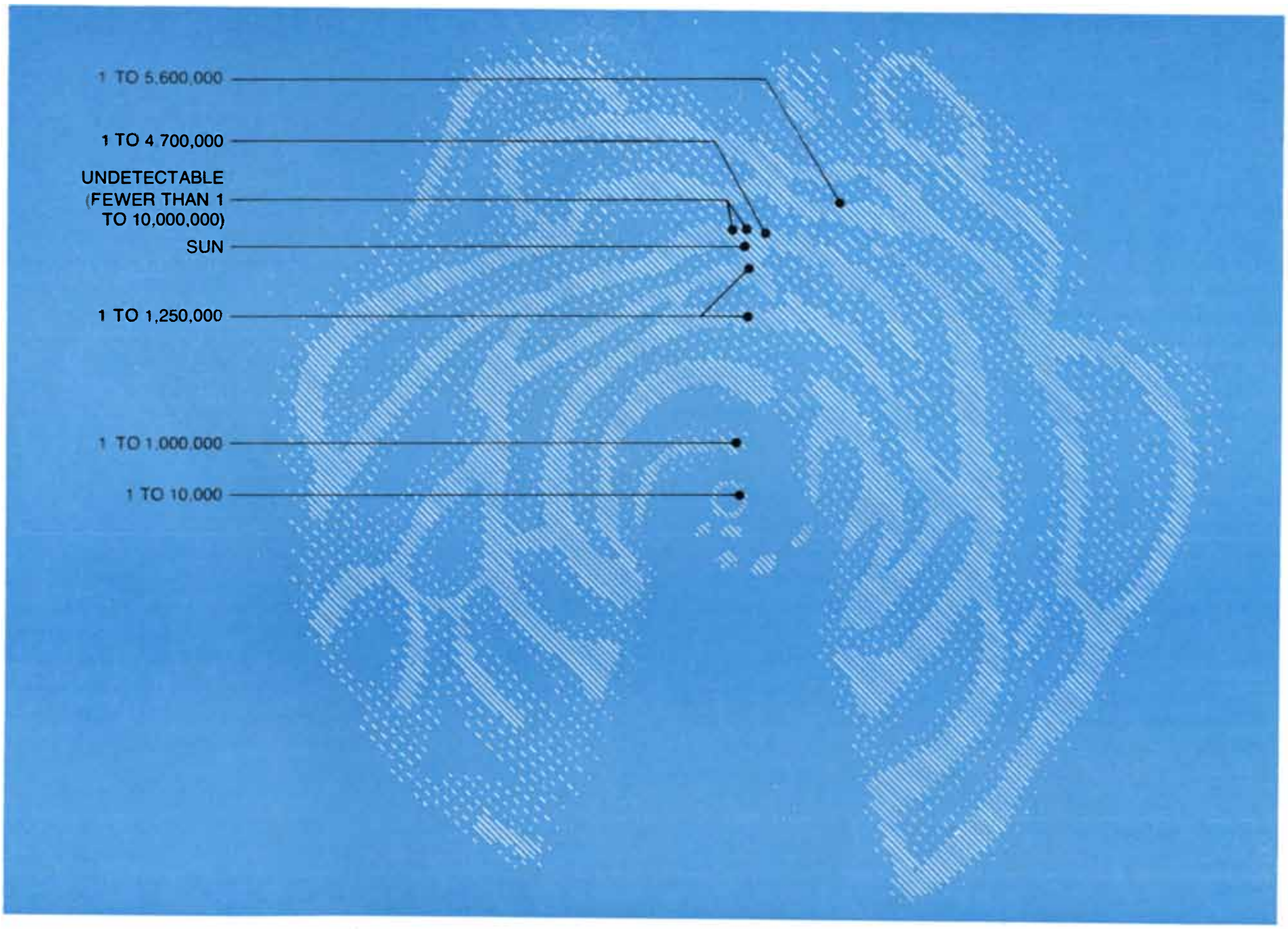
Over the past two years the galaxy has been surveyed intensively at the wavelengths characteristic of the hydroxyl radical, formaldehyde and carbon monoxide (CO). The surveys have shown that these molecules, like atomic hydrogen, are strongly concentrated toward the central plane of the galaxy in a layer that in the vicinity of the sun is less than 1,000 light-years thick. Within this disk the molecules appear to be distributed widely, reaching higher concentrations closer to the center of the galaxy. Carbon monoxide and formaldehyde are distributed in much the same way. Most of the other interstellar molecules, however, are observed only in a very few regions. Whether this distribution is because they do not exist in most regions, or because they are simply not excited in a way that would enable them to emit or absorb measurable signals, is not known.

It is now clear that the particular regions where molecules are found in interstellar space are just those regions that have a high density of particles. There are two reasons for this correlation. First, the higher the density of the interstellar medium, the greater the chance that molecules will be formed from the atoms present. The molecules are formed either by a simple collision between the atoms of interstellar gas or by their coming together on the surface of a particle of interstellar dust. Second, many of the known interstellar molecules must be excited to certain energy states in order to emit detectable radiation. These states are best produced by constant collisions with the other atoms or molecules in a gas, which will happen more frequently in a denser gas. What kinds of interstellar region possess the high densities required?

Dark Clouds and Protostars

Within 1,000 light-years of the sun there are about a dozen dark clouds that show up on photographs as holes in the background of stars. Perhaps 3,000 such clouds exist throughout the galaxy. They are typically 12 light-years in diameter, and the mass of the dust in them is about 20 times the mass of the sun. Some of the interstellar molecules—for example the hydroxyl radical, ammonia, formaldehyde and carbon monoxide—have been detected in these clouds, but most have not. From such observations it is inferred that the total mass of the gas in the clouds exceeds the mass of the dust by well over 100 times. The density of the gas is typically between a few hundred and 10,000 particles per cubic centimeter and the temperature is low: not more than four degrees K. in some clouds and perhaps as high as 25 degrees K. in the hottest ones. From the extent and mass of the dark clouds it can be calculated that they are gravitationally unstable and are collapsing to form stars.

It is believed the dark clouds slowly contract into what are called globules: small clouds of dust that show up on photographs of emission nebulas as com-



PLAN VIEW OF THE GALAXY shows the distribution of hydroxyl molecules throughout the interstellar medium. The galaxy was surveyed in the radio wavelength of the spectral line of atomic hydrogen; the map represents the way hydrogen is distributed throughout the spiral arms. The ratio of the hydroxyl radical to hydrogen increases a thousandfold between the region near the solar system and the galactic center. There one hydroxyl radical is found for every 10,000 hydrogen atoms, but surveys in other directions show fewer than one hydroxyl radical per four million hydrogen atoms or are unable to detect hydroxyl-radical absorption at all.

pletely opaque specks. Little is known about the gas content of these objects, and estimates of their mass are only guesses at a minimum figure. Recently, however, the hydroxyl radical and formaldehyde have been detected in a few globules. Preliminary analysis indicates that the temperature of the globules is even lower than that of the larger dust clouds and that the density is somewhat higher.

A globule is believed to be one of the final stages in the collapse of an interstellar cloud into a protostar: a denser concentration of dust and gas, perhaps about the size of the solar system, that is close to becoming an actual star. In the protostar stage of collapse the density and temperature near the center of the object rise sufficiently for it to radiate rather strongly in the infrared region of the spectrum. Theoretical work by R. B. Larson of Yale University shows that the protostar collapses much faster near the center than in the outer parts, so that newly born stars should be surrounded by a large envelope of dust and gas. That envelope in turn is surrounded by the remnants of the interstellar cloud. If the new star is a very luminous Type O or Type B star, it will emit enough ultraviolet radiation to ionize the hydrogen in the surrounding cloud, forming an H II region. The H II region will often be surrounded by a cloud of un-ionized gas that has not been reached by the ionizing radiation.

Where do molecules fit into this picture? It appears that they are associated with every one of the steps in the contraction of an interstellar cloud. We have seen that the hydroxyl radical, formaldehyde and carbon monoxide are observed in virtually all regions of the spiral arms and possibly between the arms. The larger dust clouds also contain ammonia. It is only when we turn to regions where it is suspected stars are being formed that we observe the remaining molecules. Here the molecules may be associated with the protostars themselves. Alternatively they may exist in the dense, un-ionized and mostly very massive clouds that surround the H II regions generated by newly formed stars.

A good example of the second kind of region is the Great Nebula in Orion. Embedded in the Orion nebula is a cluster of very hot young stars known as the Trapezium cluster. These stars are the energy source of the nebula itself. Surrounding the nebula is a large cloud of neutral hydrogen; just behind this cloud is a dense dark cloud. In its center infrared astronomers have detected several objects that appear to be protostars.

Molecules are found throughout the region of the Orion nebula. A great cloud of carbon monoxide, occupying about one degree of arc in the sky and containing some 100 solar masses of the gas, extends outward from the nebula into the surrounding cloud of neutral hydrogen. The hydroxyl radical has a similar distribution. Smaller clouds of hydrogen cyanide (HCN) and of an unidentified molecule dubbed X-ogen are seen projected against the central portions of the nebula. In the immediate vicinity of the infrared objects are found strong concentrations of formaldehyde, methyl alcohol (CH_3OH), carbon monosulfide (CS), the cyanogen radical (CN), ammonia and cyanoacetylene (HC_3N). There are also highly excited but very small clouds of the hydroxyl radical and water vapor that are emitting powerful radio signals of the maser type.

The distribution of these molecules within the Orion cloud is not accidental. Molecules exist in observable quantities only in regions that satisfy two conditions: (1) that the rate at which the molecules are formed exceed the rate at which they are destroyed, and (2) that the molecules be in a region that can be excited into the observed energy states. Regions of high density tend to satisfy both of these requirements. The hydroxyl radical (when it is observed as an absorption feature) and carbon monoxide are observed in states of low excitation and are also relatively immune to destruction by light. This explains why they are observed throughout the galaxy, and also why they are observed over a larger region in the Orion nebula than the other molecules. The molecules observed in the small central part of the Orion nebula are not observed throughout the galaxy but are seen only in a few apparently very dense areas in the vicinity of certain H II regions. In fact, it is not the Orion nebula but another H II region near the galactic center, known as Sagittarius B2, that is at present the only observed site of eight of the 26 known interstellar molecules. What makes the Sagittarius B2 cloud so unique is not its temperature (about 100 degrees K.) but its enormous densities (up to 10^8 particles per cubic centimeter) and also its size (about 20 light-years), which together make its density in the line of sight 100 or more times greater than the known density in any other interstellar cloud.

Interstellar Masers

One of the earliest indications that interstellar molecules were probably associated with protostars came from the maser emission of the hydroxyl radical and water vapor. Unlike the hydroxyl radical, which is often observed as a fairly normal absorption feature against sources that have a continuous spectrum in the radio region, water vapor is seen only in the form of powerful emission lines. Moreover, it is seen only in regions that can be smaller than the solar system, perhaps even smaller than the orbit of the earth. The emission lines of water vapor are so strong that if they were radiated from an ordinary heated body, the temperature of that body would have to be 10^{13} degrees K. Such temperatures cannot be connected with the kinetic energy of the molecules or with any temperature of radiation unless the radiation has been amplified. It is these characteristics that imply that an interstellar maser is at work.

Soon after interstellar water masers were discovered in 1968 it was noticed that they are always found in regions where the hydroxyl radical is also emitting powerful maser radiation. The equivalent radiation temperature for the hydroxyl radical is about as high as that for the water. In some of these sources the rate at which energy is radiated at just the two narrow spectral lines of the hydroxyl radical and water is as much as the energy emitted by the sun at all wavelengths. The energy must come indirectly from the gravitational collapse of the protostar; indeed, the emission acts to cool the cloud and cause it to collapse even faster.

Theories explaining such a spectacular excitation of the hydroxyl radical and water, and why they are found together, have centered on various kinds of pumping and on the effects of strong infrared radiation. These processes work well only if the densities are as high as 10^8 particles per cubic centimeter and the temperatures are several hundred to 1,000 degrees K. These conditions exist only in protostars such as the infrared sources in the Orion nebula. In fact, a rather large amount of energy (equivalent to 700 degrees K.) along with the high densities is needed to excite water molecules to radiate in the observed manner at all. This probably precludes interstellar water's ever being observed in any other type of object not having similar special conditions.

Conditions required to give rise to the maser action of the hydroxyl radical and water also exist in the outer regions of certain very cool stars. That brings us to the last category of celestial objects in which molecules are observed. A number of molecules—for example diatomic hydrogen (H_2), water, carbon monoxide, the cyanogen radical, diatomic carbon (C_2), methylidyne (CH), hydrogen cyanide (HCN) and acetylene (C_2H_2)—have long been known through their optical and infrared spectra in the atmosphere of cool red Type M stars and the stars known as Mira Ceti variables. It was not until 1968, however, that W. J. Wilson and Alan H. Barrett of the Massachusetts Institute of Technology detected powerful hydroxyl-radical signals at radio wavelengths from infrared stars not associated with H II regions. Many of these stars appear to be very young; several are Mira Ceti variables that G. Neugebauer and Robert B. Leighton of the California Institute of Technology had found earlier were strong sources of infrared radiation. Two years later maser emission from water vapor was also discovered in Mira Ceti variables; it is believed that these masers are being pumped by the strong infrared radiation.

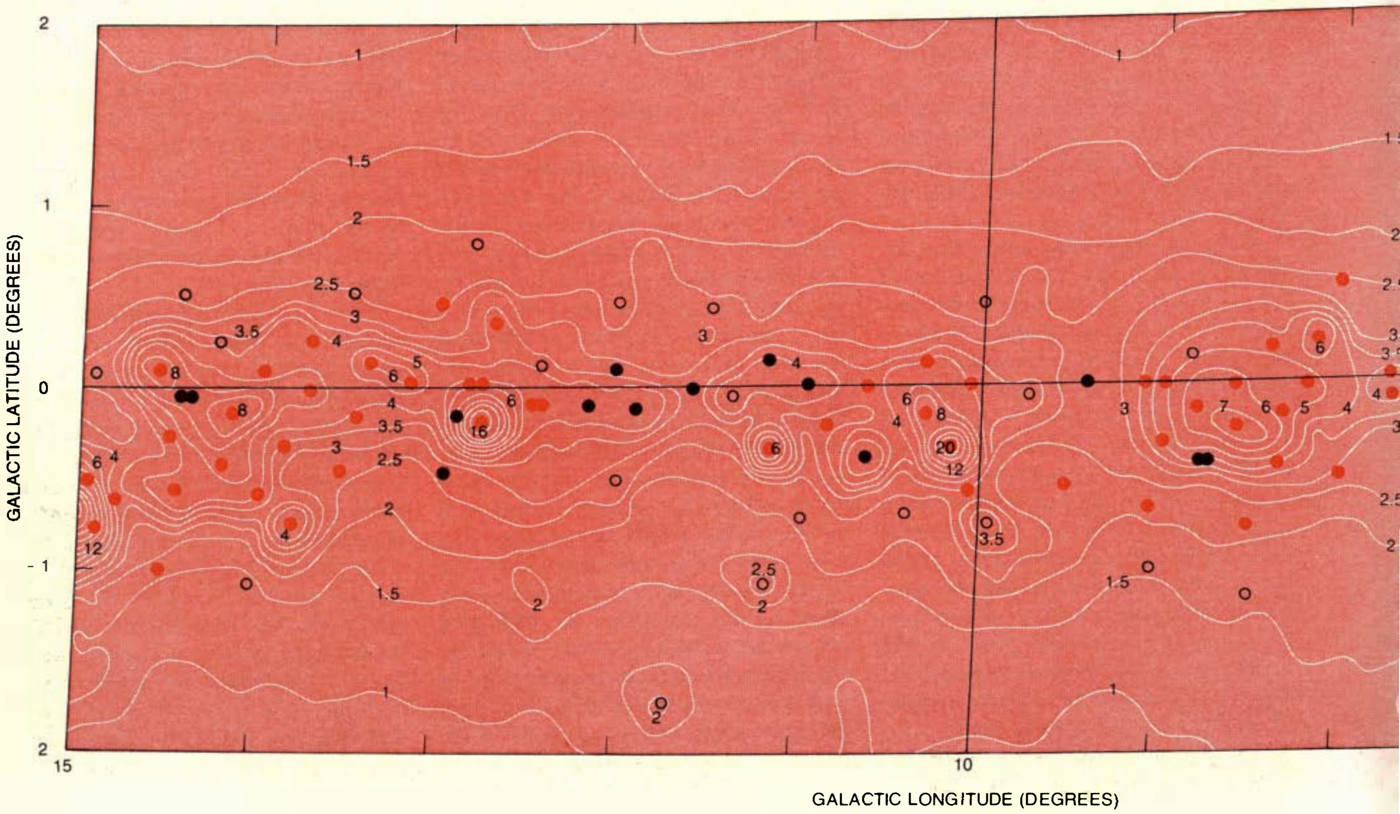
Marvin M. Litvak of M.I.T. has shown that the hydroxyl radical would emit radiation at the frequency of 1,612 megahertz (million cycles per second) under the action of infrared pumping. In the infrared sta

ly at just that frequency. This is not the case for the hydroxyl radical in protostars, which radiates predominantly at 1,665 megahertz and may be pumped by chemical processes. The emission from the hydroxyl radical and water is thought to arise in the shell of gas and dust that surrounds these stars. Observations of the shell typically show features in the spectrum that are Doppler-shifted in a way suggesting that the shell is expanding or contracting at a velocity of between 10 and 30 kilometers per second.

Most or all of the stars with the hydroxyl radical and water vapor in their outer atmosphere appear to contain more oxygen than carbon. In other types of cool star the opposite is true. The hydroxyl radical and water vapor are not found in these "carbon" stars, but carbon-containing molecules such as carbon monoxide, carbon monosulfide (CS), HCN and CN have been observed in them at both radio and infrared wavelengths. The radio observations have proved to be particularly interesting because they indicate the presence in these molecules not only of carbon 12, the commonest isotope of carbon, but also of the rarer isotope carbon 13. In these stars the abundance ratio of carbon 12 to carbon 13 is only about 4 : 1; on the earth and in most of the interstellar medium the ratio is 89 : 1.

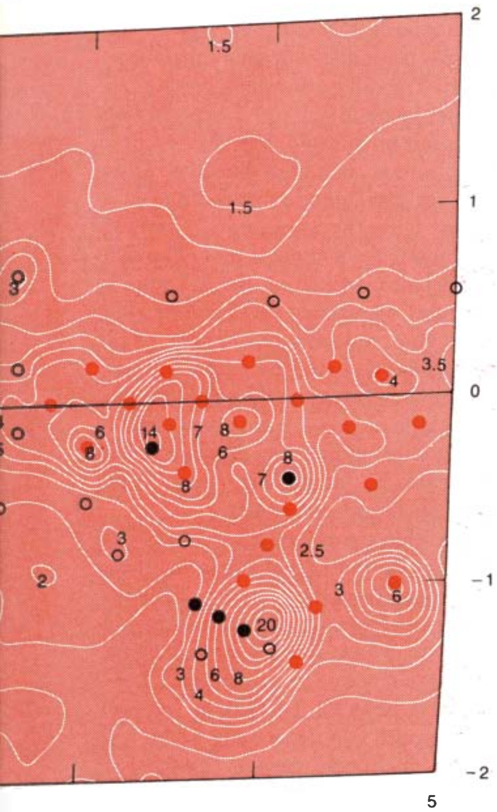
Molecules as Interstellar Probes

Direct information about the constitution, location and velocity of interstellar clouds can be gained by detecting molecules and measuring their position and the amount by which their spectral features are Doppler-shifted. Combined with other knowledge, a somewhat deeper analysis of the intensity of the signals can yield information about the physical conditions and dynamics within the clouds. The particular spectral lines that



EDGE-ON VIEW OF THE GALAXY near the galactic center has been mapped at the radio wavelength of 11 centimeters by Wilhelm J. Altenhoff of the Max Planck Institute for Radio Astronomy in Bonn. The contour lines are measures of the intensity of the radiation from various radio sources such as H II regions (regions of ionized hydrogen near hot young stars) and the remnants of supernova explosions. The radiation is in the form of a continuous spectrum, as distinguished from the "line" spectrum that characterizes atoms or molecules in which emission or absorption occurs only in certain very narrow intervals of wavelength. The numbers on the contour lines are brightness temperatures. The dots represent regions searched by the author for emission or absorption signals from hydroxyl. The black dots indicate where the hydroxyl radical has been observed as a source of powerful emission as an interstellar maser. The colored dots mark the locations where OH is seen to absorb some of the background radiation at its characteristic wavelength of 18 centimeters. Open circles are regions where no OH has been detected. Many of the regions of OH emisare the signature of molecules arise when the molecules or their constituent electrons change their state of motion. Each molecule tends to rotate around the axes of symmetry characteristic of it. Changes in rotation cause the molecule to radiate or absorb electromagnetic energy at wavelengths that typically lie in the microwave region of the spectrum: wavelengths between about one millimeter and six centimeters. The atoms within the molecule vibrate with respect to one another as well as rotate around one another. The vibrational motion can also change, causing the molecule typically to emit or absorb infrared radiation. Moreover, the electrons of the various atoms can jump from one orbit to another, and these transitions give rise to radiation or absorption in the visible or ultraviolet regions of the spectrum.

Carbon monosulfide is an example of the simplest possible molecule: a linear diatomic molecule. In its states of lowest electronic and vibrational energy, which are the only states populated in interstellar space, the possible motions of the molecule are simply movement in one direction and end-over-end rotation. The laws of quantum mechanics show that the corresponding energy states form a simple "ladder." A molecule can move up or down the ladder when it collides with another molecule or another kind of particle, or when it absorbs or emits a photon (a quantum of radiation). If left to itself, the molecule will spontaneously emit photons and cascade to the ground state, or lowest energy level, in a definitely specified time. The energy levels are labeled by the quantum number J, a measure of the rotational momentum of the molecule.



sion are also sources of strong maser emission from water vapor. As can be seen, OH is a widespread constituent of the galactic disk and is confined closely to the galactic plane (galactic latitude = 0 degrees). Carbon monoxide and formaldehyde are distributed in much the same way throughout the galaxy. Most other interstellar molecules have been observed only in very few sources.

For carbon monosulfide it takes an average of four hours for the molecule to drop from the first energy level to the ground state. The molecule spends even less time in the higher levels. The astronomer detects radiation from the higher levels, so that some mechanism must be acting to maintain the carbon monosulfide molecules in these levels, balancing their tendency to decay spontaneously. Fields of radiation could act in this way, but they would have to be fields that are much stronger than the ones that are directly observed in these clouds. A more likely mechanism of excitation is collisions with other particles.

For the kinds of particle we are considering and the temperatures involved we can estimate the density required to maintain molecules such as carbon monosulfide in excited states. Typically densities of one million particles per cubic centimeter result for the clouds of carbon monosulfide observed near H II regions. Other molecules that have energy-level schemes similar to those of carbon monosulfide include carbon monoxide, silicon monoxide (SiO), carbonyl sulfide (OCS), hydrogen cyanide and cyanoacetylene (HC_3N). Carbon monoxide has a much lower rate of spontaneous decay than the other linear molecules. Hence a lower rate of excitation by collision is required for its excited levels to be maintained. This is one reason carbon monoxide is observed over larger regions of space than other molecules, and it makes it possible to trace interstellar clouds out to regions where the density is as low as about 100 particles per cubic centimeter.

By observing different transitions of the molecules within a cloud it is possible to monitor different regions of density. The higher energy levels require higher densities for the appropriate excitation. A given spectral line has a certain width that is determined partly by the temperature of the cloud (the higher the temperature, the broader the line) and partly by turbulence or large-scale motions within the cloud. In most sources of interstellar molecules turbulence is dominant over the temperature effect. In these cases the narrower the spectral line, the smaller the net turbulence and hence the smaller the region within the cloud that is emitting the spectral line. For example, the spectral line emitted by carbon monosulfide as it drops from the second energy level to the first ($J = 2$ to $J = 1$) is seen to be emitted from two distinct ranges of velocity [*see bottom illustration on page 60*]. The line resulting from a drop from the first energy level to the ground state ($J = 1$ to $J = 0$) is emitted over a much larger range of velocities. This is interpreted as indicating that within the large cloud of carbon monosulfide there are two smaller regions of higher density. Observations of more highly excited molecules should further refine the density map.

How are the temperatures of a cloud determined from molecular observations? It turns out that molecules such as carbon monosulfide, which radiate their energy rapidly, are rather insensitive probes of temperature. Other molecules, such as carbon monoxide and ammonia, are more useful. If a molecule of carbon monoxide radiates its energy away much more slowly than it is excited by collisions, then it comes into thermal equilibrium with its surroundings. Under these conditions the intensity of the various transitions between energy levels can be exactly calculated as a function of the temperature only. Observations of the intensity of the transitions then make it possible to deduce the temperatures within the cloud. In this way carbon monoxide has yielded temperatures of between five degrees and 25 degrees K. in the large dust clouds, and typically of 100 degrees K. in the molecular clouds surrounding H II regions.

Ammonia is a unique probe of physical conditions in interstellar clouds because this molecule has several different types of energy state. The molecule is a tetrahedron with a nitrogen atom at an apex above the plane formed by three hydrogen atoms; the nitrogen atom can oscillate back and forth through the plane. As a result each level of the rotational energy, J, is split into two closely spaced levels by the oscillation. Transitions between these closely spaced levels are known as inversion transitions. They are observed at the microwave wavelength of 1.3 centimeters. For ammonia a second quantum number, K, refers to the component of angular momentum around the molecule's axis of symmetry. The quantum numbers are written to-

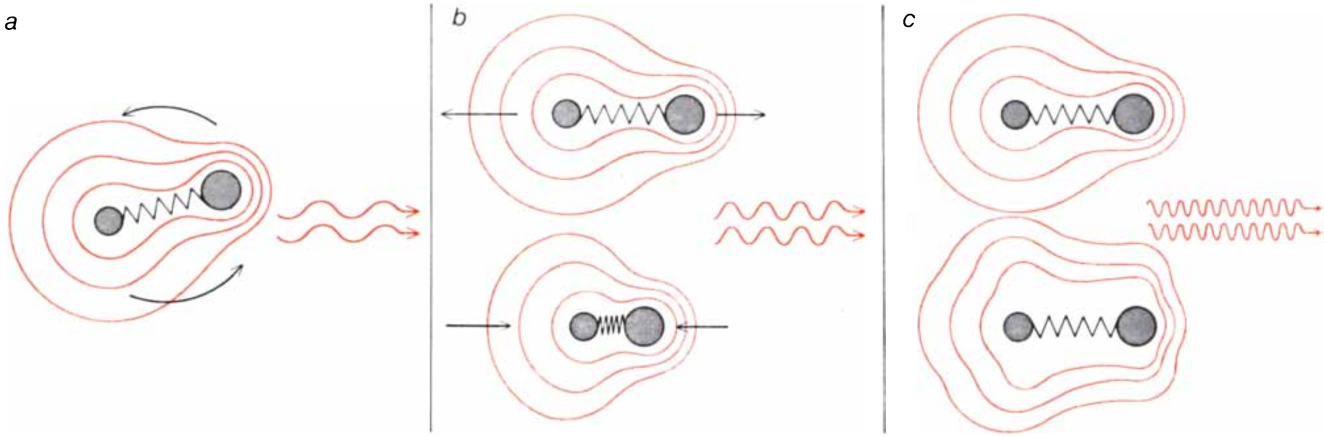
YEAR	MOLECULE	FORMULA	WAVELENGTH	NUMBER OF REGIONS	TELESCOPE
1937	METHYLIDYNE (IONIZED)	CH^+	3,958 ANGSTROMS*	88	100-INCH, M.W.O.
1937	METHYLIDYNE	CH	4,300 ANGSTROMS*	60	100-INCH, M.W.O.
1939	CYANOGEN RADICAL	CN	3,875 ANGSTROMS*	14	100-INCH, M.W.O.
1963	HYDROXYL RADICAL	OH	18.0 CENTIMETERS*	~600	84-FOOT, L.L.
1968	AMMONIA	NH_3	1.3 CENTIMETERS*	12	20-FOOT, H.C.O.
1968	WATER	H_2O	1.3 CENTIMETERS	35	20-FOOT, H.C.O.
1969	FORMALDEHYDE	H_2CO	6.2 CENTIMETERS*	~150	140-FOOT, N.R.A.O.
1970	CARBON MONOXIDE	CO	2.6 MILLIMETERS	60	36-FOOT, N.R.A.O.
1970	HYDROGEN CYANIDE	HCN	3.4 MILLIMETERS	10	36-FOOT, N.R.A.O.
1970	X-OGEN	?	3.4 MILLIMETERS	8	36-FOOT, N.R.A.O.
1970	CYANOACETYLENE	HC_3N	3.3 CENTIMETERS*	4	140-FOOT, N.R.A.O.
1970	HYDROGEN	H_2	1,060 ANGSTROMS	2	N.R.L.
1970	METHYL ALCOHOL	CH_3OH	35.9 CENTIMETERS*	3	140-FOOT, N.R.A.O.
1970	FORMIC ACID	HCOOH	18.3 CENTIMETERS	1	140-FOOT, N.R.A.O.
1971	CARBON MONOSULFIDE	CS	2.0 MILLIMETERS*	20	36-FOOT, N.R.A.O.
1971	FORMAMIDE	NH_2CHO	6.5 CENTIMETERS*	2	140-FOOT, N.R.A.O.
1971	CARBONYL SULFIDE	OCS	2.5 MILLIMETERS*	1	36-FOOT, N.R.A.O.
1971	SILICON MONOXIDE	SiO	2.3 MILLIMETERS*	2	36-FOOT, N.R.A.O.
1971	METHYL CYANIDE	CH_3CN	2.7 MILLIMETERS	1	36-FOOT, N.R.A.O.
1971	ISOCYANIC ACID	HNCO	3.4 MILLIMETERS*	1	36-FOOT, N.R.A.O.
1971	HYDROGEN ISOCYANIDE?	HNC?	3.3 MILLIMETERS	3	36-FOOT, N.R.A.O.
1971	METHYLACETYLENE	CH_3CCH	3.5 MILLIMETERS	1	36-FOOT, N.R.A.O.
1971	ACETALDEHYDE	CH_3CHO	28.1 CENTIMETERS	1	140-FOOT, N.R.A.O.
1972	THIOFORMALDEHYDE	H_2CS	9.5 CENTIMETERS	1	210-FOOT, C.S.I.R.O.
1972	HYDROGEN SULFIDE	H_2S	1.8 MILLIMETERS	7	36-FOOT, N.R.A.O.
1972	METHANIMINE	CH_2NH	5.7 CENTIMETERS	1	210-FOOT, C.S.I.R.O.

KNOWN INTERSTELLAR MOLECULES number 26 at present. They are listed in order of their discovery along with the wavelength and the telescope at which they were first detected. An asterisk by the wavelength indicates that the corresponding molecule has been observed at additional wavelengths since its discovery. Two molecules deserve special mention: the cyanogen radical (CN) was detected at the microwave wavelength of 2.6 millimeters in 1970, and carbon monoxide at the ultraviolet wavelength of about 1,400 angstroms in 1971. Only these two molecules have been observed in two entirely different regions of the spectrum. Eight of the molecules have so far been detected in just one source: the unusually dense and massive interstellar cloud Sagittarius B2, which is associated with an H II region not far from the galactic center. That hydrogen (H_2), the simplest of all interstellar molecules, should be observed in only two clouds so far is surprising because it ought to be the stable form of hydrogen in all interstellar clouds dense enough to yield other detectable molecules. The designation "100-inch, M.W.O." refers to the 100-inch reflecting telescope of the Mount Wilson Observatory; "84-foot, L.L.," to the 84-foot radio telescope at the Lincoln Laboratory of the Massachusetts Institute of Technology; "20-foot, H.C.O.," to the 20-foot radio telescope at the Hat Creek Observatory of the University of California; "140-foot, N.R.A.O.," to the 140-foot radio telescope at the National Radio Astronomy Observatory in Green Bank, W.Va.; "36-foot, N.R.A.O.," to the 36-foot radio telescope of the National Radio Astronomy Observatory located at the Kitt Peak National Observatory in Arizona; "N.R.L.," to an Aerobee rocket launched by Naval Research Laboratory; "210-foot, C.S.I.R.O.," to the 210-foot radio telescope of Commonwealth Scientific and Industrial Research Organization radio observatory located at Parkes in Australia.

gether: J,K. Ammonia can change its value of J by 1 by emitting or absorbing photons. Collisions may cause J to change by more than 1. K, on the other hand, cannot change its value by the emission or absorption of photons. Collisions between particles in a gas allow transitions only between states whose K is $3n + 1$ and $3n - 1$, or between states whose K is $3n$ (n is any integer). The species whose K is $3n + 1$ and $3n - 1$ is called para-ammonia; the species whose K is $3n$ is called ortho-ammonia.

The number of ammonia molecules in the various states of rotational energy can be directly estimated by observing the intensity of emission from the inversion transitions. Some rotational levels (such as the 2,1 level) decay in a time as short as 20 seconds to lower states (such as the 1,1 level) by spontaneously emitting a photon. To observe a spectral line from the 2,1 level as strong as some that have been detected requires a high rate of excitation by collisions to compete with the rate of deexcitation by radiation. Therefore the density must be very high. In fact, in the Sagittarius B2 ammonia source a density as high as 10^9 particles per cubic centimeter may be required. Transitions between certain other states (such as 1,1 and 2,2) occur only as a result of collision, but they occur at a high rate at such high densities. These states are thus brought into thermal equilibrium with the surrounding particles on a time scale of about a year. The ratio of molecules populating the 1,1 state to molecules populating the 2,2 state (that is, the observed ratio of intensity of the spectral lines) should be controlled only by the temperature of the colliding particles. Arguments such as this one lead to temperatures of between 20 degrees and 80 degrees K. for various clouds in the region of the galactic center.

One other temperature can be derived from the observations of ammonia. It can be obtained from the presence of the two different species of ammonia: ortho-ammonia (to which the 3,3 state belongs) and para-ammonia (to which the 1,1 and 2,2 states belong). Transitions between these species can occur only if the spin of the hydrogen nuclei (protons) is changed. That process is likely to occur only when the molecules collide with particles of interstellar dust and are broken up and reassembled. The average time between such spin-changing collisions is probably at least a million years. Thus the ratio of the intensity of the spectral line for molecules in the 3,3 state to the intensity for molecules in the 1,1 state or the 2,2 state represents some temperature inside the cloud a million



MOLECULES OR THEIR ELECTRONS can change their state of motion, giving rise to various lines in the electromagnetic spectrum. Each molecule tends to rotate around its characteristic axes of symmetry; in the case of carbon monosulfide, a simple diatomic molecule, the rotation is end over end (*a*). Changes in the rotation cause the molecule to radiate or absorb energy at wavelengths that typically lie in the microwave region of the spectrum. The atoms within the molecule vibrate with respect to each other as if they were joined by a spring (*b*). The vibrational motion can also change, causing the molecule to emit or absorb shorter-wavelength infrared radiation. Moreover, the electrons of the various atoms can jump from one energy state to another, changing the configuration of the electron orbits around the nuclei of the molecule (*c*). These transitions give rise to radiation or absorption in the visible or ultraviolet regions of the spectrum. In this case the carbon atom is to the left and the sulfur atom is to the right; the molecule is shown before the transition in the first excited state (*top*) and after the transition in the ground state (*bottom*). The contours represent the probability of finding an electron in various parts of the electron cloud. Inner contours are virtually the same for isolated atoms of carbon and sulfur as they are for the molecule; the probability of finding an electron decreases toward the outer contours.

years or so in the past. Such temperatures are found to be higher than the present temperatures, suggesting that the clouds cool as they contract.

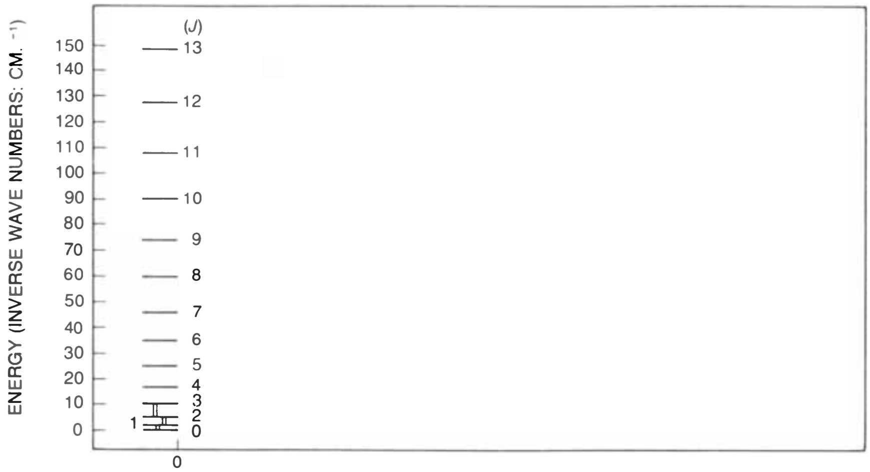
If observations of molecular lines can yield so much information about temperature and density in interstellar clouds, one would expect that they would also give an accurate measure of the abundance of the molecules. This is not the case for a number of reasons. One is that the intensity with which a cloud of molecules emits or absorbs electromagnetic energy of a given wavelength depends not only on the total number of molecules present but also on what fraction of them are in the energy states of the observed transitions. The fraction can be calculated if we know that the molecules are in thermal equilibrium with their surroundings. The majority of interstellar molecules are not in thermal equilibrium, however. The most spectacular examples of nonequilibrium are the hydroxyl-radical maser and the water maser. Another problem is that the receiving beam of a radio telescope is fairly wide, and such beam widths often fail to resolve small patches of molecular emission. The result is that the intensity of these patches is underestimated. Nevertheless, for most of the 26 molecules known it is thought the abundances that have been deduced are accurate to within a factor of 10 or so. On the other hand, the estimated amount of hydroxyl radical and of water in sources where they act as an interstellar maser may be in error by a factor of 100 or even 1,000. In spite of these uncertainties many conclusions important to the understanding of interstellar chemistry have emerged.

Chemical Composition of the Clouds

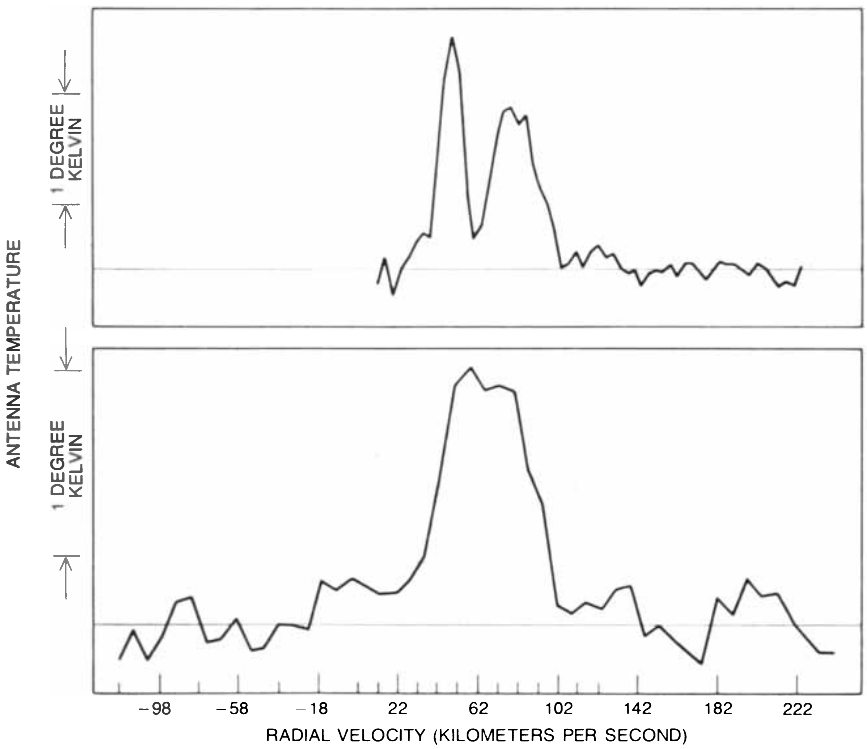
Interstellar clouds that harbor detectable quantities of molecules consist almost entirely of dust particles and molecular hydrogen. The dust particles are observed directly, since they attenuate and redden the light of background stars. Because of severe technical difficulties molecular hydrogen has been directly observed in only two clouds: very tenuous regions in front of the stars Xi Persei and Delta Scorpio that dim the background stars by about one magnitude (roughly 2.5 times). These clouds are not known at present to contain any other molecules. Theoretical considerations, however, indicate that clouds whose extinction of background starlight exceeds about 1.5 magnitudes will contain hydrogen in molecules rather than atoms. The rate at which hydrogen molecules are formed on the dust particles will exceed the rate at which they are dissociated by ultraviolet radiation. It is a general fact that molecules observed in the ultraviolet or visible portions of the spectrum are never seen in the same regions as molecules observed in the radio portion. The reason is that the rate at which any molecule can emit or absorb photons at radio wavelengths is much smaller than the rate for photons at visible or ultraviolet wavelengths. Thus many more molecules are required in the line of sight to produce detectable signals at radio wavelengths.

Any cloud containing enough molecules to be detected at radio wavelengths also contains so much dust that the background stars against which the molecules are observed as absorption features at visible or ultraviolet wavelengths are completely obscured. The result is that the molecules detected at visible or ultraviolet wavelengths (such as molecular hydrogen and CH or CH^+) are observed only in low-density clouds where the extinction of starlight is low and where the density of the particles does not exceed some 100 per cubic centimeter. On the other hand, molecules observed at radio wavelengths are seen only in much denser, darker clouds, where the density of the particles ranges from 100 to 10^9 per cubic centimeter and the extinction probably exceeds 50 magnitudes (a factor of more than 10^{20}).

In these dense clouds of molecular hydrogen and dust, molecules are observed in only trace amounts. That is because the molecules contain atoms other than hydrogen, which on a cosmic scale are much less abundant than hydrogen. For every 10,000 hydrogen atoms there are three or four oxygen atoms, two carbon atoms, one or two nitrogen atoms and one sulfur atom. Here, however, an important fact emerges. If this cosmic abundance scale (determined largely from atomic spectra observed in stars



ENERGY LEVELS of a molecule are labeled by the quantum number J, a measure of the rotational momentum of the molecule. For carbon monosulfide it takes an average of four hours for the molecule to drop from the first energy level to the ground state. The molecule spends even less time at higher energy levels. The vertical bars between the energy levels $J = 3$, $J = 2$, $J = 1$ and $J = 0$ represent transitions of carbon monosulfide that have been observed in interstellar space by radio astronomers. The unit *cm.*$^{-1}$, known as an inverse wave number, is proportional to energy and refers to the number of wavelengths per centimeter. Transitions are measured in terms of changes in energy, and the inverse wave number is a direct indication of the amount of energy involved. For example, a difference in energy of five inverse wave numbers corresponds to a transition that has a wavelength of two millimeters; the corresponding spectral line would be in the microwave region.



SPECTRA OF CARBON MONOSULFIDE in the interstellar cloud Sagittarius B2 are shown for two transitions of the molecule. A given spectral line has a certain width that is determined partly by turbulence; the narrower the spectral line, the smaller the net turbulence and hence the smaller the region within the cloud that is emitting the line. The line emitted by carbon monosulfide as it drops from the second energy level to the first (from $J = 2$ to $J = 1$) shows two distinct peaks, meaning that the line is emitted from two discrete regions each having a different radial velocity (*top*). On the other hand, the line resulting from a drop from the first energy level to the ground state ($J = 1$ to $J = 0$) is emitted over a much larger range of velocities (*bottom*). Collisions between particles account for the molecules in the higher energy state; these features indicate that within the large cloud of carbon monosulfide there are two smaller dense regions. Horizontal lines through spectra average zero-signal level around which there are some noise fluctuations.

near the sun) applies to interstellar clouds in general, then the observed molecules alone account for nearly all the carbon atoms that must be present. The observed molecules must also account for an appreciable fraction (perhaps 30 percent) of all the oxygen atoms. By way of contrast, perhaps no more than .0001 percent of the available nitrogen atoms appear in the observed molecules. Most of the nitrogen atoms are probably in the form of the unobserved diatomic molecule (N_2), which comprises most of the earth's atmosphere.

Two key conclusions follow from these facts. First, the formation of molecules per available atom in interstellar clouds must be a highly efficient process. Second, interstellar chemistry seems to favor the production of organic molecules, that is, molecules containing carbon. All the simplest organic molecules have been found in interstellar space, whereas many of the even simpler nonorganic species such as nitric oxide (NO), sulfur monoxide (SO) and the thiol radical (SH) have not been detected in spite of sensitive searches. No explanation for the fact that interstellar chemistry is predominantly organic has yet emerged.

Molecules and Background Radiation

In 1965 Arno A. Penzias and R. W. Wilson of Bell Laboratories obtained evidence that the universe is bathed in a uniform field of radiation that has a temperature equivalent to three degrees K. at microwave wavelengths. If this field could be shown to have a "black body" spectrum, that is, to have a temperature equivalent to three degrees K. at all wavelengths, the most logical explanation for it would be that it was a remnant of the primordial fireball that marked the beginning of the universe's current phase of expansion. On the other hand, large deviations from a black-body spectrum would discount such an origin. Since 1969 observations of interstellar molecules have played an important role in deciding this fundamental question.

It has been known since 1969 that interstellar formaldehyde, like the hydroxyl radical and water, is excited in a very strange way. One of its transitions, at the wavelength of six centimeters, was observed by P. Palmer of the University of Chicago, B. M. Zuckerman of the University of Maryland, L. E. Snyder of the University of Virginia and David Buhl of the National Radio Astronomy Observatory. They saw the transition as an absorption feature in certain dust clouds against a region of sky that contained no known sources of radiation with a continuous spectrum. The radiation that was being absorbed must therefore be the three-degree cosmic background. Since the observation of formaldehyde is free from many of the technical difficulties of direct measurements of the background field, it is one of the best direct confirmations that the cosmic background exists.

Measurements of the excitation of CN, CH and CH^+ have recently provided estimates of the intensity of the three-degree radiation at radio wavelengths shorter than three millimeters. No other reliable information exists in this part of the spectrum. In the rarefied clouds where those three molecules are observed the rate at which they collide with other particles is so low that their relative populations in the lowest few rotational energy levels are governed largely by the rate at which they interact with photons from the cosmic background radiation. Measurements of the relative populations, deduced from observations of two or more transitions involving their energy levels, yield the intensity of the cosmic background radiation at the wavelengths of those transitions. CN verifies that the intensity is indeed about three degrees at a wavelength of 2.63 millimeters. CH and CH^+ provide similar, although less definitive, results at the wavelengths of .56 and .36 millimeter. It seems likely that further measurements of the excitation of the various molecules now becoming known will be of great help in determining the high-frequency part of the cosmic radiation spectrum.

Isotope Abundances

On the earth and in the sun, comets and meteorites the different isotopes of the chemical elements are found in certain well-defined proportions. As I have indicated, the abundance ratio of carbon 12 to carbon 13 is 89 : 1. The ratio of oxygen 16 to oxygen 18 is 490 : 1, of nitrogen 14 to nitrogen 15, 270 : 1, and of sulfur 32 to sulfur 34, 22 : 1. In the kinds of stars that consume their nuclear fuels in processes involving carbon, nitrogen and oxygen the isotope ratios of these elements are altered. The ratio of carbon 12 to carbon 13 may be as low as 4 : 1, as we have seen, and the ratio of nitrogen 14 to nitrogen 15 may become very large. In stars that consume helium, carbon is processed in such a way that virtually all the carbon 13 is used up.

Stars continually recycle their material through the interstellar medium. They are created from the interstellar gas; they subject the gas to nuclear-burning processes, and they return the gas to interstellar space by ejecting matter or by exploding in a supernova. Therefore the isotope ratios in the solar system ought to be accidental ones resulting from nonequilibrium processes or from the mixing of gases from various sources.

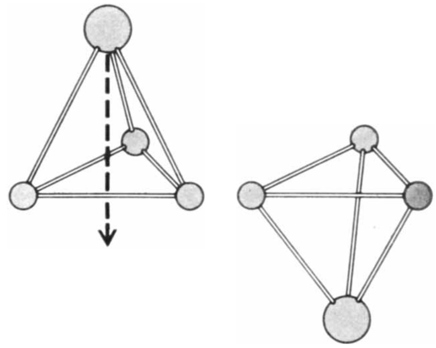
The startling result provided by the study of interstellar molecules containing different isotopes is that the interstellar isotope ratios appear to be the same as they are on the earth. The one apparent exception is in the region of the galactic center, where the ratio of carbon 12 to carbon 13 seems to be about half as large as it is elsewhere. This fact might suggest that proportionately more interstellar material has passed through carbon-burning stars in the galactic center, either because in that region there is less gas in relation to stars or because a larger fraction of the stars are of the massive carbon-burning type.

The fact that the bulk of the interstellar medium resembles the earth in isotope ratios would seem to indicate that interstellar chemistry has changed little in the five billion years since the earth was born. There are two possible explanations. Perhaps there are no regions (except for the galactic center) that are oversupplied with carbon-burning stars in relation to helium-burning stars. Alternatively, the interstellar gas must be well mixed over great distances and within a time that is comparable to the relatively short lifetime of the hottest stars.

Molecular Life Cycles

As we have seen, the discovery of the simple molecules CN, CH and CH^+ came as a surprise to astronomers. They had predicted that because it was highly improbable that two atoms would come together to form a molecule, and because any such molecules would be rapidly destroyed by the harsh interstellar ultraviolet radiation, the abundance of the molecules ought to be vanishingly small. Given the fact that the rate of destruction of the complex organic molecules is even higher and the fact that they are observed in much larger quantities than CN, CH and CH^+, it is obvious that new theories for their existence are needed.

In the past few years several mechanisms for the formation of molecules have been proposed. They break down into five major categories. First, the molecules might be built up by atoms colliding with and sticking to one another as they move around in the gaseous interstellar medium. Second, the molecules could form on the surface of dust particles out of atoms or other molecules that

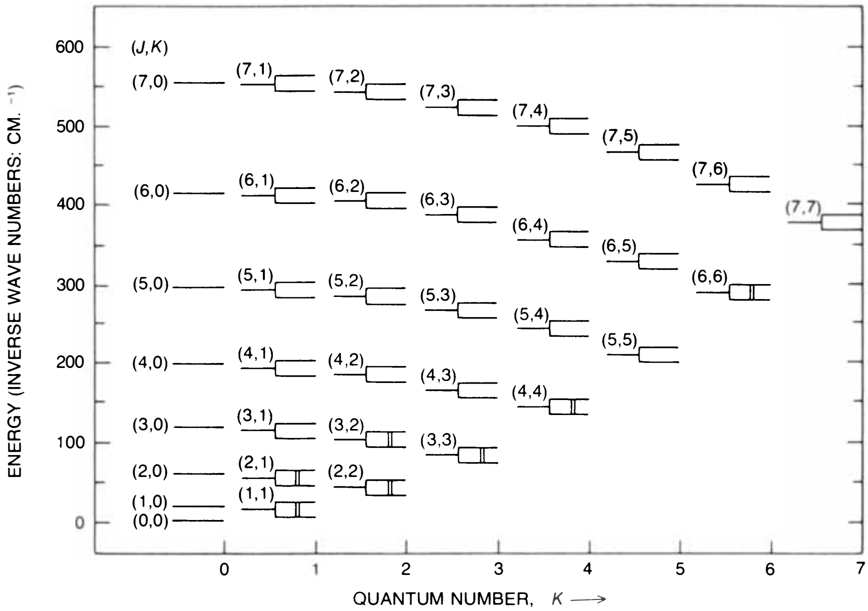


AMMONIA MOLECULE (NH_3) is a tetrahedron with a nitrogen atom at the apex above the plane formed by three hydrogen atoms. The nitrogen atom can oscillate from one side (*left*) to the other (*right*) through the plane. As a result each level of the rotational energy, *J*, is split into two closely spaced levels. Transitions between the two levels are inversion transitions and are observed at the wavelength of 1.3 centimeters.

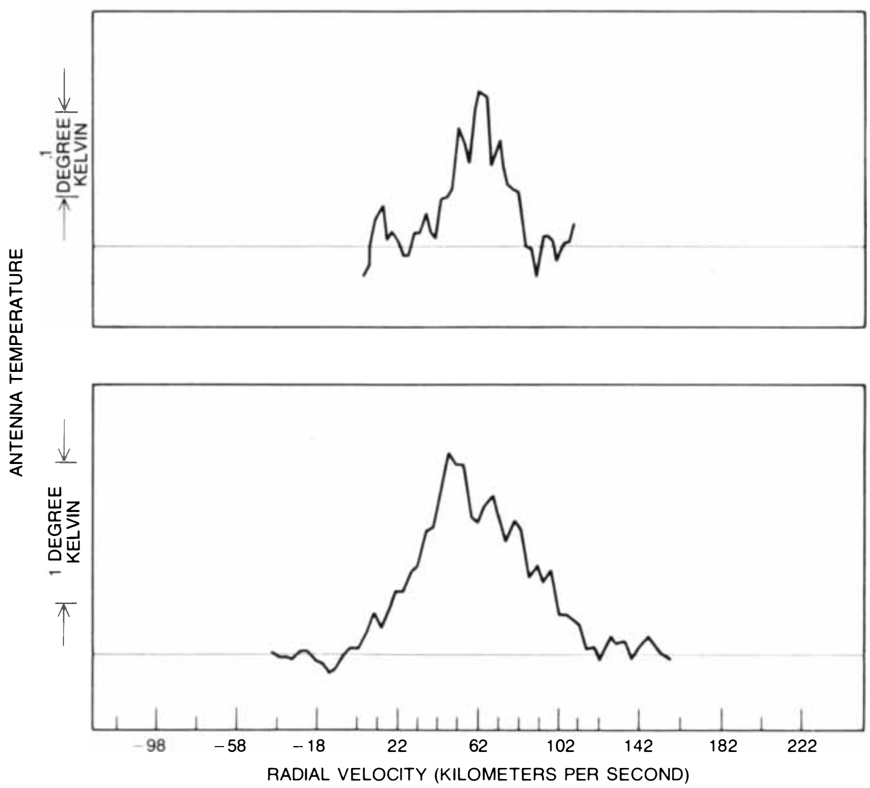
impinge on such surfaces. Third, the molecules might form from the evaporation or decomposition of dust particles when they are struck by energetic particles or photons or encounter shock waves or other heating phenomena. (This mechanism leaves open the question of how the dust particles themselves originated.) Fourth, the molecules might come into existence through collisions with atoms in the dense atmosphere of stars and might then be expelled into interstellar space. Fifth, the molecules might be formed in the dense environs of a "presolar nebula," that is, in the final phases of the collapse of a protostar into a self-luminous star.

The molecules must be produced as fast as they are removed or the lifetime of a molecular cloud would be limited, and it might be very short indeed. The two principal processes that destroy interstellar molecules are photodissociation by ultraviolet radiation and freezing out on the surface of dust particles. Recently L. J. Stief and his associates at the Goddard Space Flight Center in Greenbelt, Md., have studied the rates at which molecules are photodissociated, utilizing the results of elegant laboratory work and measurements of the intensity of interstellar ultraviolet radiation made during rocket flights. The important result is that molecules such as water, ammonia, formaldehyde and carbonyl sulfide dissociate in somewhat less than 100 years if they are unshielded by dust and are exposed to the average flux of interstellar ultraviolet radiation. This is a very short time by astronomical standards.

Dust clouds, however, are an extreme-



AMMONIA'S ENERGY LEVELS differ from those of carbon monosulfide because of the oscillation of the nitrogen atom. A second quantum number, *K*, refers to the component of angular momentum around the molecule's axis of symmetry. The two quantum numbers are written together: *J,K*. Ammonia can change its value of *J* by 1 by emitting or absorbing a photon; collisions may cause *J* to change by more than 1. *K* changes only by collisions.



SPECTRA OF AMMONIA in Sagittarius B2 yield information about the density inside the source. The number of ammonia molecules in the various states of rotational energy can be estimated by observing the inversion transitions. Some rotational levels, such as the 2,1 level (*top*), decay in a time as short as 20 seconds to a lower state, such as the 1,1 level (*bottom*), by spontaneously emitting a photon. To observe a spectral line from the 2,1 level as strong as some that have been detected requires a high rate of excitation by collisions to compete with the spontaneous rate of decay. Therefore the density must be very high. In Sagittarius B2 a density as high as 10^9 particles per cubic centimeter may be required.

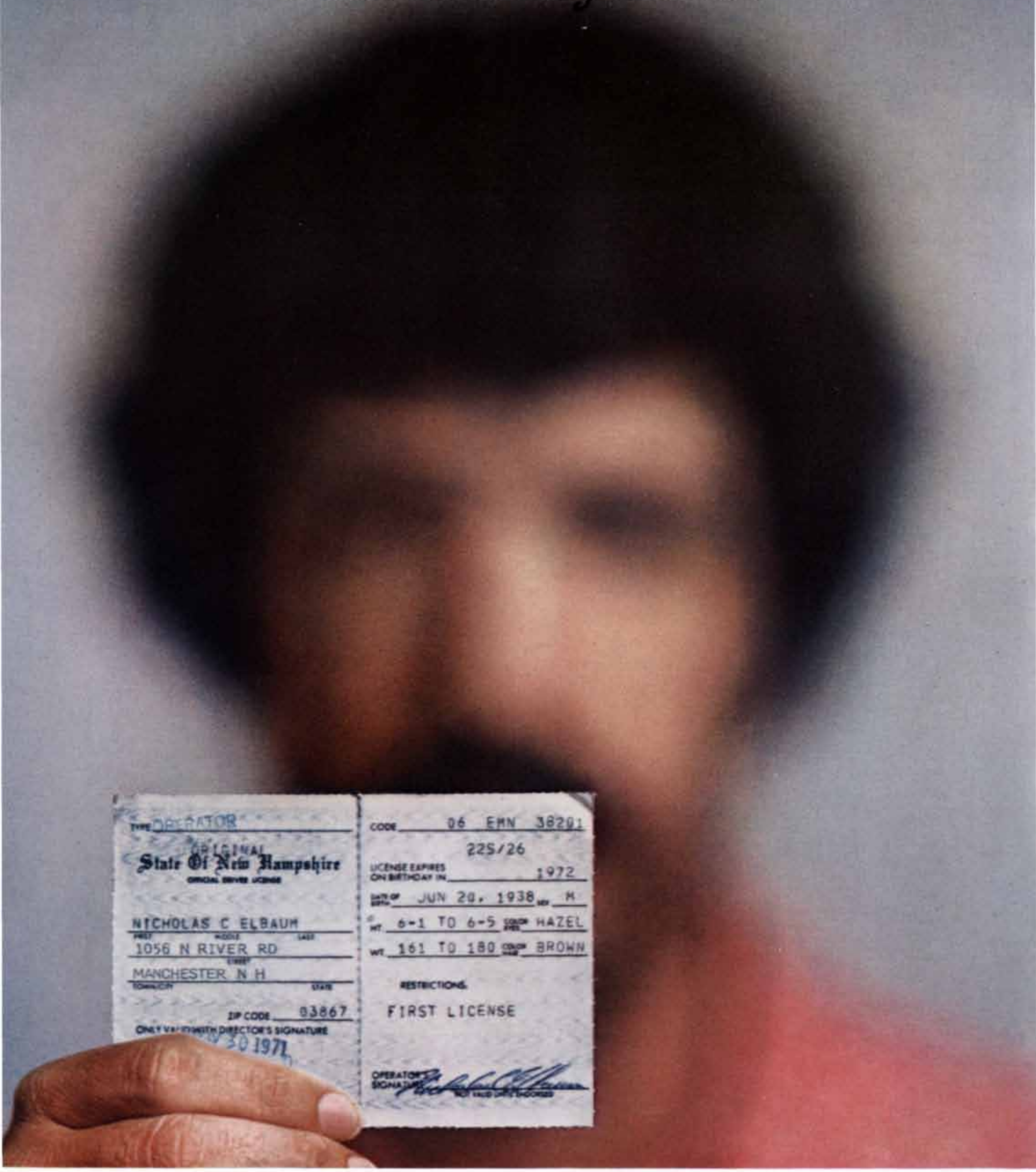
ly effective shield against ultraviolet radiation. A typical dust cloud with a diameter of one light-year and a density of 1,000 molecules of hydrogen per cubic centimeter will attenuate light by about four magnitudes and ultraviolet radiation by a factor perhaps as large as 10^{24}. As a result the lifetime of the molecules can be increased to as much as 10 million years. We must therefore conclude that interstellar molecules have been created within the dust clouds where they are now observed or that they have been transported into these clouds with some kind of protection, perhaps on the surface of dust particles.

Within the large dust clouds the density of the dust particles is such that a molecule will condense out of the gas onto the surface of a dust particle in about 100,000 years. The temperatures in these clouds are so low that it appears highly unlikely that the molecules would evaporate back into the gaseous state. A period of 100,000 years is considerably shorter than the lifetime of such clouds. Therefore, regardless of how the molecules originated, they must somehow be regenerated from the dust particles many times during the lifetime of the cloud. It is possible that this regeneration is accomplished by the few cosmic rays or photons of ultraviolet radiation that manage to penetrate the cloud or by invisible sources of infrared radiation such as protostars within the cloud.

The situation is much more critical in the dense molecular clouds associated with H II regions. Because of the much higher density of the cloud, molecules will stick to dust particles in as short a time as 100 years. The higher temperatures within these clouds, however, probably allow significant numbers of the molecules to evaporate back into the gaseous state. While the molecules are on the dust particle, surface reactions may well occur that give rise to new and more complicated molecules. Such effects might possibly explain why a wider variety of complex molecules is found near H II regions than is found in the relatively rarefied dust clouds.

It is only fair to state that there is no consensus among molecular astronomers in favor of any of the five mechanisms proposed for the formation of molecules. Perhaps no one of them fully applies. Nonetheless, there are a number of ways that these processes could act to form molecules. Some choices between the various possibilities now seem to be emerging from the observations. It does seem that the atmosphere of stars cannot be the direct source of the bulk of the interstellar molecules. Complex mole-

The identity crisis



Is the whole world "digit happy"? Think of your zip code. Area code. Driver's license. Charge cards. Bank cards.

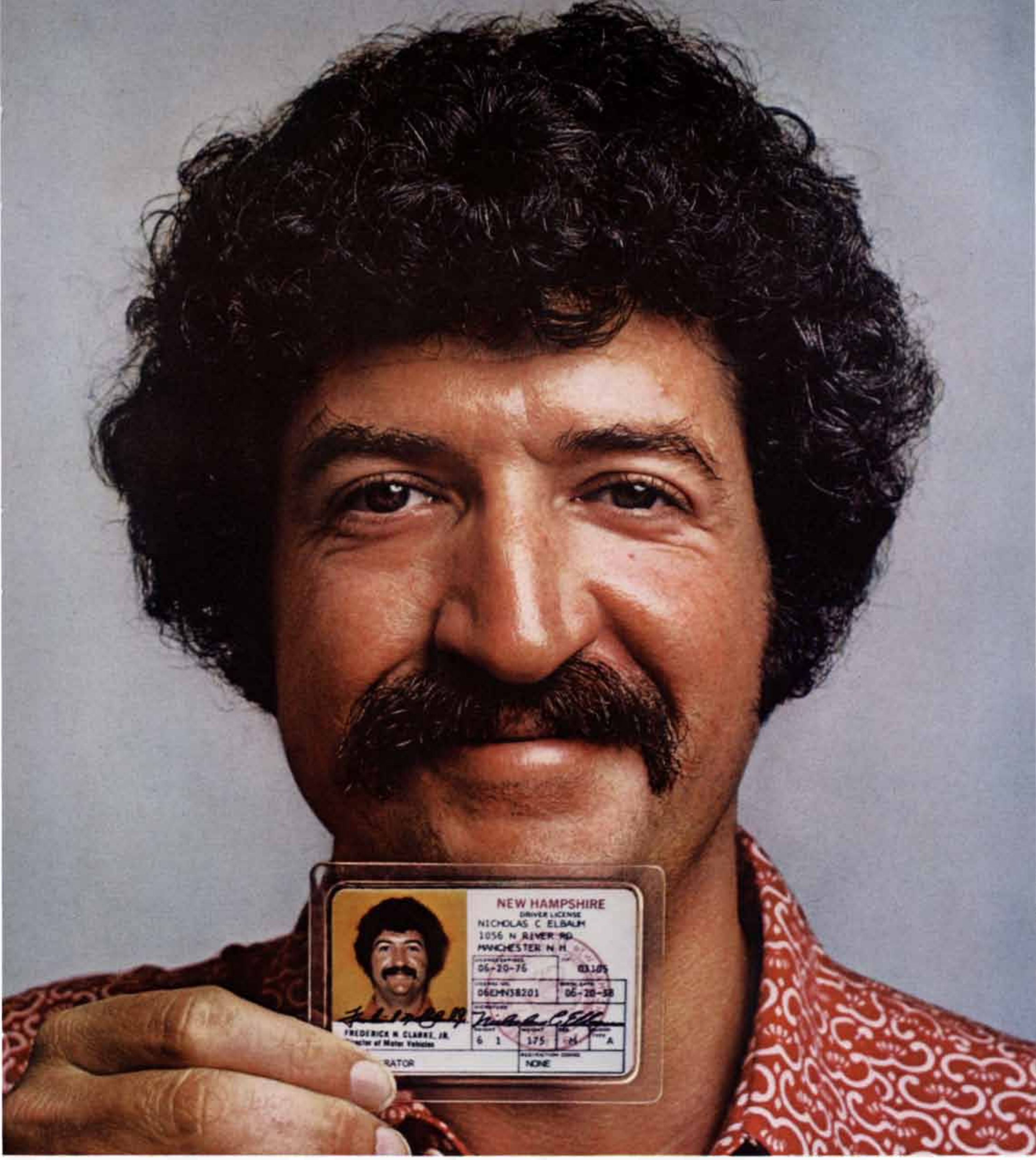
Does it ever make you feel faceless? Like just a set of numbers?

And when those cards get into the wrong hands, look what happens. Security problems. Thefts and frauds. Forged checks and phoney purchases. Even with all those numbers. Even with signatures.

But with Polaroid Portrait ID cards, you can be recognized for who you really are. On the spot. And you can have a feeling of security. Because they're virtually tamperproof. And no one else can use your face.

The result is that millions of people throughout the world carry them. And because the Polaroid ID-3 Identification System is portable and simple to operate, more and more organizations are using

and how to get a little recognition.



it. Whether businesses or banks, country clubs or colleges. For credit cards, security cards, driver's licenses or whatever. Like the license you see above. From New Hampshire, which is the latest state to use our system.

All of which means that more and more people will be showing their faces.

Why not see how the system works. For a demonstration, or just for more information, write to Polaroid Corporation, Dept. 57-244, Cambridge, Mass. 02139.

We'll show you how easy it is to help the individual get some recognition.

The Polaroid Portrait ID-3 System.

Polaroid®

Phillips helps you make plastics perform:

cules could not survive passage through the hot corona of the stellar atmosphere on their way to the interstellar medium, even if a suitable mechanism for their ejection could be found. Furthermore, the very types of stars whose atmospheres are cool enough to form molecules are known from theory (and in a few cases from observation) to have a ratio of carbon 12 to carbon 13 of between 4 : 1 and 10 : 1, as against the interstellar ratio of 89 : 1.

Some evidence, perhaps not overly conclusive, also suggests that molecules do not come solely from the breakup of interstellar dust particles. If they did, the dust particles themselves would not have been originally formed in interstellar space but would need to have been ejected from stars. Although such ejection is almost certainly possible in principle, again it is only those stars with a low ratio of carbon 12 to carbon 13 that can produce dust particles efficiently. Interstellar molecules that resulted from the breakup of these dust particles would share the low ratio of carbon 12 to carbon 13, which is contrary to the observations. In addition, molecules are not observed primarily in regions showing evidence of shock waves, energetic subatomic particles or other agencies that might decompose the dust particles by heating.

Therefore it seems that one of the first, second and fifth mechanisms, or some combination of them, is responsible for the interstellar molecules. Different processes may well occur in different types of molecular cloud. In principle, reactions on the surface of dust particles can proceed in any type of cloud, at a rate that depends on the density, temperature and chemical constitution of the particles. Still other primary mechanisms may well dominate the formation of molecules on dust particles. For example, George H. Herbig of the Lick Observatory has suggested that molecules might be created in presolar nebulas within the dense clouds surrounding H II regions. H II regions are well known to be regions of star formation, so that many presolar nebulas would be present. Such nebulas can expel both molecules and dust quite efficiently into the surrounding molecular clouds. The molecules and dust are expelled by the pressure of infrared radiation from the central protostar or by an excess of rotational momentum developed during the latter stages of the collapse of the presolar nebula. In protostars nuclear reactions have not yet begun; in such objects the isotope ratios would be the same as those observed in interstellar space.

In the less dense dust clouds, well away from H II regions, only the first and second mechanisms seem applicable to the formation of molecules. These mechanisms are not, however, without their problems. If surface reactions alone are considered, the question remains: How did the dust particles originate? Conceivably they could have been manufactured by stars elsewhere and have come to rest in clouds that slowed their passage because they were denser than the surrounding interstellar medium. Or the dust particles could have been built up over long periods of time by accretion, starting with pairs of atoms forming diatomic molecules. The theoretical details of this process are quite complicated. It appears that diatomic molecules can form at an adequate rate in this way, but subsequent collisions with the atoms in the surrounding gas would be likely to destroy the molecules already formed rather than creating new ones with more atoms. Only when the molecules already consist of at least a dozen atoms do collisions in the gas appear able to create larger molecules.

Whatever the mechanism responsible for molecules in the dust clouds may be, it seems certain that, in view of the low densities and temperatures in such clouds, the rate at which the molecules interact with the dust particles must be much lower than it is in the denser clouds near H II regions. Moreover, in dust clouds there are probably no efficient primary sources of molecules such as presolar nebulas. In these low-density clouds the more complex molecules (such as cyanoacetylene and methyl alcohol) appear to be less abundant with respect to the simpler molecules (such as the hydroxyl radical, carbon monosulfide and formaldehyde) than they are in the denser clouds near H II regions. If this suspicion is confirmed by additional observation, such a difference in chemical constitution may indicate that the mechanism of molecular production in one of the two types of source differs from the mechanism in the other.

Further Questions

The mechanisms I have discussed so far appear capable of forming interstellar molecules in the observed quantities in regions adequately shielded from destructive ultraviolet radiation. They do not, however, explain either the relative abundances of the observed molecules or the absence of certain others. What is known about the rate at which these molecules are dissociated by ultraviolet radiation indicates that there is no relation between the rates and the observed amounts. The inescapable conclusion is that very specific and very poorly understood formation mechanisms are at work. They almost certainly involve reactions on the surface of dust particles. Such reactions are likely to be important even in the clouds near H II regions where presolar nebulas may be the primary source of molecule formation; after molecules are ejected from the protostar into the surrounding cloud they must continue to interact rapidly with the dust particles. Unfortunately, little is known about surface reactions even when the surfaces are well defined. The composition of interstellar dust particles is unknown, and that composition may vary from region to region depending on local conditions. Over the years proposals for the composition of the dust particles have included graphite, graphite covered with frozen water and ammonia and methane, and a mixture of graphite, silicate and iron. It goes without saying that the particular molecules whose formation would be catalyzed on these surfaces would depend strongly on the composition of the dust particles.

If I have ended this discussion on a somewhat negative note, it is intentional. It is now clear that many scientific disciplines are needed to solve the problems of interstellar chemistry. It can be said that a new discipline has been born: astrochemistry. Our knowledge of interstellar molecules, full of gaps as it is, has already been advanced immeasurably by cooperative work in traditionally nonastronomical fields. Microwave spectroscopists have contributed information about the precise frequencies needed to search for new molecules. Photochemists have measured the probability that ultraviolet photons will destroy particular molecules. Experiments on the formation of simple organic molecules have been conducted for many years by workers interested in the problem of how life itself arises. With the discovery of at least four interstellar molecules that are considered to be necessary for the subsequent formation of biological molecules (water, formaldehyde, hydrogen cyanide and cyanoacetylene) these experiments obviously apply to the fascinating question of whether primitive life, or at least the amino acids that are the essential building blocks of it, can be detected in interstellar space.

In at least two areas further laboratory work is critically needed. One area involves identifying unknown spectral lines in interstellar space that are apparently molecular in origin. Two lines in the radio region of the spectrum are in this category; one is "X-ogen" and the other may possibly be a very unstable isomer of hydrogen cyanide that is unknown on the earth. A broad, diffuse spectral line of unknown origin has also been observed at the visible wavelength of 4,430 angstroms. These lines originate in many regions of interstellar space and may accordingly be quite basic to our understanding of cosmic chemistry. If they arise from molecules that are very unstable on the earth, as is often supposed, then an understanding of how these molecular species form and survive so well in the interstellar medium but not on the earth may bear on some fundamental aspects of molecular structure. The second area requiring much labora-

ATOM	TERRESTRIAL ISOTOPE RATIO	INTERSTELLAR ISOTOPE RATIO	INTERSTELLAR MOLECULE AND REGION
CARBON 12/CARBON 13	89 : 1	ABOUT 50 : 1 IF OXYGEN 16/OXYGEN 18 = 488 : 1	H_2CO, GALACTIC CENTER
		ABOUT 89 : 1 IF OXYGEN 16/OXYGEN 18 = 870 : 1	H_2CO, GALACTIC CENTER
		ABOUT (89 ± 15) : 1	H_2CO, OUTSIDE GALACTIC CENTER
		(82 ± 15) : 1	CH^+, ZETA OPHIUCHI
OXYGEN 16/OXYGEN 18	488 : 1	488 : 1 IF CARBON 12/CARBON 13 = 50 : 1	H_2CO, GALACTIC CENTER
		488 : 1 IF CARBON 12/CARBON 13 = 50 : 1	CO, GALACTIC CENTER
		870 : 1 IF CARBON 12/CARBON 13 = 89 : 1	H_2CO, GALACTIC CENTER
		(390 ± 100) : 1	OH, GALACTIC CENTER
OXYGEN 16/OXYGEN 17	2,700 : 1	AT LEAST 2,700 : 1	OH, GALACTIC CENTER
NITROGEN 14/NITROGEN 15	270 : 1	GREATER THAN 70 : 1	NH_3, GALACTIC CENTER
		(230 ± 70) : 1	HCN, ORION
SULFUR 32/SULFUR 34	22.5 : 1	(24 ± 5) : 1	CS, GALACTIC CENTER AND ELSEWHERE
SULFUR 32/SULFUR 33	125 : 1	GREATER THAN 100 : 1	CS, ORION

RATIOS OF ISOTOPES of various atoms in interstellar space compare well to those same ratios on the earth. When molecules other than carbon monoxide and formaldehyde are used, the comparison can be made relatively simply for oxygen, nitrogen and two isotopes of sulfur. In the galactic center, however, the very large abundances of carbon monoxide and formaldehyde confuse the interpretation. The spectral lines become "saturated," that is, the abundance is no longer proportional to the intensity of the spectral line. In this case it is possible only to obtain the ratio of carbon and oxygen together, that is, the ratio of carbon 13 and oxygen 18 to carbon 12 and oxygen 16. This dual ratio is found to be 1.8 times the terrestrial ratio. No known nuclear processes in stars affect the ratio of oxygen 16 to oxygen 18; moreover, this ratio seems to have a terrestrial value, as measured by the hydroxyl radical. It is therefore believed that the nonterrestrial value of the dual ratio of carbon and oxygen is due to a nonterrestrial value of 50 for the carbon ratio carbon 12 to carbon 13 in the region of the galactic center. Such a ratio could be produced by nuclear burning of carbon in giant stars, which may be more numerous in the galactic center. Elsewhere in the galaxy even the carbon ratio appears to be the same as that on the earth. Since all the interstellar isotope ratios seem to be near the terrestrial value, the chemistry of the interstellar medium must have been remarkably constant for a time at least equal to the age of the earth: some five billion years.

tory work is the chemistry of surfaces. Although the subject has been well investigated on an empirical basis in some kinds of catalytic processes that are important to industry, fundamental understanding of the elementary processes occurring at surfaces is largely lacking and thus predictions about them are generally not possible.

Whether astronomers are working at visible, infrared or radio wavelengths, they still must contribute vital information. The composition of the interstellar dust particles will most likely be determined from their properties of attenuation and polarization at visible and ultraviolet wavelengths and from the way they emit infrared radiation. Many molecules that are currently undetected and that must play central roles in interstellar chemistry can be observed only in the visible or infrared regions of the spectrum. Three of these molecules are molecular nitrogen (N_2), molecular carbon (C_2) and methane (CH_4).

Higher resolution is possible with infrared wavelengths than is possible with radio waves. Thus observations in the infrared will be vital in studying the fine structure of the interstellar sources where much of the activity may be going on. In the near future most new molecules will probably be detected at radio wavelengths, where the technology is more highly developed. The need for increasingly detailed study of the spatial distribution of the molecules within the sources, however, is at least as important as the discovery of new molecules. Recent observations have indicated that there is a wealth of unexpected detail in the structure of the molecular clouds. If such structure is not taken into account, serious errors may result in the estimation of the physical conditions and the relative abundances of different molecules within the clouds.

Notwithstanding these problems, which promise to maintain the field of astrochemistry in a highly excited state for some time, molecular spectroscopy has emerged as a major astronomical tool. The energy levels of molecules are typically much more closely spaced than the energy levels of individual atoms, so that molecules are ideally suited for the study of the cool regions between the stars. This fact, in addition to the ability of the longer wavelengths of the spectral lines of molecules to penetrate regions of dust and gas, has enabled us to explore the dust clouds and the early stages of star formation within them. Soon we should also be able to study the cool regions of galaxies outside our own.

THE ASYMMETRY OF THE HUMAN BRAIN

In most people the left cerebral hemisphere plays a dominant role in speech. The specialized functions of the right hemisphere are now emerging. One of these functions is the perception of melody

by Doreen Kimura

In most animals the structure of the nervous system is essentially symmetrical. In mammals the symmetry is made more striking by the prominence of the uppermost part of the brain: the cerebral hemispheres. In man, however, the two cerebral hemispheres differ greatly in their functions. It is well known that the left hemisphere plays a dominant role in speech [see "Language and the Brain," by Norman Geschwind; SCIENTIFIC AMERICAN, April, 1972]. The right hemisphere also has specialized functions, but until recently we have had less information about them because of the emphasis on studying language disorders. It turns out that the right hemisphere plays a dominant role in man's perception of his environment.

For more than a century the principal source of knowledge about the division of labor between the two cerebral hemispheres of man has been malfunctions of the brain caused by accident, surgery or disease. Although studies of intellectual impairments in patients with various kinds of brain lesion have provided much valuable information, such studies have the disadvantage that the damage may have affected not only the specific functional systems but also their interaction. In the past few years I have been involved in developing methods for studying the asymmetry of hemispheric functions in normal people.

I first became aware of the possibility that some aspects of brain function could be readily studied in normal people while I was doing research with patients at the Montreal Neurological Institute. One of the tests administered to the patients was a modification of a dichotic listening technique devised by Donald E. Broadbent of the British Medical Research Council's Applied Psychology Research Unit. His technique involves simultaneously presenting a spoken digit to one ear and a different spoken digit to the other ear. Three such pairs are usually delivered in sequence during one trial, and the subject is asked to report all the numbers he has heard. Patients with damage to the left temporal region of the brain reported fewer digits correctly than patients with damage to the right temporal region. A quite unexpected and intriguing finding was that most patients, no matter what part of the brain had been damaged, reported the words they had heard with their right ear more accurately than those they had heard with their left. The same turned out to be true for a group of normal people. There is evidence from numerous studies of pure-tone thresholds that the left and right ears do not differ in their basic capacity for detecting sounds, and so we concluded that the perceptual superiority of the right ear for words was somehow related to that ear's connections with the brain.

A peculiarity of the human nervous system is that each cerebral hemisphere receives information primarily from the opposite half of the body. Man's visual system is so arranged that vision to the right of a fixation point is mediated by the left half of the brain and vice versa. The auditory system is somewhat less crossed in that each half of the brain receives input from both ears, but the crossed connections are nevertheless stronger than the uncrossed ones. The tactual and motor systems of the brain are almost completely crossed: sensations from the left half of the body and movement of the left half of the body are served primarily by the right cerebral hemisphere and vice versa. The two hemispheres are themselves interconnected by nerve pathways. These pathways, the largest of which is the corpus callosum, play an important role in coordinating the activities of the hemispheres.

Since the auditory system is a predominantly crossed system, the neural input from the right ear to the left cerebral hemisphere should be stronger than that from the right ear to the right hemisphere. And since the left hemisphere usually contains the neural system for perception of speech, it is reasonable to suppose speech sounds presented to the right ear would have readier access to the speech-perception system. This supposition can be directly tested by observing people whose speech functions are not in the left hemisphere but in the right. Which side serves speech can be determined by the sodium amytal test, devised by Juhn A. Wada. The test involves injecting sodium amytal (a sedative) into the carotid artery of one side of the neck or the other. The drug disturbs the functioning of the cerebral hemisphere on that side for a few minutes, and if the subject's speech is disturbed as well one infers that speech is represented in that hemisphere.

Thirteen patients whom I tested at the Montreal Neurological Institute were found to have speech represented in the right hemisphere rather than in the left. The scores of these patients on the dichotic listening task were higher for the left ear. The results supported the hypothesis that the superiority of the right ear in normal subjects is due to better connections between that ear and the left (speech) hemisphere than between that ear and the right hemisphere.

We found further evidence that the superior performance of one ear on dichotic listening tasks did in fact reflect a hemispheric specialization of function. Brenda Milner of the Montreal Neurological Institute found that whereas damage to the left temporal lobe of the brain impaired comprehension of spoken material, damage to the right temporal lobe impaired the perception of certain other kinds of auditory material, particularly the discrimination of tonal quality and tonal pattern. I developed a dichotic listening task in which a headset was used to simultaneously play one melodic pattern to one ear and a different melodic pattern to the other ear. The subject was then asked to pick out the two melodies he (or she) had just heard from four melodies, each of which was played one at a time to both ears. Since melodies are processed predominantly by the right temporal lobe, normal subjects were able to pick out the melody presented to the left ear better than the one presented to the right ear.

The results were particularly exciting because they opened the way for exploration of the characteristics of verbal and nonverbal processes in the brain with relatively simple techniques. Although it has been known for more than a century that the left hemisphere is involved in speech functions, we still do not have a very clear idea of what the characteristics of those functions are. The traditional way of distinguishing them is to use a term such as "symbolic," which implies that the defining characteristics have to do with the capacity to let an event stand for something else. When we applied the dichotic listening technique, we got a rather different answer. The right ear was found to be superior for nonsense syllables and nonsensical sounds (such as recorded speech played backward or a foreign language unknown to the subject). Donald Shankweiler and Michael Studdert-Kennedy of the Haskins Laboratories in New Haven, Conn., have also applied the dichotic method to the problems of defining the characteristics of speech. They found that there was no right-ear superiority for the perception of isolated vowels but that there was such an effect for consonant-vowel syllables. It is difficult to reconcile all these findings with the notion that the left-hemisphere speech system primarily processes symbolic material. Why should vowels, which can have symbolic value, be processed equally well by both hemispheres whereas nonsense sounds such as speech played backward, which do not have a symbolic value, are apparently processed primarily by the left hemisphere?

One is forced to conclude that in auditory perception the left hemisphere is specialized for the perception of certain kinds of sound generated by the human vocal cords and vocal tract. By cutting a tape recording of natural speech into small segments Laurain King and I found that the briefest duration that yielded a right-ear superiority was about 200 milliseconds, or about the duration of an average spoken syllable: a consonant and a vowel. That size of unit seems to be necessary, although not always sufficient, for asymmetrical processing, and it supports the notion that the syllable is a basic unit in speech.

We further studied the dichotic perception of different vocal nonspeech sounds, such as coughing, laughing and crying. Instead of finding a right-ear superiority with these sounds, we ob-

DICHOTIC LISTENING TASK consists in simultaneously playing one melody to one ear and a different melody to the other ear. The subject is then asked to select these melodies from a series of four melodies that are played subsequently one at a time to both ears. The melodies are played through earphones connected to a dual-channel tape recorder. Other dichotic listening tasks involved listening to pairs of digits, words, nonsense syllables, vowels, backward speech and speech sounds such as laughing and crying.

tained a left-ear superiority. This result suggests that these sounds are processed primarily by the right hemisphere, just as melodic patterns are. Melodies hummed by another person were also identified better when heard by the left ear. Therefore if the left hemisphere distinguishes sounds by their articulatory features, these features must be rather specific.

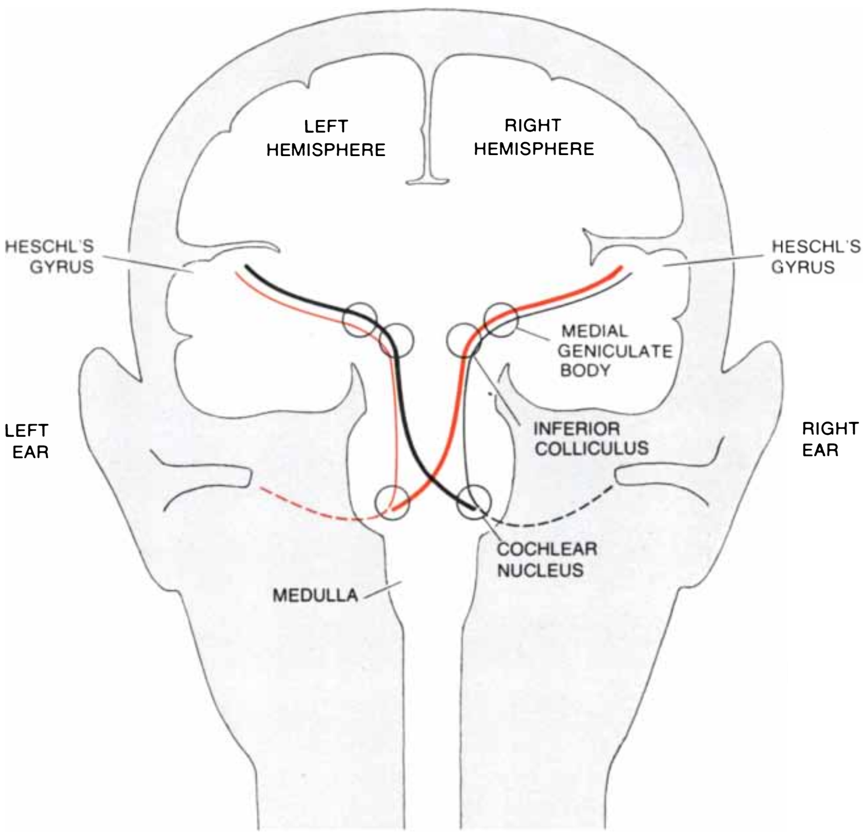
With some justification speech is generally regarded as being primarily an auditory-vocal system. Insofar as comprehension of a written word is a consequence of prior experience with the spoken equivalent, one might expect the speech system to take part in the processing of printed and written material. In actuality when reading abilities are thoroughly tested, disturbances of speech are nearly always accompanied by at least a mild disorder in reading. We set out to find whether or not there were visual perceptual asymmetries analogous to the right- and left-ear effects we had found in the auditory modality. Some earlier studies had in fact shown that words and letters are more accurately reported from the right half of the field of vision than from the left half, but the effects were explained as being caused by reading habits. In order to relate such effects specifically to the asymmetry of cerebral function, we needed to find tasks that tapped right-hemisphere visual functions as well.

Although the visual system is crossed, its connections are different from those of the auditory system. The connections are not from each eye to the opposite half of the brain but from each half of the visual field to the visual cortex on the opposite side [*see illustration on opposite page*]. Vision to the left of the point of fixation is received by the right half of each retina and the neural pathways from the right side of both retinas go to the visual cortex of the right hemisphere. Obviously the fibers from the right half of the retina of the left eye must cross the midline of the brain to get to the right hemisphere but the fibers from the right half of the retina of the right eye do not cross.

An important difference between vision and audition is that when the head is motionless, the ears are stationary but the eyes are not. Since the eyes are constantly moving, under normal viewing conditions one cannot present an image in only one visual field. In order to overcome this difficulty stimuli must be presented very rapidly during some period when the point of fixation is known. This can be done with the instrument known as the tachistoscope [*see illustration on page 74*]. The subject looks into the instrument and is asked to fixate on a designated point. While he is fixating, a visual stimulus is presented very rapidly either to the left of the point or to the right. Before the subject can make a new fixation to look at the stimulus, the stimulus has disappeared. As a result only one side of each of the retinas and only one cerebral hemisphere are directly stimulated.

In most of our tachistoscope experiments we did not use competition between two stimuli, as we had in the dichotic listening tests. In fact, our findings have been less ambiguous when only one visual field is stimulated. In normal people, then, words and letters are reported more accurately from the right visual field than from the left, a finding compatible with the right-ear superiority for the recognition of spoken sounds. That is, the recognition of visual verbal material is also more accurate when such material initially stimulates the left hemisphere.

Much of our work with visual perception has dealt with uncovering some of the specialized functions of the right hemisphere. It has been known for some time that injury to the right posterior part of the brain (the parieto-occipital region) results in the impairment of complex abilities such as drawing, finding one's way from place to place and building models from a plan or picture. In our studies with normal subjects we found evidence that the right hemisphere is also primary for some very fundamental visual processes. We find, for example, that in the simplest kind of spatial task—the location of a single point in a two-dimensional area—the right hemisphere is dominant. We tested for this ability by presenting dots one at a time either in the left visual field or in the right for a hundredth of a second. The dot was presented at various locations within a circle drawn on a plain white card. The subject then identified the location of the dot on a similar card outside the tachistoscope. The scores for correctly locating dots were higher for dots presented in the left field than for dots presented in the right field. Moreover, ascertaining the number of dots and of geometric forms was more accurate for the left field.



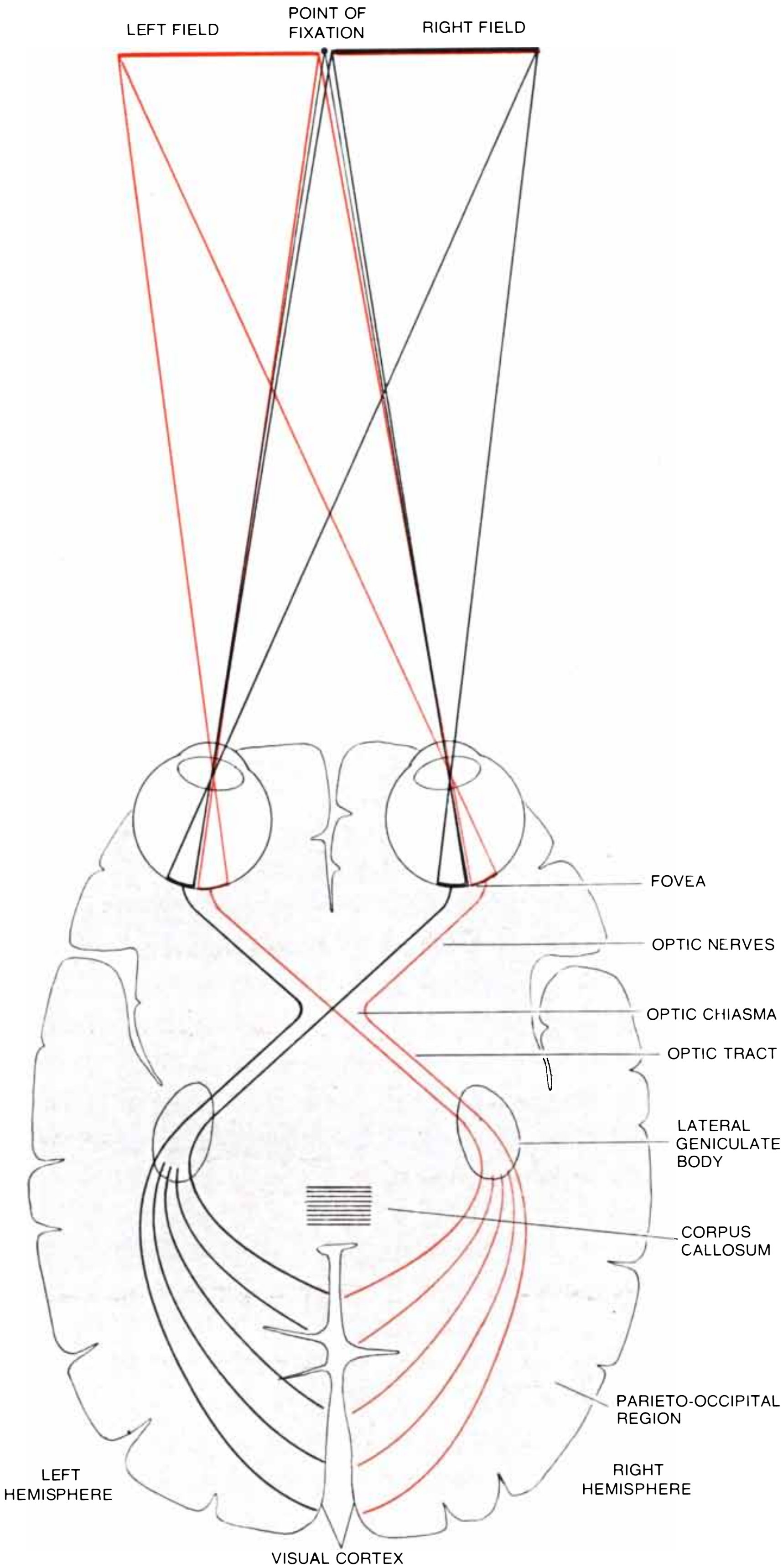
AUDITORY PATHWAYS from the ears to the cerebral auditory receiving areas in the right and left hemispheres are partially crossed. Although each hemisphere can receive input from both ears, the neural connections from one ear to the hemisphere on the opposite side are stronger than the connections to the hemisphere on the same side. When ipsilateral (same side) and contralateral (opposite side) inputs compete in the auditory neural system, it is thought that the stronger contralateral input inhibits or occludes the ipsilateral signals.

This ability is not due merely to some kind of heightened attention to stimuli on the left, since the simple detection of dots is no more accurate in one field than in the other. Detection was tested either by having the subject report whether or not a dot was present on any one trial with a fixed exposure time, or by determining the exposure time required for detection of a dot. In neither case is there a difference between the detection accuracy of the left field and of the right. It appears instead that the right hemisphere incorporates important components of a system of spatial coordinates that facilitates the location of a point in space. Of course, stimuli arriving at both visual cortexes must ordinarily have access to this system, but when input is deliberately limited to one visual field, it is possible to determine if one hemisphere has a functional advantage.

We then asked if the right hemisphere might also be important for depth perception. Locating objects in three dimensions can be mediated by one or more of several cues. Most of these cues are monocular, that is, they can be distinguished by one eye. The cues include the relative sizes of retinal images, the obscuring of one object by another or the relative speed with which two objects move across the field of vision. Another important cue to depth, binocular disparity, requires both eyes. Binocular disparity refers to the fact that because the two eyes are separated each eye receives a slightly different retinal image. The disparity between the two images can provide information about the depth of an object because nearer objects have a larger binocular disparity than farther ones.

Margaret Durnford and I initiated some studies in depth perception by attaching a classical depth-perception box to the back of the tachistoscope. The box contains a fixed vertical central rod in line with the fixation point. On each side of the central rod is a track on which another vertical rod can be moved. The movable rod is seen with both eyes for only a fraction of a second, and the subject was asked whether it was nearer than the central rod or farther. When the variable rod was in the left visual field, that is, when the information went to the right hemisphere, the reports were more accurate. Thus spatial information in the third dimension also is processed more accurately by the right hemisphere than by the left.

When the movable rod was viewed with only one eye, there was no difference in accuracy between the left field



VISUAL PATHWAYS are completely crossed, so that when the eyes are fixated on a point, all of the field to the left of the fixation point excites the visual cortex in the right hemisphere and stimuli from the right visual field excite the left visual cortex. The visual cortexes can communicate via the corpus callosum, which connects the two hemispheres.

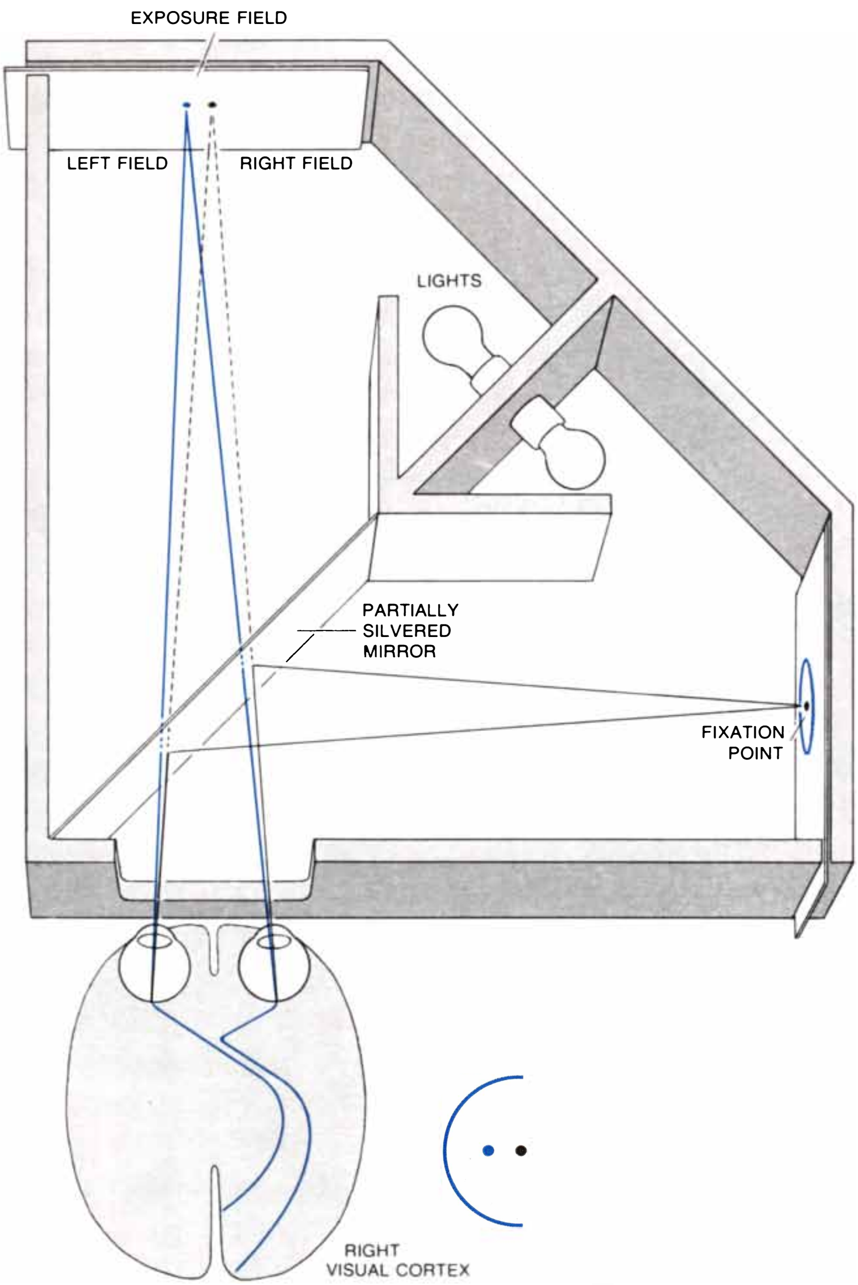
and the right. This result suggested that binocular information was processed primarily by the right hemisphere and that monocular information was processed by each hemisphere. As a further test of the specificity of hemispheric asymmetry we presented visual stimuli in which the only cue to depth was binocular disparity. When two slightly different views of a two-dimensional picture are seen separately by the two eyes by means of a stereoscope, the natural binocular disparity can be simulated. The subject reports seeing the two as being fused and as possessing the characteristics of depth that neither image alone yields.

Random-dot stereograms developed by Bela Julesz of the Bell Telephone Laboratories are ideal for this purpose, since each monocular image appears as a random array of dots. When the two images are viewed stereoscopically, a form such as a square or a triangle appears either in front of the rest of the array or behind it. A series of such stereograms with different forms was tachistoscopically presented to either the left or the right visual field, and the subject was asked to identify the form that stood out. For example, on a left-field trial one random-dot image would be presented to the left field of the left eye and its stereogram partner to the left field of the right eye. The two patterns thus presented are initially processed in the right cerebral hemisphere. A presentation to the right field, on the other hand, leads to stereoscopic processing in the left hemisphere.

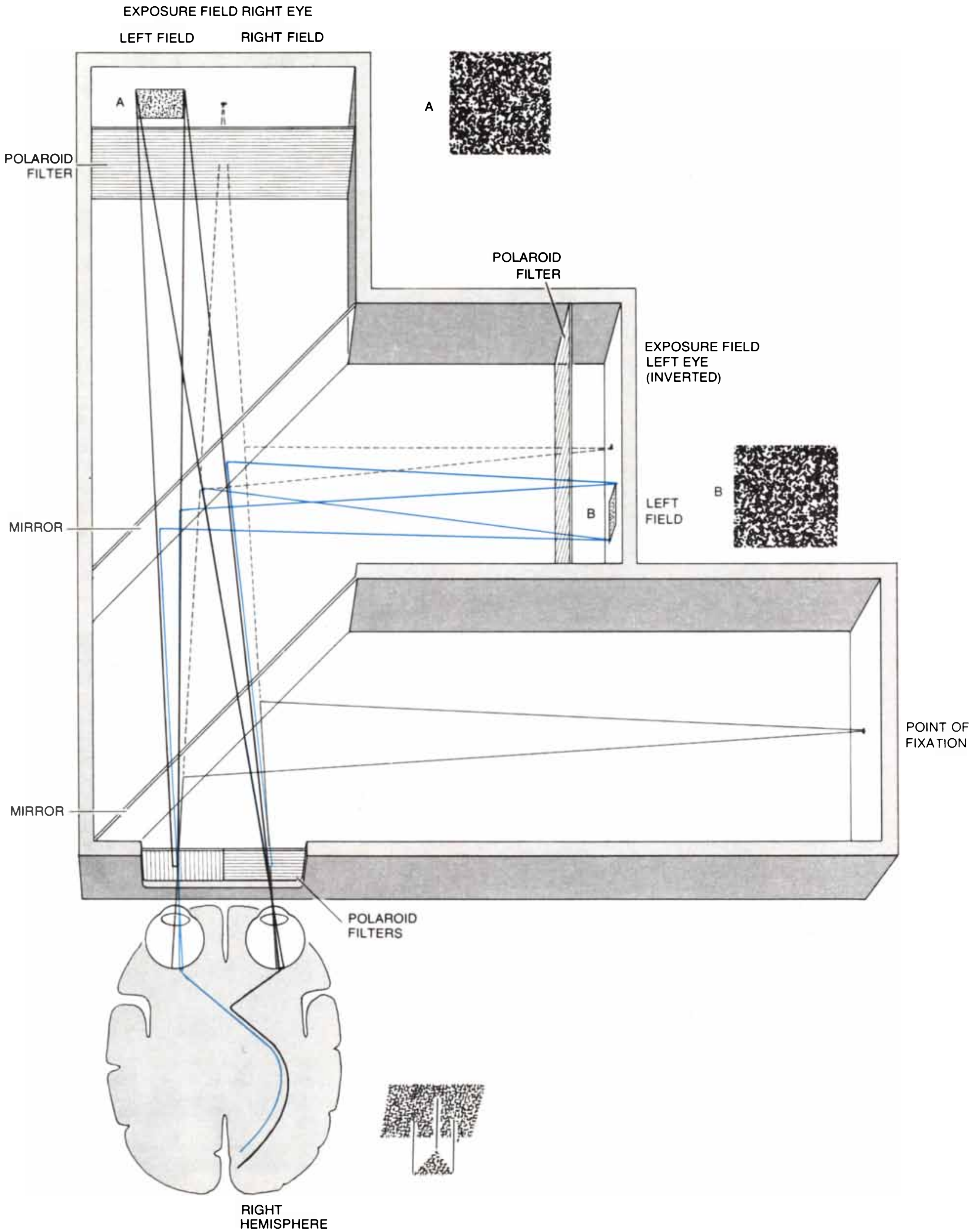
We had anticipated some difficulty in getting binocular fusion because of the brief exposure (100 milliseconds) we had to use, but somewhat to our surprise most people we tested could do the task. The identification of stereoscopic stimuli was clearly better when stimuli were presented to the left visual field, that is, when the disparate images went to the right hemisphere. It may be, then, that the right-hemisphere processing of depth information is rather specifically connected to the utilization of such binocular cues.

Clearly the right hemisphere works better than the left hemisphere in analyzing information about where objects are located in space. The probability that we were tapping functions fundamental to orientation in space encouraged us to study hemispheric specialization for still another basic visual process: the perception of the slant of a line. Very short lines were presented one at a time in either the left visual field or the right. The lines varied in slope from 15 degrees to 165 degrees in 15-degree steps. After the subject had seen the line he was asked to pick it out of a multiple-choice array of slanted lines on a sheet of paper. There was a small but consistent superiority for slope identification in the left visual field, again suggesting that this information is asymmetrically processed in the brain. David H. Hubel and Torsten N. Wiesel of the Harvard Medical School have suggested that the initial processing of the orientation of a line is done in the visual cortex. The fact that we found a hemispheric asymmetry for such perceptual tasks opens the possibility that there may be some functional asymmetry between the hemispheres in primary processes as well as in the more complex associative processes.

We do not know precisely what neural systems in the brain we are sampling with our perceptual techniques. The evidence we have suggests that with the



TWO-FIELD TACHISTOSCOPE is used for study of visual perception. When the fixation field is lighted, an observer sees a reflection of the field in the partially silvered mirror. He is asked to fixate on a point in the center of the field. Then the fixation-field light is turned off and the exposure-field light is simultaneously turned on for a few milliseconds. The image on the exposure field passes through the partially silvered mirror and is briefly seen by the observer. At the end of the exposure the fixation-field light comes on and the exposure-field light goes off. By placing the exposure image in the left or right visual field as desired the experimenter can selectively stimulate either the right or the left visual cortex.



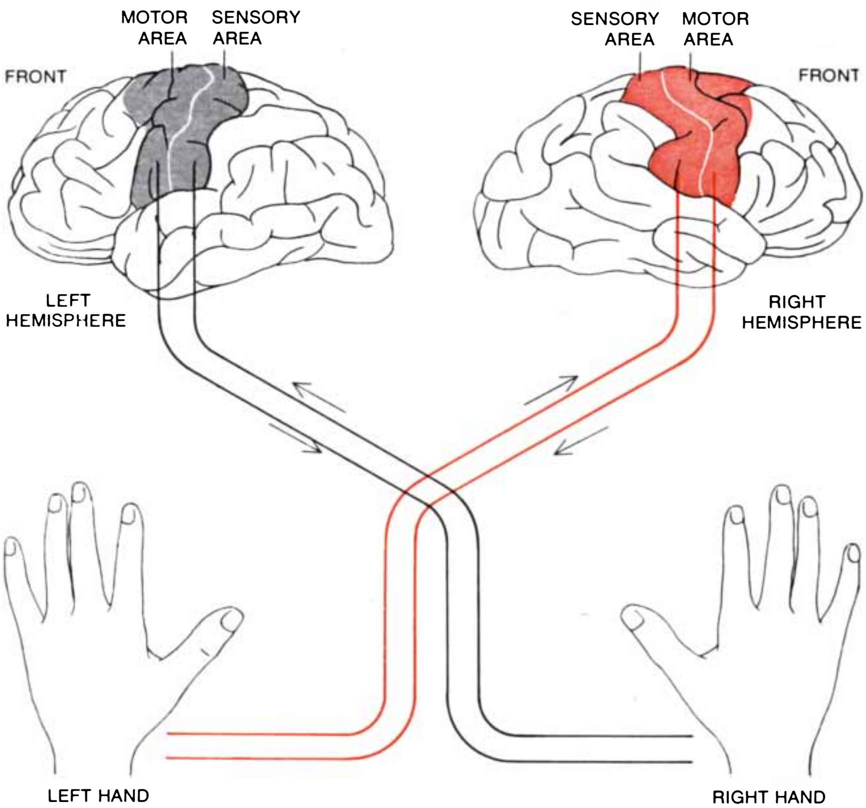
THREE-FIELD TACHISTOSCOPE is used in studies of depth perception. The subject initially fixates on a point, which he sees as a reflection from a partially silvered mirror. The two exposure fields become visible when the light on the fixation field is turned off and the lights near the exposure fields are flashed on. The Polaroid filters are arranged so that each eye receives a different image. Using a technique developed by Bela Julesz of the Bell Telephone Laboratories, the author presented one random-dot pattern to the right hemiretina of the right eye and a slightly different random-dot pattern to the right hemiretina of the left eye. In other trials the pattern was presented to the left hemiretinas. When each pattern is viewed alone, no shapes or depth can be seen, but when the two patterns are fused somewhere in the visual system, the subject in this case sees a triangle floating in front of the background dot pattern. Most people are better at identifying the stereoscopic figure when the images are flashed onto the left visual field than when they are flashed onto the right visual field, indicating that the right hemisphere may be better at processing depth information.

tachistoscopic tests we are tapping the function of regions near the striate cortex, the major visual pathway to the hemispheres, rather than more remote regions such as the temporal lobes, which also have visual functions. With the auditory dichotic tests, on the other hand, we are probably tapping the functions of the temporal lobe area.

A related point of interest is that we have not found any left-visual-field superiority for the perception of form, although we have tried in several tests. We know that damage to the right temporal lobe impairs the perception of nonsense designs, suggesting that some portions of the right hemisphere may indeed be critical for form perception. The tachistoscopic methods we used, however, do not sample the function of the systems for perception of form. Apparently the neural systems involved in spatial processing are relatively independent of those involved in form perception, a suggestion that was made many years ago by the British neurologist Gordon Holmes and that is supported by recent studies of visual perception in hamsters conducted by Gerald Schneider of the Massachusetts Institute of Technology.

Although we have concentrated on visual and auditory perception, it appears that there is an analogous asymmetry in tactual perception. Beata Hermelin and Neil O'Connor of the Medical Research Council's Developmental Psychology Unit have reported that the tactual perception of Braille dot patterns by blind people is more rapid with the left hand than with the right. Diana Ingram in our laboratory has found that when one arm is used to locate a point out of sight under a table on which the location of the point is indicated, the left arm performs more accurately than the right.

From these observations and others one can conclude that the posterior part of the right hemisphere is involved in the direct analysis of information about the external environment. The parieto-occipital area is particularly critical for the kinds of behavior that are dependent on spatial relations, whereas the temporal region takes part in processing nonspatial stimuli such as melodic patterns and nonsense designs. An important secondary process in the analysis of perceptual input is the attaching of a verbal label. We know that verbal transformation of information involves the left hemisphere, but we still have much to learn about the transfer from right-to-left-hemisphere processing and from left-to-right-hemisphere processing. For example, we can demonstrate left-visual-field superiority in spatial perception in spite of the fact that the subject ultimately gives us a verbal response that is controlled from his left hemisphere. We explain this result by saying that the primary analysis is accomplished by the right hemisphere, and the verbal response is secondary. In other situations, however, the mode of response (manual or vocal) can influence which field dominates. Hence we do not yet have a completely satisfactory explanation.

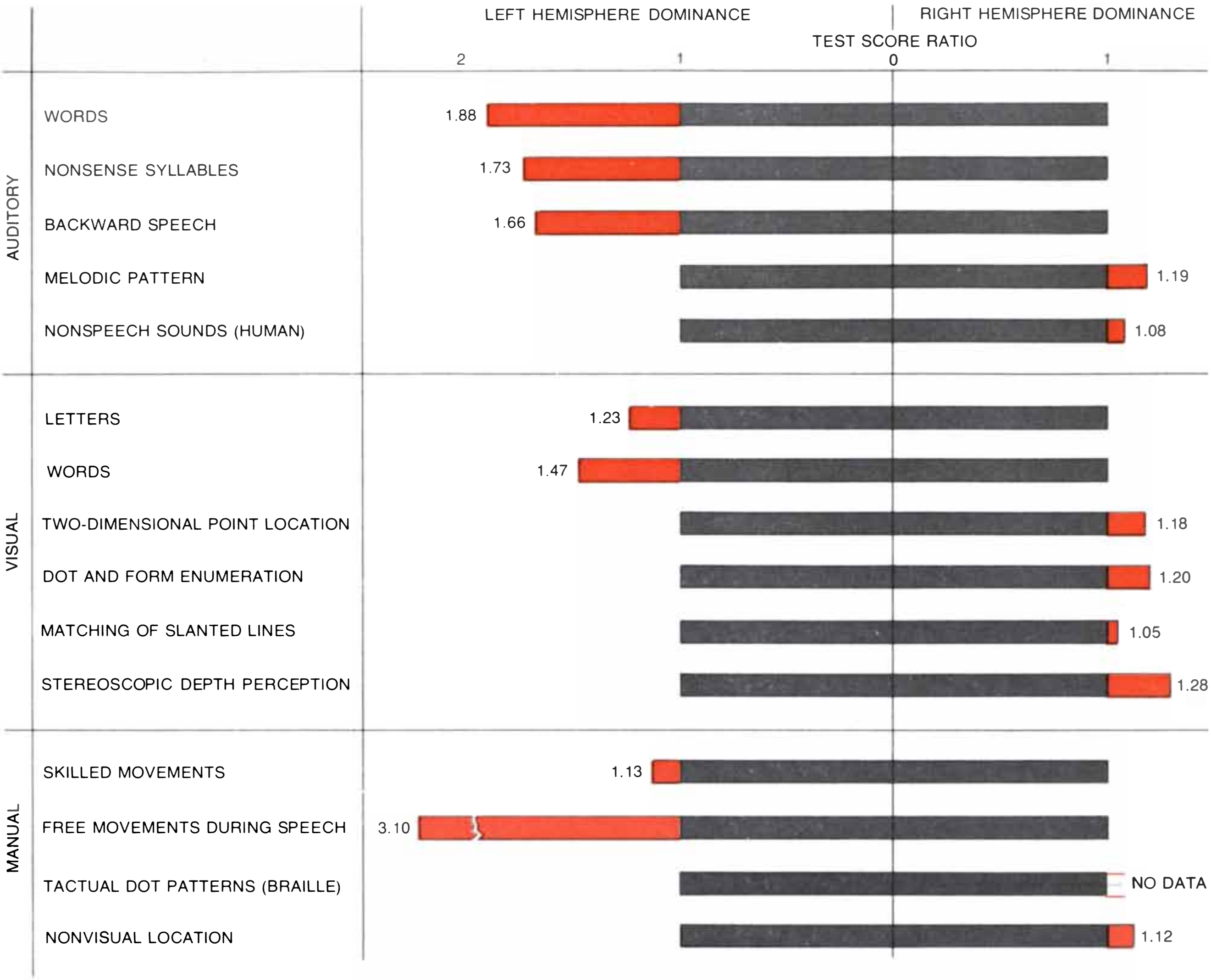


MOTOR AND SENSORY PATHWAYS from the hands are almost completely crossed, so that each hand is served primarily by the cerebral hemisphere on the opposite side.

We have seen that the left hemisphere is critical for the production and perception of certain sounds made by the human speech system. Recently my colleagues and I have obtained evidence from patients with cerebral strokes that the left hemisphere may also be essential for some types of movements of the hand. We found that patients with left-sided cerebral damage have difficulty copying a series of hand movements, whether the movements are meaningful or not. Moreover, there are reports in the clinical literature of deaf mutes who used hand movements as a means of communication and who, after suffering damage to the left hemisphere, displayed disturbances of these movements that were analogous to disturbances of speech. That the left hemisphere has a special control over some aspects of manual behavior is further suggested by the fact that most people use their right hand for many skilled acts. Although the relation between speech lateralization and hand preference is not perfect, the high incidence of both left-hemisphere control of speech and right-hand preference is probably not coincidental.

We have found further support for a relation between speech and certain manual activities by observing the hand movements of normal people while they are speaking. As everyone knows, speech is often accompanied by gestures, in which the hands are moved around freely in space without touching anything. Such movements are hardly ever seen during a nonspeech vocal activity such as humming. In both humming and speaking, however, there may be other kinds of manual activity, such as touching the body, rubbing the nose or scratching. Equally interesting is the fact that the free movements during speech are made primarily by the hand opposite the hemisphere that controls speech (as determined by means of the dichotic verbal method). If speech is controlled by the left hemisphere, as it is in most people, the right hand makes more of the free movements, whereas if speech is controlled by the right hemisphere, the left

GESTURES DURING SPEECH are made primarily by the hand that is opposite the cerebral hemisphere controlling speech. In most people speech is controlled by the left hemisphere, and the right hand of such people makes more free movements than the left hand during speech, as is shown in these pictures from a video tape recording made by the author during an experiment. The asymmetrical use of the hands during speech has been found only when the hands move freely in space without touching the body.

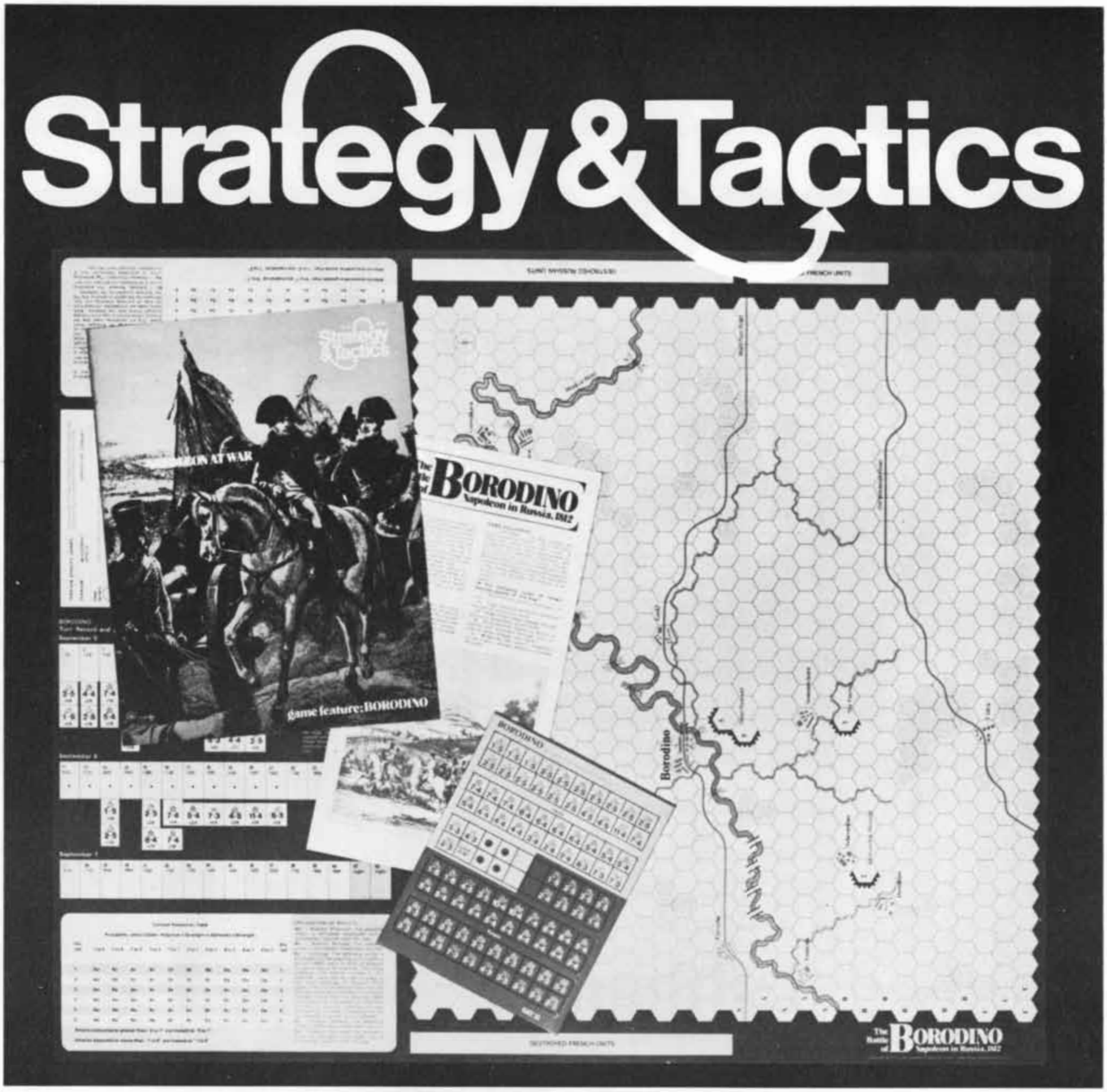


FUNCTIONAL ASYMMETRIES of the cerebral hemispheres in normal, right-handed people are found in the auditory, visual and manual modalities. Test scores for the left and right sides were converted to ratios for comparison. The ratio for left-hemisphere dominance for perception of spoken words is 1.88 : 1, whereas the ratio for right-hemisphere dominance for melodies is 1.19 : 1. These ratios are not fixed values since they vary with the type of stimulus, the kind of response required and the difficulty of the task.

hand makes more of the free movements. Curiously, this asymmetry is restricted to free movements; it does not appear in self-touching movements. These findings and the clinical findings I have mentioned suggest that there is indeed some overlap between the speaking system in a hemisphere and the system controlling certain kinds of manual activity. It may be that the left hemisphere is particularly well adapted not for the symbolic function in itself but for the execution of some categories of motor activity that happen to lend themselves readily to communication.

In our work on cerebral asymmetry in normal people we have occasionally encountered sex differences. In right-hemisphere tasks males tend to have a greater left-visual-field superiority for dot location and dot enumeration than females. We also know that males are superior to females in certain visual-spatial tasks. It may be that right-hemisphere specialization is more pronounced in males than in females, and that such specialization may sometimes be advantageous. Recently Jeannette McGlone and Wilda Davidson in my laboratory at the University of Western Ontario found some evidence in favor of this idea. They administered a standard test of spatial perception in which the subject must identify a design after it has been rotated. They found, as is usual in this task, that males usually performed better than females. The females who did particularly poorly on the test were those who showed some left-hemispheric specialization for spatial functions (as inferred from the tachistoscopic dot-enumeration test). Usually such functions are controlled by the right hemisphere.

In contrast, females tend to have greater verbal fluency than males. There is no evidence, however, that adult females are more asymmetrical in speech lateralization than males. Dichotic studies nonetheless suggest that speech lateralization may develop earlier in girls than in boys. It appears that for some intellectual functions the brains of males and females may be differently organized. Most of human evolution must have taken place under conditions where for the male hunting members of society accurate information about both the immediate and the distant environment was of paramount importance. For the females, who presumably stayed closer to home with other nonhunting members of the group, similar selection processes may not have operated. It will be interesting to discover whether or not the sex differences in verbal and nonverbal asymmetries we have uncovered with our relatively simple techniques hold true for other cultures.



The paper time machine

STRATEGY & TACTICS is a magazine. It's also a tool: a time machine that enables you to replay the crucial events — past, present, and future — that shape our lives.

Now, instead of merely reading about what's happening, you can explore and experience the alternatives and decision points through the technique of Conflict Simulation.

What is Conflict Simulation?

Conflict Simulation is a way of analyzing a political or military conflict situation. A way that is as intellectually stimulating as a game of chess, and as thorough as a written analysis.

Through the use of the Conflict Simulation (or "game") format, the conflict situation is re-created — so that you are in a position to make the vital decisions and, in the game at least, change the way things were, are, or will be.

What you get

STRATEGY & TACTICS magazine is published bi-monthly. Each issue contains:

★ A ready-to-play conflict-simulation game with a 22 x 28" playing surface, die-cut playing pieces, and complete rules.

★ An analytical article on the same subject as the game in that issue.

★ Other feature articles on historical and military subjects.

★ Game and book reviews, commentary on existing games, and discussions of subscribers' questions.

The magazine is 48 + pages long, and all material is handled in a highly organized (and easily understandable) graphic format.

Games recently published in STRATEGY & TACTICS were: *GRUNT* (ground combat in Vietnam), *LOST BATTLES* (tactical combat in Russia, 1941-44), *USN* (the war in the Pacific, 1941-43), *COMBAT COMMAND* (tactical combat in Western Europe, 1944).

We also publish a separate line of conflict-simulation games, which you will find listed in the coupon.

Free to new subscribers

NAPOLEON AT WATERLOO, history's greatest battle presented in a game-design specially created to introduce new readers to Conflict Simulation.

Send check or M.O. to
Simulations Publications Inc.
Dept. 581 44 East 23rd St., New York, N.Y. 10010

Please enter my subscription to S. & T. for:
☐ **1 Year** (6 issues)—$10
☐ **2 Yrs.** (12 issues)—$17
☐ **3 Yrs.** (18 issues)—$24
☐ Current issue $4

Send me the following Simulation Games:
☐ **Kursk** (Russia, 1943)—$6
☐ **Korea** (1950-51)—$6
☐ **Phalanx** (ancient Greece)—$6
☐ **Barbarossa** (Russia, 1941-45)—$6
☐ **Leipzig** (Napoleonic Wars, 1813)—$6
☐ **Normandy** (the D-Day Invasion)-$6

☐ Please send me your free brochure.

Name ____________________
Address ____________________
City __________ State __________ Zip __________

We can give you a good address anywhere in the world.

Your business may take you to all kinds of places. In some of them you'll know your way around.
In others you won't. In any case we'd like to put our address at your disposal. To have your mail sent to or leave a message at.
If possible, to arrange a business introduction or some sight-seeing.
We maintain offices all over the world, even in places we don't fly to. Like Belem or Katmandu, for instance. And in all of them, we're ready to help you. Certainly, we'll have a friendly smile when you get there.
Even if you didn't fly with us.

Lufthansa
the more you fly

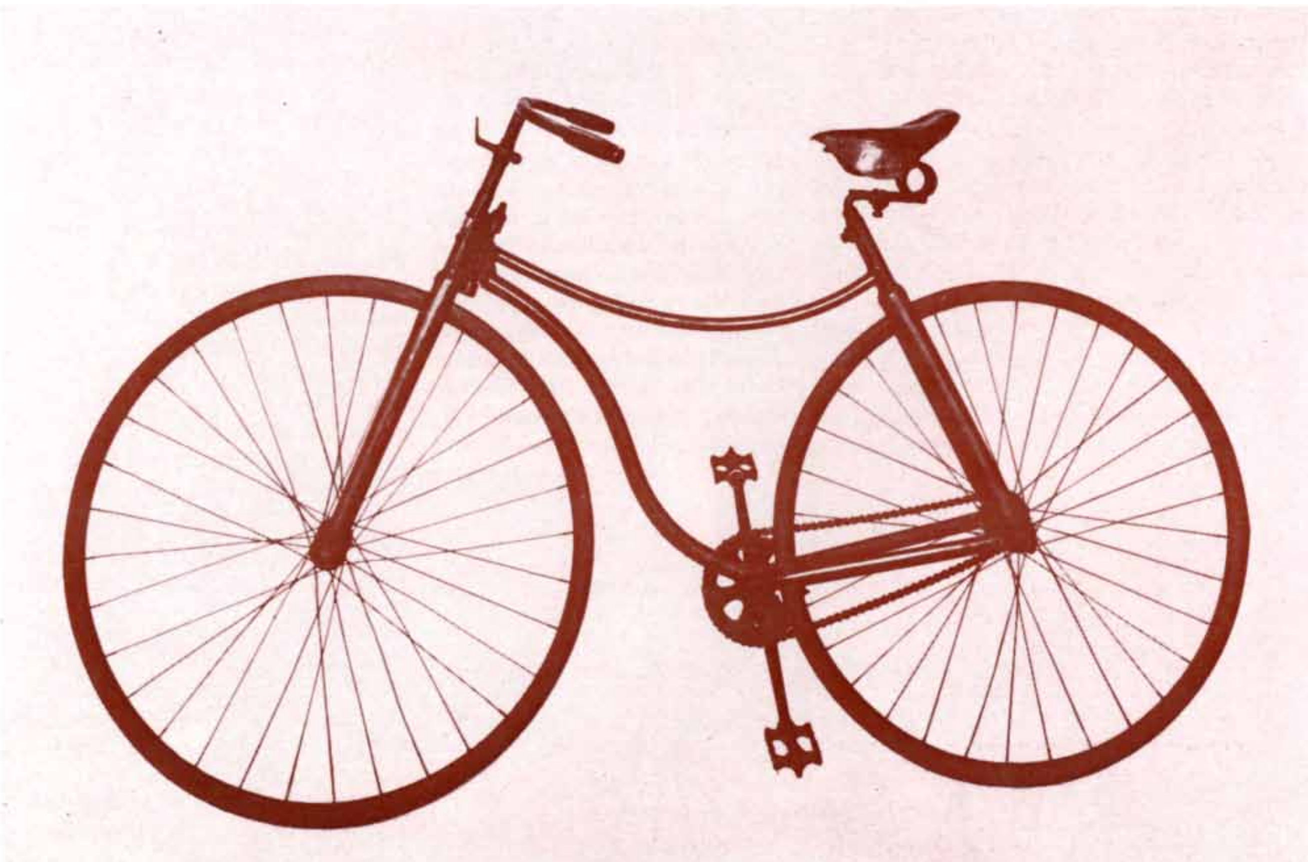
BICYCLE TECHNOLOGY

This humane and efficient machine played a central role in the evolution of the ball bearing, the pneumatic tire, tubular construction and the automobile and the airplane

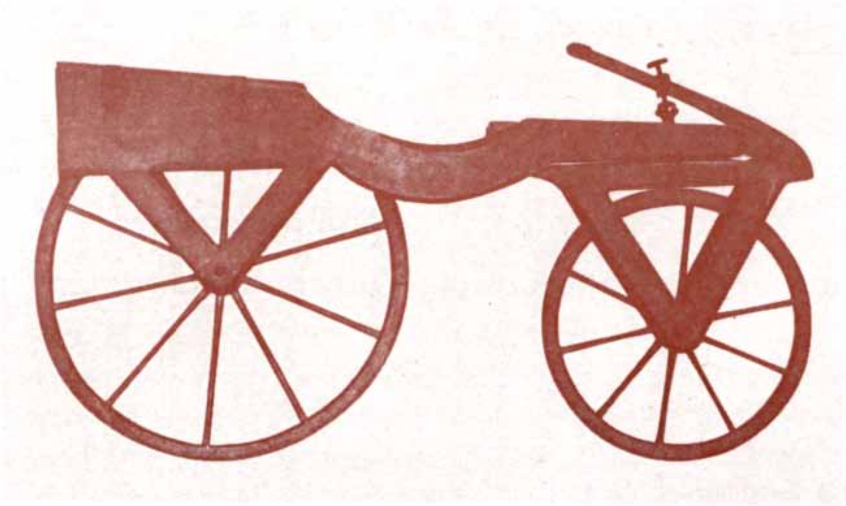
by S. S. Wilson

We tend to take the bicycle too much for granted, forgetting the important role it played in the evolution of modern technology. The first machine to be mass-produced for personal transportation, the bicycle figured prominently in the early development of the automobile. Thus in addition to its own considerable direct impact on society the bicycle was indirectly responsible for substantial social and economic changes. A remarkably efficient machine both structurally and mechanically, the bicycle continues to offer distinct advantages as a means of personal transportation in both developed and underdeveloped countries.

When one considers how long the wheel has served in transportation (more than 5,000 years), it seems odd that the first really effective self-propelled wheeled vehicle was developed only about 100 years ago. As with most epoch-making inventions, many men and many nations can claim a share in its development. The earliest legitimate claimant would be Baron von Drais de



ROVER SAFETY BICYCLE, introduced in 1885 by J. K. Starley of England, is widely regarded as marking the final development of the bicycle form. The Rover had most of the major features of the modern bicycle: rear-wheel chain-and-sprocket drive with a "geared up" transmission, ball bearings in the wheel hubs, tangentially mounted wire spokes, lightweight tubular-steel construction and a diamond-shaped frame. Unlike most modern bicycles, the Rover incorporated two curved tubes without the extra diagonal tube from the saddle to the bottom bracket; also the front forks, although sloping, were straight instead of curved. Unless otherwise noted, the old vehicles shown in the photographs used to illustrate this article are now in the Science Museum in London.



DRAISIENNE, a two-wheeled "pedestrian hobby-horse" devised between 1816 and 1818 by Baron von Drais de Sauerbrun of Baden-Württemberg, is considered the earliest forerunner of the bicycle. The vehicle, which was propelled by the feet pushing directly on the ground, was not, however, regarded as a serious means of transportation. This particular model, dating from 1817, was photographed at the city museum in Heidelberg.

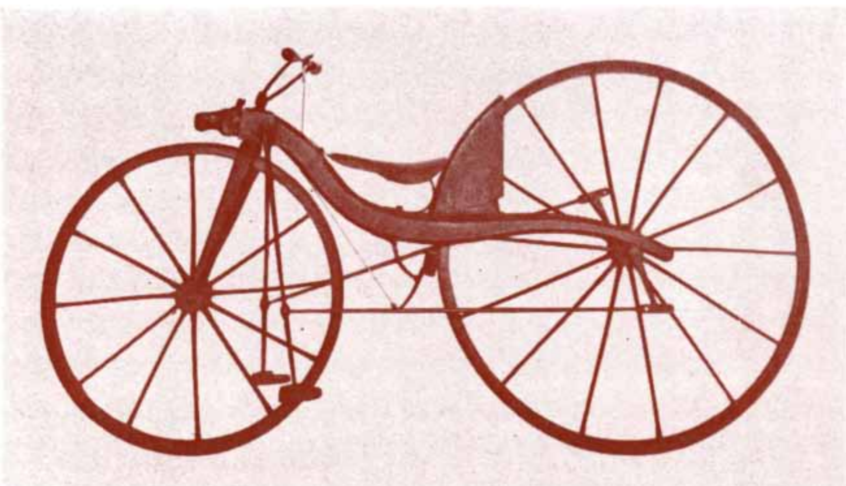
Sauerbrun of Baden-Württemberg, who between 1816 and 1818 devised the Draisienne, a two-wheeled "pedestrian hobby-horse" propelled by the feet pushing directly on the ground [*see illustration above*]. The vehicle had a brief vogue, but it was not taken seriously as a means of transportation. In 1839 Kirkpatric Macmillan, a Scottish blacksmith, succeeded in making a treadle-driven two-wheeled machine, which was copied but was never a popular success [*see illustration below*].

The first commercially important machine in this lineage was the French velocipede, developed by Pierre and Ernest Michaux in Paris in 1863 [*see top illustration on opposite page*]. This vehicle, sometimes called "the bone-shaker," had cranks fixed directly to the hub of the front wheel, like the simplest child's tricycle. As a result it suffered from the limitation of having too low a "gear ratio," to use the modern term. This meant that one turn of the pedals advanced the machine a distance equal to the circumference of the front wheel, perhaps only 10 feet. (In a modern bicycle one turn of the pedals, by means of a chain drive from a large sprocket to a small one, advances the machine 16 feet or more.) The only way to overcome this limitation while retaining the simplicity of direct drive was to use a very large front wheel. Thus the next stage in the evolution of the bicycle, the famous "high-wheeler," was characterized by a front wheel as much as 60 inches in diameter accompanied by a back wheel with a diameter of 20 inches or less.

The high-wheeler, also known as the "penny farthing" or "ordinary" machine, evolved primarily at Coventry in England, now the center of the British automobile industry. It was largely the work of one family, the Starleys. This sequence of events ensued from the accident that the Coventry Sewing Machine Company had a representative in Paris, one Rowley Turner, who brought a Michaux velocipede back to Coventry in 1868. James Starley, a self-taught engineer and inventor, was works manager at the time, and he immediately saw the possibilities of bicycle manufacture. The firm promptly became the Coventry Machinists' Company, Limited, and began to take orders for the manufacture of several hundred machines of the Michaux type for the Paris market. As it happened, the Franco-Prussian War of 1870 intervened, so that the machines were sold mostly in England. This episode led to a period of intense technical and commercial development that resulted not only in the bicycle's achieving its definitive form but also in the emergence of the motorcycle, the motor tricycle and the automobile.

Before considering these developments in detail it is worth asking why such an apparently simple device as the bicycle should have had such a major effect on the acceleration of technology. The answer surely lies in the sheer humanity of the machine. Its purpose is to make it easier for an individual to move about, and this the bicycle achieves in a way that quite outdoes natural evolution. When one compares the energy consumed in moving a certain distance as a function of body weight for a variety of animals and machines, one finds that an unaided walking man does fairly well (consuming about .75 calorie per gram per kilometer), but he is not as efficient as a horse, a salmon or a jet transport [*see illustration on page 90*]. With the aid of a bicycle, however, the man's energy consumption for a given distance is reduced to about a fifth (roughly .15 calorie per gram per kilometer). Therefore, apart from increasing his unaided speed by a factor of three or four, the cyclist improves his efficiency rating to No. 1 among moving creatures and machines.

In order to make this excellent performance possible the bicycle has evolved so that it is the optimum design ergonomically. It uses the right muscles (those of the thighs, the most powerful in the body) in the right motion (a



TREADLE-DRIVEN TWO-WHEELER was built in 1839 by Kirkpatric Macmillan, a Scottish blacksmith, for his own use. Although copied, the machine was never a popular success.

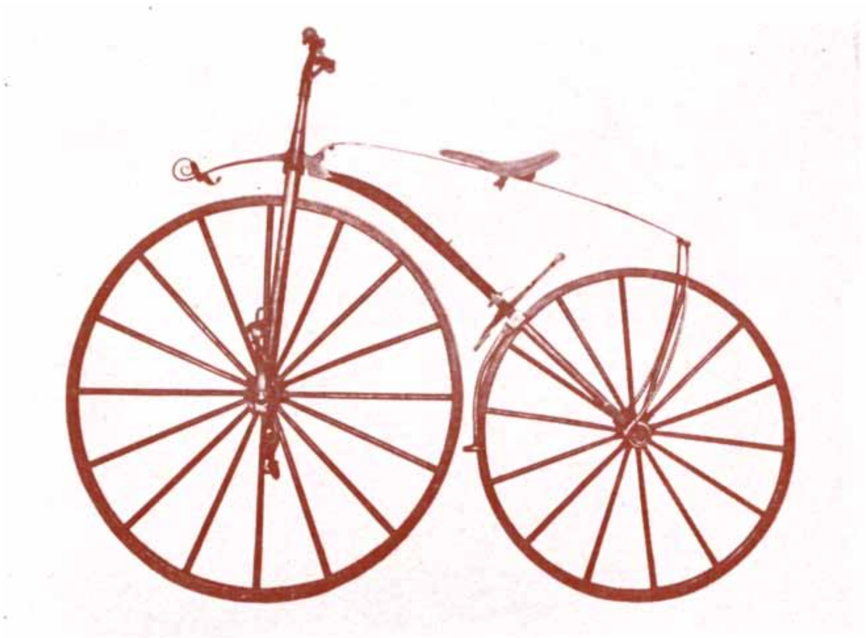
smooth rotary action of the feet) at the right speed (60 to 80 revolutions per minute). Such a design must transmit power efficiently (by means of ball bearings and the bush-roller chain); it must minimize rolling resistance (by means of the pneumatic tire), and it must be the minimum weight in order to reduce the effort of pedaling uphill.

The reason for the high energy efficiency of cycling compared with walking appears to lie mainly in the mode of action of the muscles. Whereas a machine only performs mechanical work when a force moves through a distance, muscles consume energy when they are in tension but not moving (doing what is sometimes called "isometric" work). A man standing still maintains his upright posture by means of a complicated system of bones in compression and muscles in tension. Hence merely standing consumes energy. Similarly, in performing movements with no external forces, as in shadowboxing, muscular energy is consumed because of the alternate acceleration and deceleration of the hands and arms, although no mechanical work is done against any outside agency.

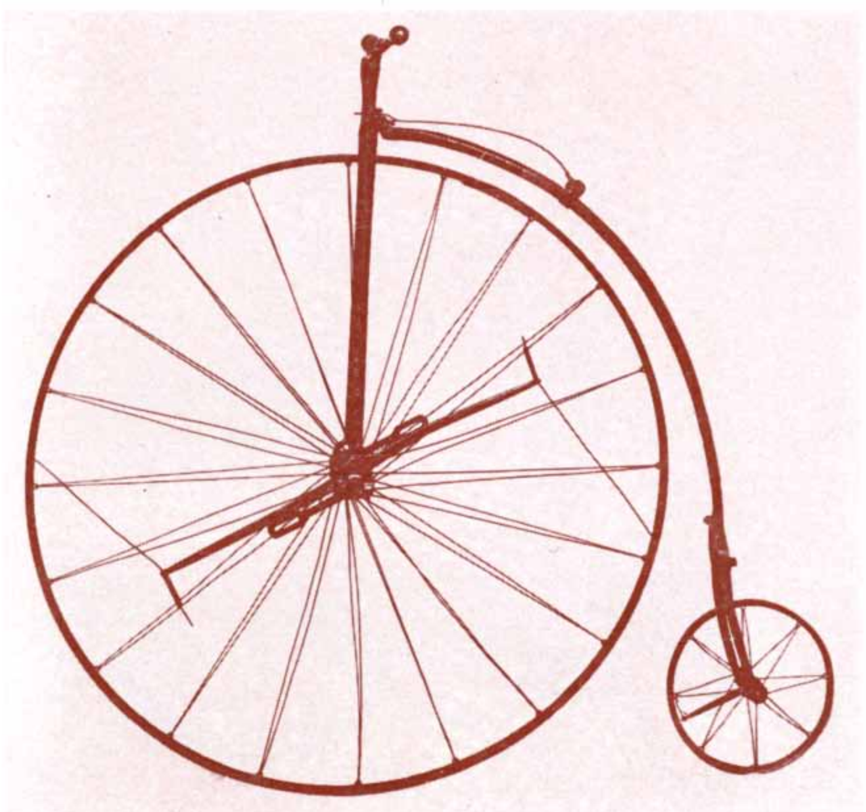
In walking the leg muscles must not only support the rest of the body in an erect posture but also raise and lower the entire body as well as accelerate and decelerate the lower limbs. All these actions consume energy without doing any useful external work. Walking uphill requires that additional work be done against gravity. Apart from these ways of consuming energy, every time the foot strikes the ground some energy is lost, as evidenced by the wear of footpaths, shoes and socks. The swinging of the arms and legs also causes wear and loss of energy by chafing.

Contrast this with the cyclist, who first of all saves energy by sitting, thus relieving his leg muscles of their supporting function and accompanying energy consumption. The only reciprocating parts of his body are his knees and thighs; his feet rotate smoothly at a constant speed and the rest of his body is still. Even the acceleration and deceleration of his legs are achieved efficiently, since the strongest muscles are used almost exclusively; the rising leg does not have to be lifted but is raised by the downward thrust of the other leg. The back muscles must be used to support the trunk, but the arms can also help to do this, resulting (in the normal cycling attitude) in a little residual strain on the hands and arms.

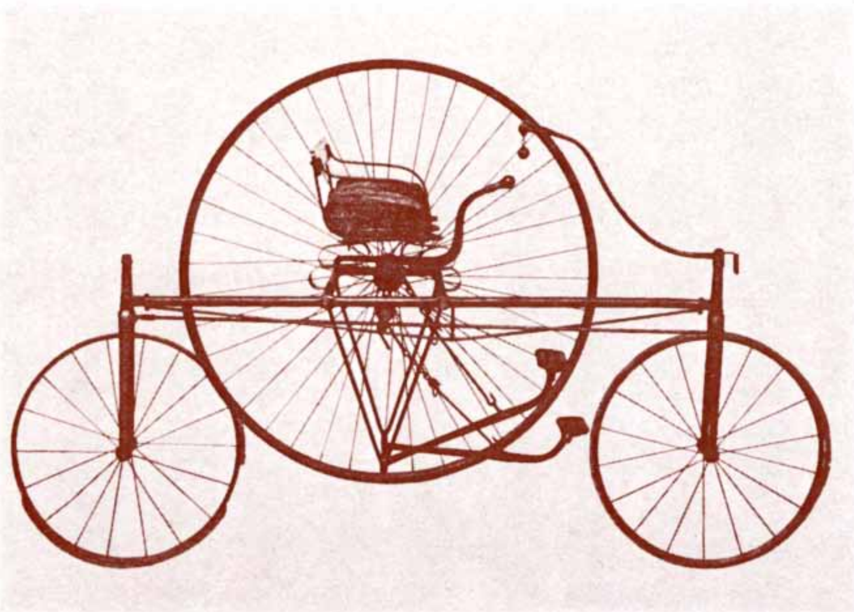
A less comfortable attitude is adopted by a racing cyclist in order to lessen wind resistance, perhaps the worst feature of



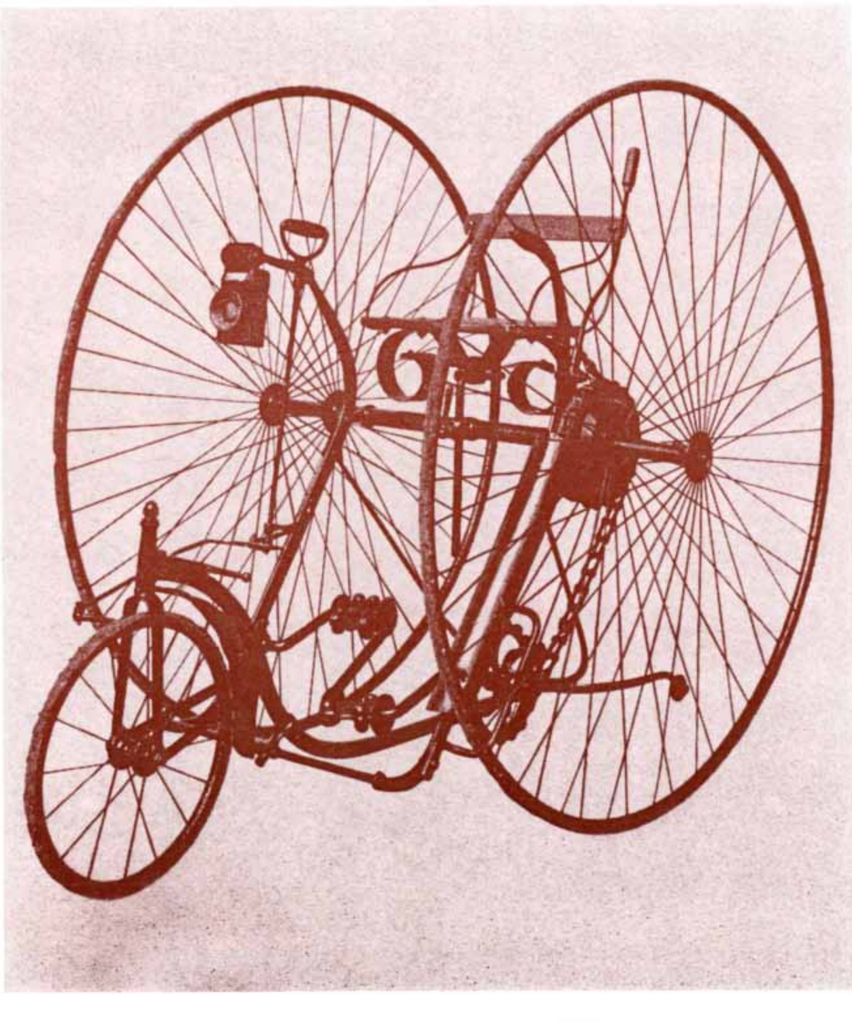
FRENCH VELOCIPEDE, produced by Pierre and Ernest Michaux in Paris in 1863, was the first commercially important machine on the way to the modern bicycle. The "bone-shaker," as the vehicle was sometimes called, had cranks with pedals fixed directly to the hub of the front wheel and hence suffered from the limitation of having too low a gear ratio.



HIGH-WHEELER, developed primarily by the Starley family, was designed to overcome the low gear ratio of the Michaux-type velocipede while retaining direct drive. The high-wheeler, also known as the "penny farthing" or "ordinary" machine, had a front wheel as much as 60 inches in diameter and a back wheel only 20 inches or less. This particular all-metal design, called the Ariel, was produced in 1870 by James Starley and William Hillman. Its spokes, which were radial, were not well adapted to resist the large torque exerted by the pedals on the hub. Hence the two extra rigid bars, each with its own adjustable spoke, were added to help transmit the torque from the hub to the rim. The definitive solution to this problem, the tangent-spoked wheel, was patented by Starley four years later.



COVENTRY LEVER TRICYCLE was designed by James Starley in 1876 as a way of circumventing the difficulties encountered in mounting the high-wheeler and then staying aloft.



ROYAL SALVO TRICYCLE, also designed by James Starley, attracted the attention of Queen Victoria, who ordered two of the machines, thereby establishing the respectability of the new fad of cycling. This tricycle, which incorporated one of the earliest differential gears, invented by Starley in 1873, was considered particularly well suited for female riders.

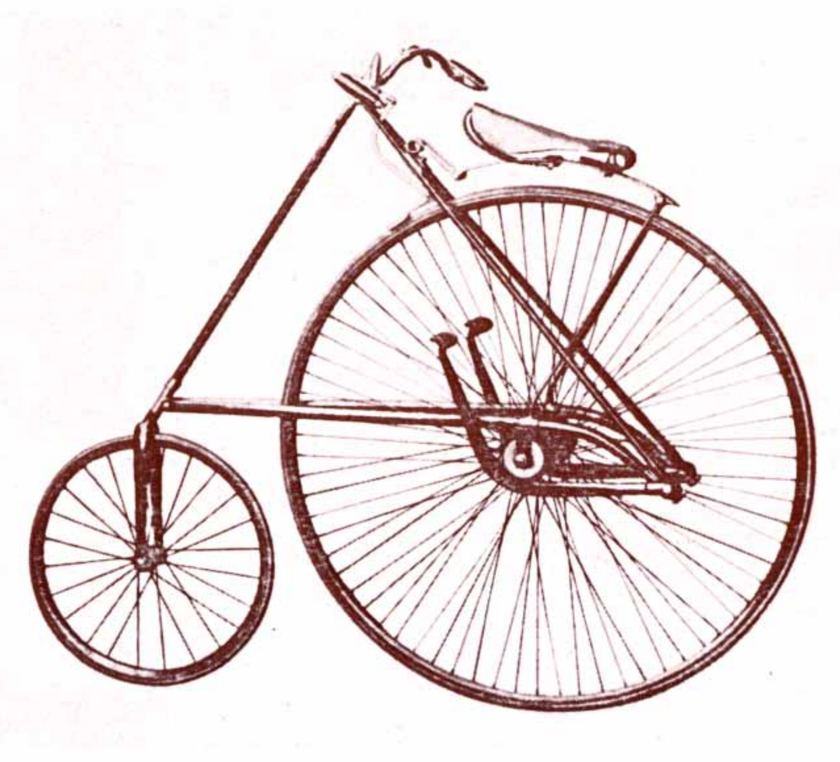
the bicycle for energy loss. Wind resistance varies as the square of the velocity of the wind with respect to the cyclist. Hence if one were to cycle at 12 miles per hour into a wind blowing at six miles per hour, the wind resistance would be nine times greater than if one were to maintain the same road speed with a following wind of six miles per hour. In practice, as every cyclist knows, one's speed can be adjusted to suit the wind conditions with a change of gear ratio to maintain an optimum pedaling speed. Apart from wind resistance the only significant form of energy loss is due to rolling resistance, which with normal-size wheels and properly inflated tires is very small on a smooth surface and is almost independent of speed.

It is because every part of the design must be related to the human frame that the entire bicycle must always be on a human scale. The lightness of construction, achieved mainly through the development of the wire-spoke wheel and the tubular frame, was dictated not only by the fact that the machine has to be pedaled uphill but also by the desirability of making it easy to lift. Since the bicycle makes little demand on material or energy resources, contributes little to pollution, makes a positive contribution to health and causes little death or injury, it can be regarded as the most benevolent of machines.

To return to the story of the bicycle's evolution, in 1870 James Starley and William Hillman (who later founded the automobile firm named after him) designed the Ariel machine: an elegant all-metal high-wheeler with wire-spoke wheels [*see bottom illustration on preceding page*]. The spokes, which were radial and could be tightened as desired, were not well adapted to resist the large torque exerted by the pedals on the hub. Four years later Starley patented the definitive solution to the problem: the tangent-spoked wheel. In this design, now universal, the spokes are placed so as to be tangential to the hub in both the forward and the backward direction, thus forming a series of triangles that brace the wheel against torque during either acceleration or braking. The usual number of spokes in a modern bicycle is 32 in the front wheel and 40 in the back; they are of uniform thickness or else butted (thickened toward both ends) for greater strength with lightness.

The major flaws of the high-wheeler bicycle were the difficulty in mounting the machine and then staying aloft. To overcome these dangers tricycles were developed during the 1870's. Again

Starley took the lead with his Coventry lever tricycle of 1876 [*see top illustration on opposite page*]. Another notable attempt to make the high-wheeler safer was the Star bicycle made by the Smith Machine Company of New Jersey in 1881 [*see illustration at right*]. That machine had the small wheel in front and used a system of levers to drive the large wheel at the rear. The Star bicycle had some success, but the tricycle was much more suitable for women riders, one of whom attracted the attention of Queen Victoria; she ordered two of Starley's Royal Salvo tricycles, met the designer and presented him with an inscribed watch. Nothing could have been more effective in establishing the respectability of the new craze of cycling. It made it possible for well-brought up young ladies to get out and away from the stuffiness of their Victorian homes and led to such new freedoms as "rational dress," a trend led by Amelia Jenks Bloomer. It is not too farfetched to suggest that the coincidence of cycling with the gradual spread of education for women played a significant part in the early stages of women's movement toward political and economic equality.



STAR BICYCLE, built by the Smith Machine Company of Smithville, N.J., in 1881, was another notable attempt to make the high-wheeler safer. The machine had a small wheel in front and used a system of levers, drums and straps to drive the large wheel at the rear.

The Starleys were also responsible at about this time for a technical innovation in one of their machines that was to be of major importance for the automobile. That development arose from a difficulty encountered with a side-by-side two-seater machine in which James Starley and one of his sons, William, each pedaled a driving wheel. The result of young William's superior strength was a spill into a bed of nettles for his father. While recovering, James thought up the idea of the differential gear (actually a reinvention), which spreads the effort equally between the wheels on each side and yet allows the wheels to rotate at slightly different speeds when turning a corner. William Starley himself went on to be a prolific inventor, having 138 patents to his name, many for bicycles, when he died in 1937.

Two major developments of 1877 were the introduction of the tubular frame and ball bearings. In neither case was the concept altogether new, but it was its widespread adoption for the bicycle that brought each technique to fruition and universal use. This pattern was to be repeated later with the pneumatic tire and other innovations.

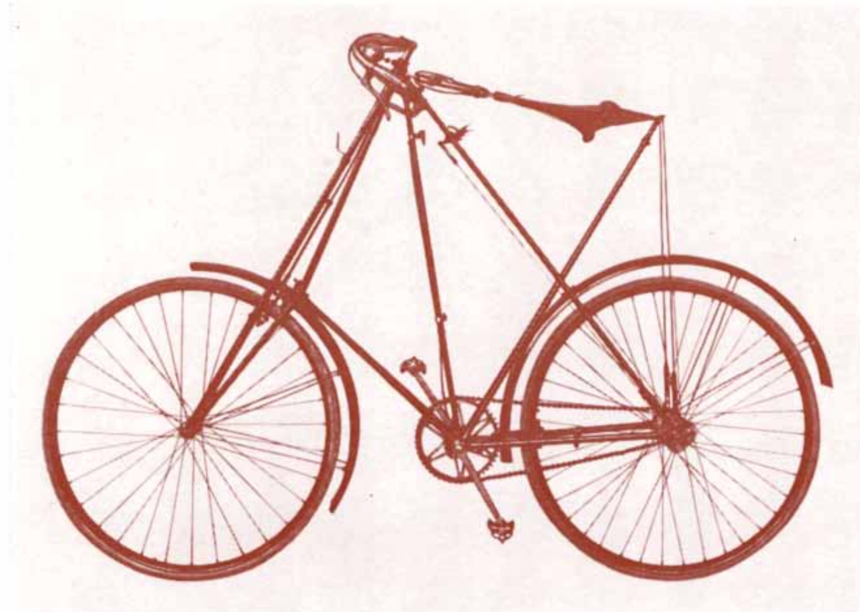
The thin-walled tube of circular cross section is a most efficient structural member; it can resist tension or compression, bending, torsion or the combination of stresses that are exerted on the frame of a vehicle. Although for bending in a particular plane an *I*-section joist may be more efficient, if the bending load can be applied in any plane, then the thin tube is to be preferred. It is for this reason that tubes are used as a strut or compression member in which failure could occur by elastic instability, or bowing. For torsion there is no better section, hence the tube in the typical main transmission shaft of an automobile. The stem of the bamboo plant is an excellent example of the properties of a hollow tube, as is attested by its use in the Far East for buildings, bridges, scaffolding and so on. Indeed, bicycles have been made of bamboo.

For the smallest stresses the design of a structure should be such as to result in tension or compression, not bending or torsion. That is the principle of the "space frame," which is used in bridge trusses, tower cranes and racing cars. Such construction is not practical in a bicycle, and so a compromise emerged in the classic diamond frame. In such a frame the main stresses are taken directly, even though there are bending stresses in the front fork and torsional stresses in the entire frame as the rider exerts pressure first on one pedal and then on the other. On a high-wheeler the rider feels these forces through the handlebars.

An alternative to the diamond frame that first appeared in 1886 and has recently reappeared is the cross frame. It consists of a main tube, extending from the steering head above the front wheel directly to the rear axle, crossed by a second tube from the saddle to the bottom bearing carrying the pedals. This is a simple arrangement, but it relies entirely on the strength and stiffness of the main tube, unless further members are added to obtain a partial triangulation of the frame.

Here an improvement is to increase the cross sectional area of the main tube, thereby obtaining the full benefit of thin-tube construction. This principle, used in most airplane fuselages since the 1930's, is described as monocoque or stressed-skin construction. Some recent motorized bicycles ("mopeds") have gone further and incorporate an enlarged main tube that forms the gasoline tank, a principle applied also in the construction of certain modern racing cars, which have a tubular fuel tank on each side of the driver. One advantage of the cross-frame bicycle is that it is equally suitable for men and skirt wearing women.

At least one early example of a fully triangulated frame achieved some success: the Dursley-Pedersen, which used small-diameter tubes in pairs and weighed only 23 pounds [*see illustra-*



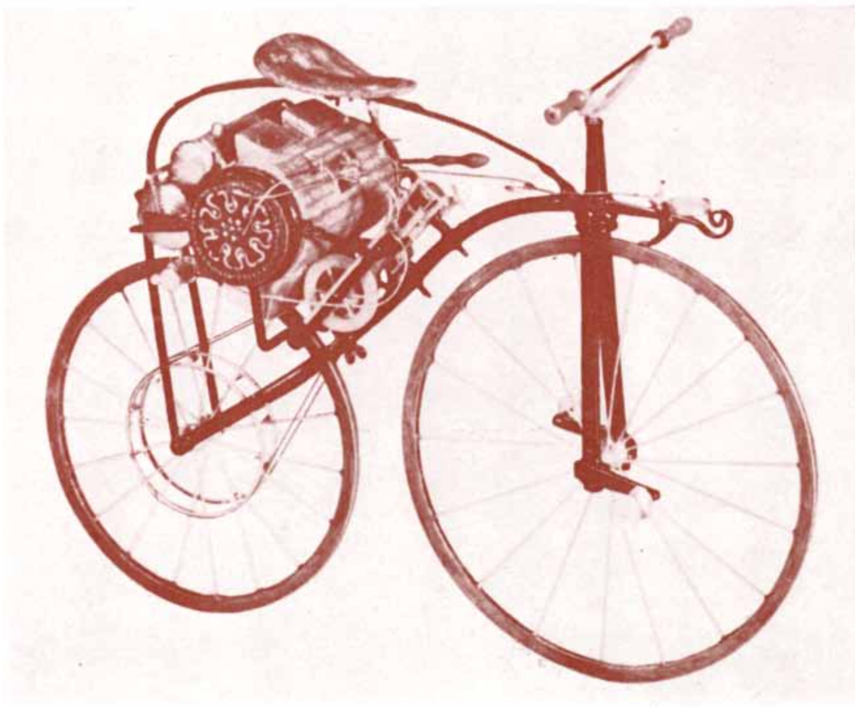
FULLY TRIANGULATED FRAME was the most distinctive feature of the Dursley-Pedersen bicycle, which achieved some success in England in the 1890's. This machine used small-diameter tubes in pairs and was remarkably light, weighing in at only 23 pounds.

tion above]. Some remarkably low weights were achieved at a very early date, mainly for racing machines. For example, a Rudge "ordinary" of 1884 weighed only 21.5 pounds. For purposes of comparison, a modern racing bicycle weighs about 20 pounds and a typical modern tourer weighs about 30 pounds. Steel tubes are still the normal choice, but light alloys, titanium and even plastic reinforced with carbon fibers have been used. The steel tubes, usually of a chrome-molybdenum or a manganese-molybdenum alloy, are butted and then brazed into steel sockets to form a complete frame. The result is a structure able to carry about 10 times its own weight, a figure not approached by any bridge, automobile or aircraft.

The use of a roller to reduce friction finds its ultimate development in ball bearings and roller bearings. The main bearings of a modern bicycle have at least 12 rows of balls, all rolling between an inner cone and an outer cup. All the bearing surfaces are hardened steel for resistance to deformation and wear. If such bearings are lubricated and properly adjusted, their life can be surprisingly long. Even if they are neglected, they continue to function for a long time, and all their parts can easily be renewed if necessary.

Another important advance, the adoption of chain-and-sprocket drive to the rear wheel, was made by Harry J. Lawson in 1879. The following year Hans Renold produced the definitive form of the bicycle chain, the bush-roller chain, which combines the virtues of long life, efficiency and low weight. At first sight the design seems to have little subtlety but a closer look reveals just how significant its various features are [*see illustration on page 88*]. The progenitor was the pin chain, or stud chain, in which the pins bear directly on the sprocket teeth and the link plates swivel on the studs at each end of the pins. In such a chain there is undue wear and friction both at the teeth and at the holes in the plates. An improvement devised by James Slater in 1864 was the bowl chain, or roller chain, in which friction and wear of the sprocket teeth was reduced by rollers on the pins, but wear of the plates on the studs was still too great. Renold's design, by the addition of hollow bushes to spread the load over the entire length of the pin, overcame this final shortcoming and led to the foundation of the precision-chain industry. The bush-roller chain spread from bicycles to textile machinery and other power-transmission applications, in competition with the Morse "silent chain." Bush-roller chains replaced belt drives for motorcycles, and they served as the main drive for the rear wheels of automobiles until they were replaced by shaft drive. Today the bush-roller chain drive remains virtually the universal choice for automobile camshafts, although it is threatened now by the toothed-belt drive. The demands of increased power, greater speed and longer life for chain drives meant that the makers were pioneers in metallurgy, heat treatment, lubrication and production.

The satisfactory solution to the problem of an efficient drive giving a "step up" ratio of any desired value made possible the final evolution of the bicycle to



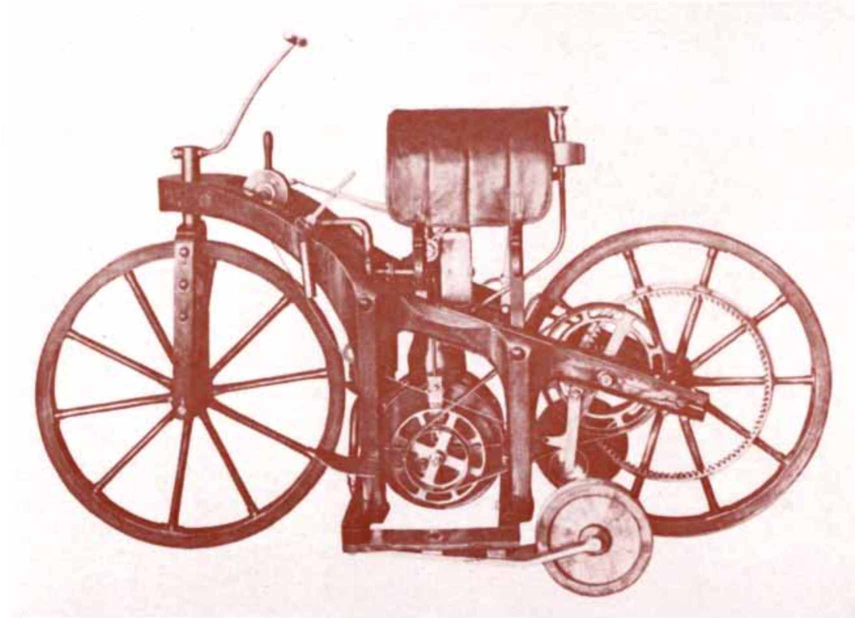
STEAM-DRIVEN BICYCLE was built in 1869 by Pierre Michaux using a Perreaux steam engine. Although earlier steam-driven road vehicles had proved too heavy and cumbersome, the technology of the lightweight bicycle seemed at this time to offer new possibilities. The later evolution of the internal-combustion engine ended this line of development.

its modern form. This last step was achieved mainly by J. K. Starley, a nephew of James Starley. James, honored in Coventry as the "father of the bicycle industry," had formed a partnership with William Sutton in 1878 to produce tricycles. In 1885, however, J. K. independently brought out his famous Rover safety bicycle [*see illustration on page 81*]. This machine can be regarded as the final development of the bicycle form. From that form the bicycle has not departed, in spite of a recent attempt to use a spring frame in conjunction with small wheels. The Rover bicycle had a diamond frame, incorporating two curved tubes without the extra diagonal tube from the saddle to the bottom bracket. The front forks were straight, although sloping. The slope was used to give a self-centering action to the steering. The later development of curved front forks such that the line of pivoting of the steering head meets the ground at the point of contact with the tire results in a reduction of steering effort because side forces do not tend to turn the handlebars.

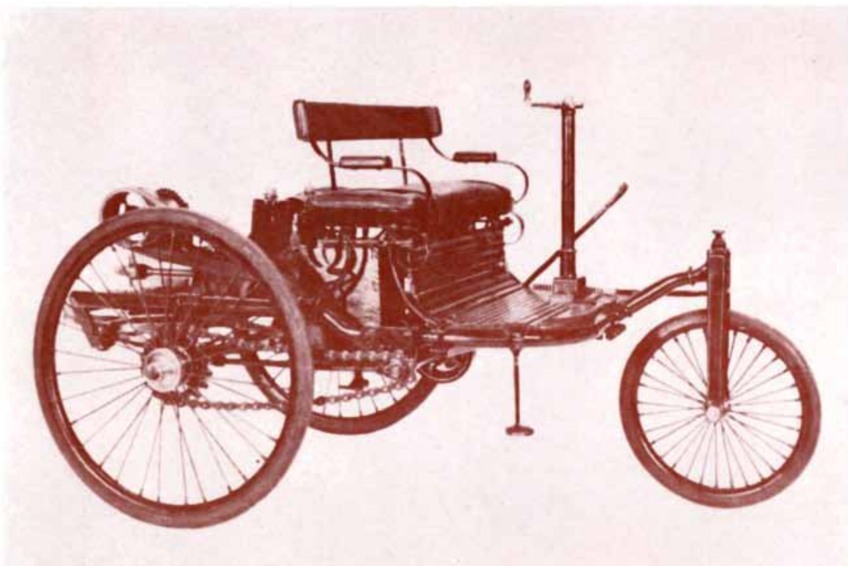
The appearance of the Rover safety bicycle started a boom in bicycles that quickly established them as an everyday means of transport, as a sport vehicle and as a means of long-distance touring. For Coventry and other Midlands cities there was a trade boom that lasted until 1898. Then there was a disastrous slump, largely because of the financial manipulations of one Terah Hooley. The bicycle spread all over the world, and it was adapted to local needs in such machines as the bicycle rickshaw of the Far East. Among the early manufacturers in the U.S. were the Duryea brothers, builders of the first American automobile in 1893. By 1899 there were 312 factories in the U.S. producing a million bicycles a year.

Before this time, however, there was one more signal development: the pneumatic tire. This feature had actually been patented as far back as 1845 by R. W. Thomson, a Scottish civil engineer, to reduce the effort needed to pull horse-drawn carriages. It failed to establish itself until it was reinvented in 1888 by John Boyd Dunlop, a Scottish veterinary surgeon practicing in Belfast. This time success was rapid, owing to the enormous popularity of the bicycle and to the obvious superiority in comfort and efficiency of the pneumatic tire over the solid-rubber tire. Further developments came quickly. Charles Kingston Welch of London produced the wire-edged tire in 1890, and at almost the same time William Erskine Bartlett in the U.S. produced the bead-edged tire. Then David Mosely of Manchester patented the cord-construction tire, consisting of layers of parallel cords rather than a woven fabric, which gave rise to undue internal friction. Even the tubeless tire, which was not used for automobiles until the 1940's, was invented for bicycles in the late 1890's. E. S. Tompkins, a student of the pneumatic tire, has written: "This richness of invention in the very earliest years arises from the fact that the cycling enthusiasts were young folk and they had a young and enquiring approach to the design of tyres for their machines."

A further result of the popularity of



DAIMLER MOTOR BICYCLE, the forerunner of the modern motorcycle, was designed and built by Gottlieb Daimler in 1885. The vehicle was equipped with a single-cylinder internal-combustion engine. The two smaller jockey wheels retracted when the machine was under way. The original Daimler machine, shown in this photograph from the Bettmann Archive, is now in the restored Daimler workshop at Bad Cannstatt near Stuttgart.



BENZ MOTOR TRICYCLE, the forerunner of the modern automobile, also made its appearance in 1885. This first attempt of Carl Benz incorporated such features as electric ignition, effective throttle control, mechanical valves, horizontal flywheel and even a comfortably upholstered seat. In tests through 1886 the vehicle was developed to deliver a reliable nine miles per hour. The original Benz car is now in the Deutsches Museum in Munich; this photograph shows a replica of the original in the Daimler-Benz Museum in Stuttgart.

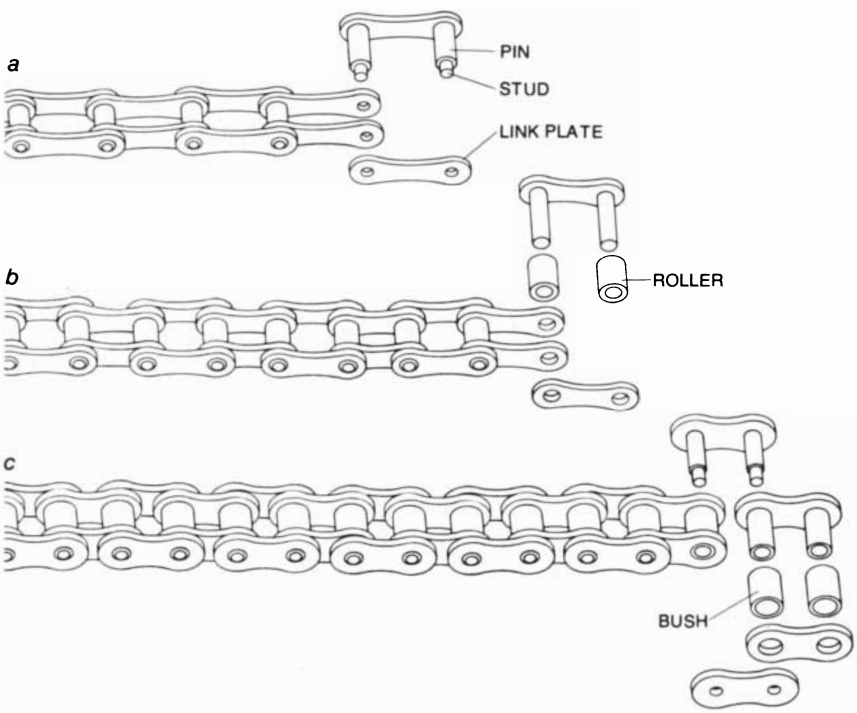
bicycles was the demand for better roads. In Britain the Cyclists' Touring Club was founded in 1878 and the Roads Improvement Association in 1886, leading to successful pressure for better roads. A similar movement in the U.S., founded by Colonel Albert A. Pope, a pioneer bicycle manufacturer of Boston, was also influential. The bicycle quite literally paved the way for the automobile.

How the bicycle led directly to the automobile is described by the inventor Hiram P. Maxim, Jr., in *Horseless Carriage Days* (published in 1937): "The reason why we did not build mechanical road vehicles before this [1890] in my opinion was because the bicycle had not yet come in numbers and had not directed men's minds to the possibilities of independent long-distance travel on the ordinary highway. We thought the railway was good enough. The bicycle created a new demand which was beyond the capacity of the railroad to supply. Then it came about that the bicycle could not satisfy the demand it had created. A mechanically propelled vehicle was wanted instead of a foot-propelled one, and we now know that the automobile was the answer."

Looking backward after another 36 years, one is tempted to ask whether the automobile is really as good an answer as it once appeared to be. Steam-propelled road vehicles had been tried earlier, but they had failed to establish themselves because they were heavy and cumbersome. The technology of the lightweight bicycle seemed to offer new possibilities, and a steam-driven motorcycle was actually built in France in 1869 by Pierre Michaux, using a Perreaux engine [*see bottom illustration on page 86*]. The internal-combustion engine was successfully applied in 1885 by Gottlieb Daimler to a velocipede of the Michaux type [*see top illustration on preceding page*] and by Carl Benz to a lightweight tricycle [*see bottom illustration on preceding page*]. From then on development was rapid, particularly in France and later in the U.S.

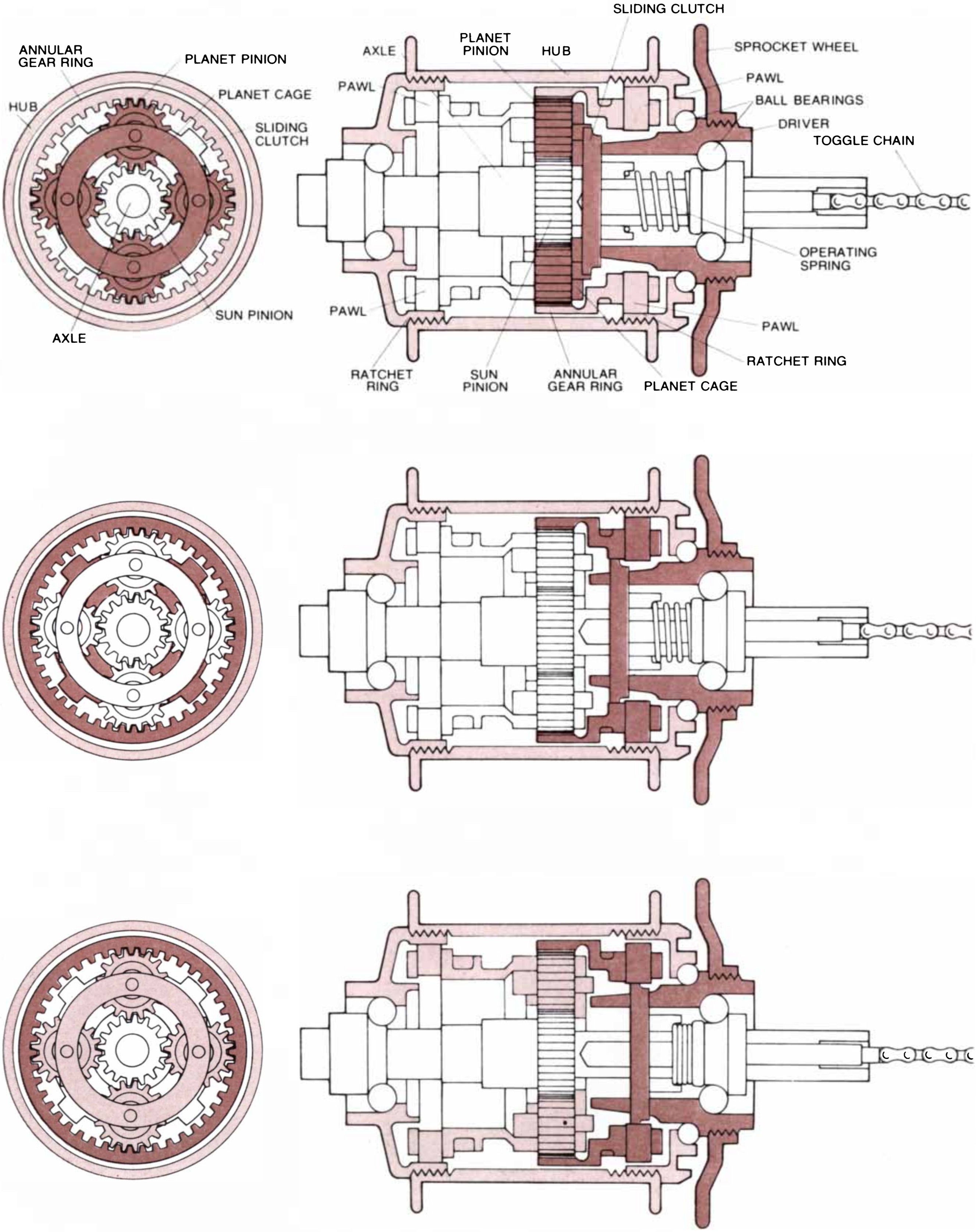
Many of the pioneer automobile manufacturers started as bicycle makers, including Hillman, Colonel Pope, R. E. Olds, Henry M. Leland and William Morris (later Lord Nuffield). Meanwhile in Coventry the firm of Starley and Sutton became the famous Rover Company, producing an electrically driven tricycle in 1888, a motorcycle in 1902 and an automobile in 1904. Both Morris and Rover are now part of the British Leyland Motor Corporation, the largest British automobile company. Henry Ford's first car used bicycle wheels and chains, as did other early vehicles of the "motorized buggy" type. The Wright brothers were bicycle makers, and the early flying machines benefited considerably from the lightweight and efficient design features evolved so successfully for bicycles.

Perhaps the most interesting modern application of the technology of the lightweight bicycle to flying is in man-powered aircraft. There were attempts to develop such aircraft in Germany and Italy before World War II, and recently much interest has been aroused in England by the offer of a prize of £10,000 by Henry Kramer for the first man-powered flight over a figure-eight course around two poles half a mile apart. Several designs have succeeded in flying for distances of up to some 1,000 yards in a straight line; the best flight so far is one made by Flight Lieutenant John Potter of the Royal Air Force in a monoplane designed by Christopher Roper [*see illustration on page 91*]. This aircraft, with a wingspan of 80 feet and a weight of only 146 pounds, clearly shows the debt owed to the bicycle in the design of efficient and lightweight machinery for the production and transmission of power. The Kramer prize has not yet been won, but it represents a goal that will no doubt eventually be achieved.



DEVELOPMENT OF BICYCLE CHAIN went through several stages before arriving at the definitive form, the bush-roller chain invented by Hans Renold in 1880. The progenitor was the pin chain, or stud chain (*a*), in which the pins bore directly on the sprocket teeth and the link plates swiveled on the studs at each end of the pins. In such a chain there is undue wear and friction both at the teeth and at the holes in the plates. A subsequent improvement was the bowl chain, or roller chain (*b*), in which friction and wear of the sprocket teeth was reduced by rollers on the pins, but wear of the plates on the studs was still too great. Renold's invention of the bush-roller chain (*c*), by the addition of hollow bushes to spread the load over the entire length of the pin, overcame this final shortcoming.

Production engineering also owes much to the sudden demand created by the bicycle for precision parts in quantity. The average bicycle has well over 1,000 individual parts. Admittedly nearly half of these parts are in the chain, but the rest of them call for high standards of pressing and machining, and the methods worked out for producing them represented a big step forward. A comparable demand for automobile parts had to await the Ford assembly line. Perhaps the most ingenious process used in producing bicycle parts is the procedure for forming the complicated bottom bracket of the frame, which has four tube sockets and a threaded barrel to receive the outer races for carrying the pedals. The bottom bracket is made in a series of press operations, a method evolved by the Raleigh Cycle Company of Nottingham in 1900. This company, which is now by far the largest bicycle



STURMEY-ARCHER HUB GEAR commonly used in British touring bicycles is shown here in both transverse section (*left*) and longitudinal section (*right*). In both cases the driving elements are indicated in dark color and the driven elements in light color. In this three-speed system a single epicyclic gear train is used in such a way that when the bicycle is in high gear (*a*), the cage is driven by the sprocket and the wheel is driven by the annulus. When the bicycle is in middle gear (*b*), the drive is direct to the wheel. When it is in low gear (*c*), the sprocket drives the annulus and the cage drives the wheel at a reduced speed. If the "sun" wheel (which is fixed to the stationary axle) has the same number of teeth as the "planet" wheels, and the annulus gear has three times this number, then high gear will have a step-up ratio of 4/3 compared with direct drive and low gear a step-down ratio of 3/4.

manufacturer in Britain, was founded by Sir Frank Bowden in 1887.

Other accessories that made their appearance during the 1880's and 1890's include the freewheel mechanism, originally introduced as an aid to mounting, since with the old "fixed" drive some form of mounting step was needed. The abandonment of a fixed drive created the need for better brakes. Originally bicycle brakes were of the simple "spoon" type, pressing on the front tire. Rim brakes came in later, first the stirrup brake, which acts on the underside of the rim, then the caliper brake, which presses on the flat outer sides of the narrow rim used in racing bicycles and lightweight touring machines. The original purpose of the caliper brake was to make it easier to change the wheel without disturbing the brake. Such a brake is operated by a cable in which an inner flexible wire in tension is contained within a flexible outer tube in compression. This system provides a most effective means for the remote operation of a mechanism such as a brake, for which it was first developed. It has since been widely used for such purposes as operating the clutch and throttle mechanisms of motorcycles and the control surfaces of airplanes.

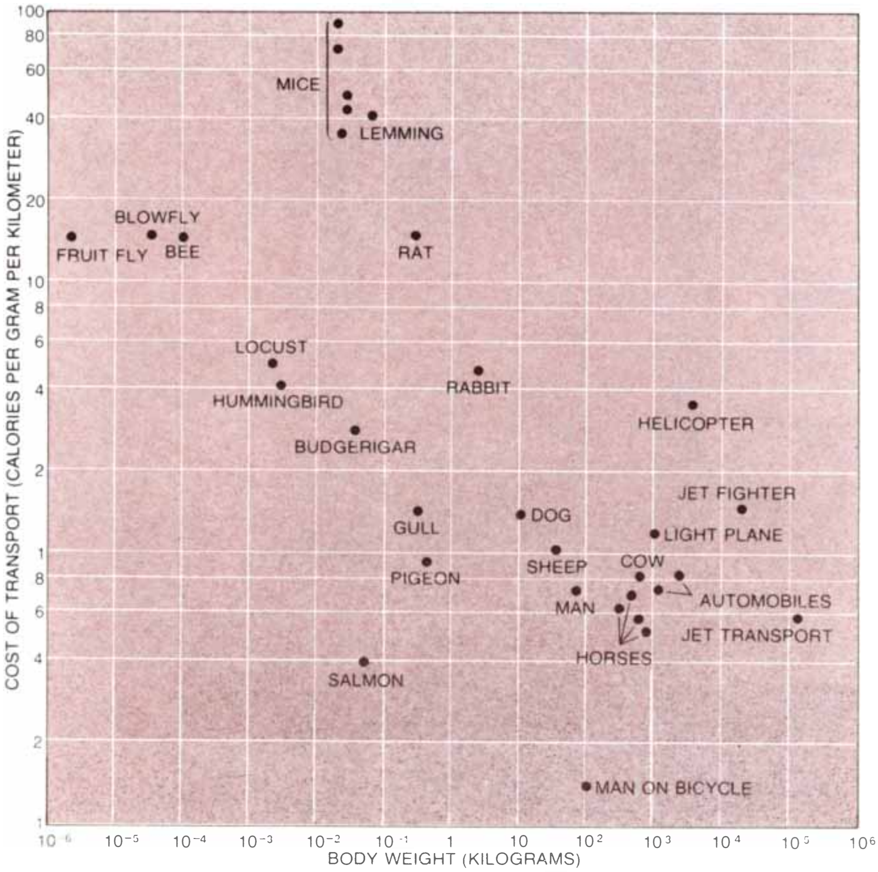
Two other types of brake made their appearance later. One is the coaster brake, or back-pedaling brake, which is particularly popular in the U.S. The other is the hub brake, or drum brake, of the type used in automobiles and motorcycles. Both types have the advantage of remaining effective in wet conditions.

The desirability of being able to change the gear ratio of a bicycle to provide for pedaling uphill or against a head wind is fairly obvious. This gear ratio is still defined by the diameter of the equivalent wheel of the original high-wheelers. Thus a wheel 27 inches in diameter driven by sprockets of 44 and 18 teeth has a "gear" of $(27 \times 44)/18$, or 66 inches. The idea of having different sizes of sprockets on the rear wheel led to the ***dérailleur*** change-speed gear of 1899, in which the rider can transfer the chain from one sprocket to another while pedaling. The necessary variation in the length of the chain on the tight side of the drive is accommodated by the use of a spring-loaded jockey pulley on the slack side of the chain. This type of gear is light and efficient but needs proper adjustment and lubrication if it is to last.

The alternative form of change-speed gear is the Sturmey-Archer hub gear [*see illustration on preceding page*]. In this system a single epicyclic gear train is used in such a way that when the bicycle is in high gear, the cage is driven by the sprocket and the wheel is driven by the annulus. When the bicycle is in middle gear, the drive is direct to the wheel, and when it is in low gear, the sprocket drives the annulus and the cage drives the wheel at a reduced speed. If, for example, the "sun" wheel (which is fixed to the stationary axle) has the same number of teeth as the "planet" wheels, and the annulus gear has three times this number, then high gear will have a step-up ratio of 4/3 compared with direct drive and low gear a step-down ratio of 3/4. A five-speed version is available incorporating two epicyclic gear trains of different ratios. The advantage of hub gears is compactness and the fact that the mechanism is well protected against dust and damage.

A large number of firms grew up to supply bicycle parts or accessories. One firm in particular, Joseph Lucas Limited, owes its early prominence to the manufacture of the first successful bicycle lamp, an oil-burning wick lamp rejoicing in the name of the King of the Road Cycle Hub Lamp. Later this company made acetylene lamps and then electric lamps, so that they were in a good position to develop the market for automobile lights and other electrical equipment for automobiles. An early form of an electric generator for bicycles was invented by Richard Weber of Leipzig in 1886. The modern hub generator, which requires a minimum of additional effort on the part of the rider, was first produced by Raleigh in 1936.

Bicycle manufacture is still a big business, accounting for a worldwide production of between 35 and 40 million vehicles per year. The leading manufacturing country is still the U.S., with about six million bicycles per year, followed closely by China, with about five million. By any standards these figures demonstrate the importance of the bicycle. If one examines the extent to



MAN ON A BICYCLE ranks first in efficiency among traveling animals and machines in terms of energy consumed in moving a certain distance as a function of body weight. The rate of energy consumption for a bicyclist (about .15 calorie per gram per kilometer) is approximately a fifth of that for an unaided walking man (about .75 calorie per gram per kilometer). With the exception of the black point representing the bicyclist (*lower right*), this graph is based on data originally compiled by Vance A. Tucker of Duke University.

which bicycles are in use today, one finds that in most of the world they play a role far more significant than that of the automobile. China with its 800 million inhabitants relies heavily on the bicycle for the transport of people and goods. So do the countries of Southeast Asia and Africa. Even the U.S.S.R., with only about 1.5 million automobiles, has an annual production of 4.5 million bicycles. Europe and North America are therefore in a minority in relying so heavily on the automobile. The true cost of doing so is becoming increasingly evident, not only in the consumption of resources but also in pollution and other undesirable effects on urban life.

For those of us in the overdeveloped world the bicycle offers a real alternative to the automobile, if we are prepared to recognize and grasp the opportunities by planning our living and working environment in such a way as to induce the use of these humane machines. The possible inducements are many: cycleways to reduce the danger to cyclists of automobile traffic, bicycle parking stations, facilities for the transportation of bicycles by rail and bus, and public bicycles for "park and pedal" service. Already bicycling is often the best way to get around quickly in city centers.

Two important factors must gradually force a reappraisal of the hypertrophic role the automobile plays in Western life. The first is the undoubted diminution of fossil-fuel resources and the accompanying increase in fuel prices. The second is the sheer inequity in per capita energy consumption between automobile-using and non-automobile-using countries. In these days of universal communication such a situation will appear more and more inequitable and a source of resentment. It is inconceivable that 800 million Chinese will ever become consumers of energy on the per capita scale of 200 million Americans, and the end result must be a gradual reduction of energy consumption in the U.S. To this end the bicycle can play a significant part and thereby become a great leveler.

For the developing countries the bicycle offers a different set of opportunities. With the continuing spread of bicycles from cities to towns and villages go the accompanying mechanical skills and essential spare parts. Thus bicycle technology serves the purpose of technical education on which the peoples of these countries can build in the same way that we in the developed countries did only 70 to 100 years ago. There is evidence of such a process at work. The Chinese are replacing the wooden wheel on their traditional wheelbarrow with a bicycle wheel, thereby making it both easier to push and kinder to road surfaces. Threshing and winnowing machines have been designed that incorporate bicycle bearings and chain drives. It is this kind of do-it-yourself, village-level technology that offers the best route to self-improvement, a route far more plausible than any form of large-scale aid from outside. The power of the bicycle with respect to the power of the most sophisticated modern technology is perhaps best shown in Indochina, where the North Vietnamese have used it as a major means of transport. The Japanese who captured Singapore in World War II also traveled largely by bicycle. Nonetheless, the bicycle remains essentially a peaceful device, and we do not need to include it in strategic-arms-limitation negotiations. We might do well, however, to go all out in encouraging its use. If one were to give a short prescription for dealing rationally with the world's problems of development, transportation, health and the efficient use of resources, one could do worse than the simple formula: Cycle and recycle.

RECENT APPLICATION of bicycle technology to flying is in the design of man-powered aircraft, a goal that has attracted a great deal of interest in England lately following the offer of a prize of £10,000 by Henry Kramer for the first man-powered flight over a figure-eight course around two poles half a mile apart. The best flight so far is one made by Flight Lieutenant John Potter of the Royal Air Force in a monoplane designed by Christopher Roper. The Roper aircraft, shown in the photograph with Potter in the driver's seat, has a wingspan of 80 feet and a weight (without Potter) of only 146 pounds. In actual flight an aerodynamic nose canopy completely encloses the rider and drive mechanism.

THE MIGRATIONS OF THE SHAD

The largest member of the herring family, much prized as a food fish, moves between the sea and its river spawning grounds with remarkable precision. Its main guide appears to be temperature

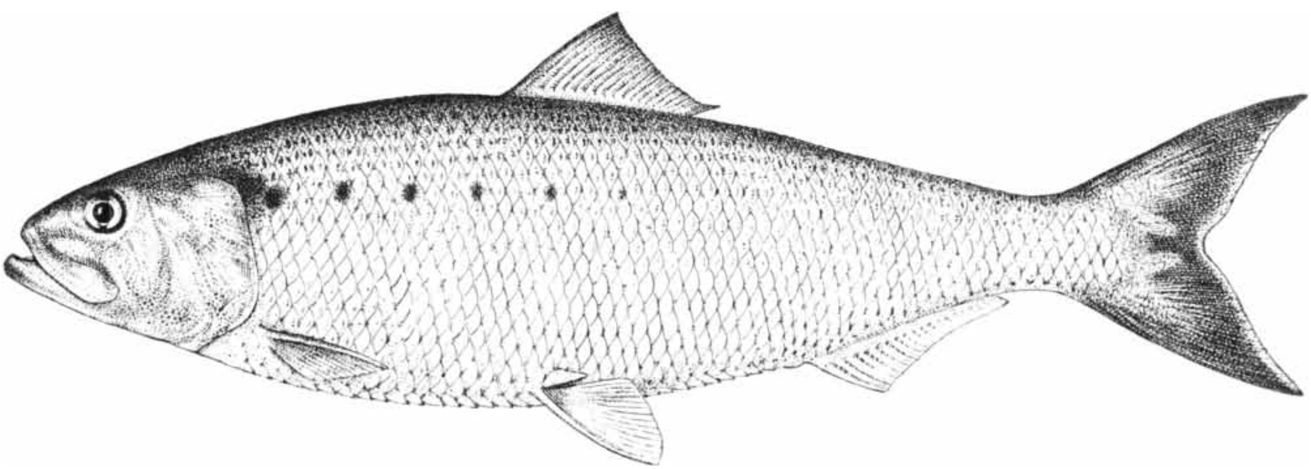
by William C. Leggett

In many rivers, particularly those along the Atlantic Coast of North America, the annual "shad run" is a dramatic biological event. Within a space of a few weeks shad by the tens of thousands come in from the sea and move upriver to spawn. The timing of these migrations differs from river to river, being earliest in the south and moving progressively toward the north; in the St. Johns River of Florida, for example, the peak run is in January, whereas in the St. John River of Canada it is in June. What governs the timing of the movement of the shad into the rivers? The question is of both biological and practical interest: in North America the shad is an important seasonal food fish, and it is also highly prized by sport fishermen. The answer lies largely in the shad's preference for specific water temperatures in both the rivers and the sea. In these days of human intervention in such temperatures the question has added significance.

The American shad (*Alosa sapidissima*) is the largest member of the herring family, attaining an average length of from 40 to 50 centimeters (16 to 20 inches) and an average weight of from 1,400 to 2,200 grams (three to five pounds). Together with such fishes as the salmon, the sturgeon and the striped bass, the shad has the life cycle termed anadromous: the young, born in fresh water, remain there for a time (a matter of a few months for the shad) and then descend to the sea, where they remain until they attain sexual maturity and return to the home stream to spawn. By far the greatest part of a shad's life is spent at sea, since it takes from three to six years for the young fish to attain sexual maturity and since the adults that survive the annual spawning run return promptly to the sea.

In spite of the substantial commercial and recreational value of the shad, little was known of the fish's movements until recently. The state of knowledge at the turn of the century was indicated by C. H. Stevenson of the U.S. Fish Commission, who wrote: "It was formerly considered that the entire body of shad wintered in the south and started northward at the beginning of the year... sending a detachment up each successive stream, this division, by a singular method of selection, being the individuals that were bred in those respective streams.... But zoologists now recognize [that] the young shad hatched out in any particular river remain within a moderate distance off the mouth of that stream until the period occurs for their inland migration... entering the rivers as soon as the temperature of the water is suitable."

In the early part of the century a few



AMERICAN SHAD, which bears the formal name *Alosa sapidissima*, is the largest member of the herring family. An adult male is shown here. The fish attains an average length of from 40 to 50 centimeters, or 16 to 20 inches, and an average weight of from 1,400 to 2,200 grams, or three to five pounds. It was originally native to the Atlantic Coast of North America, but in 1871 the Sacramento River was stocked with shad from the East Coast, and the fish has since come to range widely along the Pacific Coast.

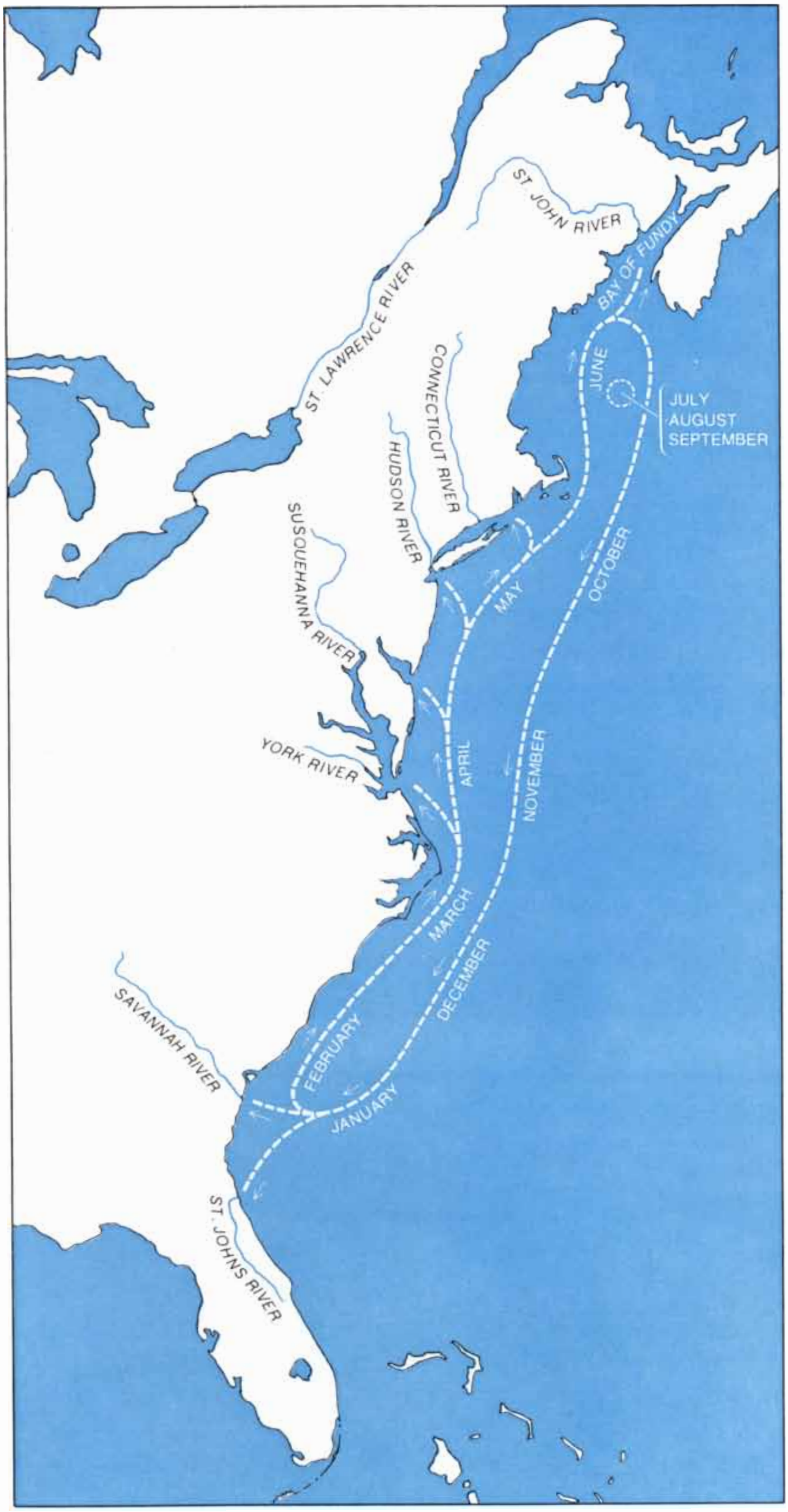
shad were found (from tagging studies) to have moved over considerable distances at sea. Such movements were generally regarded as atypical. It was not until 1958 that Gerald B. Talbot and James E. Sykes of the U.S. Bureau of Commercial Fisheries showed long-range movement to be characteristic of the species.

Talbot and Sykes analyzed recoveries from more than 17,000 shad that had been tagged at several locations along the Atlantic Coast. They found that the distribution of recaptured fish exhibited a regular pattern in both time and space. The tag returns indicated that adult shad native to rivers from Chesapeake Bay to Connecticut migrated northward after spawning, congregating during July, August and September in the Gulf of Maine and the Bay of Fundy, together with shad from Canadian rivers.

In October and November the shad disappeared from the Gulf of Maine and the Bay of Fundy. Evidence of where they went was sparse. Tag returns indicated, however, that shad were off Massachusetts in October and November and between Long Island and North Carolina in February and March. On the basis of these returns Talbot and Sykes concluded that during the winter months shad move slowly southward to the Middle Atlantic region of the coast. In the spring they migrate in the direction of the river in which they were spawned, as indicated by the fact that tag recoveries were progressively closer to the home river as the season advanced.

The tag data taken as a whole showed that the shad migrated extensively, with annual movements exceeding 2,400 miles. The factors governing the movements, however, remained unknown until I began in 1965 a study of the population dynamics and migratory behavior of the shad native to the Connecticut River. The main purpose of the work was to assess the effect on the shad population of the discharge of heated water into the river from the nuclear generating station of the Connecticut Yankee Atomic Power Company at Haddam Neck, 15 miles above the mouth of the river. As so often happens, the work has produced a considerable body of new information quite unrelated to the problem at hand, including a clearer understanding of the migrations of shad.

As part of the study more than 32,000 adult shad have been tagged and released. More than 5,000 of them have been recovered in the river and close to 100 more were found at various places along the Atlantic Coast. As one might



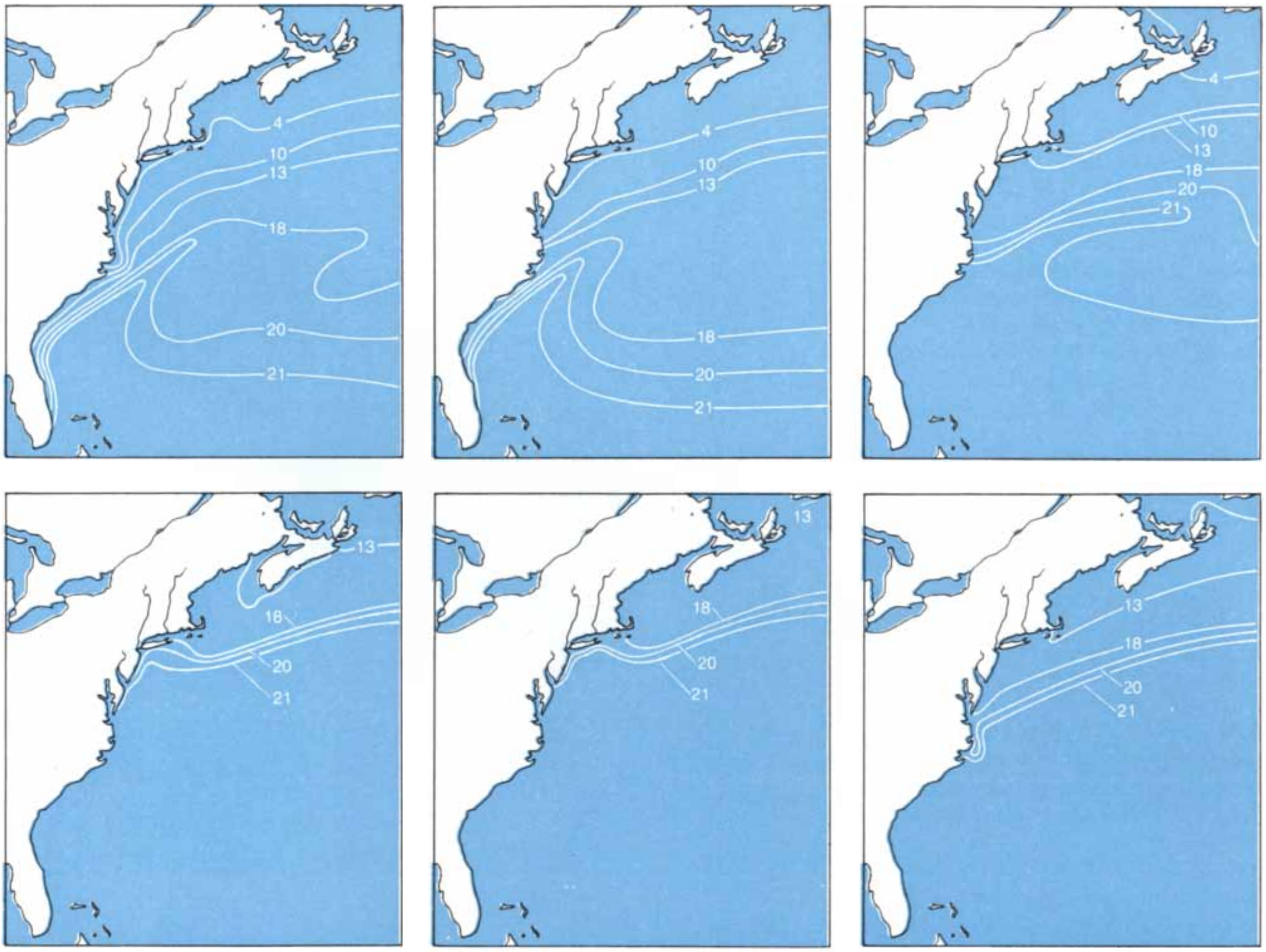
ATLANTIC MIGRATIONS of the shad are depicted according to the time of year when they are found or believed to be in each location. Also shown are several of the rivers that have shad runs on a large scale. The effect of the temperature dependence of the shad's movements at sea is to bring the fish near its home river in time for the spawning run.

have predicted on the basis of the earlier evidence, the recoveries at sea ranged from North Carolina to New Brunswick. What was not anticipated was the remarkable precision in the timing of the movements as revealed by these recoveries.

In the Connecticut River from 1966 to 1972 the date when the first shad was captured varied by only five days, the earliest capture being on April 7 and the latest on April 12. By the last week of April large numbers of shad were entering the river to spawn, and this movement continued until the middle of June. Shad tagged and released in the river in April and May, while they were moving upriver to spawn, consistently reappeared in coastal waters off Rhode Island and Massachusetts in the last days of July and the first part of August. By late August they were in the Gulf of Maine and the Bay of Fundy, where they remained until late September or early October. Annually there were no further recoveries until the second week of March, when tagged shad began appearing along the coast of North Carolina. By the first week of April they had moved north to the vicinity of Virginia and Chesapeake Bay, and by the second week of April they were off Delaware and New Jersey. Each year the timing of recoveries from specific locations was highly consistent, suggesting that some factor exerted a strong influence on the marine movements of the shad.

The first clue that this factor was water temperature came not from marine investigations but from observations of the movements of shad in fresh water. Although the idea that temperature was a factor affecting the timing of the spawning migrations was far from new, firm evidence in support of the idea had been lacking. I was able to compare records of the weekly catch of shad with the mean weekly temperature of the Connecticut River over a period of 11 years. The comparison showed that few, if any, shad entered the river when the temperature of the water was below four degrees Celsius; above that temperature the number of shad increased steadily until the water was at about 13 degrees C. At higher temperatures the catch declined.

Further evidence was provided by records kept at a fish lift associated with a hydroelectric dam at Holyoke, Mass. The records showed how many shad were lifted over the dam each day and what the water temperature was. During 15 years the temperature at which the peak movement occurred was in a range of five degrees C., from 16.5 to 21.5 degrees. (The temperature is higher than it is during the peak migration at the mouth of the river because Holyoke is some 86 miles upriver and thus receives the main body of shad almost four weeks later.)



SEASONAL TEMPERATURES of the water along the Atlantic Coast of the U.S. are portrayed by means of isotherms, which show where a given temperature is found at a given time. The upper three maps are for January, March and May respectively and the lower ones are for July, August and October. Normally the seasonal distribution of the shad is closely associated with water temperature in the range from 13 to 18 degrees Celsius. The shad move northward in the spring and southward in the autumn within this range.

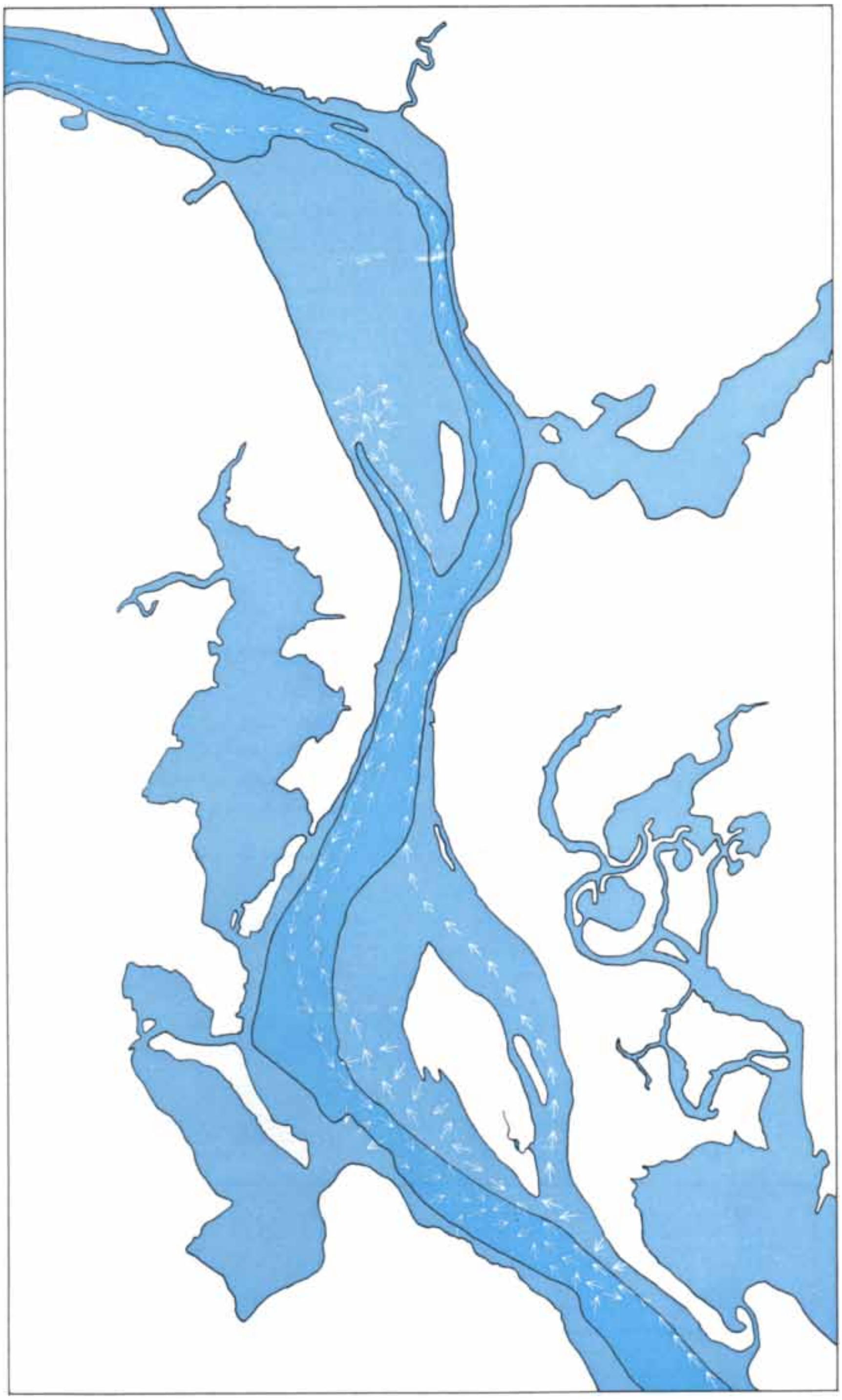
Although the data from the Connecticut River were strongly suggestive, I needed information from other rivers in order to make generalizations about the influence of temperature on the movements of shad. Accordingly I began in 1967 to assemble information from the York River of Virginia and the St. Johns River of Florida. Work had been done in the York River earlier by William H. Massmann and Anthony L. Pacheco, who at the time were associated with the Virginia Fisheries Laboratory. The data from the York River revealed a condition essentially identical with that in the Connecticut River.

In the St. Johns River the correlation between temperature and the timing of the migration was close but negative, that is, the shad moved in when the river cooled to their preferred range rather than when it warmed up. In the St. Johns River the water temperature seldom goes below 13 degrees C. even in the winter, and it is usually above 20 degrees C. from March through November. Few shad entered the river when the water temperature was above 20. As it fell below this temperature in November and December the number of shad increased. The peak movement occurred at the seasonal low of 13 to 16 degrees C., which was the same as the level that coincided with the peak movements in the rivers to the north.

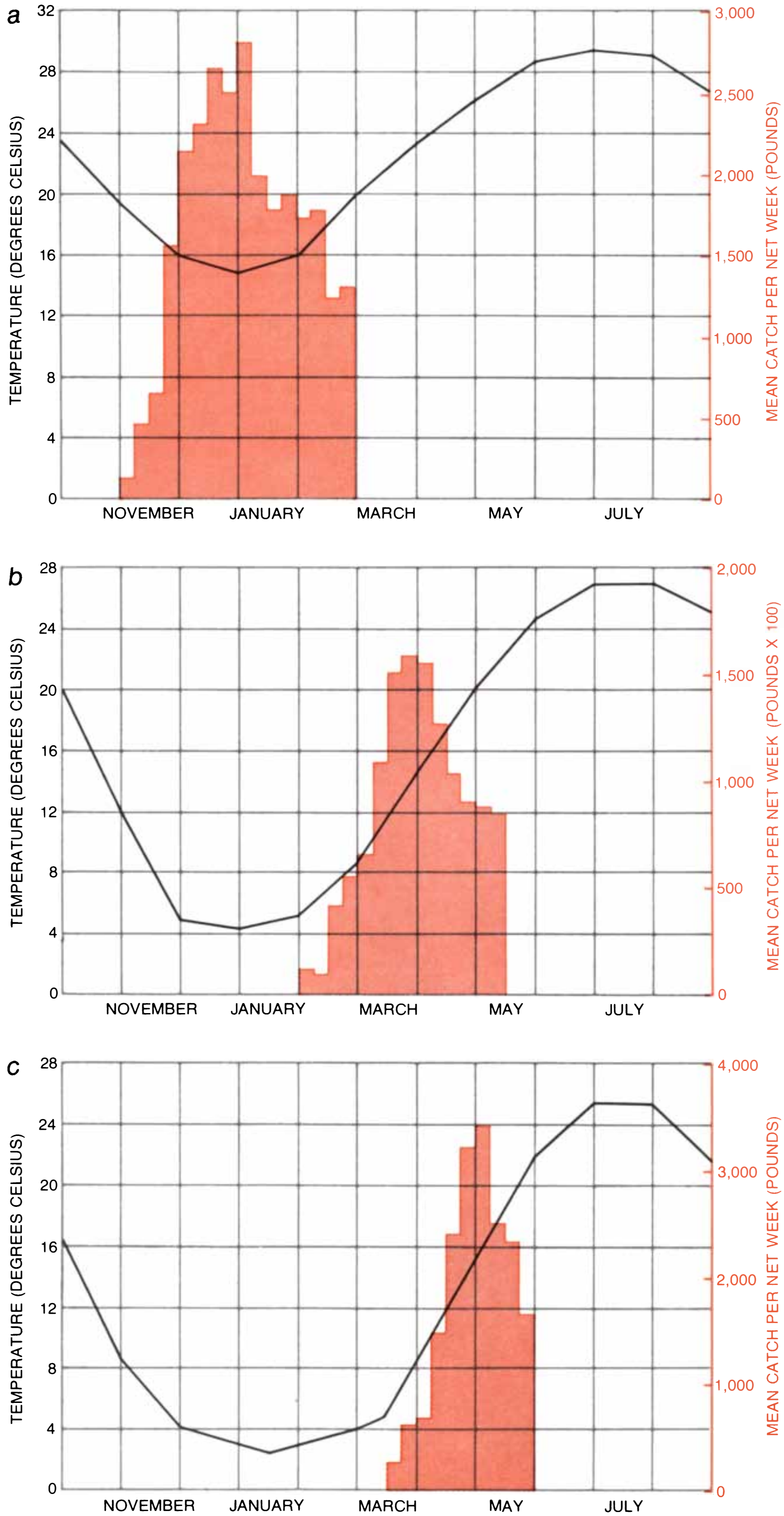
These findings prompted my colleague Richard R. Whitney of the University of Washington to conduct a similar study on the Pacific Coast. He found that in 78 percent of the years from 1938 to 1969 the peak movement of shad at the Bonneville Dam on the Columbia River occurred in the temperature range from 16.5 to 19 degrees C. In 26 of the 32 years the temperature range within which 90 percent of the shad appeared at the Bonneville fish ladders varied by no more than four degrees. (Shad from Eastern rivers were stocked in the Sacramento River in 1871; by 1880 they ranged widely along the Pacific Coast.)

Having established conclusively that water temperature had a strong influence on the timing of the spawning runs of shad, Whitney and I turned to the question of whether or not temperature affected their movements at sea. We plotted the coastal recoveries from the Connecticut River taggings against the seasonal position of the portion of coastal waters where the temperature was between 13 and 18 degrees C. The correlation was again close. It is particularly noteworthy that during August and September the only area on the Atlantic Coast of North America having the preferred water temperature is the Gulf of Maine, which is precisely where the shad are at that time of the year. Moreover, the northern and southern limits of the range of the species in coastal water correspond closely to the range of acceptable temperature both in the rivers and in the ocean along the Atlantic Coast. It seems clear that the preference of the shad for a certain range of temperature leads to an orderly and precisely timed movement of the entire Atlantic Coast population southward in the fall and winter and northward in the spring and summer as the coastal water temperatures change with the seasons.

Examining the significance of this



MEANDERING BEHAVIOR of shad at the interface of salt water and fresh water is typified by a single fish that was tracked in the Connecticut River by means of an ultrasonic transmitter that had been attached to the fish. Area is about six miles above river mouth. Shad on a run to spawn apparently meander while making physiological adjustments to fresh water.



SHAD CATCH, represented by the colored bars, varies seasonally from river to river but in each case reaches a peak when the water temperature (*black curve*) is around 15 degrees C. In the St. Johns River of Florida (*a*) the peak catch is in January, when the river has cooled enough for the shad to enter for their spawning run. In the York River of Virginia (*b*) and the Connecticut River (*c*) the shad enter when the river has warmed up enough.

temperature dependence, one is led to the view that it bears on the conditions needed to ensure maximum survival of eggs and larvae. In 1924 the Canadian biologist A. H. Leim published the results of studies showing that a minimum water temperature of 12 degrees C. was required for spawning but that optimum survival of shad eggs occurred at approximately 17 degrees. Subsequent experimental studies have shown that hatching and survival of shad are at a maximum when the water temperature is between 15.5 and 26.5 degrees C. The temperature-regulated oceanic migrations of shad bring individual populations to their home rivers at times when the river temperatures are approaching the level that is best for spawning.

Water temperature appears also to affect the survival rate of adults. Studying returns from my tagging program in the Connecticut River, I found that shad entering the river to spawn early in the season, when temperatures were low, had a significantly higher rate of return to the river in subsequent years than shad tagged at higher temperatures. This finding seems to be related to the observation that in general the shad populations from rivers south of Chesapeake Bay lack repeat spawners and that north of the bay the proportion of repeat spawners increases with latitude.

To understand this phenomenon one must realize that shad do not eat while they are in fresh water. All the energy for the spawning run comes from stored reserves in the body. As a result a shad loses some 40 to 50 percent of its weight during its brief residence in fresh water. Since in shad as in most other cold-blooded animals the metabolic rate increases as the environmental temperature increases, shad migrating at 18 degrees C. would use more stored energy than shad migrating at eight degrees C. It seems likely that the additional energy required for migrating at higher temperatures is one of the reasons that shad spawning in southerly rivers do not survive to return another year. Brian Glebe of my laboratory at McGill University is investigating this aspect of the migrations of shad.

The absence of repeat spawning in the shad populations of Southern rivers would result in marked reductions in reproductive potential if reproductive adaptations were not developed to compensate for their loss. My studies of the reproductive ecology of the species indicate that Southern shad populations have adapted by earlier maturity and greater fecundity. In the St. Johns River the average age at maturity for female shad is 4.3 years and the average fecundity of females spawning for the first time is 412,000 eggs. In the Connecticut River the corresponding figures are 4.8 years and 263,000 eggs. Moreover, the variations in maturity and fecundity are gradual and in direct relation to corresponding gradual increases in the frequency of repeat spawning with latitude. The net effect of these changes in reproductive characteristics is to maintain a similar lifetime production of eggs in all the populations of the Atlantic Coast.

Our interest in the shad has not been limited to defining the path of their migrations and determining the environmental factors that govern the timing of their marine and freshwater movements. Since 1967 we have also been studying the migratory behavior of shad during the final stages of their approach to their home river from the sea and during their progress upriver. In these investigations we have employed ultrasonic tracking, which is a relatively new technique in fisheries research whereby the movements of individual fish can be followed from boats and shore stations by means of small ultrasonic transmitters attached to the fish.

Our work in Long Island Sound, which is under the direction of one of my graduate students, Julian J. Dodson, is aimed at determining how shad recognize that they are in the area of their home river and how they locate the river mouth. We have now tracked some 60 shad during their approach to the river from the sea; most of them were followed continuously for at least 24 hours. The observed movements of the fish were compared with data collected simultaneously on the temperature and depth of the water, the intensity of the light, wind and wave action, salinity, tidal currents and the location of the sun. So far only salinity, temperature and tidal currents appear to play a major part in directing the movements of shad in the final stages of the marine migration.

Shad enter Long Island Sound from the east, so that in order to reach the Connecticut River they must progress some distance westward. In doing so they exhibit a complex behavior. The open-water orientation of shad appears to be guided mainly by tidal currents: the fish tend to orient into the current. Since the current in Long Island Sound reverses direction approximately every six hours because of the tide, the shad change their orientation by about 180 degrees every six hours too. We observed that they swim faster when oriented toward the west than when oriented to the east. Facing east against an incoming tide they swam at a speed about equaling the speed of the current and so made little net headway; facing west against an outgoing tide they exceeded the speed of the current and moved west.

It is noteworthy that the orientation was as precise at night as it was during the day, whereas the adjustment of speed was less precise at night. This observation suggests the existence of a compass sense to aid in orientation and the need for visual contact with the bottom or some other reference point in order to adjust swimming speed to the velocity of the current.

Reduced salinity at the surface and increased water temperature are among the more obvious indicators of the presence of water from the Connecticut River in Long Island Sound. Reduced salinity can occasionally be detected well east of the Connecticut River near the approaches to Long Island Sound from the Atlantic Ocean. Tracked shad were observed to respond to both indicators by swimming faster. It may be that the shad are actually responding to chemical substances characteristic of Connecticut River water. Indeed, this is the more likely hypothesis, inasmuch as changes in temperature and salinity alone would not be adequate clues for recognizing a specific river. Perhaps the detection of these identifying clues during the northward migration in the spring triggers the appropriate behavior pattern leading to the locating of the home river.

We have also carried out experiments in which we blinded some shad, blocked the olfactory capsule of others and performed both operations on still others. All three groups exhibited less ability in orientation than the unimpaired shad, the anosmic group being most seriously affected. Both the blind fish and the anosmic ones exhibited a reduced ability to find the river, and none of the shad that were both blind and anosmic homed on the river. The experiments indicate that the visual and olfactory systems figure importantly in the ability of the shad to orient and to home.

Robert A. Jones of the Connecticut Department of Environmental Protection and I have collaborated closely in the studies of the behavior of shad during the river stage of the spawning migration. We have now followed the movements of some 230 individual shad in fresh water for periods of up to 100 hours. One phenomenon we observed was a distinct change in the behavior of the shad at the interface of salt water and fresh water. Because the Connecticut River has a narrow mouth and a large discharge, the salty part of the river is a well-defined wedge that extends only a few miles upstream. The upstream movements of shad in the saltwater section were quite direct. Near the upper edge of the saltwater intrusion, however, the movements of the fish were characterized by extensive meandering that continued for a day or two. During that time the fish meandered up and down the river as the wedge of salt water advanced and retreated with the tides, and their average swimming speed was about half what it was both before and after the period of meandering. At the end of this period they resumed their steady progress upriver, swimming now in water that was entirely fresh. We believe the meandering behavior is associated with physiological changes involved in the transition from salt water to fresh water. This hypothesis is supported by our finding that shad experience considerable osmotic stress when they are transferred quickly from salt water to fresh water.

The steady upriver migration of shad in fresh water reveals still another behavioral change. The Connecticut River is tidal in its lower 45 miles; as a result the direction of flow can reverse completely as much as 25 miles inland. Once the shad are in the river they abandon the strict countercurrent behavior characteristic of their marine movements and proceed upriver toward the spawning areas without regard to the direction of flow.

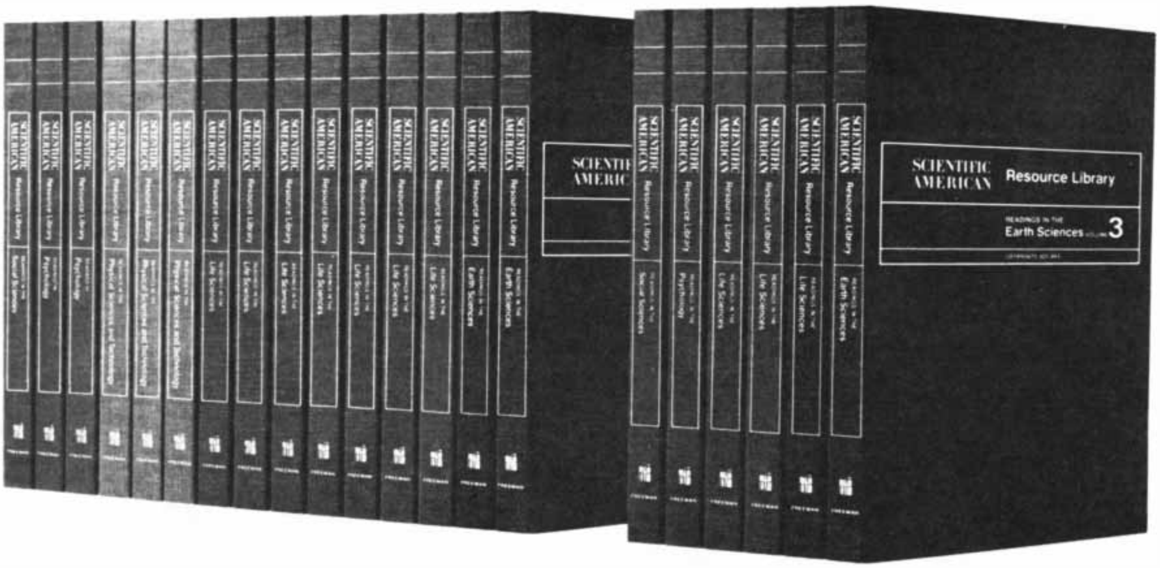
In fresh water the shad follow the natural and dredged channel of the river, making only occasional movements into shallow water. This orientation to channels results in remarkably consistent migration routes in any given area of the river. How the shad establish the location of the channel is unclear. The speed of the current is slightly higher in the channel than in the shallow areas, and it may be that the difference provides an orienting clue.

One of the major obstacles the shad confront as they move upriver is the series of nets (about 40 of them) put across the stream by commercial fishermen seeking shad. Our tracking studies showed that the shad have an impressive ability to detect and avoid the nets. Typically the shad would move close to a net before sensing its presence and then would swim parallel to it until they reached the end, where they would turn upstream again. This behavior was observed in both daylight and darkness, suggesting that senses other than vision may be involved.

So far we have not observed any effect of the "plume" of heated water from the atomic power plant on the shad we have tracked as they moved through that section of the river. Perhaps the reason is that at Haddam Neck the channel runs along the west bank of the river, whereas the thermal discharge is on the east bank, so that the shad, moving as usual in the channel, are about as far away from the plume as they could possibly be. Even if any are near the plume, lack of knowledge about the depth at which shad swim leaves us uncertain about whether they swim under the plume or go through it without being affected. We intend to pursue the matter with equipment that shows how deep the fish is swimming and what the temperature of the water is at that point. In view of our finding that shad have a narrow range of preference for temperature, it is important to ascertain whether or not a permanent increase in the temperature of a river or a section of it resulting from the thermal plume would affect the shad run.

As is often the case in investigations of animal behavior, our findings have raised as many challenging questions as they have answered. Much remains to be learned about the path shad take during their southward migration in the fall and winter and of the conditions they encounter. We are also interested in knowing whether temperature is the only controlling factor in the timing of the ocean migrations or whether other factors such as food supplies that may be governed by seasonal cycles of temperature are also involved. Much remains to be learned about the sensory systems involved in orientation and homing and about the nature of the environmental clues that guide the ocean and river migrations. We must also gain a better understanding of the effects of man-made changes in the physical and chemical characteristics of rivers on the migratory and homing behavior of the shad. Fortunately recent advances in technology are providing new and powerful tools with which we can hope to attack these questions fruitfully.

GARUDA is a part-bird, part-human deity who is worshiped by Hindus and Buddhists, as well as by the followers of Bon, the pre-Buddhist religion of Tibet.



Announcing the first expansion of the SCIENTIFIC AMERICAN Resource Library

Six new volumes have just been added to what *The New York Times Book Review* has called "The Intelligent Layman's Guide to the Golden Age of Science."* They are:

READINGS IN THE EARTH SCIENCES, Volume 3
Offprints 875-899, $10.00

READINGS IN THE LIFE SCIENCES, Volumes 8, 9, and 10
Offprints 1121-1247, $30.00

READINGS IN PSYCHOLOGY, Volume 3
Offprints 516-542, $10.00

READINGS IN THE SOCIAL SCIENCES, Volume 2
Offprints 638-675, $10.00

The preexisting volumes are:

READINGS IN THE EARTH SCIENCES
With an Introduction by
William H. Matthews III, Lamar State College of Technology
Two volumes, Offprints 801-874, $20.00

READINGS IN THE LIFE SCIENCES
With an Introduction by Marston Bates, the University of Michigan
Seven volumes, Offprints 1-199; 1001-1120, $70.00

READINGS IN THE PHYSICAL SCIENCES AND TECHNOLOGY
With an Introduction by
Isaac Asimov, Boston University School of Medicine
Three volumes, Offprints 201-326, $30.00

READINGS IN PSYCHOLOGY
With an Introduction by Harry F. Harlow, the University of Wisconsin
Two volumes, Offprints 401-515, $20.00

READINGS IN THE SOCIAL SCIENCES
With an Introduction by Harry F. Harlow, the University of Wisconsin
One volume, Offprints 601-637, $10.00

COMPLETE LIBRARY
Twenty-one volumes in all, $210.00 the set

*"Their basic contents are hundreds of articles which explain simply, vividly, and interestingly what modern science is all about . . . most of the vast riches that have appeared monthly since the late 1940's in the *Scientific American,* this country's and perhaps the world's outstanding forum for communication between scientists and the intelligent public." —*The New York Times Book Review,* April 26, 1970

W. H. FREEMAN AND COMPANY
660 Market Street, San Francisco, CA 94104
58 Kings Road, Reading, England RG 1 3AA

The Origins of Number Concepts

Experiments with children indicate that they first become aware of numbers in terms of ordered sequences and only later in terms of quantities. Such findings may give rise to a newer "new math"

by Charles J. Brainerd

How do we first learn concepts of number? The question has been a topic of debate among philosophers and mathematicians at least since the time of Pythagoras in the sixth century B.C. Today it continues to be important not only to scholars but also to society as a whole; witness the triumph of the "new math" in the elementary schools of America during the past decade. The fact remains that comprehensive studies of the origin of number concepts in young children are relatively rare. I have recently participated in investigations at the University of Alberta that have developed some new information on the subject. Before I describe these studies it will be useful to review the main theories of the origin of number concepts.

Pythagoras believed that what we call the positive integers or natural numbers (1, 2, 3 and so on) were god-given entities that formed the ultimate foundation both of mathematics and of the universe. The Pythagoreans' own discovery of such "incommensurable" quantities as the ratio between the diameter of a circle and its circumference ultimately dispelled the belief that the universe was built on natural numbers. That the natural numbers provided the foundation of mathematics, however, persisted as an article of faith among mathematicians until well into the 19th century.

By then the attitude toward the natural numbers had begun to change. The centrality of natural numbers was no longer considered an accepted fact but was viewed as a conjecture that required rigorous proof. The proofs usually took the form of a stepwise derivation of such well-known number systems as rational, real and complex numbers from the natural numbers themselves. Two examples are the attempts of Karl Weierstrass and Richard Dedekind to "arithmetize" mathematical analysis. Both scholars derived real numbers—the combined set of all rational and irrational numbers that is employed in most classical mathematics—from the natural numbers. A third example is the proposal of Leopold Kronecker to found all mathematics on the natural numbers. This Kronecker attempted to accomplish solely with "finitary" methods, that is, methods invoking neither nonfinite entities nor proofs involving more than a finite number of steps.

Still other mathematicians, in particular those who were conversant with contemporary advances in symbolic logic, put forward the suggestion that, far from being god-given, natural numbers were constructions of the human mind. The three most famous propagators of this suggestion were Gottlob Frege, Giuseppe Peano and Bertrand Russell. Obviously a theory was needed that would trace the rise of the natural numbers from some more basic notion or notions, but how was such a theory to be constructed? If most or all of classical mathematics had evolved from the natural numbers, it was improbable that the required theory could be devised entirely within the bounds of classical mathematics.

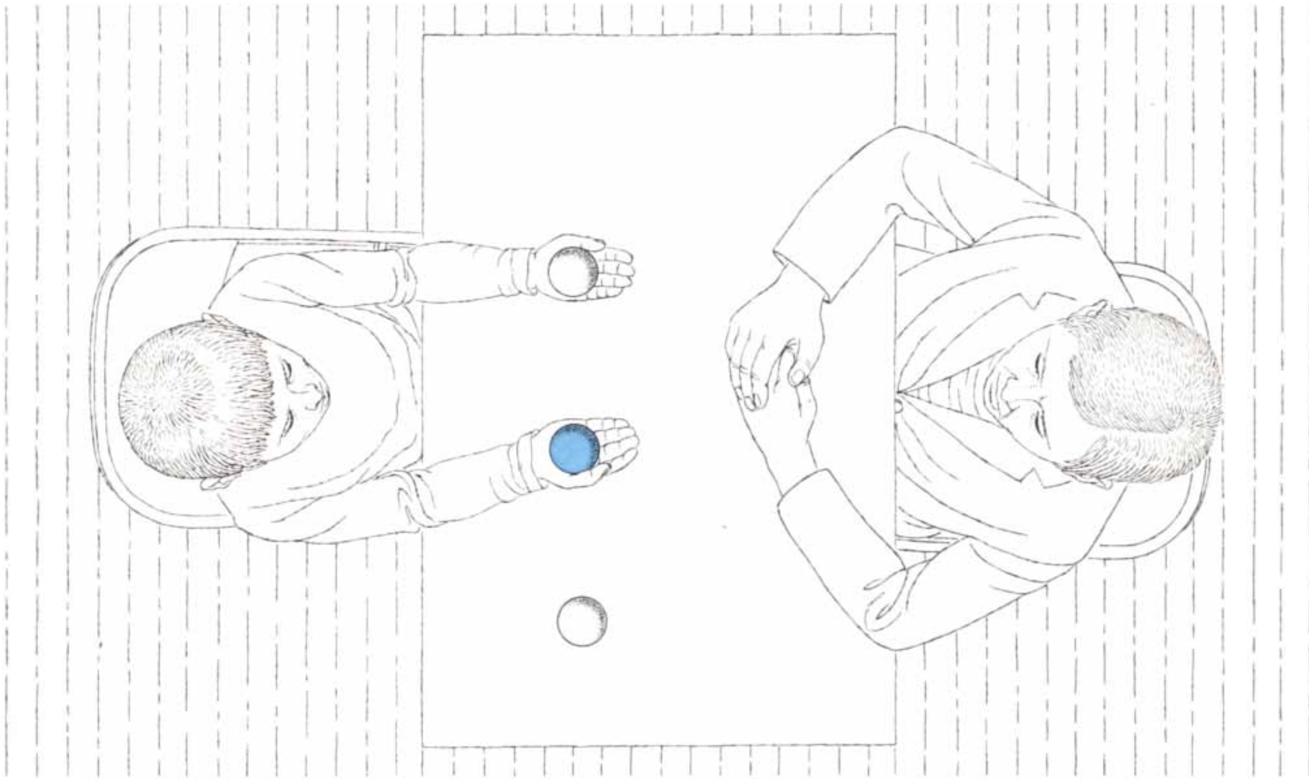
First Frege, then Peano and finally Russell turned to symbolic logic as a potential source of the fundamental no-

CARDINATION

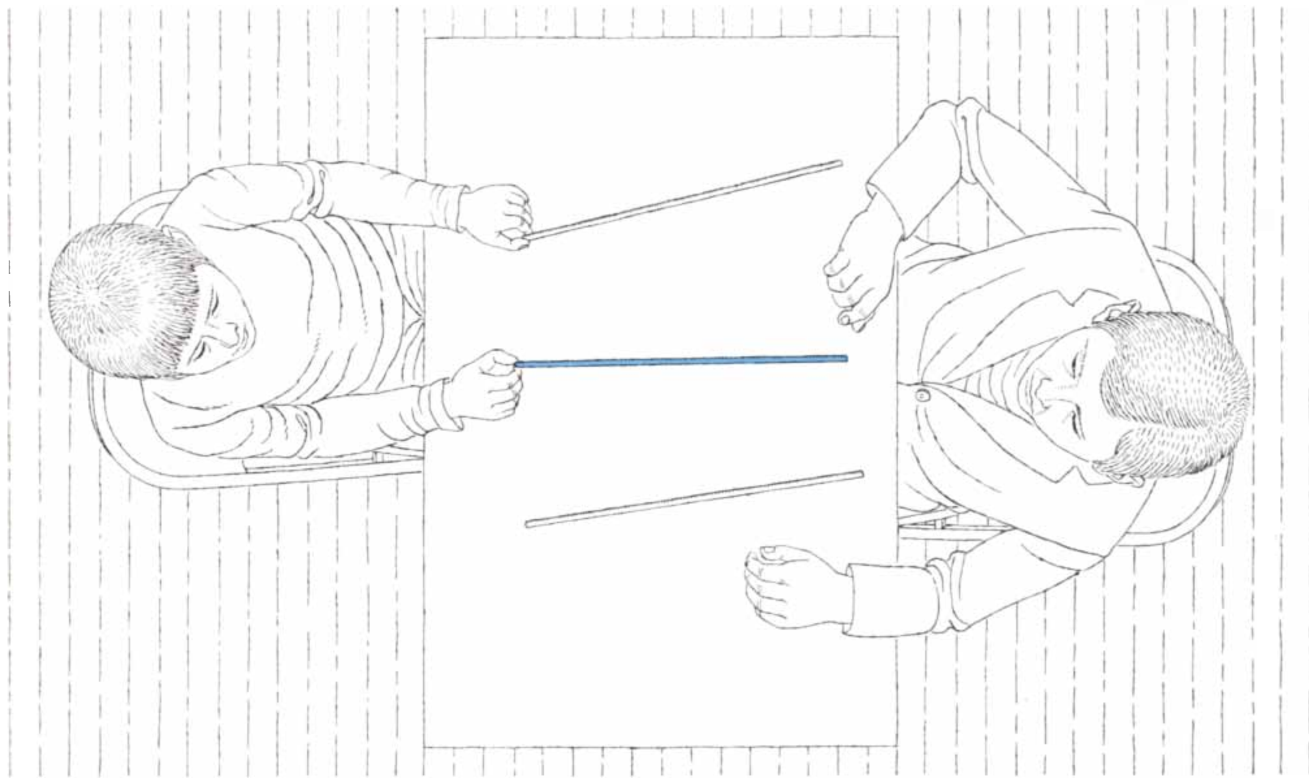
CARDINATION determines the number of objects in a set by matching rather than by counting. In this example the objects correspond in number to the five fingers of a hand; thus the set's cardinal number is 5.

ORDINATION

ORDINATION determines the number of objects in a set by counting "One, two, three" and so on rather than by matching. In this set, as in all finite sets, the last ordinal number, 5, is the cardinal number of the set.



CONCEPT OF ORDINATION in terms of "heavier than/lighter than" was tested by asking the subjects to lift clay balls that were identical in size but different in weight. In some tests the heaviest ball was at the right end of the row and the lightest at the left; in other tests the order was reversed. The comparisons in all the tests were with a third ball (*middle*), which was intermediate in weight.



SECOND ORDINATION TEST, involving the relation "longer than/shorter than," presented the subjects with three lengths of wooden dowel; the longest stick was half a centimeter longer and the shortest stick half a centimeter shorter than the third stick, which was always placed between the other two. As with the weight test, the order of presentation was randomly left to right or right to left. Subjects compared long and short sticks with intermediate stick (*color*) but never directly compared long stick with short.

tions necessary for a theory of natural number. Frege was the first of the three to publish a specific theory. In *Die Grundlagen der Arithmetik,* which appeared in 1884, he proposed that the natural numbers could be reduced to the notion of "class" and the operation of "correspondence," by virtue of which classes are quantified. According to Frege, each natural number n was to be regarded as a "superordinate class" whose members, "subordinate classes," each contain precisely n elements. Given two subordinate classes, A and B, the two are said to be members of the same superordinate class, that is, instances of the same number, if and only if a one-to-one correspondence can be established between their respective elements. If instead the correspondence is many to one, then A and B are said to be instances of different numbers.

In essence Frege's theory states that the series of natural numbers presents a general problem of quantification, but that the general problem can be reduced to the more restricted notion of "cardination," or quantifying classes. The commonest example of cardination is the matching of things.

Frege's cardinal theory remained generally unknown until Russell rediscovered it in 1901. Russell subsequently published the cardinal theory, with full acknowledgement to Frege, in his *Principles of Mathematics* (1903), in his joint work with Alfred North Whitehead, *Principia Mathematica* (1910–1913), and in his *Introduction to Mathematical Philosophy* (1919).

Between the time Frege first published the cardinal theory and the time Russell rediscovered it Peano developed a second theory about the natural numbers. This theory first appeared in his work *Formulaire de Mathématiques* (1894) in the form of five axioms that I shall slightly reword here. First, 1 is a natural number. Second, any number that is the successor of a natural number is itself a natural number. Third, no two natural numbers have the same successor. Fourth, the natural number 1 is not the successor of any natural number. Fifth, if a series of natural numbers includes both the number 1 and the successor of every natural number, then the series contains all natural numbers.

In essence Peano's theory places the natural numbers in an ordinal relation or, in the language of symbolic logic, a "transitive, asymmetrical relation." If we are willing to stipulate that the relation R that obtains between every nonidentical pair of natural numbers be an ordinal relation, then the complete series of natural numbers can be constructed stepwise with the aid of the rule of mathematical induction. Like Frege's cardinal theory, Peano's states that the series of natural numbers presents a general problem of quantification. Unlike Frege's theory, however, Peano's ordinal theory reduces the general problem to the more restricted notion of quantifying transitive, asymmetrical relations, or ordination. The commonest example of ordination is the counting of things.

Just which of the two theories, the cardinal or the ordinal, is mathematically preferable is a question that has never been answered to everyone's satisfaction. Reasonable objections can be lodged against both. For example, the cardinal theory is subject to the celebrated paradox, discovered by Russell in 1901, concerning the class composed of all those classes that are not members of themselves. Briefly, if Y is the class of all those classes that are not members of themselves, then it can be shown that Y both is and is not a member of itself.

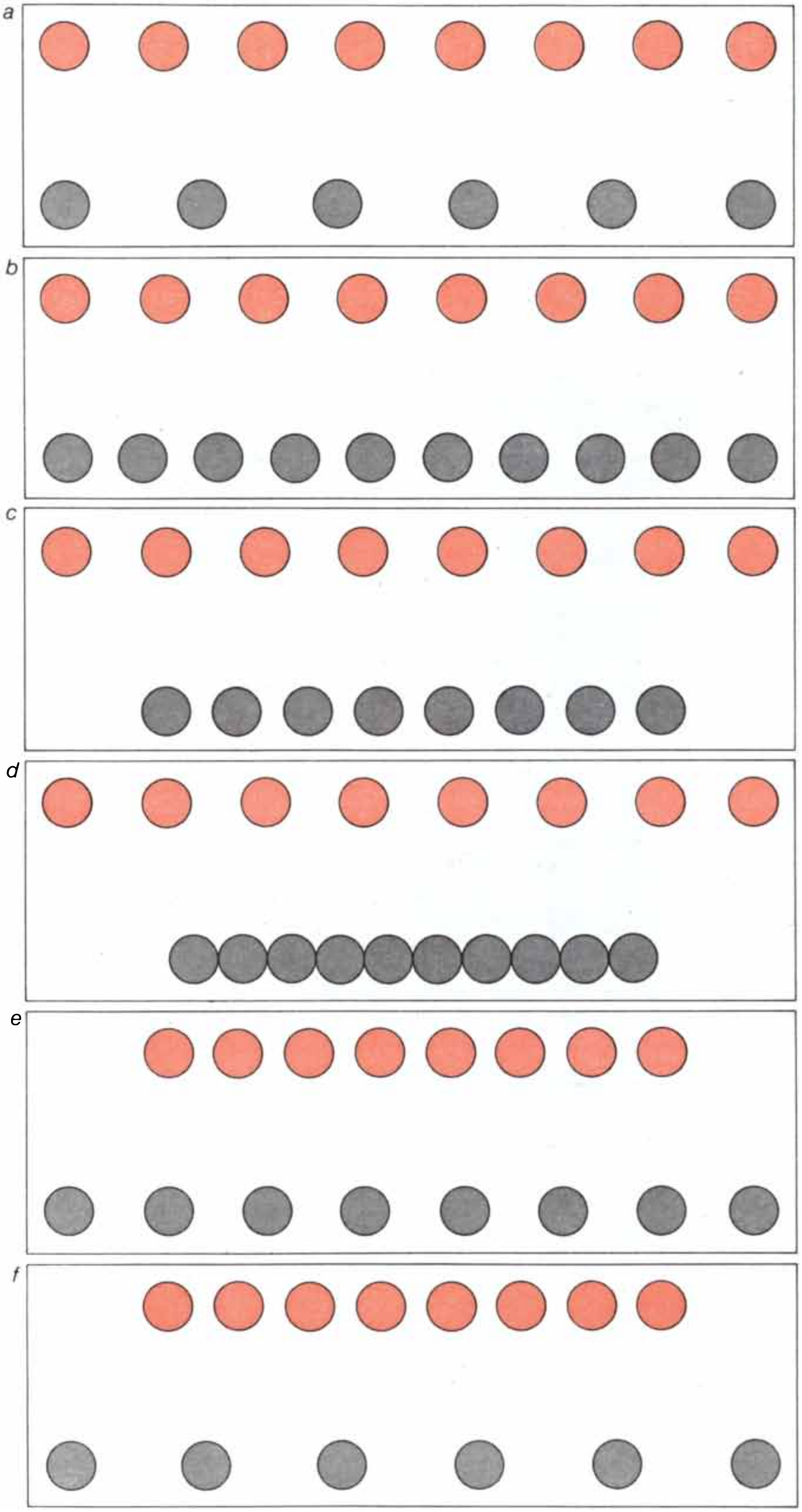
With respect to the ordinal theory, as Russell pointed out, whereas Peano's five axioms obviously are satisfied by the series of natural numbers, they are equally satisfied by other number systems. For example, the rational fractions (1, 1/2, 1/3, 1/4 and so on) satisfy the axioms, as will any series of mathematical or empirical entities that has a beginning, no repetitions and no end and is such that every entity can be reached in a finite number of steps. In short, the domain of application of the ordinal theory is much wider than the series of natural numbers.

Because there is no universally accepted mathematical basis for choosing between the cardinal and the ordinal theories the choice becomes a subjective matter. Typically the choice is determined by one's degree of sympathy with one or another of three modern schools of mathematical thought: logicism, formalism and intuitionism. Those who lean toward logicism favor the cardinal theory, a choice that is natural enough when one considers that the codiscoverers of the theory, Frege and Russell, were the principal founders of logicism. Those whose sympathies are with formalism lean toward the ordinal theory; the fact that Peano's axioms seem to denude the number concept of innate "meaning" probably explains this preference. As for intuitionists, they have in effect returned to the Pythagorean position that the natural numbers must be accepted without further analysis as the foundation of mathematics. They deny that the natural numbers are the invention of mathematical minds and offer a "psychological" thesis: The series of natural numbers is an innate intuition, present at birth in all normal members of the human species.

Nonmathematical scholars tend to view with profound indifference the tortures that mathematicians suffer over such basic issues as the nature of number. They have learned from centuries of hard experience that the mere fact that the foundations of some mathematical system or concept are not secure need not deter them from employing the system in their work. On the contrary, mathematical notions whose foundations have been matters of continuous debate have often yielded the most mileage; the notion of an infinitesimal is perhaps the best-known example.

An infinitesimal can be defined as a quantity that, although it is infinitely small, is nevertheless greater than zero. The calculus, as it was originally formulated by Newton and Leibniz, involved infinitesimals; Newton called them fluxions. Intense disagreement followed over the question of whether such a curious item as an infinitesimal could be viewed as a legitimate mathematical concept. By the latter half of the 19th century the debate seemed to have been finally silenced by Weierstrass' success in reformulating the calculus without infinitesimals. Within the past decade, however, Abraham Robinson's work on nonstandard analysis has resurrected the infinitesimal [see "Nonstandard Analysis," by Martin Davis and Reuben Hersh; SCIENTIFIC AMERICAN, June, 1972]. Throughout this long period the calculus of infinitesimals has been regularly applied in the physical and biological sciences without regard to its foggy mathematical status and without damage to the disciplines concerned.

Just because neither the sciences nor society has suffered much as a result of scholarly indifference toward (more precisely ignorance of) the debate over infinitesimals, it should not be assumed that the debate over the origins of the number concept can be ignored with equal safety. Unlike the infinitesimal, number is not the exclusive property, or even largely the personal property, of the mathematician. Number has been a concept of social importance since the dawn of recorded history. The significance to society of number and number-related skills has increased tremendous-



CONCEPT OF CARDINATION was tested by asking the subjects to decide, without counting, which of six matched sets of dots (*a–f*) were in one-to-one correspondence and which in many-to-one correspondence. The top and bottom row of dots in each set differed in color. If both rows contained the same number of dots, the dots were spaced so that the rows were not of equal length. Similarly, when one row contained more or fewer dots than the other, dots were spaced so that rows were of equal length or the row with fewer dots was longer.

ly with the rise of industrial civilization.

In most Western nations today children receive considerable exposure to number concepts soon after they begin to speak. More or less haphazard at first, the exposure becomes simultaneously more intense and more systematic with the onset of formal education. During the first few years of elementary school roughly 50 percent of the curriculum is normally given over to inculcating the natural numbers and methods of manipulating them. The child learns to add, subtract, multiply and divide the natural numbers and finally to derive other number systems from them. We expect our children to possess by pubescence a numerical competence much higher than that of an educated Greek or Roman adult of two millenniums ago. We have even gone so far as to develop labels that imply mental turpitude on the part of those otherwise normal children who fail to attain the standards of numerical competence that we deem desirable, for example "learning disability" and "underachievement."

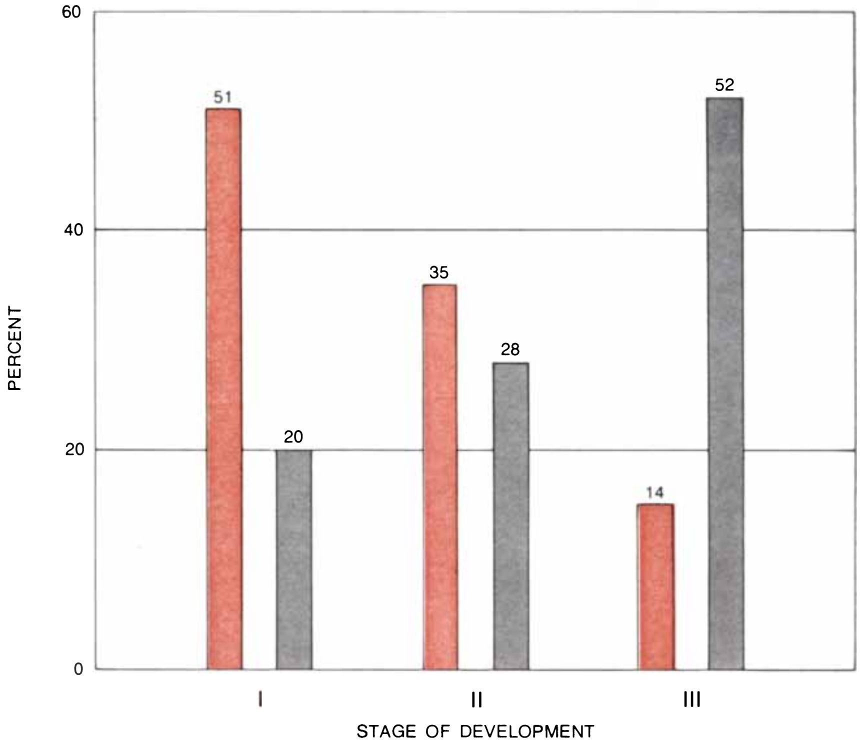
Considering the emphasis our society puts on numerical competence, it might be expected that extensive information would long ago have been collected on how the concept of number naturally unfolds in the course of a child's mental growth. The truth is that we possess very few hard facts about the emergence of numerical ideas in children's thinking. It is the paucity of such data that gives social relevance to the mathematical debate on the origins of the number concept I have just outlined. In what follows I shall describe some recent studies in the area of developmental psychology that were designed to investigate the relation, if any, between the actual emergence of numerical ideas in young children and the hypothetical patterns of emergence suggested by the two principal theories of the origin of number concepts.

When the mathematician says that some concept Y (for example natural number) is "founded on" or "based on" some other concept X (for example cardination or ordination), he means that X is more fundamental in a logical sense than Y. When the developmental psychologist makes the same statement, he means that X plays a necessary part in the emergence of Y in the thinking of the growing child. From the psychologist's viewpoint such a statement leads to two major predictions that can be tested experimentally. The first prediction is that X will invariably appear earlier than Y in the course of the child's mental development. The second is that improvements in Y will be a consequence of improvements in X.

Different research strategies are used to examine these two predictions. The "X before Y" prediction calls for normative, or average-performance, studies. The normative study consists of four steps. First, tests are devised that measure the presence of the concepts in question in the child's thinking. Then the children to be tested are drawn from the range of ages when the concepts in question are normally acquired. The third step is testing for the presence or absence of the concepts. The fourth is comparison of the test results to determine the presence or absence of interconcept correlations of the predicted "X before Y" kind.

The "better X means better Y" prediction is tested by what is known as a transfer experiment, a method that is slightly more complicated than the normative method. Seven steps are involved; the first three are the same as the first three steps in a normative study. Taken together, they constitute the "pretest" phase of the transfer experiment. The information collected in the pretest phase is then used to divide the subjects into experimental groups and control groups. In the fifth step the experimental groups receive training designed to improve their performance in one or another of the concepts in question. After that the tests administered in the pretest phase are given to the experimental and control groups. The seventh and final step compares the test results both to determine the degree of improvement that resulted from training and to find whether or not training in one specific concept tends to "transfer" and improve performance with respect to another concept for which no training was provided.

Both normative and transfer experiments appear to be simple, but a major pitfall awaits the unwary investigator. In constructing the tests whereby the subjects' grasp of various concepts will be assessed, one must guard against loosely defined methods of assessment that bear only a vague or intuitive resemblance to the concepts that are to be measured. For example, in the present instance it was essential that the tests concerning the concepts of cardination and ordination be explicitly deduced from the mathematical conceptions of these notions. If that had not been done, the findings of the experiments could only have been suggestive rather than conclusive. Although the pitfall may seem obvious, it was not avoided in most earlier investigations of the growth of children's number concepts, not excluding Jean Piaget's historic studies in the late 1930's and early 1940's.

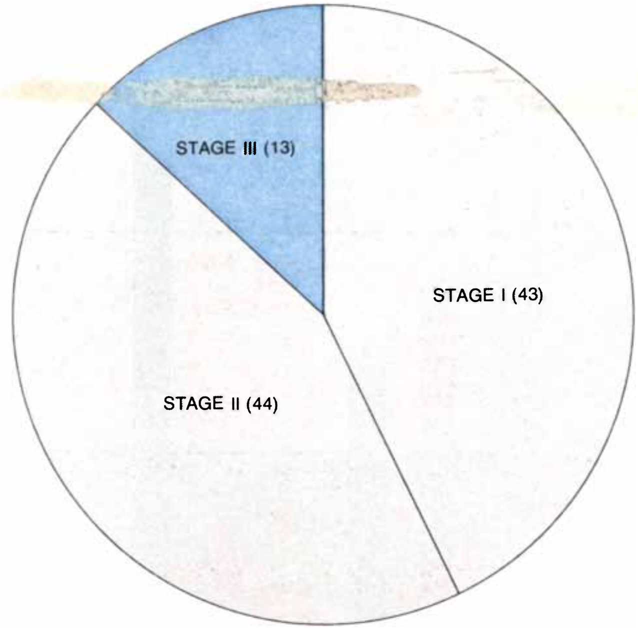


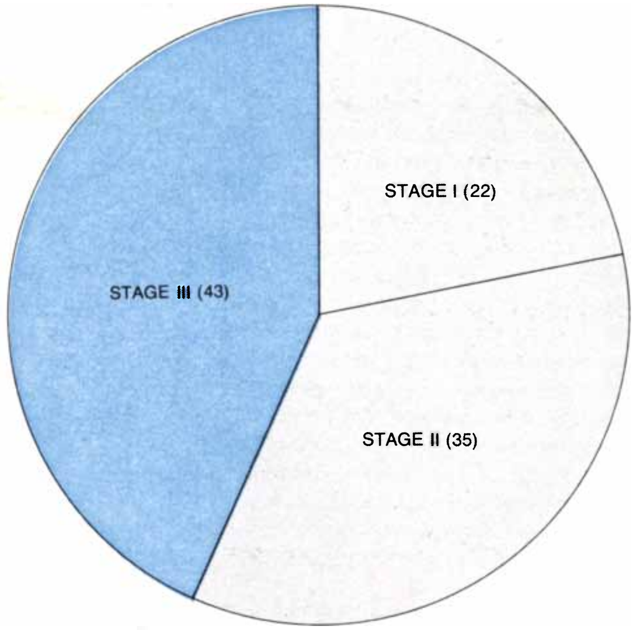
GRASP OF CONCEPTS, as revealed by the test scores, placed each of 180 subjects in one of three stages of development. The children at Stage I were totally incapable of either cardination (*color*) or ordination (*gray*). Those at Stage II were capable of many-to-one cardination (*color*) and of left-to-right ordination (*gray*) but failed to grasp one-to-one and right-to-left relations. Children at Stage III grasped cardination (*color*) and ordination (*gray*) fully. As the percent of children at each of the stages indicates, the concept of ordination appeared in the children's thinking long before the concept of cardination.

At the University of Alberta we recently undertook two large-scale studies of the growth of three arithmetic elements in the thinking of young children: the concept of cardination (the "set-matching" of Frege and Russell), the concept of ordination (the sequential quantification of Peano) and the children's competence in manipulating natural numbers. The studies were designed to answer three related questions of the "X before Y" kind: Which of the two comes first, cardination or ordination? Which comes first, ordination or natural-number competence? Which comes first, cardination or natural-number competence?

The first of our studies was concerned exclusively with the question of whether cardination or ordination came first. The subjects were 180 Canadian children between five and seven, the range of ages when the average child first acquires the concepts of cardination and ordination and develops some competence in manipulating natural numbers. The tests for ordination assessed the children's capacity to quantify two familiar transitive, asymmetrical relations: "heavier than/lighter than" and "longer than/shorter than."

In the first instance each child was shown three clay balls in a row on a tabletop. All three balls were the same size; the balls at each end of the row were the same color [*see top illustration on page 102*]. Although the balls were identical in size, they were different in weight. In some tests the heaviest ball would be at the left end of the row and the lightest at the right end; in other tests the order was reversed. In each case the child was asked to compare the weights of the left and middle balls and then the weights of the middle and right balls. When that had been done, the child was questioned about the quantitative relation between the balls at each end of the row. The child had been given no information about this relation, so that a grasp of the ordination con-





SHARP CONTRAST between the elementary grasp of cardination and of ordination is apparent in this pair of charts. Of 93 children with a top-level grasp of ordination (*left*), only 13 percent (*color*) simultaneously possessed a top-level grasp of cardination. Ninety-two children (*right*) had no comprehension at all of cardination; 43 percent (*color*) nonetheless had a top-level grasp of ordination.

cept was necessary if the questions were to be answered correctly.

In the second instance three lengths of quarter-inch dowel were arranged in a row from left to right in front of the child. The three sticks appeared to be identical in length, but actually the sticks at each end of the row, which were painted the same color, were respectively half a centimeter longer and half a centimeter shorter than the middle stick. As with the balls, the short stick was sometimes placed at the left and sometimes at the right [*see bottom illustration on page 102*]. The child was asked to compare the lengths of the left and middle sticks and the lengths of the middle and right sticks and was then questioned about the quantitative relation between the sticks at the ends of the row. Again, if the questions were to be answered correctly, ordination was necessary.

The two ordination tests were administered to the 180 children. It was at once apparent that the group comprised children at three distinctive stages of development. The children at what I shall call Stage I were totally incapable of ordination. Those at Stage II were capable of spatial ordering but not of true quantitative ordination. This is to say that they answered correctly when the order of lightest to heaviest or shortest to longest was from left to right but could not do so when the order was reversed. The children at Stage III were capable of true quantitative ordination and answered correctly regardless of the order of presentation of the objects.

Because cardination is the quantification of classes on the basis of the correspondence between their respective classes, the cardination test was designed to assess the children's capacity to quantify pairs of classes that displayed either one-to-one or many-to-one correspondence. The test consisted of six problems [*see illustration on page 104*]. In each problem the child was shown two parallel rows of dots; the dots in the top row were red and the dots in the bottom row were blue. The minimum number of dots in a row was six and the maximum number 10. When each row had the same number of dots as the other (one-to-one correspondence), the dots were spaced so that the rows were nonetheless of unequal length. Similarly, when one row had more or fewer dots than the other (many-to-one correspondence), the dots were spaced so that the rows were the same length or the row with fewer dots was the longer. This arrangement guarded against the child's making his judgment on the basis of gross perceptual cues. To further ensure that the six pairs were quantified only by correspondence the child was directed not to count the dots. (Counting, of course, is an ordinal operation.)

All 180 children were given the cardination test. As with ordination, three stages of development were found. The children at Stage I were totally incapable of cardination. The children at Stage II were capable of quantifying the many-to-one correspondences but not the one-to-one correspondences. The children at Stage III were capable of cardination on the basis of both kinds of correspondence.

Our most important finding came from comparing the results of the two tests. The difference in the numbers of children at each of the three stages of development had already suggested that ordination appears in a child's thinking long before cardination. For example, in ordination 93 of the 180 children were functioning at Stage III and 24 at Stage I, whereas in cardination only 15 children were at Stage III and 92 at Stage I [*see illustration on preceding page*]. When the same child's performances in both tests were compared, the "*X* before *Y*" position of ordination became even more apparent. Only 12 of the 93 children functioning at Stage III of ordination were also at Stage III of cardination. Of the 92 children functioning at Stage I of cardination, however, 40 were functioning at Stage III and 32 more at

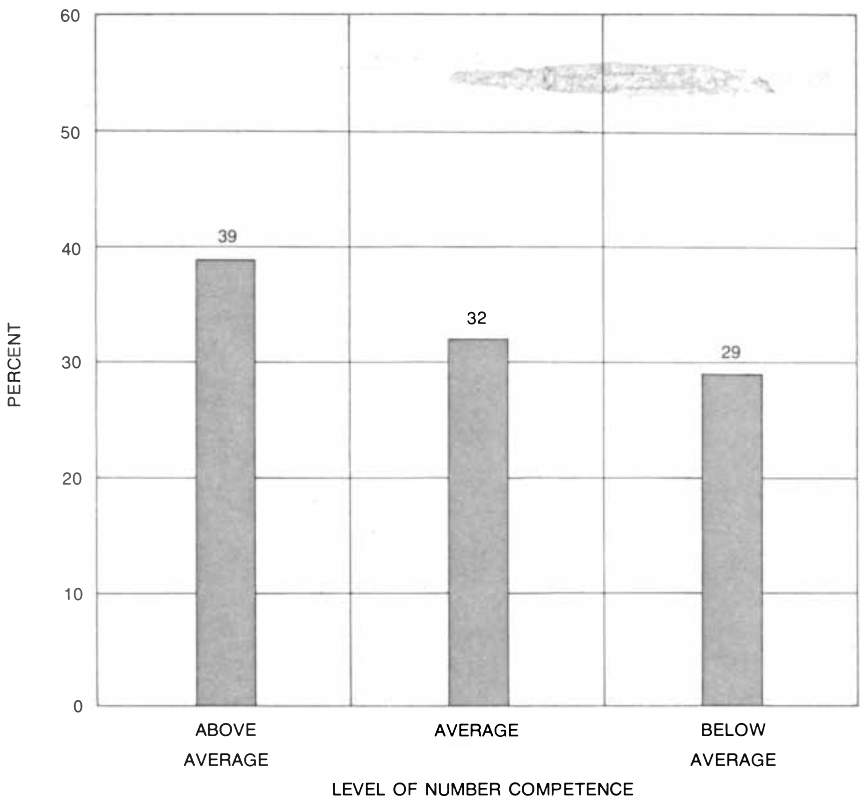
Stage II of ordination [*see illustration on opposite page*]. All things considered, one could not ask for a clearer answer to the question of whether ordination or cardination is the first concept to appear in a child's thinking.

For our next normative test we selected a second group of 180 schoolchildren, divided into equal numbers of kindergarten pupils (five to six years old) and first-graders (six to seven years old). The children were given the ordination and cardination tests described above and in addition were given a two-part test of natural-number competence. The first part of the test assessed the children's capacity to add the first four natural numbers; it consisted of the 16 problems $1 + 1 = ?$, $1 + 2 = ?$ through $4 + 4 = ?$. The second part of the test assessed the capacity to subtract natural numbers of 8 or less whose difference corresponds to one of the first four natural numbers. It too consisted of 16 problems: $2 - 1 = ?$, $3 - 1 = ?$ through $8 - 4 = ?$. The 90 six-to-seven-year-olds were given both parts of the test but only the first part was given to the 90 children aged five to six, who had not yet been introduced to subtraction in kindergarten.

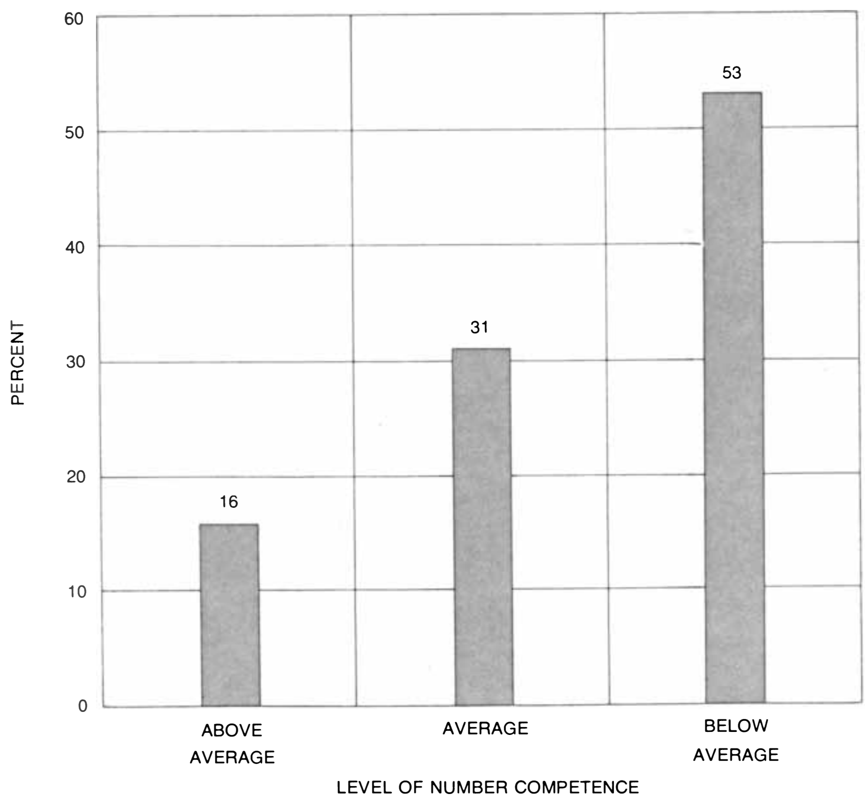
The number-test items were presented simultaneously in verbal and written form. For example, for the first problem the child was given a sheet of paper with $1 + 1 = ?$ printed on it and was simultaneously asked, "How many apples are one apple and one apple?" If the child wrote or spoke the correct answer, or did both, the problem was scored as having been solved.

After the test scores had been collected we consulted with the superintendents, principals and staffs of several elementary schools to determine what they considered to be below average, average and superior numerical skill in children of this age level. On the basis of these consultations we established three levels of performance. Twelve to 16 correct answers were scored as superior, six to 11 correct answers average, and zero to five correct answers below average. We sorted our subjects accordingly into three levels of numerical skill.

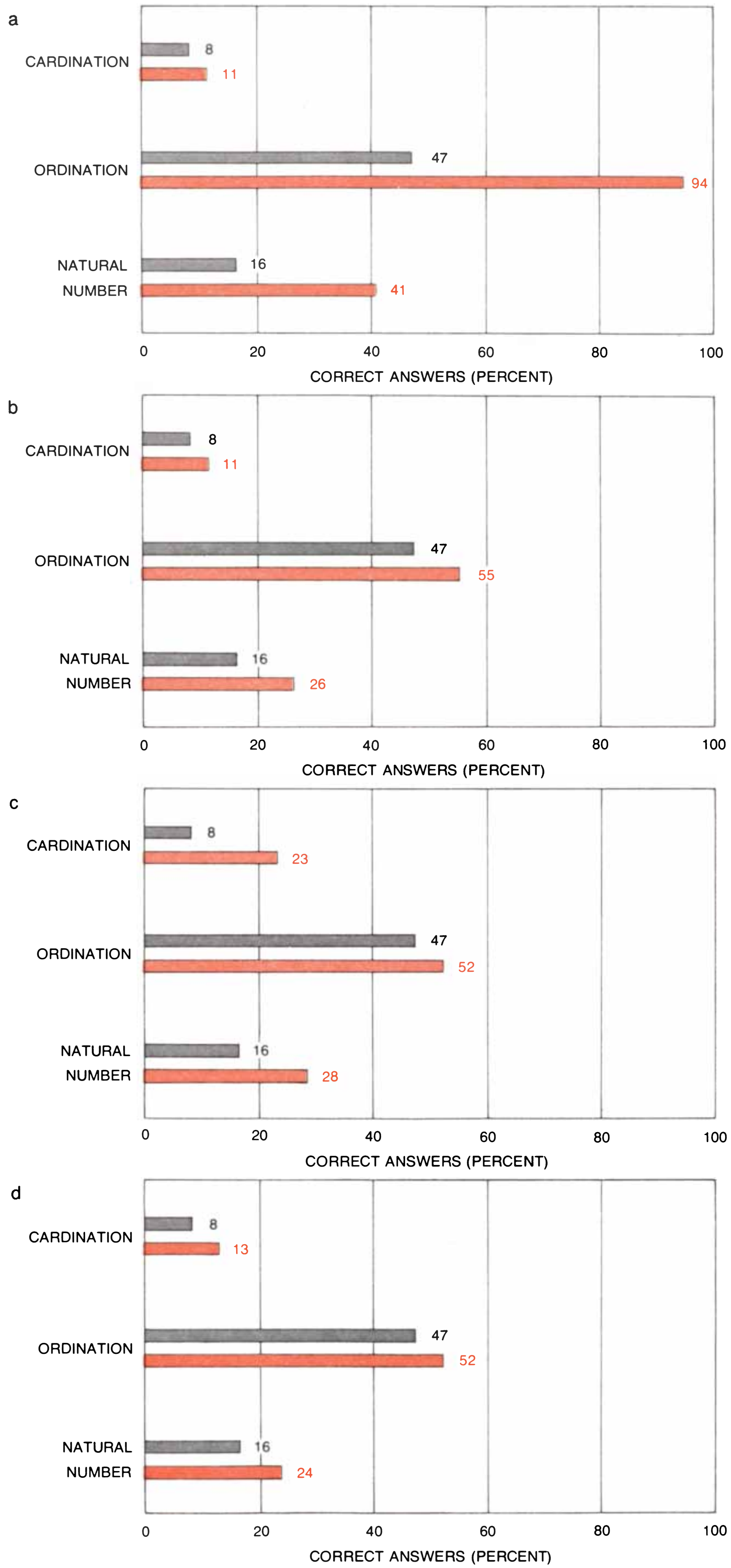
Analysis of the study showed first of all that the test results replicated our earlier findings about ordination and cardination. Among the second group of 180 children ordination emerged in the same three stages as it had among the first 180. Cardination emerged in the same three stages as before, and ordination emerged long before cardination. The next question was whether ordina-



GRASP OF ORDINATION proved to be associated with superior number competence among 180 additional elementary-school subjects. Of 119 children with a top-level grasp of ordination, nearly three-quarters showed above-average or average number competence.



GRASP OF CARDINATION proved not to be associated with competence in using numbers on the part of the same 180 subjects. Of the 95 children with no comprehension of cardination, nearly half nonetheless showed an above-average or average number competence.



tion preceded or followed natural-number competence in order of emergence. To find the answer we compared each child's stage of development in ordination (stages I, II or III) with the child's degree of natural-number competence (below average, average or superior). Of the 180 children 119 were functioning at Stage III, the top level of ordination, 39 were at Stage II and 22 were at Stage I. Of the 119 Stage III children 46 displayed the highest level of number competence, 38 displayed average competence and 35 were below average. The comparison demonstrated to our satisfaction that ordination precedes natural-number competence in order of emergence.

A third question remained: What is the first to emerge, natural-number competence or cardination? We compared each child's level of cardination with the child's degree of natural-number competence. Of the 180 children only 33 were functioning at Stage III, the top level of cardination. Thirty-eight were functioning at Stage II and 95 at Stage I. Of these 95, 15 displayed superior number competence and 30 average competence. The remaining 50 children displayed the lowest level of natural-number competence. The comparison satisfied us that natural-number competence emerges before cardination.

Taken together, our two normative studies showed an invariant sequence in the growth of the children's concept of number. Ordination was the first to emerge, followed by natural-number competence and then by cardination. The sequence strongly suggests that natural-number competence is founded on

TRANSFER EXPERIMENT (*left*) tested whether training in cardination and ordination would affect a child's competence in using natural numbers. A test group of 240 children was divided into four sets of 60, each set equally competent in cardination, ordination and use of numbers (*gray bars*). The sets were then paired; the children in sets *a* and *c* were specially trained in ordination and cardination respectively and the children in sets *b* and *d*, as controls, received only routine testing. After eight training or testing sessions all four sets were given a final test (*colored bars*). The training in ordination nearly doubled the scores of the children in set *a*; a transfer effect is also apparent in the improvement in natural-number scores for set *a*. The training in cardination nearly tripled the scores of the children in set *c*. The children's scores were nonetheless low and their training had no significant effect on natural-number scores.

a prior understanding of ordination and not on a prior understanding of cardination.

It is possible to interpret the sequence established by our normative tests in other ways. For example, it could be asserted that ordination, natural-number competence and cardination are acquired in relative isolation from one another and that the observed sequence does not necessarily imply any cognitive dependence among these concepts. It was for this reason that we went on to our third experiment: the transfer experiment that would test the validity of our "better X means better Y" prediction. Our subjects were 240 children between the ages of five and six; each child was seen once a week over a 10-week period. During the first meeting the child was given the three tests employed in the earlier normative studies. The children were then divided into four groups of 60 each. The groups were carefully matched with respect to the children's test performances; this matching ensured that any differences in performance that might later emerge would not be due to initial differences among the groups.

Two of the groups were now selected as an experimental group and a control group for training in ordination and the remaining two as an experimental group and a control group for training in cardination. Over the next eight weeks both experimental groups were given a weekly test of the kind employed in the normative studies. The training was reinforced by simple feedback. This is to say that each child was told whether or not the answers to the test questions were correct; there were roughly 15 minutes of feedback training in each of the eight sessions. The control groups were given the same kind of weekly test but without feedback. No number tests were given during the eight weeks.

During the 10th and final testing session all four groups were given the same tests for ordination, number performance and cardination that had been administered during the first session, and the results of the final tests were subjected to statistical analysis. The analysis revealed five major findings [*see illustration on opposite page*]. First, the ordination performance of the 60 children in the first experimental group was clearly superior to that of the 60 children in the matching control group, demonstrating that ordination definitely improves as a result of feedback training. Second, the cardination performance of the children in the second experimental group was also superior to that of the control group, showing that cardination too improves as a result of training.

Our third finding was that the average improvement in ordination performance was much greater than the average improvement in cardination, indicating that cardination is more difficult to acquire than ordination. Our last two findings concerned number performance. We found that the children in the experimental group trained in ordination were superior in number performance to the children in the matching control group. The number performance of the children in the experimental group trained in cardination, however, was not significantly superior to that of their control group. These findings confirmed what had been suggested by our normative studies: The conception of natural number derives from a prior understanding of ordination and not from a prior understanding of cardination.

It is instructive to consider the question of how children might best be taught numerical skills in the light of these findings. Whether we like it or not, number competence is socially prescribed in modern civilization. One might even say that number competence has come to have significant survival value. For example, some of the most important decisions made by the average person during a lifetime involve number-related ideas such as money. Yet it is obvious that if we do not design mathematical instruction to agree with the hard facts of mental growth, we run the risk of making early mathematics education unduly difficult for children. Even worse, we run the risk of discouraging a large number of the minds we are attempting to instruct.

During the 19th century and most of the 20th it was standard practice in the nations of the West to begin the formal mathematical education of young children with the natural numbers. It was this practice that Russell decried in his introduction to the revision of *Principles of Mathematics* published in 1937. Russell argued that the practice was derived from the discredited Pythagorean assumption that the natural numbers are unanalyzable entities that we simply must accept as being given. It is safe to assume that he would have preferred a curriculum wherein the introduction of natural numbers is postponed until the child has been taught the rudiments of the logic of classes and, in particular, taught the quantification of classes through the correspondence of elements. The Pythagorean assumption was well entrenched in educational circles, however, and Russell's argument was ignored for nearly a quarter of a century.

In the past decade the change that Russell called for has begun to take place. Most parents, however unaware they may be of the details, remember the rallying cry of the period: "New math." In the "new math" the logicist theory that natural numbers are derived from the quantification of classes was at last put into educational practice. Indeed, it would be more accurate to refer to the "new math" as the cardinal approach to early mathematics instruction, whereby the child is first introduced to the central idea of the logic of classes and the natural numbers are introduced later as by-products of the process of quantifying classes.

Today, although pockets of resistance remain, the cardinal approach dominates educational circles in North America almost as completely as the Pythagorean approach did when Russell spoke against it. Most major publishers of elementary textbooks either have adopted or are in the process of adopting some version of the cardinal approach, and continued resistance to the change has become very difficult.

Taking into account the implications of our normative and transfer experiments, it seems dangerous to rest content with either the old Pythagorean approach or the new cardinal approach. The order of emergence of the various concepts in children's thinking is ordination first, number second and cardination third. Moreover, ordination training tends to improve the number competence of young children, whereas cardination training does not. The conclusion that a new "new math" emphasizing ordinal notions is called for seems inescapable.

How might such an approach be instituted? The first concepts to be introduced should be ordination and ordination-like ideas such as the logic of relations. Later the natural numbers could be introduced as by-products of the quantification of transitive, asymmetrical relations. Still later cardination could be introduced as a generalization of the natural numbers. Such an approach to early mathematics education would provide a much better fit with what is now known about the emergence of number concepts in the young child's mind than either the Pythagorean or the cardinal approach.

MATHEMATICAL GAMES

The calculating rods of John Napier, the eccentric father of the logarithm

by Martin Gardner

In his celebrated *Budget of Paradoxes* Augustus De Morgan defines a "graphomath" as a person ignorant of mathematics who tries to describe a mathematician. As an example, he quotes from the second chapter of Sir Walter Scott's novel *The Fortunes of Nigel,* in which David Ramsay, a whimsical clockmaker and amateur mathematician, swears "by the bones of the immortal Napier!"

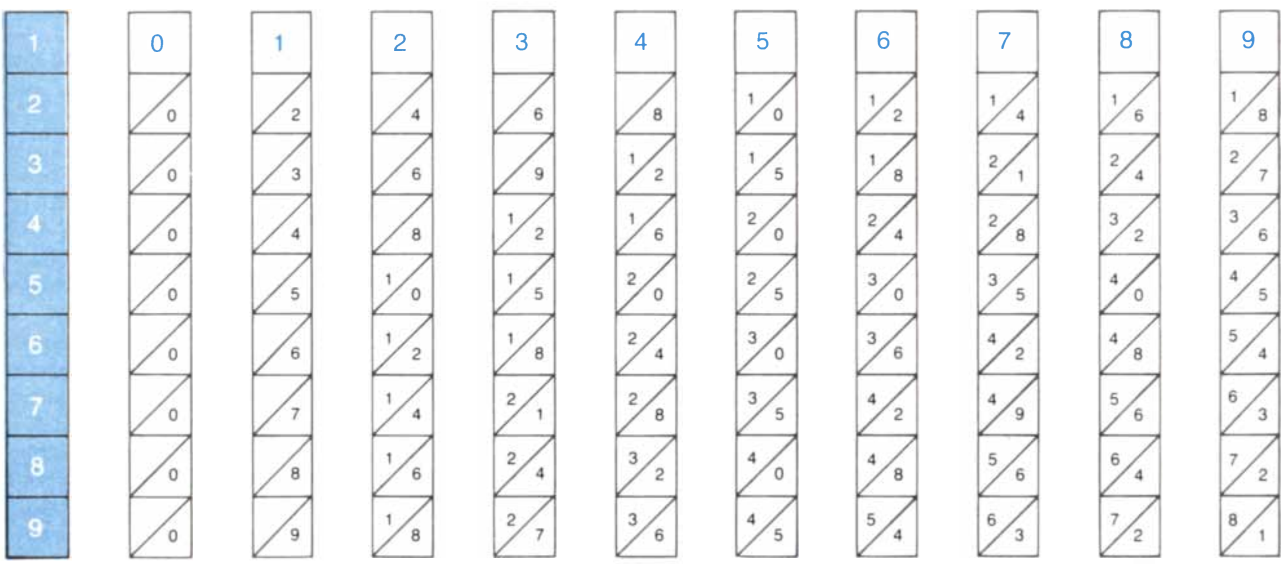
It is hard to tell from the passage whether Scott actually was uninformed or whether he merely intended Ramsay to make an ignorant or a joking remark. In any case, "Napier's bones" have nothing to do with the skeletal remains of Baron John Napier (1550–1617), the Scottish mathematician who discovered logarithms and who was the first important mathematician of Britain. The phrase refers to a set of numbered rods that Napier invented for doing multiplication. We shall discuss his method later, but first some remarks about Napier himself.

His father, Sir Archibald Napier, master of the Scottish mint, was just 16 when John was born. And John was a mere 13 when he entered the University of St. Andrews. He left the university without getting a degree, took over the family castle and estates at Merchiston (now part of Edinburgh), married and had one son and one daughter, was widowered, remarried, then continued the symmetry with five sons and five daughters. The Protestant Reformation in Scotland had started at about the time John was born, and while a youth at St. Andrews he became a passionate Protestant with a compulsion to explicate biblical prophecy. In 1593 he published what he always considered his masterpiece (much more important than logarithms), the full title of which was:

"A Plaine Discovery of the whole Revelation of Saint Iohn: set downe in two treatises: The one searching and proving the true interpretation thereof: The other applying the same paraphrastically and historically to the text. Set foorth by John Napier L. of Marchistoun younger. Whereunto are annexed certaine Oracles of Sibylla, agreeing with the Revelation and other places of Scripture. Edinburgh, printed by Robert Walde-grave, printer to the King's Majestie, 1593. Cum privilegio Regali."

It was the first major Scottish work on the Bible and one of the most thorough attempts ever made before or since to explore the symbolism of the Apocalypse. It is ironic that today, when many college students seem more interested in the Second Coming than in current politics, there is no available reprint of Napier's treatise. It was enormously influential in its day, with 21 editions in English and numerous European translations.

Perhaps the main reason the book is out of print is that Napier made a slight miscalculation about the end of the world. He had been strongly influenced by the religious speculations of Michael Stifel, a German algebraist who proved that Pope Leo X was the anti-Christ by rearranging the Roman numerals in LEO DECIMVS to make DCLXVI, or 666, the notorious "mark of the Beast." Where did Stifel get the X? From Leo X and from the fact that LEO DECIMVS has 10 letters. What happened to the M? He left that out because it stood for *mysterium*. Stifel predicted that the world would end on October 3, 1533. Napier perceived that this was a mistake. He decided that it was the Pope of 1593 who was the actual anti-Christ. God had ordained that exactly 6,000 years would elapse between the earth's creation and its destruction. Since there was some uncertainty about the exact date of creation, Napier set the



Rabdology, or "Napier's bones"

end of the world as being between 1688 and 1700.

Napier begins his book by apologizing for having written it in a language so base as English, and concludes it by appealing to the Pope as follows:

"In summar conclusion, if thou o *Rome* aledges thyselfe reformed, and to beleeue true Christianisme, then beleeue Saint *John* the Disciple, whome Christ loued, publikely here in this Reuelation proclaiming thy wracke, but if thou remain Ethnick in thy priuate thoghts, beleeuing the old Oracles of the *Sibyls* reuerently keeped somtime in thy *Capitol:* then doth here this *Sibyll* proclame also thy wracke. Repent therefore alwayes, in this thy latter breath, as thou louest thine Eternall salvation. *Amen.*"

"Strange," comments De Morgan in his *Budget,* "that Napier should not have seen that this appeal could not succeed, unless the prophecies of the Apocalypse were no true prophecies at all."

After clearing up the apocalyptic mysteries, Napier turned his ingenuity toward ways of defending Scotland against a threatened invasion by Catholic Spain. His 1596 document was titled *Secrett Inventionis, proffitabill and necessary in theis dayes for defense of this Iland, and withstanding of strangers, enemies of God's truth and religion.* It describes three inventions: mirrors for setting fire to enemy ships (shades of Archimedes!), a machine gun and a metal chariot (that is, a tank) housing soldiers who could fire through holes in the sides.

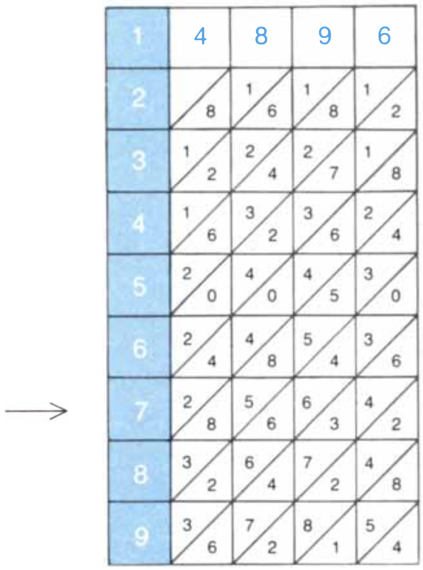
Napier's next book, the Latin title of which begins *Mirifici Logarithmorum Canonis Descriptio*... (*A Description of the Marvelous Rule of Logarithms*...), appeared in 1614. This was the book in which Napier explained logarithms, called them logarithms (a term he coined) and gave the world its first log table. It has often been pointed out that if exponents had then been in common use, logarithms would have immediately been recognized as a great toil saver, but Napier conceived of them without reference to exponents at all. This is not the place to explain how he arrived at logs the hard way by considering the relation of an arithmetic series to a geometric series. Napier's logs rested awkwardly on a base slightly less than 1. (They were not natural logarithms based on e, even though these later became known as Napierian logarithms.) The London geometer Henry Briggs quickly realized that 10 was the most convenient base for logarithmic calculations in the decimal system, and Napier at once agreed. It is said that when the two men first met at Merchiston Castle (where Briggs remained for a month), they admired each other for 15 minutes before either spoke a word.

Navigators and astronomers, notably Johannes Kepler, found the base-10 logs (or common logarithms as they are now called) invaluable, and years of drudgery were devoted by Briggs and others to preparing better and better log tables. (Today it is faster to compute a log all over again on a pocket electronic calculator—it takes only a microsecond—than to look it up in a book!) In Napier's posthumous work *Mirifici Logarithmorum Canonis Constructio*... (1619) he explained how he calculated his original logs. In doing so he made systematic use for the first time in history of a decimal point, placing it above the baseline and using it exactly as it is used today in England.

Two of the most amusing of many anecdotes about Napier are recounted by Howard W. Eves in his delightful *In Mathematical Circles.* Because a neighbor's pigeons were flying onto Napier's estate and eating grain, Napier told his neighbor that he would impound the birds as payment. The neighbor replied that Napier was welcome to any pigeon he could catch alive. Napier scattered brandy-soaked peas over his grounds and the pigeons were soon staggering about in such a stupor that he had no trouble collecting all of them in a sack.

It was a time when almost everyone in Scotland (including Napier) believed in astrology and black magic. One day Napier called his servants together and told them that his black rooster had the occult power to tell him which servant had been stealing from the estate. One at a time each servant was asked to enter a dimly lighted room and stroke the bird's back. As Napier had anticipated, only the guilty person, fearing exposure, would not do as asked. Napier had covered the rooster's black feathers with soot, and so only the guilty servant emerged with clean hands.

The age was also one of intense interest in calculating. The average person did arithmetic on his fingers, but more skillful mathematicians took great delight in completing tedious computations. Napier's hobby was to find ways to simplify such work. Logarithms were, of course, his best invention, but in 1617 (the year he died) he brought out a little book called *Rabdologia* that explained three other methods of calculating. The book's title was his name for the first method, one that soon became known as "Napier's bones" because it used rods that often were made of animal bone.



4,896 × 7 = 34,272

The reader is urged to make a set of Napier's bones by labeling 11 strips of heavy cardboard (or Popsicle sticks, tongue depressors or any other available wooden strips) as shown in the illustration on the opposite page. The index rod is not essential, but it makes it easier to locate desired rows. Each of the rods has a digit at the top. Below the digit, from the top down, are the products when that digit is multiplied successively by numbers 1 through 9. The set of bones obviously is nothing more than a multiplication table cut into strips so that it can be manipulated manually, with a zero strip added to serve as a place holder.

The procedure is ridiculously simple. Suppose you wish to multiply 4,896 by 7. Rods topped with 4, 8, 9 and 6 are placed side by side with the index rod on the left [*see illustration above*]. Only row 7 (the multiplier) is considered. Write down 2, the last digit of the row, as the final digit of the product. The product's next digit (working to the left on both rods and paper) is obtained by adding the next pair of digits (the diagonally adjacent digits inside the little parallelogram) of the row. They are $4 + 3$, and so put down 7 as the second digit from the end of your product. The sum of the next pair ($6 + 6 = 12$) is more than 9, therefore write 2 as the third digit of the product and carry 1. The next pair, 5 and 8, add to 13, but you are carrying 1, so that the sum is 14. Put down 4 and again carry 1. The last digit of the row is 2. Two plus 1 is 3, so that 3 is the final digit (on the left) of your product.

You have now obtained the correct answer, 34,272, by using only simple addition. Of course, if you know your multiplication table through the 9's, you can do it just as easily without the rods. In Napier's day, however, the ordinary person's ability to calculate was feeble and so the rods became an instant success.

To multiply 4,896 by a larger number, say 327, it is necessary to obtain three partial products and add them in the usual way. In other words, write down 34,272 (the product of 4,896 and 7), then put below it the products obtained from rows 2 and 3, jogging them to the left in the standard manner,

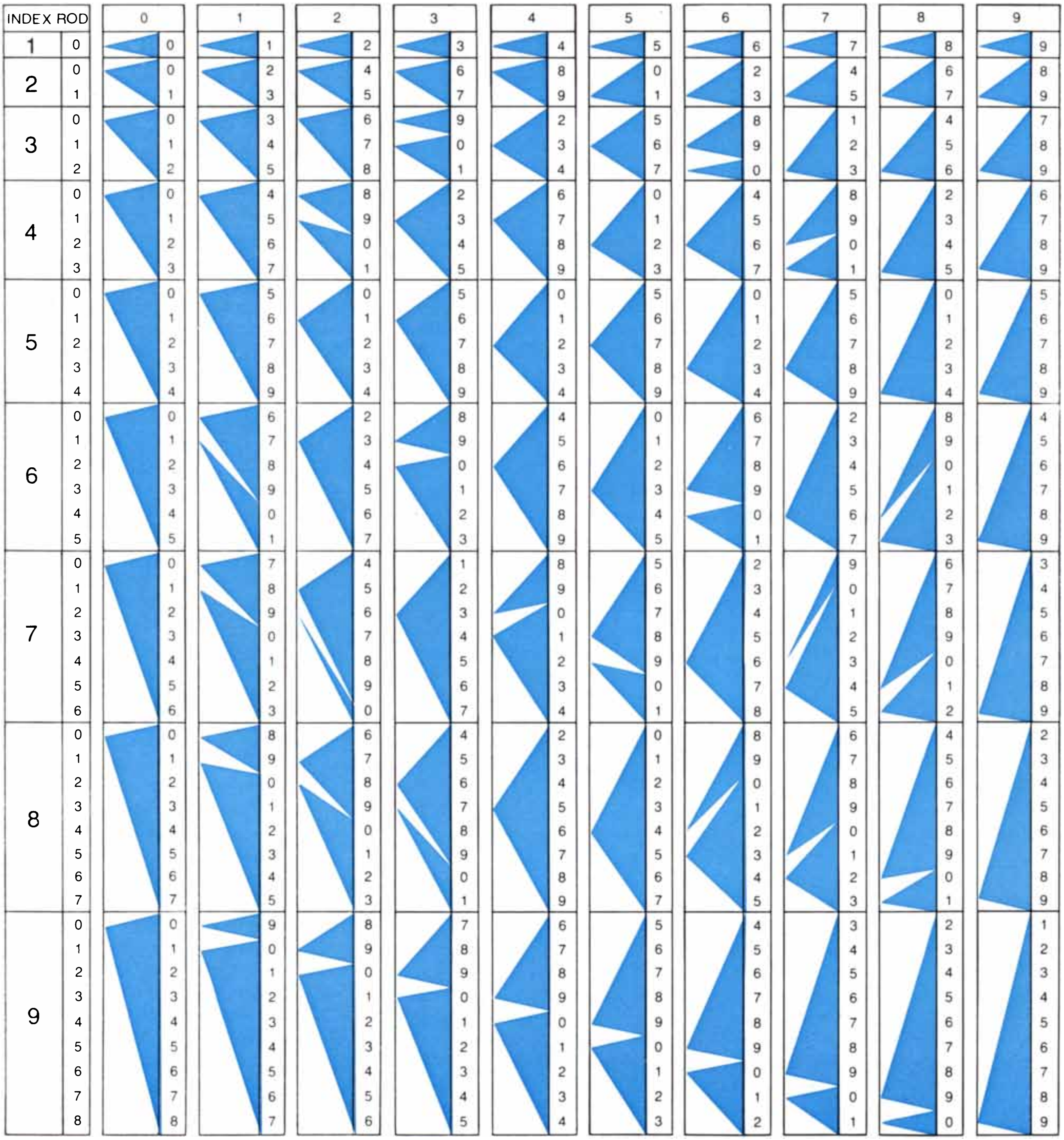
```
 34272
 9792
14688
```

then add to obtain the final product.

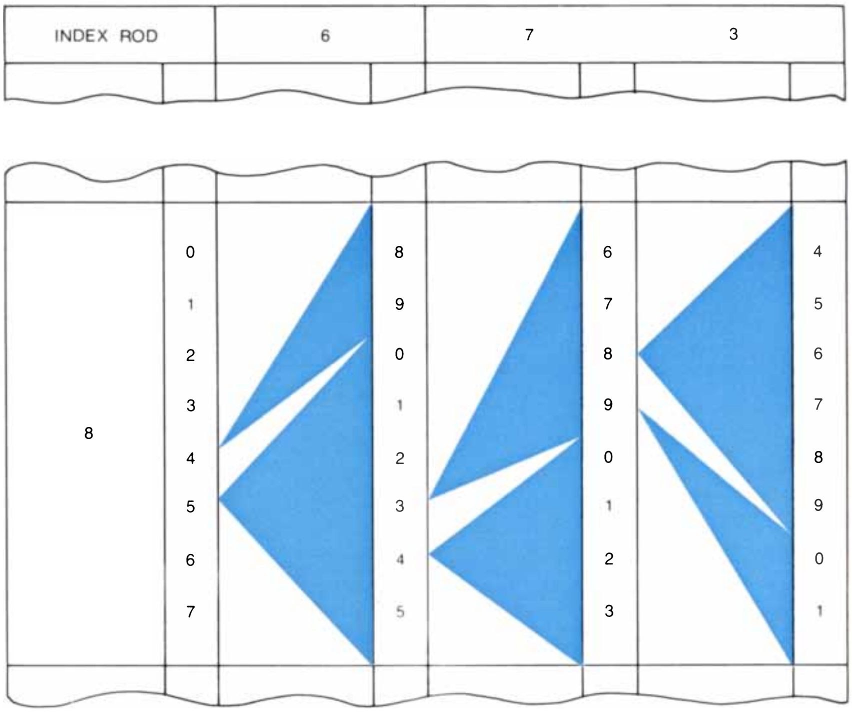
The rods are of little use unless you have more than one set because a multiplicand may contain duplicate digits. Napier's rods had square cross sections, each face of a rod corresponding to one of the strips in our cardboard set. He arranged the four columns so that the top digits on opposite sides of each rod added to 9. The following are the quadruplets of Napier's set of 10 bones:

0, 1, 9, 8	1, 3, 8, 6
0, 2, 9, 7	1, 4, 8, 5
0, 3, 9, 6	2, 3, 7, 6
0, 4, 9, 5	2, 4, 7, 5
1, 2, 8, 7	3, 4, 6, 5

It is clear that such a set of 10 rods



Henry Genaille's calculating rods



673 × 8 = 5,384

can handle all multiplicands of 10 digits or fewer that are possible to form with the rods, but many multiplicands cannot be formed, so that it was advisable to own more than one set. As a little puzzle in combinatorics, can the reader determine the largest multiplicand one set of Napier's bones will form such that all smaller multiplicands can also be formed by the set? As a second exercise, find the corresponding largest multiplicand for two sets of Napier's bones. (Both answers will be given next month.)

Napier's rods can be used for division too, but the process is more trouble than it is worth. In short division you must select rods that form the desired dividend on the row for the digit divisor, then read off the quotient from the top. If the dividend cannot be formed, form the largest number you can that is less than the dividend, then subtract that number from the dividend to obtain the remainder. In long division the rods can be used for determining the successive products of the divisor and each digit in the quotient.

The charm of Napier's rods lies in their simplicity. If we are willing to complicate them a bit, however, we can eliminate the bother of having to carry 1's in our head. The cleverest way of doing this was invented about 1890 by Henry Genaille, a French civil engineer. The picture of these rods is almost self-explanatory [*see illustration on opposite page*]. They work exactly like Napier's except that the product is read directly from right to left. Start with the digit at the top right of the desired row. The next digit is the one to which the colored triangle (at the left of the previous digit) points. From now on move from each digit into the colored triangle directly at its left and go to the digit to which it points. For example, to multiply 673 by 8 start with 4 at the top right [*see top illustration at right*] and see how easily you can move to the left through the chain of triangles to obtain the product, 5,384.

Both Napier's bones and Genaille's rods are marvelous teaching devices because it is not hard to see why they work, and when you do, you obtain valuable insight into the multiplication procedure. If you have difficulty understanding why Genaille's rods operate, you can find it explained in "Genaille's Rods: An Ingenious Improvement on Napier's," by B. R. Jones, in *The Mathematical Gazette,* Volume 48, February, 1964, pages 17–22, the article from which our illustrations were taken.

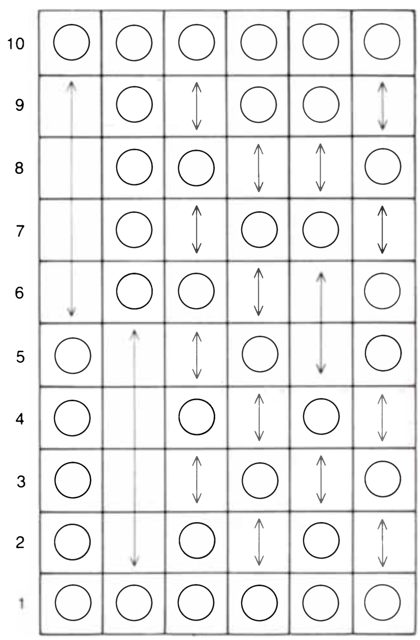
The second calculating method in *Rabdologia* had to do with arranging metal plates inside a box and is too complicated and impractical to explain here. But Napier's third method, which he regarded as being primarily an amusement, requires only a chessboard and a supply of counters. By moving the counters as one does rooks or bishops, one can do addition, subtraction, multiplication, division and square roots, and all in the binary system.

Last month's exercise problem for applying Richard M. Karp's elevator algorithm requires a minimum of 72 unit trips to sort the 45 people in the nine-floor building. The solution to the second elevator problem is given as the answer to Exercise 16 in section 5.4.9 of Donald E. Knuth's *The Art of Computer Programming: Sorting and Searching* (Addison-Wesley, 1973). The best solution known requires 12 stops as follows:

1. 123456 to 2.
2. 112334 to 3.
3. 224456 to 4.
4. 135566 to 5.
5. 112334 to 6.
6. 224456 to 2.
7. 222444 to 4.
8. 222222 to 2.
9. 555666 to 5.
10. 666666 to 6.
11. 111333 to 3.
12. 111111 to 1.

The solution to Kobon Fujimura's elevator problem, which is a disguised version of an old combinatorial problem of block design, is shown below.

Many readers will be pleased to know that Harry Lindgren's classic and beautiful work *Geometric Dissections* (1964) is now back in print as a Dover paperback with the title *Recreational Problems in Geometric Dissections and How to Solve Them.* The book has been revised and enlarged by Greg Frederickson, who has managed to break many of the old dissection records.



An answer to the elevator problem

THE AMATEUR SCIENTIST

Infrared for the amateur: infrared diode lasers and an infrared filter

Conducted by C. L. Stong

Of the numerous kinds of lasers that have been developed during the past decade, the smallest and least costly to assemble is the diode laser. The active element, a sandwich of semiconducting material the size of a pinhead, emits an intense beam of infrared radiation when it is energized by a direct current. The invisible beam holds promise as a carrier of various kinds of signals. It can also be employed for echo ranging and intrusion alarms and for demonstrating such aspects of wave behavior as refraction, diffraction, reflection and interference.

Diodes of gallium arsenide that are designed for generating pulses of coherent radiation have become available in recent months at prices that enable the experimenter to assemble a working laser for less than $30. Several diode lasers, which emit beams that range in peak power from four to 70 watts, have been built by Harry L. Latterman of Mesa, Ariz. He explains the apparatus and some experiments that can be done with it as follows:

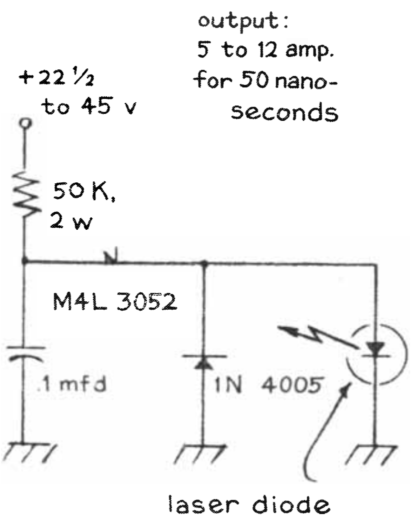
"The diode lasers I have made consist of three subassemblies: the diode and its mounting; an electronic circuit that generates pulses of direct current, and a source of power. The smallest of my lasers, which can be made in one evening, emits 200 pulses per second. Each pulse persists for 50 nanoseconds (billionths of a second) and reaches a maximum intensity of four watts. When the laser is operated by a dry-cell battery, it is self-contained and portable. The apparatus can be mounted in a protective housing smaller than a pack of cigarettes.

"The active part of a typical diode is a crystal consisting of three or more distinct layers of semiconducting materials sandwiched between electrically conducting films of metal. The layered structure accounts for the electrical characteristics of the diode. All crystals consist of atoms bound together in a lattice by forces associated with electrons in the outermost orbits of the constituent atoms—the valence electrons that bind atoms together. In ordinary crystals the valence electrons are fully occupied. In effect they act as tie rods that constitute the structure of the lattice.

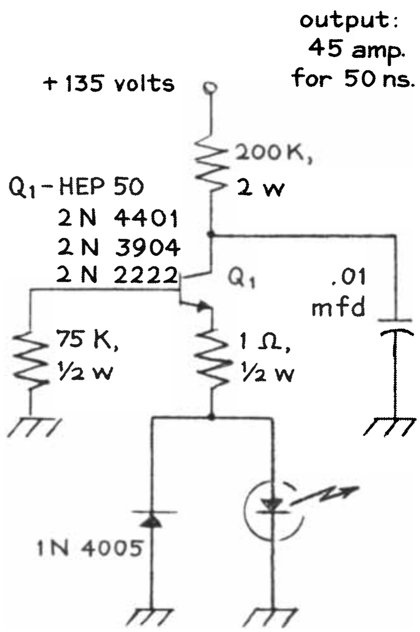
"A semiconducting material known as the *n* type can be made by incorporating in the lattice a few atoms that have one valence electron more than the number that can fit into the lattice structure. The surplus electron can be detached from its parent atom rather easily. It can then migrate through the crystal as a carrier of current. Conversely, a semiconductor material known as the *p* type can be made by incorporating a few atoms that have one valence electron less than the number that can be accommodated by the lattice structure. In this case a bond is missing from the crystal structure.

"An adjacent valence electron can become detached from its bonding site and drop into the missing bond. A hole then exists at the vacated site. In terms of its electrical behavior the hole has the properties of an elementary charge. Because it is deficient in negative charge, however, it acts as a positive charge. Crystals so made are known as *p*-type.

"The crystal of a laser diode can be made with three layers of semiconducting material: a layer of *n*-type silicon, a layer of *p*-type gallium arsenide and a third layer of *p*-type gallium arsenide that also contains atoms of aluminum. The interface between *n*-type and *p*-type layers is known as a *p*-*n* junction. Some easily detached electrons in the *n* region migrate across the interface and drop into holes of the *p*-type material. Conversely, some holes in the *p*-type material cross the junction and wander randomly in the *n*-type material. The *p*-type material thus accumulates negative charge, and the *n*-type



Circuit for 10-ampere pulses



Circuit for 45-ampere pulses

material similarly accumulates positive charge.

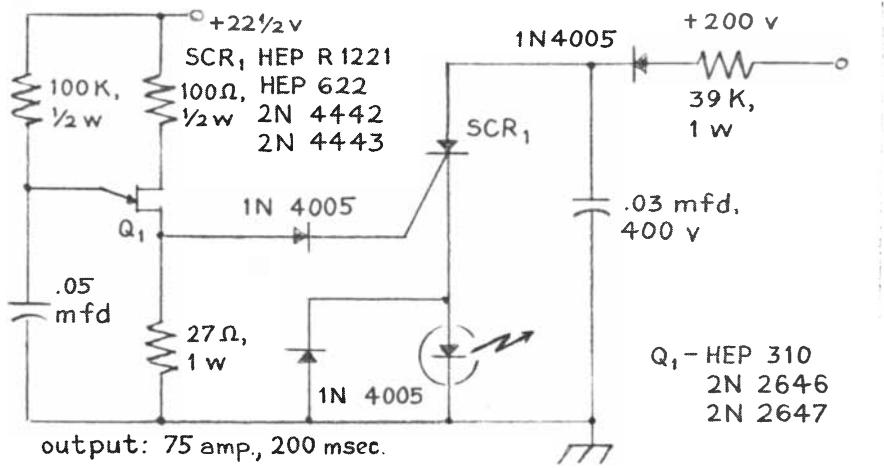
"The charges ultimately become strong enough to halt the migration of both holes and electrons. A potential barrier is then said to exist across the junction. The barrier can be modified by connecting opposite faces of the crystal to an external source of voltage. A negative potential applied to the *n*-type layer causes electrons to flow across the junction, migrate through the *p*-type material and return to the external source. The diode is said to be connected in the forward direction. Positive charge, as represented by the holes, migrates in the opposite direction. The action stops if the polarity of the external source is reversed.

"A positive charge applied to the *n*-type material simply increases the potential barrier. Hence the diode conducts current in only one direction. It is analogous to a hydraulic check valve and can be used to convert alternating current into unidirectional current, which is direct current.

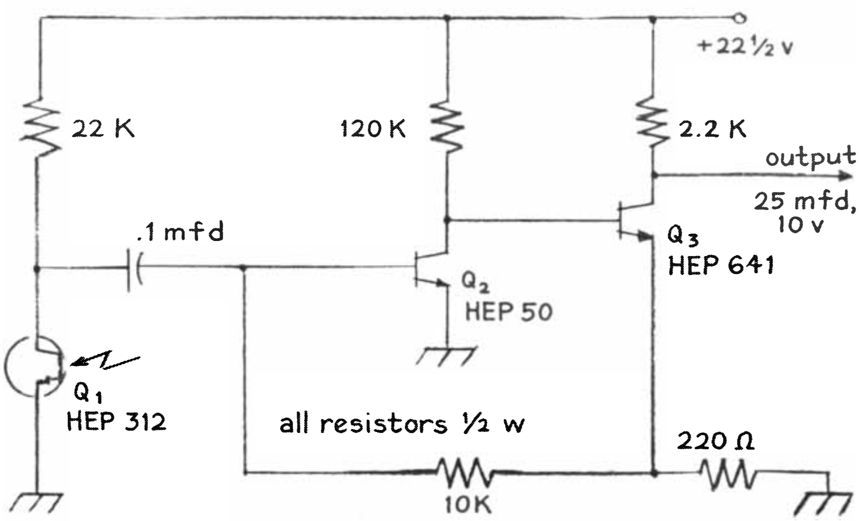
"When the diode is in its quiescent state, before voltage from the external source is applied, the electrons have minimum energy. At room temperature the crystal is in a state of continuous but gentle vibration that causes electrons to wander aimlessly through the lattice. The interesting action begins when a forward potential of about 1.2 volts from an external source is applied across the *p-n* junction. A current of electrons then flows through the diode. Collisions occur between the moving electrons and the electrons that are normally bound in the lattice structure.

"Some collisions are so violent that electrons emerge from the encounter with a discrete increment of additional energy. Energy so acquired is retained for a time. The excited electron is unstable, however, and it soon returns to a lower energy state spontaneously by emitting a photon, or quantum of radiation, that carries away the acquired energy. Many electrons, acting independently, participate in the activity as long as the diode conducts current.

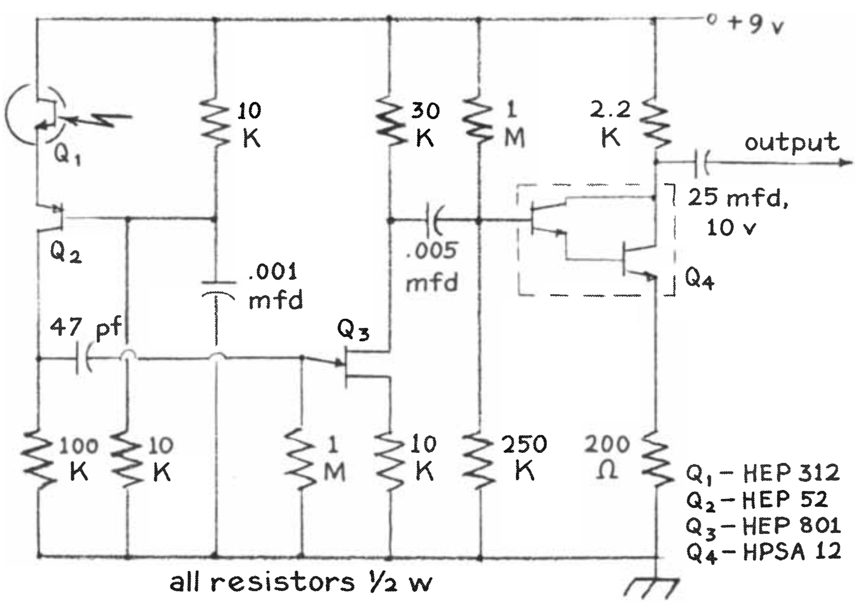
"The photons are emitted randomly in direction and in time. Hence the radiation is incoherent, analogous to the wave pattern created by tossing a handful of pebbles into a pool of still water. It is similar to the light emitted by a neon sign, and for the same reason. If a photon is pictured in the mind's eye as a short train of identical waves, coherent radiation can be pictured as a wave train consisting of two photons that unite in



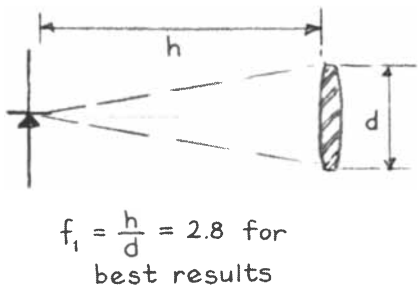
Circuitry employed by Harry L. Latterman for 75-ampere pulses



Circuit for detecting infrared rays



High-sensitivity circuit for infrared detection



Lens for infrared

lockstep and proceed in the same direction.

"Coherent radiation can be generated in various ways. For example, it appears when a photon encounters and interacts with an electron that is on the verge of emitting surplus energy in an amount equal to the energy carried by the stimulating photon. The emitted photon joins the stimulating photon and the two proceed through space together.

"In order to generate coherent radiation, lasers of all kinds provide two essential conditions. First, the laser must maintain an adequate supply of excited electrons. Second, the excited electrons must be trapped inside a resonant optical cavity that consists of a pair of facing mirrors.

"In the case of the diode laser an adequate population of excited electrons is generated by connecting a source of power to the diode in the forward direction to provide current in excess of a certain minimum threshold value. At currents below this value the diode emits incoherent radiation; such a diode is called a light-emitting diode, as distinct from a diode laser. The mirrors that form the optical cavity of the diode laser are merely the square-cut edges of the crystal. The edges function as mirrors because the index of refraction changes abruptly at the interface between the crystal and the air.

"Electrons that are excited in the region of the *p-n* junction migrate into the transparent *p*-type material. Laser action begins when a photon is spontaneously emitted. The liberated photon bounces back and forth between the mirrors, interacting with the population of excited electrons and thereby stimulating the emission of additional photons.

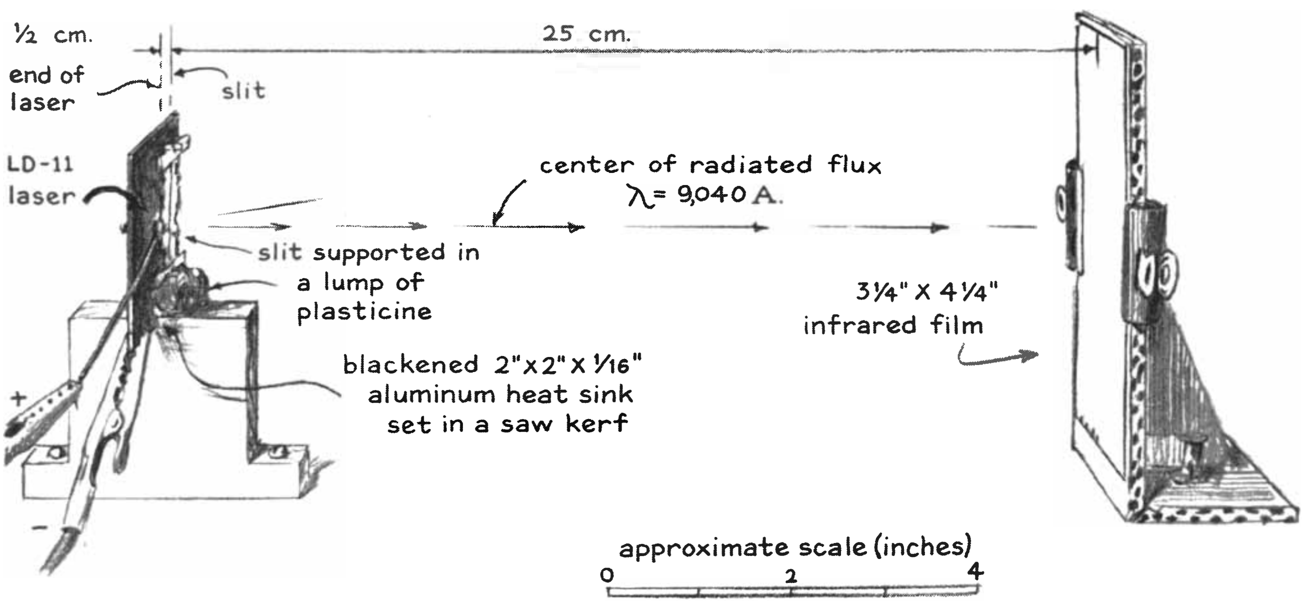
"The interface between the *p*-type gallium arsenide and the *p*-type gallium aluminum arsenide, which is known as a heterojunction, serves to confine the excited electrons and also to reduce the reabsorption of energy. In effect, the heterojunction improves the efficiency of the device. A portion of the accumulating radiant energy escapes through the mirrors in the form of a coherent beam, which is the output of the laser.

"The minimum current required to generate coherent radiation by the laser diodes now on the market ranges from about 10 to 80 amperes. The exact threshold is specified by the manufacturer, as is the peak current rating. The diode can be destroyed by current in excess of the peak value.

"The crystal in the smallest of my diodes is almost invisibly minute. A dozen such crystals could fit easily into the volume of a pinhead. Nonetheless, the diode is rated at a peak current of 10 amperes, which is equivalent to a current density in the diode on the order of 100,000 amperes per square centimeter! The problem of driving the laser at the required current without vaporizing the crystal is solved by using short pulses of current and by mounting the diode on a metal base that dissipates the liberated heat.

"The three pulsing circuits I shall describe are designed for generating peak currents of from five to 75 amperes that persist for intervals ranging from 25 to 250 nanoseconds. One or another of the three circuits will work with currently available diodes. The circuits draw current from the source less than 1 percent of the time. A duty cycle much greater than 1 percent can damage the diode. The average drain on the power supply amounts to only a few milliamperes. The diodes are shipped in a protective housing that resembles a machine screw about half an inch long with a flat, cylindrical head a quarter of an inch in diameter.

"The infrared beam is emitted through a circular window of clear plastic in the exposed end of the housing. The diode can be fastened to a heat sink with the screw, which also serves as one of the electrical terminals. The other terminal, a short length of rectangular wire, is brought out of the housing. A



Single-slit apparatus for demonstrating diffraction two-inch square of sheet aluminum at least a sixteenth of an inch thick makes an adequate heat sink.

"A circuit that develops a peak current of up to 10 amperes for operating the smaller laser diodes consists of a resistor, a capacitor, a four-layer diode, a protection diode and the laser diode. Essentially it is an oscillator of the relaxation type. As the capacitor gradually accumulates charge through the resistor, voltage rises across the four-layer diode. At a critical potential the resistance of the four-layer diode falls abruptly. The capacitor discharges through the laser, after which the cycle repeats. The resulting pulse of current persists for about 50 nanoseconds, which is equivalent to a frequency of 20 megacycles. Hence the leads between the diodes and the capacitor should be made as short as possible.

"Magnetic fields develop around the leads when the laser conducts current. At the conclusion of the pulse the fields collapse, inducing current in the wiring. As a result a reverse potential can appear across the diode. A reverse potential that exceeds a certain value can destroy the laser diode. The circuit includes a conventional diode that acts as a protective device, limiting reverse potentials to a safe value. Do not omit it. Peak current through the laser diode can be adjusted through the range from about five to 10 amperes by applying from 22½ to 45 volts to the circuit [*see illustration at left on page 114*].

"Laser diodes of intermediate power require peak pulses of current of up to 45 amperes. Currents of this magnitude can be developed by charging a capacitor to a potential of 135 volts. This potential is substantially higher than the breakdown voltage of the laser diode. Therefore an independent switch must be used for connecting the charged capacitor to the laser diode. The switching function can be accomplished by an appropriate transistor.

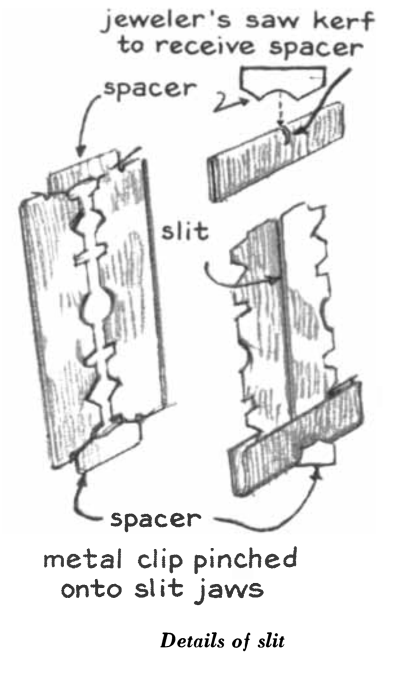
"A practical circuit is depicted by the accompanying diagram [*at right on page 114*]. A capacitor is charged through a 200,000-ohm resistor by the 135-volt source. As the capacitor accumulates charge, potential rises across the 75,000-ohm resistor in the base circuit of the transistor. When the capacitor reaches full charge, the potential of the base reaches a value that initiates conduction in the collector-emitter circuit of the transistor. The capacitor then discharges through the transistor and the laser diode. A conventional diode, connected across the laser diode, limits the reverse potential and thus protects the laser diode. The circuit generates approximately 500 pulses per second. Each pulse persists for about 50 nanoseconds.

"Current pulses of up to 75 amperes are generated by charging a capacitor of .03 microfarad to a potential of 200 volts [*see top illustration on page 115*]. The charged capacitor is connected to the laser diode by a special switch in the form of a controlled silicon rectifier, a solid-state device that is capable of conducting relatively large currents. The switch is turned on and off by an oscillator, whose active element is a transistor of the unijunction type.

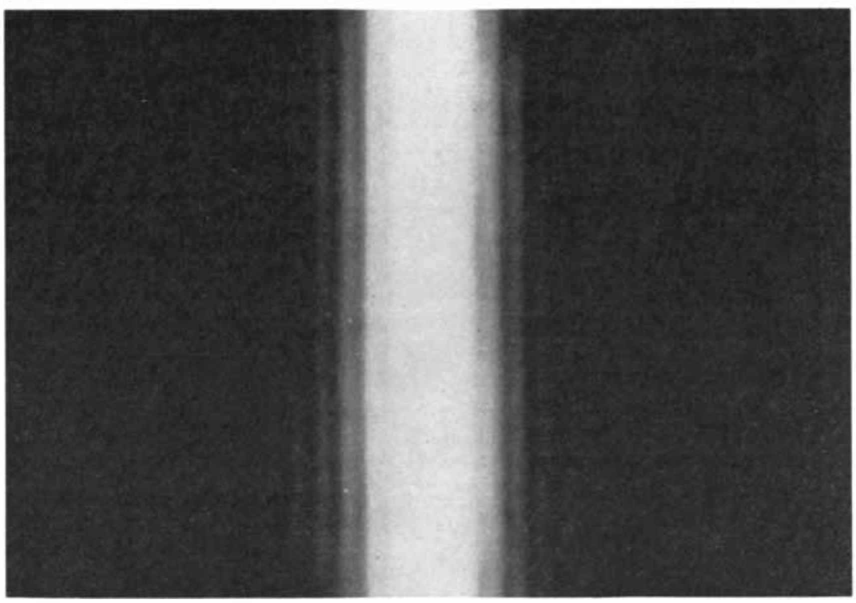
"The oscillator includes a 100,000-ohm resistor through which a .05-microfarad capacitor accumulates charge. Potential across the capacitor rises as charge accumulates and is applied to the base of the unijunction transistor. As full charge is approached the transistor conducts. The resulting discharge develops a voltage across the 27-ohm resistor in the transistor circuit. The voltage is applied through a diode to the gate terminal of the controlled silicon rectifier.

"Meanwhile the .03-microfarad capacitor in the laser-diode circuit has accumulated full charge from the 200-volt source. When the triggering voltage is applied to the gate of the controlled silicon rectifier, the rectifier conducts and the .03-microfarad capacitor discharges a 75-ampere pulse through the laser diode. Conventional diodes protect the laser diode from excessive reverse voltage. The pulse persists for about 200 nanoseconds.

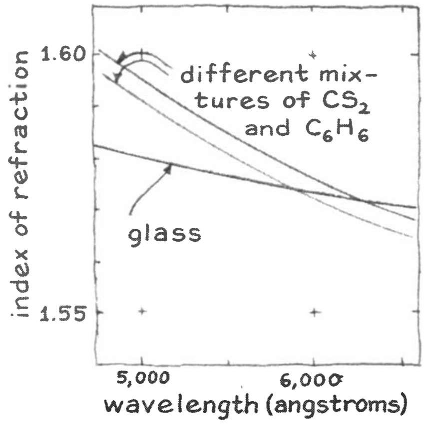
"When I assemble any of the three circuits, I usually measure the pulse before connecting the costly laser diode in the circuit. The diodes are electrically equivalent to a resistor of about .1 ohm. I make up this resistor by connecting 10 one-ohm resistors in parallel. The voltage that appears across the resistor during a pulse is measured by an oscilloscope that can respond to a frequency of at least 200 megacycles. The current is



Details of slit



Diffraction pattern from a single slit



Refractive indexes for filter

calculated by Ohm's law: Current equals voltage divided by resistance.

"Several techniques are available for detecting the invisible beam of the lasers. The wavelength is quasimonochromatic and reaches peak intensity at about 9,000 angstroms. It turns out that the sensitivity of silicon phototransistors such as the type designated HEP 312 also peaks at this wavelength. I pick up the beam with this transistor and boost the resulting output signal by either of two simple amplifiers [*see middle and bottom illustrations on page 115*].

"The output of the amplifiers can be used for driving a power amplifier or a set of earphones or for triggering apparatus of other kinds. I use the two-stage amplifier in most experiments. The more sensitive three-stage amplifier is useful for detecting faint signals, particularly for picking up the beam at a substantial distance from the laser diode.

"The lasers emit the beam at a diverging angle of approximately 20 degrees. The angle can be reduced to about six minutes of arc by placing the diode at the focus of a simple lens with a focal ratio of $f/2.8$ or less. The aperture of the lens need not be large. An $f/2.8$ lens of half-inch aperture placed 1.4 inches from the diode is equivalent to and just as effective as a one-inch lens placed at a distance of 2.8 inches. Small, simple lenses of good quality are available from suppliers such as the Edmund Scientific Co. (Barrington, N.J. 08007) for about $1.

"The emission can also be photographed with conventional infrared film that is available from dealers who stock a full line of photographic supplies. To photograph the wave nature of the radiation, place the edge of a safety razor blade at a right angle to the beam and about one centimeter from the diode so that the edge intercepts about half of the beam. In a darkened room mount a sheet of infrared film in line with the blade and the diode at a distance of about 25 centimeters. The emulsion side of the film should face the laser. A laser diode of the M4L 3052 type will expose the film adequately in about two seconds. The accompanying photograph [*preceding page*] depicts the resulting diffraction pattern. A classical diffraction pattern can be similarly photographed by projecting the beam through a narrow slit, which can be improvised from a pair of razor blades.

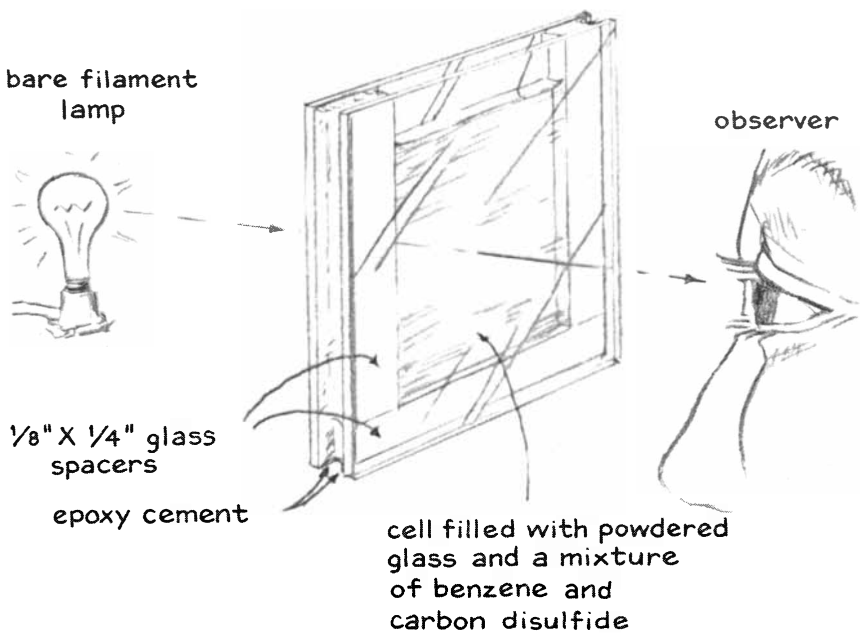
"The cost of laser diodes varies with their size. Those of the smallest size have been advertised recently by dealers in surplus materials at a price of $5. The following companies can give additional information and prices concerning new diodes: Laser Diode Laboratories (205 Forest Street, Metuchen, N.J. 08840), Marketing Manager, Electro-Optical Devices, RCA Industrial Tube Division (New Holland Pike, Lancaster, Pa. 17604) and Texas Instruments Incorporated (P.O. Box 5012, Dallas, Tex. 75222).

"Finally, a word of warning: *The emission of all lasers, including that of diode lasers, is hazardous.* Infrared emission is particularly insidious because it is invisible and can be reflected by almost any smooth surface, including objects of tarnished metal and other materials that bear little resemblance to conventional mirrors. The beam can be transmitted safely for substantial distances only if it is directed through a metal pipe or a like enclosure. At shorter distances, up to a meter, the entire apparatus should be enclosed in an opaque box that will absorb the unwanted energy. Never operate the unshielded laser in the presence of other people."

In the latter part of the 19th century the Danish physicist Christian Christiansen devised a remarkable filter for isolating a narrow band of frequencies in the infrared portion of the spectrum as well as single colors of visible light. The device is unusual because it employs only clear transparent materials and is one of the few kinds of filter that work in the infrared.

The filter consists essentially of a flat cell of glass that contains a slurry of powdered glass or a comparable substance in a fluid that has a matching index of refraction at some point in the optical spectrum. Powdered glass in air scatters light in all directions. It appears white and opaque just as snow does. The same scattering effect is observed when the particles are surrounded by a liquid of a different refractive index, such as water.

It is possible to immerse the particles in a fluid of matching refractive index. The slurry then appears clear. This stratagem is used for identifying certain kinds of glass. For example, Pyrex brand 7740 glass has a refractive index of 1.474. So does a solution that consists of 16 parts by volume of methyl alcohol in



Elements of Christiansen filter

84 parts of benzene. A piece of Pyrex vanishes when it is dipped in this solution, but glasses of a different refractive index look the way they do when they are immersed in water.

In the Christiansen filter a solution is compounded that matches the refractive index of the glass at a desired point in the spectrum but differs increasingly at all other points. The resulting filter is transparent to light of the selected wavelength and relatively opaque to all other wavelengths. A demonstration cell can be made with two sheets of ordinary glass, three glass strips about an eighth of an inch thick and a few dabs of epoxy cement. The strips serve as spacers at the sides and bottom of the cell [*see bottom illustration on opposite page*].

Seal the edges with epoxy. To demonstrate the filtering action with white light, fill the cell with powdered borosilicate glass. Make the powder by breaking the glass into small pieces, which can be done safely by putting a large piece of glass in a bag of heavy paper and striking it with a hammer. Put the fragments in a mortar, a few at a time, and grind them to fine powder with a pestle. Avoid inhaling the powder.

For about five minutes stir the powder into a cleaning solution of dilute hydrochloric acid, made by adding one volume of acid to three volumes of distilled water. Let the powder settle, pour off the acid and wash the particles in three changes of distilled water. Spread the powder on filter paper to dry. Transfer enough dry powder to the cell to fill it approximately three-quarters full. Add carbon disulfide to immerse about two-thirds of the glass particles.

Stir the mixture gently to release trapped bubbles of air. Place a 100-watt incandescent lamp about 10 feet beyond the cell and examine it through the glass slurry. With a pipette add benzene to the cell a few drops at a time. Stir the solution gently but thoroughly. Examine the lighted lamp through the cell after each addition of benzene. Eventually the lamp will appear dark red. Continue to add benzene. The colors of the transmitted rays will progress through the spectrum as benzene is added.

To convert the filter for infrared, clean the cell and refill it with a slurry of magnesium oxide in carbon tetrachloride. The transmission maximum of this combination is at a wavelength of 9,000 angstroms. The fumes of all three of the required liquids are highly toxic and must not be inhaled. The filter should be prepared in a fume hood and used only in a well-ventilated room.

BOOKS

The evolution of the computer, a new life of Einstein and pioneers of cosmography

by Philip Morrison

A COMPUTER PERSPECTIVE, by the office of Charles and Ray Eames. Edited by Glen Fleck. Harvard University Press ($15). THE COMPUTER FROM PASCAL TO VON NEUMANN, by Herman H. Goldstine. Princeton University Press ($12.50). The big dollar bills in circulation in Theodore Roosevelt's day have long been superseded, yet they have a geometrical immortality: the ubiquitous IBM punch card of today retains pretty closely the dimensions of the old notes. Herman Hollerith, "the first 'statistical engineer,'" chose that size for his cards when he moved into the commercial market a machine system he had devised for the 1890 U.S. Census. (The choice saved money in parts for his equipment.) The evidence is visual: on one page of period photographs in *A Computer Perspective* is Hollerith himself (high-collared and a little stern), the shop floor of his factory, one of his advertising leaflets, the ceiling-high file drawers of an insurance firm (evidently a prospective customer) and a few punch cards surrounding a dollar bill. With some 150 such carefully arranged pages of pictures and with densely printed captions and full documentation, the gifted Eames design office has pursued the history of computers. (The book is the offshoot, rather than the direct record, of a recent exhibition.) It takes up the story with Charles Babbage and leaves off with the untidy, tube-filled racks that were the computers of 20 years ago, fully modern in general function and design, almost all of them the progeny of the wartime ENIAC. The story since that time is told rather superficially in a few pages of text; computers are no longer individuals with names but a phylum of many species, rapidly evolving under selection pressures, products of an industry with an income of some $10 billion a year in the U.S. alone.

The Eameses' method is both penetrating and polished. They present artifacts and photographs and graphic materials of every kind that touch the main actors and their handiwork, with a happy facility for the evocation of context. They see the computer as arising out of the flowing together of three distinct realizations of mathematical relations: logical automata, statistical machines and calculators—these last including both digital-arithmetic machines and the more complex analogue devices of past and present. The immediacy of the visualization cannot fail to fascinate anyone interested in these high matters, and the thoroughness of the work means that the reader can expect fresh knowledge even when he knows the plot.

Here is Florence Nightingale, graphing in what she called "coxcombs," shaped like wind roses, the "causes of mortality in the army in the east." Out of the repetitive handling of a great many simple numbers the stimulus for machine calculation quite directly arose in a competition for handling the 1890 Census. Another page shows the 1893 Millionaire, a brass box of gears from Zurich that was the first commercially successful multiplying calculator, with its trolley-like controller. It spread through banks and laboratories for 40 years. Requiring "one turn of the crank" for each figure in the multiplier, the Millionaire made its progress most audible. (This reviewer multiplied for long, noisy hours with such a machine in the 1930's.)

Here are the first pages of the papers of Alan Turing and of Emil Post, one from Cambridge and one from C.C.N.Y., each demonstrating independently in 1936 that a long, long taped succession of machine-like "primitive acts" can solve any logical problem. The same five simple skills (mark, erase and so on) were specified for Turing's ideal machine and for Post's "worker." Here is D. R. Hartree's first differential analyzer, built out of "about £20 worth of Meccano parts" on the model of Vannevar Bush's integraph at M.I.T. It produced atomic wave functions. In 1936 Konrad Zuse built a relay calculator in his parents' living room in Berlin; by the end of the war he had independently found the way to the floating decimal, program control by tape and binary arithmetic.

Claude Shannon in an M.I.T. thesis and George Stibitz of Bell Laboratories working on his kitchen table had gone the route to binary arithmetic, reducing computing to logic quite independently in the same years. It is all here, the people and their handwriting, their books, their tape reels, an entire museum in the flat. One cannot overlook Ada Augusta, the Countess of Lovelace, whose expressive and detailed contemporary account of the Babbage proposals provides the best insight into that pioneer of computing. Her successors are here too: Adele Goldstine and Grace Hopper, who respectively programmed ENIAC and Mark I in the World War II years, breaking the trail that is now a busy highway.

The Computer from Pascal to von Neumann is more conventional, a narrative history rather than a paginated museum. Herman Goldstine is himself a pioneer of the computer, a wartime mathematician-officer at the Army's Aberdeen proving ground, where the heavy demand for lengthy ballistic tables, done anew for each new projectile, had long stimulated a farseeing leadership in the search for new aids to calculation. Dr. Goldstine writes with disarming candor and good humor. The first portion of his book briefly treats the entire story of computers (although the period from the Antikythera machine to Babbage whirs by in eight pages and three photographs). The level here is that of a general chronicle, but there are a couple of attractive digressions of a more demanding mathematical kind, such as one that tells the story of the "Gibbs phenomenon." This strange behavior of a systematic approximation procedure that is fundamental to mathematical physics came to light at the turn of the century in the operation of a specialized analogue computer made for A. A. Michelson. Willard Gibbs clarified the issue, which is still labeled with his famous name although it turns out that one Henry Wilbraham had fully explained it in print 50 years earlier.

After 1941 the stance of the author changes. No longer do we read a cheerful light gloss over the men and events of centuries. We follow the experience of Dr. Goldstine as he plays his part, mathematical, personal, managerial, in the events that led from Aberdeen by way of the Moore School of Electrical Engineering at the University of Pennsylvania to the Institute for Advanced Study and to the book's effective close about 1952.

The tale is now closely detailed, citing memoranda, letters, committees, actions hoped for and feared, contracts, rivalries. The principals are numerous, but three men stand out from the rest; theirs is the main thread of event and idea. First of all is John von Neumann. He was a mathematician of the widest interests and most powerful productivity, a rotund person of humor and wit, of glittering acuteness and glowing energy of mind; so worldly, so closely coupled to the "military-industrial complex" that he was a member of a dozen acronymic consultant boards to the Air Force, Navy, Army, Atomic Energy Commission and White House in the very years he was president of the American Mathematical Society. He enters the narrative in the wartime summer of 1944 most characteristically: "On the railroad platform in Aberdeen... along came von Neumann.... I approached this world-famous figure, introduced myself, and started talking.... The conversation soon turned to my work... the development of an electronic computer capable of 333 multiplications per second."

From about that time until von Neumann's death in 1957 no problem held sway in his mind as much as the computer, its logic, its realization and its use. Soon enough he had helped the ENIAC group move on (ENIAC had yet to light its 18,000 cathodes) to a scheme called EDVAC. His draft report of 1945 contained the first worked-out stored program, an illustration for his general account of the logical design problems of computers. His clarity "crystallized thinking in the field of computers as no other person ever did."

The pioneers, though, were John W. Mauchly, physicist, and J. Presper Eckert, Jr., engineer. It was Mauchly who in 1941 at Aberdeen foresaw the indispensability of electronic techniques—no clocking relays or turning wheels—for fast automatic computation. It was Eckert, chief engineer of the ENIAC and EDVAC projects that were established at Pennsylvania, who pressed the right solutions to a dozen different problems. By 1946 the Pennsylvania team had quarreled over patents and split irrevocably. Von Neumann led Goldstine to the Institute for Advanced Study; Eckert and Mauchly went into business by themselves, "the first commercial enterprise" of the electronic computer industry. The postwar world of digital computers arose entirely from the work of this group.

The von Neumann path led to his work in the public domain, his IBM consultantship and the Institute's computer, which came into operation about 1952. Eckert and Mauchly, denied patents after von Neumann's public disclosures, went into manufacture and sold fully electronic computers before 1951. It was these two who had the first magnetic drums and the first compiler language, and who above all began to cater to widespread business use of the large-scale digital computer. On the other hand, it was the elaborate applied-science uses (the prototype being computer weather forecasting) that won the Institute's enthusiasm. By 1953, with the famous 701, "the day of the computer" had dawned at IBM. Now the giant company makes about 70 percent of all U.S. computing machinery.

Newton held that "Philosophy is such an impertinently litigious lady that a man had as good to be engaged in Law suits as have to do with her." It is not surprising that the issues of priority surrounding the computer generated "very great heat"; they were joined at the seedtime of a flowering industry. The general reader cannot be a judge, certainly not by the canons of patent law. Both of these books are sponsored by IBM and both are scrupulously fair, although they certainly tend against the proprietary uniqueness of ENIAC and thus add legitimacy to the status quo. The parable of the talents may be enigmatic, but it is founded on shrewd observation: Unto every one that hath shall be given, and he shall have abundance. At the same time the electronic digital computer surely owes a real debt to the past, like almost every development of science and technology. In the world as it is those swift, silent vacuum tubes at the Moore School of Electrical Engineering are entirely obsolete; in the very years when they were being lighted, the imperfect crystals of germanium and silicon were first growing in the continuous furnaces of history.

ALBERT EINSTEIN: CREATOR AND REBEL, by Banesh Hoffmann, with the collaboration of Helen Dukas. The Viking Press ($8.95). "The letter killeth, but the spirit giveth life." This brief, loving, intimate and often poetic volume transmits the spirit of Einstein and his work with a fidelity easily lost in the noise of the enormous documentation that has begun to mask that wry, profound and simple man. Banesh Hoffmann is a mathematician and relativist who worked with Einstein for years in direct and successful collaboration, and Helen Dukas was Einstein's knowing and devoted secretary from the last Berlin years until his death in the hospital in Princeton.

The pair bring two gifts to their readers. The mathematician, an expert pedagogue to the nonmathematical, has constructed a generally quite happy set of simple analogies and parallels to make plain the meaning of the main junctures in Einstein's life of science, whereas his partner offers her close testimony over decades in illumination of what the man meant and felt. Citations are ample and frequent, and the photographs combine the classically familiar with the fresh and unique. Consider Einstein's family and himself in silhouettes he cut out in 1919, or the postcard sent from Paris by two old men—whom we see in an earlier photograph with Einstein when all three friends had been poor students in Bern 50 years before, before the quantum, before the end of simultaneity, before the holocaust, before the bomb.

The book begins with an acute glimpse of Einstein by way of his own autobiographical notes of 1949 (in *Albert Einstein: Philosopher Scientist*, edited by Paul A. Schilpp), an essay so scrupulously intellectual, although both candid and clarifying, that it mentions no private matters at all. Those notes tell of the sense of wonder he felt as a boy when he first saw a magnetic compass, but they do not mention "that his father, who had shown him the compass, was named Hermann." This book at hand is not so austere. It presents a full enough narrative of his life, now and again dwelling on the issues that his mind found central, those that indeed changed our world once and for all.

The text is without algebra; all the explanations are aimed squarely at the lay reader. Although they have limitations, they broadly succeed. Indeed, we see only three pages of algebra, all in holograph reproductions of Einstein's own *x-erei*. One is lecture notes written in Berlin in 1918, the page marked "Canceled because of revolution"; one bears on the reverse an unpublished quatrain on Newton in Einstein's German hand;

the third, curiously touching, is the last this "hardened *x*-brother" ever calculated, with long strings of tensor elements firmly laid down by that hand at 76, during his terminal illness. There is painfully hard work at the foundation of the lofty conceptual structures of the century's greatest physicist.

The year of wonder, 1905, is well treated. That was the year of the five papers submitted to the *Annalen der Physik*, four before July, by a technical examiner in the Patent Office, a young physicist of obviously deep ability. One of the papers, which formed his Ph.D. thesis, derived Avogadro's number from diffusion in liquids; it was the least important, and it appeared in 1906. One earned a Nobel prize; it insisted that the quanta of light suggested by Planck be taken seriously, that these particles made sense out of two fluorescence effects and certain curious experiments of Philipp Lenard (who ended up virulently Nazi) on the ejection of electrons from metal surfaces by light.

That summer Einstein wrote to a friend, a teacher of mathematics in Schaffhausen: "I promise you four papers in exchange.... The first... is very revolutionary." One of the other two explained quantitatively the random motion of dust specks under the microscope—the famous Brownian motion—and provided the best evidence for the reality of molecular motion. (Marian von Smoluchowski did the same, perhaps a little less in detail, independently a few months later.) The fourth was the first seminal paper on relativity. "Einstein was not yet done with 1905." He sent off a three-page fifth paper in late September. It showed that emitted light carries away mass from any source, and by 1907 he had it succinct and clear: $E = mc^2$. "Imagine the audacity of this step: every clod of earth, every feather, every speck of dust becoming a prodigious reservoir of entrapped energy."

Those with some taste for the *x*-language may find this book here and there at too simplified a physical-mathematical level, but the general reader can gain from it an honest "indication of the man" both by external event and by inward idea. We owe him that kind of attention. Beyond this volume the prepared seeker for Einstein can go next to the letters with Max Born, a lifelong exchange that was published in 1971 (and reviewed here) and then to the autobiographical notes. Hoffmann and Dukas have cleared a secure starting place.

"But we are shirking what cannot be shirked: on 6 August 1945 an atomic bomb was exploded over Hiroshima. Einstein's secretary heard the news on the radio. When Einstein came down from his bedroom for afternoon tea, she told him. And he said, '*Oh weh*,' which is a cry of despair whose depth is not conveyed by the translation 'Alas.' "

A NAVIGATOR'S UNIVERSE: THE LIBRO DE COSMOGRAPHÍA OF 1538, by Pedro de Medina. Translated and with an introduction by Ursula Lamb. The University of Chicago Press ($18.50). A COMMENTARY ON THE DRESDEN CODEX: A MAYA HIEROGLYPHIC BOOK, by J. Eric S. Thompson. American Philosophical Society ($25). Let us now praise famous Copernicus, not by any further study of his life or works—surely enough has been said already—but by seeking a richer grasp of what the human mind knew of the rounds of the heavens before his revolution.

While the church servant Copernicus lived out his days around the cathedral in Frombork, a small town on a nearly landlocked arm of the Baltic, another inland port across Europe mushroomed in size and wealth. That was Seville, through whose Casa de Contratación all goods and passengers from the Americas had by royal command to clear. All Spanish ships from the New World came up the difficult channel of the Guadalquivir to that city. The riches of the Americas poured into Seville for all Spain and all Europe.

Keen swords, strong faith, stout ships and brave, ruthless men were necessary to that trade. The fact remains that they were insufficient. The sea was trackless, and "how to fix a place on the globe" and "how to lay a course across the oceans to a distant landfall" were no less essential. Those skills were not yet common. They were based on theory: the theory of the earth sphere and the stars, the compass and the geometry of arc and triangle. Therefore there grew up in Seville the scientific office of the pilot major, which licensed and controlled the "pilots and masters and the scientific and nautical equipment employed" in the trade with the Indies.

One Pedro de Medina was granted a license in 1538 "to make such charts and instruments for sale in Seville." It was a new career for a man of 45 whose work had been that of private tutor to a noble house, who for 20 years had himself been a student of nautical science. "Cosmography is the description of the world," begins his book, by which he sought to become a licensed "cosmographer to his majesty." The profession was well established; there were university chairs, frequent calls to law court and royal councils—perhaps to part the globe by fixing a line and so to call the "score in the game of kings," as Dr. Lamb puts it.

Medina was a skillful tutor and a diligent examiner; he became a formal teacher of pilots and helped to build up an academic curriculum for that crucial profession. He even wrote a guidebook to imperial Spain. Throughout Europe his textbooks spread. His masterwork was his *Arte de navegar,* a technical, professional work meant for practical pilots and used for a century in several languages. ("In 1871 a copy of the Dutch translation . . . was found preserved in the ice on the track of William Barents' third voyage of 1596 from Spitsbergen.")

The Medina work that has now been reproduced in black-and-white facsimile (so that unfortunately we lose all the red initials and rulings and the four colors of the diagrams of the 32 points of the compass) was his first; it was among the materials he submitted when he sought his cosmographer's license. It was never printed; the copy we have here is a manuscript written in a clear Gothic hand, bound in velvet (perhaps he had hoped the emperor, Don Carlos, to whom it was dedicated, would see it) and held since 1817 in the Bodleian Library at Oxford. It is an introductory work of theory rather than practice, written with clarity and brevity in a sensible manner, in the form of a dialogue. An educated layman and a pilot put 82 questions to the cosmographer, who answers them all. The translation of the Spanish text completes the handsome book.

The tenor of the material is one of broad background, interesting and sometimes utilitarian questions concisely put and neatly answered, less directly useful day by day at sea than as a framework in which landlubber and ship's officer alike might set what they heard and experienced. The picture is matter of fact, confident, on the whole rather simple today, reading a little like the geography book of a generation back. The globe of the earth is the center of the work, literally and metaphorically. Its poles, zones, antipodes, weather, mountains and valleys are discussed. Why is the sea salt? The sun's heat evaporates the water. What color has the sky? None; at night we see only stars, and for the rest "the eye produces its own color." How do the antipodes walk? "Between their feet and ours there is the round body of the earth." What is the rainbow? Some say it is the image of the sun in a watery cloud. The circuit of the North Star around the pole, which provided both a star clock and a needed correction to the measurement of latitude, is carefully described with charts and mnemonics ("Those four parts we compare to the body of a man"). Four centuries ago Polaris was "sometimes 3° below [the pole]"; today the slow precession has brought it to within one degree.

The sky has changed noticeably since Pedro de Medina, but the modern view of the world was already his, less perhaps in substance than in style. The elements and the spheres are still with him, very Greek, but they are rather distant and theoretical. Planets are of little interest; Copernicus was not needed to seize the Indies. Even in our time navigation needs only a geocentric theory, sober practical judgment and a firm hold on geometry; Pedro would still be at home teaching navigators. Yet his works fell behind the growth of skill, and by the time our cosmographer grew old his texts were somewhat outmoded. The great convoys had few pilots and brought back little new data. "Routine had overtaken exploration" and "Spain's colonial enterprise had aged with him." He remained skeptical about the deviation of the compass from geographic north, although late in life he recommended that pilots on the Indies run carry two compasses, one corrected for variation and one not, "so that many useful things could be found out." The editor and translator does not see our universe in that of the cosmographer, but another reader may find it there, a little strange but recognizable in approximation; it is the same well-lighted place before it was so elaborately redone.

A Commentary on the Dresden Codex also presents a facsimile manuscript, this time in color but now from a world dimly seen, not simple at all and necessarily viewed through an editor-translator's intellectual apparatus of remarkable ingenuity and complexity. For the manuscript is Mayan. It probably reached Europe (we do not know for sure) with the gifts of Moctezuma sent by Cortés to his king in 1519 at the bright dawn of Spain's power in the Indies. It is a strip about 3.5 meters long and 20 centimeters wide, made of a paper derived from the pounded inner bark fibers of a wild species of fig, coated with a sizing of fine lime. The piece is screen-folded into 39 leaves, each about nine centimeters across. Each was painted neatly on both sides with fine brushes, often in polychrome gold, red, brown, blue, yellow and black; the work of eight artists has been distinguished. The work has been in the royal library in Dresden since the director bought it in Vienna (where Charles V normally resided) in 1739. It was seriously damaged by water before it was acquired, so that almost every page has lost a triangular area; water damage was again severe during World War II. The facsimile presented here therefore consists of black-and-white photographs of a chromolithographic edition published in 1892, with surprinted overlays in color made for this edition and with some restorations based on comparison with a reproduction painted and published in the 1830's and 1840's.

Only three such samples of Mayan writing before Columbus exist, and the Dresden Codex is "artistically... far superior to the Madrid and Paris codices." Its miniatures of gods, beasts and men and its rows of glyphs and tally digits (base 20) are strangely beautiful. The experts read the eclipse tables and look at the drawings of incense burners and throwing sticks and arrive at a date for the Codex some six or seven centuries back.

What this volume provides, besides the visual wonder of the codex, is a close look at the struggle to interpret Mayan writing. The author, dean of the students of this marvelous enigma, has presented his case in rather full detail, with plenty of self-conscious critique of his colleagues and himself. There is a good deal we are not sure about. He sees mere reckoning aids, used by the priests to extend their predictions, where some writers (even recent ones) have seen tables for the motions of Mars or Jupiter! Nevertheless, all seem to agree—and the general reader is pretty well convinced by the coherence of the evidence—that a half-dozen pages in this codex are in fact tables of the fearful helical risings of Venus, ingeniously corrected to achieve an accuracy that amounts to an error of one day in 6,000 years. It is to be remarked that this extreme accuracy requires a plausibly argued assumption of a scribe's error at one point, but an accuracy that runs to mere days in 500 or 600 years is directly readable in the tables.

Much of the text describes and interprets the codex line for line and page by page. It is not manageable by a general reader, but taking a sample of the arguments, pencil in hand, will help such a reader become familiar with the successes of this science and with its frailties. The overall text, with a long account of Mayan lore and the colonial literature, makes it plain that the codices we have—all of which are priest-astronomer's tables, accompanied by the divinatory conclusions, generally pretty gloomy—were not at all the only written documents of the Maya.

It is hard to avoid the conclusion, although it is perhaps too naïve, that Mayan sky lore was about at the stage of the cuneiform tables we find from Babylon, two or three millenniums older. This is no spatial world like that of the navigators at all. In space the Mayan world was only the blue sky over the green plain. It was a world in time, great cycles beating incessantly, now in phase, now out. From such rhythms wisdom, experience and no doubt revelation could foretell the fates of crops and men. So it was along the Yellow River and in the valley of the Euphrates long before. "Life on celestial and terrestrial planes moved as a complex machine without controls by gods or men." The dance of the 20 digits past the 13 months went on.

There is more to do, much more, although Professor Thompson, who wears his learning most engagingly, holds with some pride that we have the meaning of most of the codex glyphs, which make up more or less mnemonic records. They were written to aid understanding of the all-important cyclic machine of the heavens, and to supply key interpretations that a priest could expand from a few dozen glyphs into a holiday message or a sermon. The three codices are all we know, however, and they are not typical. "Why, O why did not Diego de Landa take with him to Spain two or three Maya codices to illustrate his treatise?" The bishop is defended by our author. True, Landa writes that he burned "a great number of books of their letters ..., which... caused them great pain," but there are no other records of book destruction. Landa had found "evidence of the sacrifice of children... by the... converted Maya," and he had to eliminate "the old pagan structure" if such sacrifice were to be ended. We cannot judge him by 20th-century attitudes.

One fears that the unity of humankind for good and evil is all too plain. Under every sky we record the dance of the sky beings, burn and ignore strange books and still sacrifice the children.

RATTLESNAKES: THEIR HABITS, LIFE HISTORIES, AND INFLUENCE ON MANKIND, by Laurence M. Klauber. University of California Press ($50). When the first edition of this large, lighthearted and learned two-volume monograph appeared in 1956, this journal wrote of it: "Says everything on the subject that an amateur or professional snake fancier or snake dreader would ever want to know." One can repeat that firmly for the second edition, which indeed is not very much changed. Mr. Klauber completed his important changes in the taxonomic and distribution chapters before his death in 1968. The work has long been out of print; the publishers have done well to reissue the improved version as it left his hand. It is a model of a complete natural history. Its biology is based on critical reading, on letters and interviews and on lifelong field experience. Nor does it neglect the practical—bites and their treatment, control and utilization of the snakes—or the large and fascinating body of folklore, ancient and modern. Giant rattlers? "I should say that rattlesnakes ought to be placed in the mythical category at 10 feet, thus allowing for some exaggeration." Be wary of stretched skins and exhibits with rattler-labeled pythons and boas. Photographs of a rattler striking a rabbit? Well, it surely happens, but the picture so widely reproduced may "attain the durability of first-rate folklore." The stuffed beasts in the easily recognizable pose are found in a museum in San Antonio.

"That the rattlesnake bears an especial enmity toward man is mythical. It seeks only to defend itself.... It could not, through the ages, have developed any... since the first human being any rattlesnake may encounter is usually the last."

GALAXIES, by Harlow Shapley. Revised by Paul W. Hodge. Harvard University Press ($10). Concise, personal and exciting, this book originally written during World War II can stand as a memorial in these columns to its author, who died in October, 1972, full of years, among the liveliest and most engaged minds in a generation of American scientists. This third edition is no relic; it has been brought quite up to date, losing little of the old flair, by a well-known and active worker in the field, which is what Shapley would have welcomed. About a third of the text is new since Shapley's reworking of 1961 (which brought radio in for the first time), and the photographs have a strong new component. The text remains Shapley's in the large, with his warmth and energy together with a degree of Harvardocentrism. It remains a very good start on the subject, strongest in its account of the nearby galaxies. Professor Hodge has added an excellent sketch of the newly recognized violent events in galaxies, such as the quasars. Only the snorting, ring-tailed anomalies that H. C. Arp of the Hale Observatories has recently brought to everyone's attention are invisible here.

INDEX OF ADVERTISERS

MARCH 1973

BIBLIOGRAPHY

Readers interested in further reading on the subjects covered by articles in this issue may find the lists below helpful.

THE PROSPECTS FOR A STATIONARY WORLD POPULATION

REFLECTIONS ON THE DEMOGRAPHIC CONDITIONS NEEDED TO ESTABLISH A U.S. STATIONARY POPULATION GROWTH. Tomas Frejka in *Population Studies,* Vol. 22, No. 3, pages 379–397; 1968.

UN TAUX D'ACCROISSEMENT NUL POUR LES PAYS EN VOIE DE DÉVELOPPEMENT EN L'AN 2000: RÊVE OU RÉALITÉ? Jean Pichat-Bourgeois and Si-Ahmed Taleb in *Population,* No. 5, pages 957–974; September–October, 1970.

THE GROWTH AND STRUCTURE OF HUMAN POPULATIONS: A MATHEMATICAL INVESTIGATION. Ansley J. Coale. Princeton University Press, 1972.

POPULATION AND THE AMERICAN FUTURE. United States Commission on Population Growth and the American Future, 1972.

THE FINE STRUCTURE OF THE EARTH'S INTERIOR

INTRODUCTION TO THE THEORY OF SEISMOLOGY. Keith E. Bullen. Cambridge University Press, 1963.

THE DENSITY DISTRIBUTION NEAR THE BASE OF THE MANTLE AND NEAR THE EARTH'S CENTER. Bruce A. Bolt in *Physics of the Earth and Planetary Interiors,* Vol. 5, pages 1–11; 1972.

OBSERVATIONS OF PSEUDO-AFTERSHOCKS FROM UNDERGROUND EXPLOSIONS. Bruce A. Bolt and A. Qamar in *Physics of the Earth and Planetary Interiors,* Vol. 6, pages 100–200; 1972.

THE VISUALIZATION OF GENES IN ACTION

PORTRAIT OF A GENE. Oscar L. Miller, Jr., and Barbara R. Beatty in *Journal of Cellular Physiology,* Vol. 74, No. 2, Part 2, Supplement 1, pages 225–232; October, 1969.

ELECTRON MICROSCOPIC VISUALIZATION OF TRANSCRIPTION. O. L. Miller, Jr., Barbara R. Beatty, Barbara A. Hamkalo and C. A. Thomas, Jr., in *Cold Spring Harbor Symposia on Quantitative Biology,* Vol. 35, pages 505–512; 1970.

VISUALIZATION OF BACTERIAL GENES IN ACTION. O. L. Miller, Jr., Barbara A. Hamkalo and C. A. Thomas, Jr., in *Science,* Vol. 169, No. 3943, pages 392–395; July 24, 1970.

MORPHOLOGICAL STUDIES OF TRANSCRIPTION. O. L. Miller, Jr., and Aimée H. Bakken in *Acta Endocrinologica,* Supplementum 168, pages 155–173; 1972.

VISUALIZATION OF RNA SYNTHESIS ON CHROMOSOMES. O. L. Miller, Jr., and Barbara A. Hamkalo in *The International Review of Cytology,* Vol. 33, pages 1–25; 1972.

INTERSTELLAR MOLECULES

INTERSTELLAR MOLECULES AND DENSE CLOUDS. D. M. Rank, C. H. Townes and W. J. Welch in *Science,* Vol. 174, No. 4014, pages 1083–1101; December 10, 1971.

INTERSTELLAR MOLECULES. B. E. Turner in *Radio Astronomy,* edited by G. L. Verschuur and K. I. Kellermann. Springer-Verlag, Inc., in press.

INTERSTELLAR MOLECULES AND THE INTERSTELLAR MEDIUM. Edited by M. A. Gordon and L. E. Snyder. Springer-Verlag, Inc., in press.

THE ASYMMETRY OF THE HUMAN BRAIN

FUNCTIONAL ASYMMETRY OF THE BRAIN IN DICHOTIC LISTENING. Doreen Kimura in *Cortex,* Vol. 3, No. 2, pages 163–178; June, 1967.

SPATIAL LOCALIZATION IN LEFT AND RIGHT VISUAL FIELDS. Doreen Kimura in *Canadian Journal of Psychology,* Vol. 23, No. 6, pages 445–458; December, 1969.

HEMISPHERIC SPECIALIZATION FOR SPEECH PERCEPTION. Michael Studdert-Kennedy and Donald Shankweiler in *The Journal of the Acoustical Society of America,* Vol. 48, No. 2, Part 2, pages 579–594; August, 1970.

RIGHT HEMISPHERE SPECIALIZATION FOR DEPTH PERCEPTION REFLECTED IN VISUAL FIELD DIFFERENCES. Margaret Durnford and Doreen Kimura in *Nature,* Vol. 231, No. 5302, pages 394–395; June 11, 1971.

BICYCLE TECHNOLOGY

RIDING HIGH: THE STORY OF THE BICYCLE. Arthur Judson Palmer. E. P. Dutton & Co., Inc., 1956.

HANDBOOK OF THE COLLECTION ILLUSTRATING CYCLES. C. F. Caunter. Her Majesty's Stationery Office, 1958.

WHEELS WITHIN WHEELS: THE STORY OF THE STARLEYS OF COVENTRY. Geoffrey Williamson. Geoffrey Bles, 1966.

THE MIGRATIONS OF THE SHAD

ATLANTIC COAST MIGRATIONS OF AMERICAN SHAD. G. B. Talbot and J. E. Sykes in *Fishery Bulletin,* Vol. 58, pages 473–490; 1958.

FISH MIGRATION. F. Harden Jones. St. Martin's Press, Inc., 1968.

WATER TEMPERATURE AND THE MIGRATIONS OF AMERICAN SHAD. W. C. Leggett and R. R. Whitney in *Fishery Bulletin,* Vol. 70, No. 3, pages 659–670; July, 1972.

THE BEHAVIOR OF ADULT AMERICAN SHAD (*ALOSA SAPIDISSIMA*) DURING MIGRATION FROM SALT TO FRESH WATER AS OBSERVED BY ULTRASONIC TRACKING TECHNIQUES. Julian J. Dodson, William C. Leggett and Robert A. Jones in *Journal of the Fisheries Research Board of Canada,* Vol. 29, No. 10, pages 1445–1449; October, 1972.

THE ORIGINS OF NUMBER CONCEPTS

THE CHILD'S CONCEPTION OF NUMBER. Jean Piaget. Humanities Press, 1964.

CONCEPT DEVELOPMENT. John H. Flavell in *Carmichael's Manual of Child Psychology,* edited by Paul H. Mussen. John Wiley & Sons, Inc., 1970.

MATHEMATICAL AND BEHAVIORAL FOUNDATIONS OF NUMBER. Charles J. Brainerd in *Journal of General Psychology,* in press.

MATHEMATICAL GAMES

THE ART OF NUMBRING BY SPEAKING-RODS: VULGARLY TERMED NEPEIR'S BONES. W. Leybourn. London, 1667.

NAPIER TERCENTENARY MEMORIAL VOLUME. Edited by Cargill Gilston Knott. Longmans, Green and Company, 1915.

THE AMATEUR SCIENTIST

PROCEDURES IN EXPERIMENTAL PHYSICS. John Strong in collaboration with H. Victor Neher, Albert E. Whitford, C. Hawley Cartwright and Roger Hayward. Prentice-Hall Inc., 1938.

GALLIUM ARSENIDE LASERS. Edited by C. H. Gooch. John Wiley & Sons, Inc., 1969.

A NEW CLASS OF DIODE LASERS. Morton B. Panish and Izuo Hayashi in *Scientific American,* Vol. 225, No. 1, pages 32–40; July, 1971.

Buick's new Opel Manta.
2,000,000,000 miles in Europe. $2,817 in America.

Manta. $2,816.50 for the plush Manta Luxus with all the standard features listed on the price sticker below. Give up a little plush and Manta is $2,622.50. Or go sporty with Manta Rallye—$2,805.50. All prices are Manufacturer's Suggested Retail Prices, including dealer new vehicle preparation charges. Optional equipment, State and local taxes and destination charges are additional.

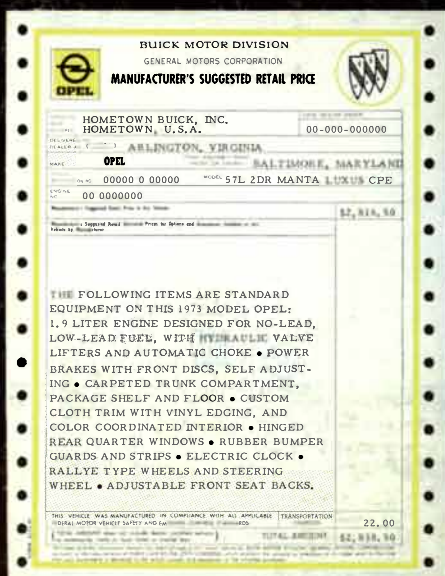
Manta. The personal-size sport coupe proven in more than 2 billion miles of actual driving. Not on test tracks or proving grounds but on European roads.

Manta. With rack and pinion steering for twisting Alpine curves. A 1.9 liter engine for responsive autobahn driving. Disc brakes up front. And a grand-touring ride you'll just have to feel.

Manta. Luxurious. Economical. Comfortable. Contour bucket seats that recline.

Interiors that even include corduroy fabric in the Luxus model.

The Opel Manta, Manta Rallye, Manta Luxus.
Really ready for America.


BUICK MOTOR DIVISION
GENERAL MOTORS CORPORATION
MANUFACTURER'S SUGGESTED RETAIL PRICE
OPEL

HOMETOWN BUICK, INC.
HOMETOWN, U.S.A.
00-000-000000
ARLINGTON, VIRGINIA
MAKE OPEL
BALTIMORE, MARYLAND
00000 0 00000
MODEL 57L 2DR MANTA LUXUS CPE
ENGINE NO 00 0000000
$2,816.50

THE FOLLOWING ITEMS ARE STANDARD EQUIPMENT ON THIS 1973 MODEL OPEL: 1.9 LITER ENGINE DESIGNED FOR NO-LEAD, LOW-LEAD FUEL, WITH HYDRAULIC VALVE LIFTERS AND AUTOMATIC CHOKE • POWER BRAKES WITH FRONT DISCS, SELF ADJUSTING • CARPETED TRUNK COMPARTMENT, PACKAGE SHELF AND FLOOR • CUSTOM CLOTH TRIM WITH VINYL EDGING, AND COLOR COORDINATED INTERIOR • HINGED REAR QUARTER WINDOWS • RUBBER BUMPER GUARDS AND STRIPS • ELECTRIC CLOCK • RALLYE TYPE WHEELS AND STEERING WHEEL • ADJUSTABLE FRONT SEAT BACKS.

TRANSPORTATION 22.00
TOTAL AMOUNT $2,838.50

German precision imported by General Motors.
Sold and serviced by 2200 Buick dealers.