SCIENTIFIC AMERICAN

THE WRIGHT "FLYER"

$1.50

July 1979

To create the ideal light, we had to improve on the original

To create eye-pleasing light, man has always looked to the sun for guidance.

At Westinghouse, we did it the other way around. We looked at the eye. And made such a dramatic discovery that we had to look away from the sun and start anew.

The sun and conventional lamps produce light made from all colors of the spectrum. But we found that our eyes respond best to light made from only three colors: red/orange, green and blue/violet. So that's the light our Ultralume™ fluorescent lamps produce.

Under Ultralume, colors appear richer. Textures and fine details seem clearer than ever before.

Ultralume is the first fluorescent that delivers the light our eyes prefer to see. Because once again, Westinghouse has looked at an old problem in a new light.

Westinghouse

A powerful part of your life.

Established 1845 SCIENTIFIC AMERICAN July 1979 Volume 241 Number 1

ARTICLES

DEPARTMENTS

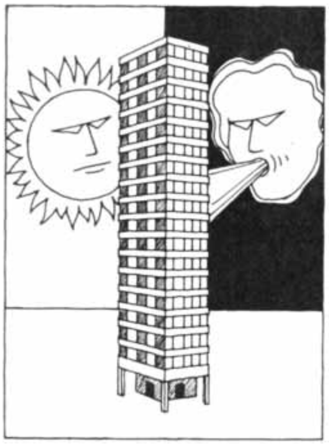
SCIENTIFIC AMERICAN (ISSN 0036-8733), PUBLISHED MONTHLY BY SCIENTIFIC AMERICAN, INC., 415 MADISON AVENUE, NEW YORK, N.Y. 10017. COPYRIGHT © 1979 BY SCIENTIFIC AMERICAN, INC. ALL RIGHTS RESERVED. PRINTED IN THE U.S.A. NO PART OF THIS ISSUE MAY BE REPRODUCED BY ANY MECHANICAL, PHOTOGRAPHIC OR ELECTRONIC PROCESS, OR IN THE FORM OF A PHONOGRAPHIC RECORDING, NOR MAY IT BE STORED IN A RETRIEVAL SYSTEM, TRANSMITTED OR OTHERWISE COPIED FOR PUBLIC OR PRIVATE USE WITHOUT WRITTEN PERMISSION OF THE PUBLISHER. SECOND-CLASS POSTAGE PAID AT NEW YORK, N.Y., AND AT ADDITIONAL MAILING OFFICES. AUTHORIZED AS SECOND-CLASS MAIL BY THE POST OFFICE DEPARTMENT, OTTAWA, CANADA, AND FOR PAYMENT OF POSTAGE IN CASH. SUBSCRIPTION RATE: $18 PER YEAR, U.S., ITS POSSESSIONS AND CANADA; $22 PER YEAR, ALL OTHER COUNTRIES. POSTMASTER: SEND ADDRESS CHANGES TO SCIENTIFIC AMERICAN, 415 MADISON AVENUE, NEW YORK, N.Y. 10017.

We're helping to build a quieter,

When our new DC-9 Super 80 enters service next year, it will be the quietest commercial jetliner ever to fly. So quiet that the high-noise area around airports will be cut in size to just one-fifth of what it is with today's comparable aircraft.

Quiet, and efficient, too. Airlines flying the Super 80 will pay the lowest operating costs per seat mile of any jetliner in its class.

Watch for the Super 80. You'll see it long before you hear it.



The best way to conserve valuable energy resources in a building is to start, literally, from the ground up. Which is exactly what architects are now doing—with help from our MCAUTO® Division.

Working with our computer programs, design engineers use weather data, peak demand estimates, lighting and other building and environment features to simulate the proposed building's heating and cooling requirements. These needs are then factored into the final building plans. The result is good building design combined with full energy efficiency.

more energy efficient,

THE COVER

The painting on the cover depicts part of the first powered, man-carrying airplane: the "Flyer" of 1903, designed and built by Wilbur and Orville Wright (see "The Origins of the First Powered, Man-carrying Airplane," by F. E. C. Culick, page 86). The view is from the rear of the craft and slightly to the left of center. The rear-mounted 8.5-foot propeller, one of two on the plane, was a crucial factor in the Wrights' success. They had to design the propellers themselves when they found that the published information on marine propellers was largely irrelevant. The propellers were so efficient that the Wright brothers were able to fly their airplane even with a fairly modest gasoline engine, which, like everything else, they built themselves. To the right of the propeller can be seen the footrest for the pilot, who flew prone, and the hip cradle with which he was able to warp the trailing edges of the wings up on one side and down on the other to execute a turn, a maneuver that the Wrights were the first to succeed at. Beyond the wings is the horizontal control surface. The forward location of this surface was a distinctive feature of the early Wright planes.

THE ILLUSTRATIONS

Cover painting by Ted Lodigensky

Page	Source
18	Ben Rose
20	Scott Kim
36–38	Alan D. Iselin
39	Albert Miller
40–41	U.S. Air Force
42–45	Alan D. Iselin
46–47	Albert Miller
49	Werner Henle
50	M. Anthony Epstein, University of Bristol Medical School (*top*); Ilil Arbel (*bottom*)
51–52	Ilil Arbel
53	Gunnar Bjursell, University of Aarhus
54–58	Ilil Arbel
59	Werner Henle
61	Hale Observatories (*top*); Eric E. Becklin and Gerry Neugebauer, California Institute of Technology and Hale Observatories (*bottom*)
62–63	George V. Kelvin
64	Hale Observatories
65–70	George V. Kelvin
86–87	Dan Todd
88–93	Library of Congress
94–99	Dan Todd
100	Dan Todd (*top*), Library of Congress (*bottom*)
103	Bunji Tagawa
104	Bunji Tagawa (*top*); Peter A. Lawrence, Medical Research Council, Cambridge (*middle and bottom*)
105–110	Bunji Tagawa
113–120	Walken Graphics
123	Simon Conway Morris
124	Simon Conway Morris (*top*), Tom Prentiss (*bottom*)
125–127	Tom Prentiss
128	Palaeontological Association, London (*top and bottom right*); E. Schweizerbart'sche, Stuttgart (*middle and bottom left*)
129	Albert Miller
130	Royal Society of London (*top and bottom*); Palaeontological Association, London (*middle*)
131	Ellis L. Yokelsen (*top*), Simon Conway Morris (*bottom*)
133	Simon Conway Morris and H. B. Whittington (*top*), Simon Conway Morris (*bottom*)
137	U.S. Air Force
140–151	Jerome Kuhl
152–160	Michael Goodman

No-Fault Radar

If you've ever been caught by radar or if you own a radar detector, please read this important message.

JS&A has never offered a radar detector.

As our president put it, "A radar detector is a flagrant anti-police device that does nothing but permit abuse of our traffic laws."

Although many devices were presented to JS&A, none were acceptable. Despite all of our efforts, our president stood firm, "Our company will not, under any circumstances, sell radar detectors."

For three years we saw radar detectors–some good, some bad–but because of our president's policy, we were unable to offer a single unit. We saw the units go to both X and K bands; we saw the police develop radar jamming devices; and we saw the FCC prohibit these jamming devices. We followed with great envy as other companies sold thousands of them while JS&A stood firm on its decision not to sell them.

In January of I979, our president was travelling on an interstate highway at 55 miles per hour. Other cars were passing him.

As he approached the top of a hill, he neglected to pay attention to his speedometer. As he rolled down the hill his speed increased to 63 MPH. At the bottom of the hill was a police radar trap.

He was apprehended and charged with exceeding the speed limit by eight miles per hour. He was taken to a Justice of the Peace who was in the barber shop, so our president had to wait until he finished. Finally there was a quick trial and a fine was paid.

Our president was four hours late. He felt that he was treated like a common criminal despite his good driving record and he lost very valuable time.

ATTITUDE CHANGES

This small incident created an entire change in his attitude. Our president saw for the first time that even law-abiding citizens are subject to the inequities of radar justice. He saw that the law-abiding citizen must also be protected from the abuses of radar power when unfairly used.

And when he studied the entire situation, our president realized something very frightening for all motorists. Many police departments have quotas imposed on them to realize either federal or state funds. They must issue a certain amount of tickets to qualify. With more and more motorists using radar detectors and CB's, police must strictly enforce speed limits to reach their quotas. Now, even law-abiding motorists, who might make a slight mistake, are more vulnerable to speeding violations.

NEW MEMORANDUM

In a recent memorandum our president stated, "Due to the changing nature of police radar, JS&A may offer radar detectors as part of its program if presented within the quality image of our company and if the product represents a truly unique radar detector product."

With the green light to find a radar detector, our product selection group was prepared. They had brochures from practically every manufacturer in the world. And they eventually selected what even our president thought was the most professional and well-designed unit available.

HIDDEN ANTENNA

Manufactured by a company called Chicago Radar, the unit consists of two parts–one that is hidden behind your grill, and the other under your dash. There's nothing on top of your dash board to indicate that you've got a radar detector and the system is difficult for anybody to steal. The unit under your dash is attached with a self-adhesive Velcro material so there's no screws or installation to worry about.

The control unit has two lights–one to indicate that the unit is on, and the other to indicate that your car is under radar surveillance. There is also an audible alarm that will sound. But at night, when the light is all that you need, you can switch off the audible sound.

The control unit plugs into your cigarette lighter. The radar antenna is placed behind your grill. Just pull into any service station and the mechanic can easily install the entire system. The Velcro material and mounting brackets are all provided.

AMPLIFIED SENSING

The antenna is one of the keys to the unit's high performance. Instead of the square-shaped dish antennas, the Chicago Radar version is a round cylinder. It tends to sense the radar signals sooner and around curves and hills because of its unique design.

The unit responds to both police radar bands X and K and uses all solid-state computer technology in its design.

We urge you to test our selection of what we feel to be the nation's finest radar detector. Order one from JS&A. When you receive it, drive to your nearest service station or CB dealer and have them install your unit. The antenna installs with just a few brackets and the control unit attaches under your dash with the Velcro material.

Then use it for 30 days. During that time, count the number of radar traps you encounter. On the 30th day, turn off your unit as you travel. See how naked and unprotected you feel.

40 DAYS PROOF

If for any reason you are not completely satisfied, just return your unit within our 40-day trial period and we'll gladly send you a prompt and courteous refund.

To order your system, send **$179.95** plus $3.00 for postage and handling to the address shown below. (Illinois residents, please add 5% sales tax.) Credit card buyers may call our toll-free number below. By return mail, you'll receive the complete system, all cables, Velcro material, instructions and a 90-day limited warranty.

The patented unit is precision crafted by Chicago Radar–one of the most respected names in radar detection systems. JS&A is America's largest single source of space-age products–further assurance that your modest investment is well protected.

We firmly support our police departments and their efforts, but if they are encouraged to use radar to maintain quotas, the law abiding consumer has no choice but to protect himself. Start today. Order your unit now at no obligation.

JS&A® PRODUCTS THAT THINK

Dept.SA One JS&A Plaza
Northbrook, Ill. 60062 (312) 564-7000
Call TOLL-FREE 800 323-6400
In Illinois Call .... (312) 564-7000

LETTERS

Sirs:

In your February issue you published an article by Congressman Les Aspin, "The Verification of the SALT II Agreement." The name and, more important, the reputation of your magazine would lead your readership to conclude that the Aspin article is both analytical in approach and accurate in fact. Because in substantial degree it is neither, I ask that the contents of this letter be made known to your readers.

Aspin should as a minimum have presented data pertinent to his topic. That topic is verification that the Soviets do not exceed the force levels authorized by a treaty extending through 1985. Where the U.S. force is largely static, the capabilities of the U.S.S.R. force are being steadily enhanced. Yet Aspin's main chart depicts the "Russian strategic arsenal" at some point in the recent past whereas what counts is the totals they are authorized to have and to which all intelligence agrees they are building. For instance, that chart shows the number of deployed launchers for the SS-18 ICBM, the largest and most accurate in the Soviet inventory, as 170. The relevant total is the 308 or 326 (if operational launchers within Soviet test sites are included) the Soviets will have when the ongoing conversion program is complete. The U.S., incidentally, is authorized to have none of these large missiles. This distortion is compounded by showing the number of MIRV's per SS-18 as "1 or 8," whereas the missile is in fact being armed with 10 individual warheads. As a result Aspin understates by about 100 percent the inventory of MIRV warheads the Soviets will have by the terms of the treaty.

The SS-18 faux pas is but one example of efforts to sell SALT II by the utilization of irrelevant or misleading data. For instance, Aspin states in the caption for the referenced chart that the U.S.S.R. has no delivery systems, other than those shown, capable of reaching targets in the U.S. What about the submarine launched Shadduck cruise missile? What about the Backfire bomber?

In Aspin's general description of U.S. verification systems he states that over-the-horizon radars can tell one missile from another. Wrong. Over-the-horizon radars can tell you that a missile has disturbed the ionosphere but provide little technical data. It is primarily a warning system. Aspin states that we have infrared satellites that can detect camouflaged silos and virtually prevent Soviet hiding of additional strategic weaponry. Wrong. Infrared can reveal the difference between real foliage and camouflage nets. It cannot detect missiles stored under roofs or underground.

Aspin asserts that the Soviet SS-20 mobile missile now being deployed cannot be upgraded to ICBM range without further testing that we would detect. Wrong. The SS-20 is an SS-16 ICBM less one stage. The SS-16 has already been thoroughly tested and has been deployed in a nonmobile mode in older SS-13 silos. The Soviets need only strap on the third stage, or even more simply remove one of the three SS-20 warheads, and the missile is capable of striking U.S. cities. Assuring the U.S. Government that this is not happening is not within intelligence capabilities.

Mr. Aspin admits that the Soviet Backfire strategic bomber (which like the SS-20 is not even counted in SALT II) now has the capability to strike the United States *with or without* refueling. He then states, however, that the unilateral (nontreaty) assurances of the Soviets will limit the "strategic value" of the bomber. Mr. Brezhnev may agree to limit production of the bomber to current rates, but so far he has refused to state what those rates are. Thus U.S. intelligence is in essence being required to verify Soviet good intentions, an impossibility by all logic.

With regard to cruise missiles Aspin asserts that even if the Soviets cheat by adding cruise missiles to their bombers, "it is doubtful that more than about 120 [Soviet] bombers would be available before the expiration date of the treaty." But that is equal to the number of aircraft the U.S. expects to equip with long-range cruise missiles, and moreover the Soviet launch vehicles will be more modern. As *Aviation Week* authoritatively reported, the Soviets have already flown a subsonic-speed replacement for the Bear bomber and a supersonic replacement, comparable to the B-1, for the Bison bomber. Moreover, the Soviets have already been detected testing cruise missiles on their uncounted Backfire bomber, of which they will have 400 or more by the expiration date of the treaty.

Finally, Aspin's entire treatise on verification is silent on the numerous attempts by the Soviets to circumvent the provisions data through the encryption of telemetry, and the disastrous blows dealt to our verification capabilities by the Soviet acquisition of a top-secret manual on our KH-11 satellite and the loss of our monitoring stations in Iran.

The fact of the matter is that given the damage sustained by U.S. intelligence capabilities over the past few years we cannot today verify the much simpler SALT I treaty, let alone the vastly more complex SALT II treaty.

It should be noted that the U.S. arms controllers themselves signaled the nonverifiability of SALT II when they asked in December that the Soviets desist from the practice of encrypting their telemetry. This practice denies us reliable information on the technical characteristics of Soviet missiles, in particular the number of MIRV's per missile. The Soviets balked and counterproposed that henceforth there be no encryption of that portion of the test data relevant to verification—the determination to be made unilaterally by the testing side. (Astonishingly, this suspect Soviet offer is being supported by the Administration.) The question that arises is: If SALT II is "fully verifiable," why did they try to get the Soviets to cease encryption of telemetry?

Lack of verifiability is only one of several major flaws of SALT II, and certainly not the most significant. It only makes a bad treaty worse. Indeed, it can be argued that the Soviets would be foolish to cheat since the treaty is so heavily advantageous to Moscow as it stands.

GEN. RICHARD G. STILWELL

Coalition for Peace through Strength
Washington

Sirs:

General Stilwell's letter is filled with misleading statements. The chart in my article, depicting Russian strategic forces, refers to currently deployed forces, not to those of 1985. Not all SS-18 missiles have 10 warheads. Two of the four versions are fitted with a single warhead; the others have eight or 10. Under SALT counting rules they are all presumed to have 10—something of a

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

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

Offprint correspondence and orders should be addressed to W. H. Freeman and Company, 660 Market Street, San Francisco, Calif. 94104. For each offprint ordered please enclose 40 cents.

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

Name ____________________

New Address ____________________

____

____

Old Address ____________________

____

____

disadvantage to the Russians, I would think.

I did not include the Shadduck cruise missile as a weapon capable of striking U.S. targets because its range is only 250 miles, it has poor accuracy and it can be launched only from the surface (a dangerous place for a Russian submarine to be only 250 miles off American shores). It clearly is not a strategic weapon by anyone's definition, except perhaps General Stilwell's.

General Stilwell is incorrect in his assessment of our military capabilities. Over-the-horizon radars can observe the propulsion characteristics of missiles (which is one way of distinguishing among them), and infrared devices can distinguish among their rocket-exhaust plumes. Note that General Stilwell completely ignores all the other detection devices, of which there are many, that my article discusses in detail.

He is also wrong about infrared satellites. As long as the ground immediately around an object in question has a temperature or radiative emission different from that of the surrounding terrain, an underground silo will stand out clearly.

General Stilwell has been given poor information on the SS-16. The missile has been tested only once since 1975, and that test was a failure.

If the Russians deploy a new heavy bomber, as General Stilwell suggests, then it must be counted as a "strategic delivery vehicle," thus limiting the number of other vehicles the Russians can deploy. If they deploy Backfire bombers with long-range cruise missiles, they too must be counted under SALT II. If there are no "functionally observable differences" between these Backfires and others, then all Backfires must be counted and the Russians must proportionately reduce the number of bombers, ICBM's or SLBM's. This is fine with me.

I was silent on the controversies over telemetry and the KH-11 because they arose after my article had gone to press. Telemetry conveys important technical information but very little that is related to SALT II; the little it does relay can be detected by other means. The KH-11 scandal somewhat degrades our capability to get some technical information, but virtually none of this would be pertinent to the provisions of SALT II. Unfortunately, this cannot be discussed in detail without getting into highly classified information. On the basis of what I know as a member of the House Armed Services and Intelligence committees, however, I believe the Administration is right in stating that telemetry encryption and the KH-11 disclosure do not seriously degrade our capability to verify compliance with a SALT II treaty.

LES ASPIN

House of Representatives
Washington

NEW! THE UNDER-THE-SEAT KLUGE BAG™

Lots of room in very little space.

Companion to America's best selling flight bag, the new Kluge Under-the- Seater incorporates every element and design feature these seasoned travelers required.

> "A close to perfect piece of luggage. I still find it hard to believe I can pack everything I need for a six day trip in a bag that goes under the seat." Ron Polacco, President, Travel Technology, Inc.
> Ron Palacco

Over 20,000 businessmen have ordered our famous, original Kluge Bag™. And inundated us with requests for an under-the-seater with the same roominess, versatility and thoughtful attention to detail.

Here, then, is the compact, under-the-seater you'd design yourself.

The only bag that lets you travel light and still take along everything you could possibly need — and more. A bag that slips under the seat but still gives you all the advantages of an easy-to-pack, wrinkle free garment bag or valet.

One Night or Four or More

A highly distinctive, exceptionally handsome, extremely durable bag you can carry on your shoulder —for hands free convenience and because even an overnighter can be a nuisance before you get to the gate or find a taxi.

A bag you'll use for one night or two and when you'll be away four nights or five. Because it holds not just one suit but two . . . plus a full complement of shirts, socks, and underwear . . . plus an extra pair of shoes . . . even your robe and slippers . . . and the things you'll need if you're going to play tennis, use the pool, want to relax in a sweater and slacks.

Removable Portfolio

All this, and you still haven't come to some of the best features. Like the portfolio that's removable. Something almost everyone seemed to want—and couldn't find.

A roomy, well-made case you'd expect to spend $30 or $40 for all by itself. It not only eliminates your having to carry a separate attache, it also enables you to pack all those extra files and samples you pick up in the portfolio pouch, while you keep the portfolio itself with you.

An Extra Easy Access Outside Compartment

There's another compartment, on the other side, which is perfect for last minute items, for things you need while you're traveling, like a carton of cigarettes, reading material, a sweater to wear on the plane, or for presents you pick up at the airport.

Raincoat/Umbrella Carrier

Just above this zippered section you'll find adjustable straps to carry your raincoat and/or an umbrella. It seems hard to believe but up until now nobody made an airline bag that recognized the fact that the weather could change or that you often travel from a place where it's cold to where it's warm, or vice versa.

More Bonus Features

The new Kluge Bag, though, pays attention to all these details. So, there's a fitted toiletries compartment. . . a place to hang your ties and an extra belt. . . even a special hook to hang the bag in a closet.

Only $45.00!

Simply, there's no other under-the-seater that's anything like this. Beautifully made of top-quality, heavy-duty, linen-like rayon, the material most often used in expensive luggage today because it's as strong as it is light, and retains its beauty through years of use and abuse, the Under-the-Seat Kluge Bag is available in natural canvas color with rich brown piping and trim or a striking combination of black with brown trim.

Most extraordinary of all, though, is the price. At $90 or $100, which is how much you'd probably have to spend in a store, the Under-the-Seat Kluge Bag would be an excellent value. At **$45.00** it's absolutely unbeatable.

A price that's possible only because we're one of the largest mail merchandisers in the U.S.—able to commit for an entire manufacturing run, and to eliminate salesmen, distributors, retailers and all their costs by selling direct.

So, the Deluxe Under-the-Seater, with beautiful belting leather trim and matching zippers is an even more unusual value at just $75.00. Shipping and handling for either bag is just $3.60 extra.

30 Day No-Risk Trial

Some of Hollywood's biggest stars, television's top performers, chairmen of the Board of America's largest corporations fly with a Kluge Bag. Like them, you must find your Kluge Bag the finest, most useful and truly convenient piece of luggage of its kind on the market today. An extraordinary value as well as the best Under-the-Seater available anywhere. Or return it within 30 days for a complete refund—no questions asked.

Call 800-621-5554

Illinois residents call 800-972-5858
(In operation 24 hours a day, 7 days a week)

With this understanding, call the toll free number above to order with any national credit card. Send your check to Douglas Dunhill at the address below. Be sure to specify natural or black . . . standard or deluxe. (Illinois residents include the sales tax.)

You'll have the only under-the-seat flight bag with its own portfolio and room for easy, wrinkle-free packing of everything you need for almost every trip you take. Order now without obligation.

Douglas Dunhill
INC. AFFORDABLE QUALITY

Dept. 47-2577
Ten Douglas Dunhill Drive, Oak Forest, IL 60452

© Douglas Dunhill Inc. 1979

50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

JULY, 1929: "More than 80,000 miles of main and gathering oil lines now provide a great underground transportation system between oil fields in various parts of the country and ship terminals, railway loading racks and refineries. The longest of the oil trunk lines extends from Texas to Bayonne, N.J., a distance of nearly 1,700 miles. Many large cities hundreds of miles from the source of supply are receiving natural gas through pipelines from 16 to 24 inches in diameter, and a 1,200-mile pipeline project now under construction will send natural gas from the Louisiana fields to Birmingham, Atlanta, Chattanooga and other southern cities. The mileage of the main natural-gas lines in the U.S. almost equals our 81,000 miles of manufactured-gas lines. This is particularly significant in view of the fact that most of the larger natural-gas projects have been built within the past two or three years, whereas the manufactured-gas lines have been extended gradually since the first American coal-gas plant was built in Baltimore in 1812."

"The sun sends out a much wider spectrum and more ultra-violet light than we observe on the earth. Ozone, as if it were a piece of metal, screens off the short-wavelength end. We can measure how much ultra-violet is absorbed by a layer of ozone of one or two or five millimeters' thickness, and by studying the bright and the dark bands in the spectrum we can compare this with the amount the atmosphere absorbs. The result is that there must be at about 40 kilometers' (25 miles) altitude a layer of ozone equivalent to three millimeters (an eighth of an inch) at sea level. The ultra-violet spectrum is being studied at Arosa in Switzerland, at the Meteorological Department of Oxford University, at the University of Montpelier in France (where the sky is very clear) and at many other localities. The first years of study have already given some interesting results. In all the places in the Northern Hemisphere there is a marked variation in the course of the year of the amount of ozone present. The thickness of the layer is greatest at the beginning of the spring and decreases during the summer."

"Florida, the land of fruits and flowers, is fast attaining quantity production of the alligator pear, or avocado as it should more properly be called. Southern California is not behind in this respect. Thousands of acres in both states are bearing full crops of this singular fruit, which is in substance a vegetable and is perhaps the most nutritious tree product. Many have eaten avocados but do not like them at first. They are somewhat insipid and seem to lack flavor, even though they are nutritious. The answer probably lies in the acquisition of taste. Did you like your first raw oyster or your first olive?"

"Everybody knows the end products of scientific investigation because everybody uses them, but few know the stages by which these end products have developed. Many see only what is on top. They accept, for example, the blessings of our knowledge of germs and germicides without knowing how Pasteur discovered germs as the chief source of disease; they use a large variety of chemical compounds without knowing the meaning of chemistry; they admire skyscrapers, great bridges and other complicated structures not knowing how Galileo laid the firm foundation of the laws of forces without which these structures could not be safely erected. Even today only a few carry on the advance of science. There are millions of otherwise intelligent people whose outlook on the world is essentially that of the Middle Ages. A large number confuse science with pseudo-science and are easy game for all kinds of quackery purporting to cure bodily ills. Astrology is more popular than astronomy. The people of a sovereign state attempt to settle by referendum the question of whether man is a product of organic evolution; for them to attempt a popular verdict on the Einstein theory would be no less absurd."

SCIENTIFIC AMERICAN

JULY, 1879: "The grand fact in the history of iron for the past 20 years has been the rapid replacement of wrought iron by the softer varieties of steel. The earlier experiments of Bessemer and others were directed to the production of wrought iron by a quicker and easier process than 'puddling' for the removal of the carbon of cast iron. These experiments were unsuccessful so far as the complete removal of carbon and production of a fairly soft, pure iron was concerned, but they proved the feasibility of manufacturing steel directly from cast iron with far less expenditure of time, fuel and labor than was called for by the older methods. The new processes, moreover, carried with them the great advantage of operation at a temperature so high that the metal was completely fused and hence was much more uniform than when masses were consolidated by welding. Gradually it became apparent that the true function of the newly invented methods was the production of a new material, containing carbon in relatively small amounts, combining the toughness and workable qualities of wrought iron with the great tensile strength, rigidity, some of the hardness and particularly the homogeneity due to the practicable fusibility of steel. The 'mild steel' that is the essential result of the recent processes of manufacture is not the 'tool steel' that not many years ago was pictured in everyone's mind in connection with the term steel, but it is a substance available for a far wider range of uses and applicable upon a far greater scale."

"M. Janssen is still working hard at his interesting solar observatory at Meudon. Gradually but surely he is perfecting his method of observing the sun by photography. M. Janssen stands alone in his researches in solar photography. His last important discovery was the fact that when a picture is taken of the sun with the briefest exposure, there is a vast amount of important detail to be seen. In some portions of the surface the sun appears covered with round granules; in others the granules are comet-shaped, or elongated, implying motion. M. Janssen still uses his revolver apparatus, which he first used to record the transit of Venus, and the successive views obtained within two minutes of each other have already given valuable results. In a word, it is found that two images of the sun taken with but a short interval between them differ materially from each other. The phenomena we have chosen to designate as sunspots apparently change much more rapidly than is supposed. Thus we have ample proof that changes are for ever going on in the sun."

"The new steamer *Arizona,* which passed Sandy Hook at half-past five the afternoon of June 17, arrived at Queenstown at 20 minutes past seven the morning of June 25. The actual running time was seven days nine hours 23 minutes, the fastest trans-Atlantic trip on record. It is one and a half hours less than the time of the *Britannic* in August, 1877."

"The Fall River spinners' strike began as threatened June 25. By the action of about 800 operatives the industry of 15,000 has been arrested. The strike will cost the mill hands about $100,000 a week in loss of wages. The leaders in the strike are described as operatives recently from England, particularly from Blackburn and Preston. It is further affirmed that no person prominent in these difficulties was an inhabitant of Fall River in 1871. One man is particularly described. He is a weaver who can get no employment in any of the mills on account of his bad reputation for causing trouble. He makes a business of agitating. A few men of this sort have been allowed by their fellow operatives to stop their work and stop their wages."

THE AUTHORS

KEVIN N. LEWIS ("The Prompt and Delayed Effects of Nuclear War") is a fellow of the Arms Control Project at the Center for International Studies of the Massachusetts Institute of Technology, a consultant to the Rand Corporation and a Ph.D. candidate in defense policy in the M.I.T. department of political science. He received his B.S. in operations research from Yale University in 1977 and his master's degree from M.I.T. this past spring. Lewis' current research interests include the structuring of general-purpose forces, the economic impact of nuclear warfare and air-defense technologies.

WERNER HENLE, GERTRUDE HENLE and EVELYNE T. LENNETTE ("The Epstein-Barr Virus") are research virologists at the Children's Hospital of Philadelphia. The Henles are also professors of virology in pediatrics at the University of Pennsylvania School of Medicine. Both were born in Germany, and they met as medical students at the University of Heidelberg. Because Werner Henle had a Jewish grandfather he emigrated from Nazi Germany to the U.S. in 1936; Gertrude Henle followed him in 1937. They have been on the research staff of Children's Hospital and the Pennsylvania School of Medicine ever since. This year the Henles were given the second annual Bristol-Meyers Award for Cancer Research in recognition of their work on the Epstein-Barr virus. Lennette was born in China and was raised in Vietnam and Okinawa. Her family then moved to California, where she did her undergraduate work in physics at the University of California at Berkeley. After obtaining her Ph.D. in molecular biology from Washington University she learned techniques of virology at the California Department of Health laboratories.

THOMAS R. GEBALLE ("The Central Parsec of the Galaxy") is a senior research fellow at the Hale Observatories of the California Institute of Technology. Born and raised in Seattle, he studied at the University of Washington, the University of Amsterdam and the University of California at Berkeley, where he received his bachelor's degree in 1967. He then continued his studies at Berkeley, obtaining his Ph.D. in physics in 1974 with a dissertation on astronomical infrared spectroscopy. Following a postdoctoral year at Berkeley, Geballe spent two years at the University of Leiden, returning to the U.S. to accept a fellowship at the Hale Observatories.

F. E. C. CULICK ("The Origins of the First Powered, Man-carrying Airplane") is professor of applied physics and jet propulsion at the California Institute of Technology. Born in New Hampshire, he studied at the Massachusetts Institute of Technology both as an undergraduate and as a graduate student, receiving his Ph.D. in aeronautics in 1961. He then accepted a postdoctoral fellowship at Cal Tech, where he has remained ever since. His research interests include gas lasers, fluid mechanics, combustion and the aerodynamics of the Wright brothers' 1903 "Flyer." Culick is a private pilot and shares the ownership of an aircraft he uses for business and pleasure. He writes: "Last summer my 15-year-old son and I flew around the U.S. in two weeks in celebration of the 75th anniversary of the Wrights' first powered flights. We made 48 landings, only one in each of the continental states, and covered 8,400 miles in 66.5 hours of flying. Kitty Hawk was of course one of our stops."

ANTONIO GARCÍA-BELLIDO, PETER A. LAWRENCE and GINES MORATA ("Compartments in Animal Development") are developmental biologists. García-Bellido is director of the Institute of Genetics of the Spanish Research Council in Madrid. He obtained his Ph.D. from the University of Madrid in 1962 and did postdoctoral work at the universities of Cambridge and Zurich and at the California Institute of Technology. Lawrence is on the staff of the cell-biology division of the Medical Research Council Laboratory of Molecular Biology in Cambridge, England. He did his undergraduate work in zoology at Cambridge and went on to do graduate work in insect development in the laboratory of V. B. Wigglesworth. From 1965 to 1967 he was a research fellow at the University of Virginia and Case Western Reserve University. Lawrence then joined the genetics faculty at Cambridge, moving in 1969 to the Laboratory of Molecular Biology. Morata is on the staff of the Institute of Genetics in Madrid. He received his undergraduate degree in 1968 and went to work in García-Bellido's laboratory the following year. After obtaining his Ph.D. in insect development in 1973, he spent a year as a fellow in the department of genetics at the University of Oxford and two years at the Laboratory of Molecular Biology, where he met Lawrence.

KENNETH A. JOHNSON ("The Bag Model of Quark Confinement") is professor of physics at the Massachusetts Institute of Technology. He did his undergraduate work at the Illinois Institute of Technology and received his Ph.D. from Harvard University in 1955. After two years as a lecturer at Harvard he went to the Institute for Theoretical Physics in Copenhagen as a postdoctoral fellow. In 1958 he joined the M.I.T. faculty.

SIMON CONWAY MORRIS and H. B. WHITTINGTON ("The Animals of the Burgess Shale") are paleontologists who have studied the rich harvest of fossils from the Burgess Shale locality in British Columbia. Conway Morris is lecturer in the earth sciences at the Open University in Milton Keynes, England. Born in London, he did his undergraduate work at the University of Bristol and obtained his Ph.D. from the University of Cambridge in 1976. In 1975 he was elected research fellow of St. John's College, Cambridge, the first paleontologist to be elected since L. S. B. Leakey. In his spare time Conway Morris collects antiquarian geology books and navigates English canals. Whittington is Woodwardian Professor of Geology at Cambridge and head of the department of geology. He was educated at the University of Birmingham, where he received his Ph.D. and lectured from 1945 to 1949. He then joined the geology faculty of Harvard University, where until 1966 he was curator of invertebrate paleontology at the Museum of Comparative Zoology. In 1966 and 1967 he joined parties from the Canadian Geological Survey in collecting fossils at the Burgess Shale locality.

DAVID REGAN, KENNETH BEVERLEY and MAX CYNADER ("The Visual Perception of Motion in Depth") are experimental psychologists at Dalhousie University. Regan is Killam Research Professor and director of the Center for Research in Sensory Psychology and Medical Physics. He studied physics at the Imperial College of Science and Technology, which gave him his Ph.D. in 1964, and obtained his D.Sc. from the University of London in 1974. After teaching physics at London for a few years, he worked in the research department of communication at the University of Keele, and then in 1976 he moved to Canada. Regan's major research interests are the application of visual and auditory psychophysics to the diagnosis and management of brain damage and to aviation. Beverley is a postdoctoral fellow. He received his Ph.D. in 1975 from the University of Keele and then worked on visual displays of image intensifiers in the department of electrical engineering at the University of Nottingham. Beverley moved to Dalhousie in 1977. Cynader is associate professor of psychology. He obtained his bachelor's degree in psychology at McGill University and then did graduate work on the visual system of cats and monkeys at the Massachusetts Institute of Technology. After receiving his Ph.D. in 1972, he moved to Dalhousie.

MATHEMATICAL GAMES

Douglas R. Hofstadter's "Gödel, Escher, Bach"

by Martin Gardner

This sentence no verb.
—DOUGLAS R. HOFSTADTER,
Gödel, Escher, Bach: an Eternal Golden Braid

Every few decades an unknown author brings out a book of such depth, clarity, range, wit, beauty and originality that it is recognized at once as a major literary event. *Gödel, Escher, Bach: An Eternal Golden Braid,* a hefty (777 pages) volume just published by Basic Books, is such a work. The author (and the illustrator and typesetter) is Douglas R. Hofstadter, a young computer scientist at Indiana University who is the son of the well-known physicist Robert Hofstadter.

What can Kurt Gödel, M. C. Escher and Johann Sebastian Bach have in common? The answer is symbolized by the objects shown in the photograph on page 18 and in the photograph on the book's jacket. In each photograph two wood blocks floating in space are illuminated so that their shadows on the three walls meeting at the corner of a room form the initials of the three surnames Gödel, Escher and Bach. More precisely, the upper block casts "*GEB,*" the heading of the book's first half, and the lower block casts "*EGB,*" the heading of its second half. The letters *G, E* and *B* may be thought of as the labels for three strands that are braided by repeatedly switching a pair of letters. Six steps are required to complete a cycle from *GEB* (through *EGB*) back to *GEB.*

Dr. Hofstadter (his Ph.D. is in physics from the University of Oregon) calls such a block a "trip-let," a shortened form of "three letters." The idea came to him, he explains, "in a flash." Intending to write a pamphlet about Gödel's theorem, his thoughts gradually expanded to include Bach and Escher until finally he realized that the works of these men were "only shadows cast in different directions by some central solid essence." He "tried to reconstruct the central object, and came up with this book."

Hofstadter carved the trip-let blocks from redwood, using a band saw and an end mill. The basic idea behind this form is an elaboration of the classic puzzle that asks what solid shape will cast the shadows of a circle, a square and a triangle. Can a trip-let be constructed for any set of three different letters? If the letters may be distorted sufficiently, the answer is yes, and so to make the problem interesting some restrictions must be imposed. To begin with, the letters (preferably uppercase) must all be conventionally shaped, and they must fit snugly into the three rectangles that are the orthogonal projections of a rectangular block. In addition the solid must be connected, that is, it must not fall apart into separate pieces. It is not easy to determine except by trial and error whether such a trip-let can be made for any three given letters. As it turns out, some trip-lets are not possible. This problem suggests exotic variations, for example n-tup-lets that project n letters; four-dimensional tup-lets that project solid trip-lets that in turn project plane shadows of letters; solids that project numbers, pictures or words, and so on. (For this description of trip-lets I am indebted to Hofstadter's friend Scott Kim, who worked closely with him on many aspects of the book.)

What reality does Hofstadter see behind the work of his three giants? One aspect of that reality is the formal structure of mathematics: a structure that, as Gödel's famous undecidability proof shows, has infinitely many levels, none of which are capable of capturing all truth in one consistent system. Hofstadter puts it crisply: "Provability is a weaker notion than truth." In any formal system rich enough to contain arithmetic true statements can be made that cannot be proved within the system. To prove them one must jump to a richer system, in which again true statements can be made that cannot be proved, and so on. The process goes on forever.

Is the universe Gödelian in the sense that there is no end to the discovery of its laws? Perhaps. It may be that no matter how deeply science probes there will always be laws uncaptured by the theories, an endless sequence of wheels within wheels. Hofstadter argues eloquently for a kind of Platonism in which science, at any stage of its history, is like the shadow projections on the wall of Plato's cave. The ultimate reality is always out of reach. It is the Tao about which nothing can be said. "In a way," Hofstadter writes at the end of his preface, "this book is a statement of my religion."

For laymen I know of no better explanation than this book presents of what Gödel achieved and of the implications of his revolutionary discovery. That discovery concerns in particular recursion, self-reference and endless regress, and Hofstadter finds those three themes vividly mirrored in the art of Escher, the most mathematical of graphic artists, and in the music of Bach, the most mathematical of the great composers. The book's own structure is as saturated with complex counterpoint as a Bach composition or James Joyce's *Ulysses.* The first half of the book serves as a prelude to the second, just as a Bach prelude introduces a fugue. Moreover, each chapter is preceded by a kind of prelude, which early in the book takes the form of a "Dialogue" between Achilles and the Tortoise. Other characters enter later: the Sloth, the Anteater, the Crab and finally Alan Turing, Charles Babbage and the author himself. Each Dialogue is patterned on a composition by Bach, and in several instances the mapping is strict. For example, if the composition has n voices, so does the corresponding Dialogue. If the composition has a theme that is turned upside down or played backward, so does the Dialogue. Each Dialogue states in a comic way, with incredible wordplay (puns, acrostics, acronyms, anagrams and more), the themes that will be more soberly explored in the chapter that follows.

There are two main reasons for Achilles and the Tortoise having been chosen to lead off the Dialogues. First, they play the major roles in Zeno's paradox (the topic of the book's first Dialogue), in which Achilles must catch the Tortoise by escaping from an infinite regress. Second, they are the speakers in an equally ingenious but less familiar paradox devised by Lewis Carroll. In Carroll's paradox, which Hofstadter reprints as his second Dialogue, Achilles wishes to prove Z, a theorem of Euclid's, from premises A and B. The Tortoise, however, will not accept the theorem until Achilles postulates a rule of inference C, which explicitly states that Z follows from A and B. Achilles adds the rule to his proof, thinking the discussion is over. The Tortoise then, however, jumps to a higher level, demanding another rule of inference D, which states that Z follows from A, B and C, and so it goes. The resulting endless regress seems to invalidate all reasoning in much the same way that Zeno's paradox seems to invalidate all motion.

On Photographing a Small Step

It is unearthly. The landscape underfoot. It is soft and powdery.

A silver beast stands there. His head is a shiny globe. His huge frankfurter fingers hold a small box that talks.

Click. Whir. Click. Whir.

A man is taking the first pictures on the moon. He takes them with a camera that gives him startling detail. It is a Hasselblad.

It is the first of twelve taken to the moon on six landings. And it is still there today.

Photograph by Hasselblad courtesy of NASA

Hasselblads have taken tens of thousands of photographs during millions of miles of travel in space.

In all, 60 Hasselblads on 27 flights with 6 astronauts have completed 3566 orbits of earth and 363 of the moon.

Every Hasselblad proved reliable on every flight under every condition.

Today, 10 years after, Hasselblad celebrates mankind's giant leap with a commemorative issue of a limited edition (earth version) of the 500EL/M motor drive camera that took those historic photographs.

The only medium format camera on earth or in space that can equal it, is another Hasselblad.

For even in an age of space, much of a year goes into building each Hasselblad; much of it crafted by hand.

100 Hasselblads from one

No Hasselblad can be any better than any other Hasselblad; only more suited to your purposes.

Hasselblad is a prodigious system of 4 cameras, 22 lenses, 8 viewfinders, 9 film magazines, and over 300 other accessories. And there is no obsolescence.

Simply by choosing the right pieces, you can *custom build* over 100 distinct Hasselblad cameras for almost any photographic application.

From a man's first steps on the moon to a child's first steps on earth.

An incredible motor drive

The 500 EL/M is the only hand held medium format camera with an integrated motor unit.

So unlike cameras with optional motor drive attachments, it is built to withstand the punishing demands of motorized operation.

And the possibilities are infinite.

There is a dial set in the body that will order the camera to fire in any of five different ways.

From single shot to speeded up continuous firing at the rate of one shot every 8/10th of a second.

That means you can walk away from the camera, devoting all your concentration to your subject, while firing with a remote cable.

Plug in the intervalometer, for instance, and you can trigger the camera automatically. At any moment you want, from 2 seconds to 14 minutes.

Or trigger up to four cameras simultaneously with a command unit.

Or, fitted for remote radio control, fire from a half mile away.

And the standard film frame on a Hasselblad is 2¼ inches square, *almost 4 times the area* of a 35mm frame. So the ability to hold sharpness and contrast is phenomenal. (For actual size of film frame, see box at lower right.)

The eyes of a Hasselblad

A Hasselblad can see with any one of 22 lenses; all multi-coated (except for two special purpose lenses) in focal lengths from 30mm to 500mm with angular fields from 180 degrees to 9 degrees.

There are lenses for sports, aerial, architectural, and fashion photography. Portrait, landscape, medical, underwater and news photography. Wildlife, laboratory, industrial and child photography.

A lavish brochure on the 500 EL/M and other Hasselblad cameras is available free if you write: Braun North America, Dept. SA7H, 55 Cambridge Parkway, Cambridge, MA 02142—a division of The Gillette Company and exclusive marketer of Hasselblad in the U.S.

There is even the 105mm Zeiss UV-Sonnar f4.3 which can photograph the ultraviolet portion of the electromagnetic spectrum with costly quartz elements that can detect radiations unseeable by the human eye.

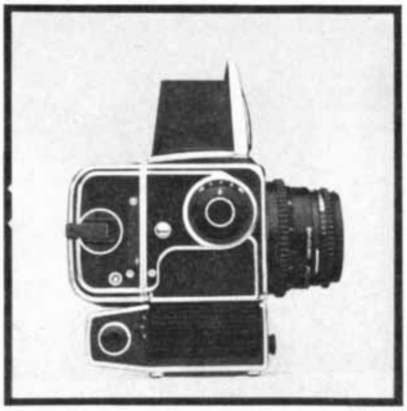
There are also lenses that adjust the aperture fully automatically as the light changes.

Even if you are not yet into these more exotic applications but simply recording travel and everyday happenings, it's nice to know that your camera produces the very best images in the medium format.

So, in effect, there is nothing you can't see with a Hasselblad. And photograph. From what is invisible to the naked eye to all that is visible in the universe.

Is it any wonder then that NASA came to Hasselblad for the camera which was to photograph space? And that top professional and serious amateur photographers still rely on Hasselblad.

HASSELBLAD®

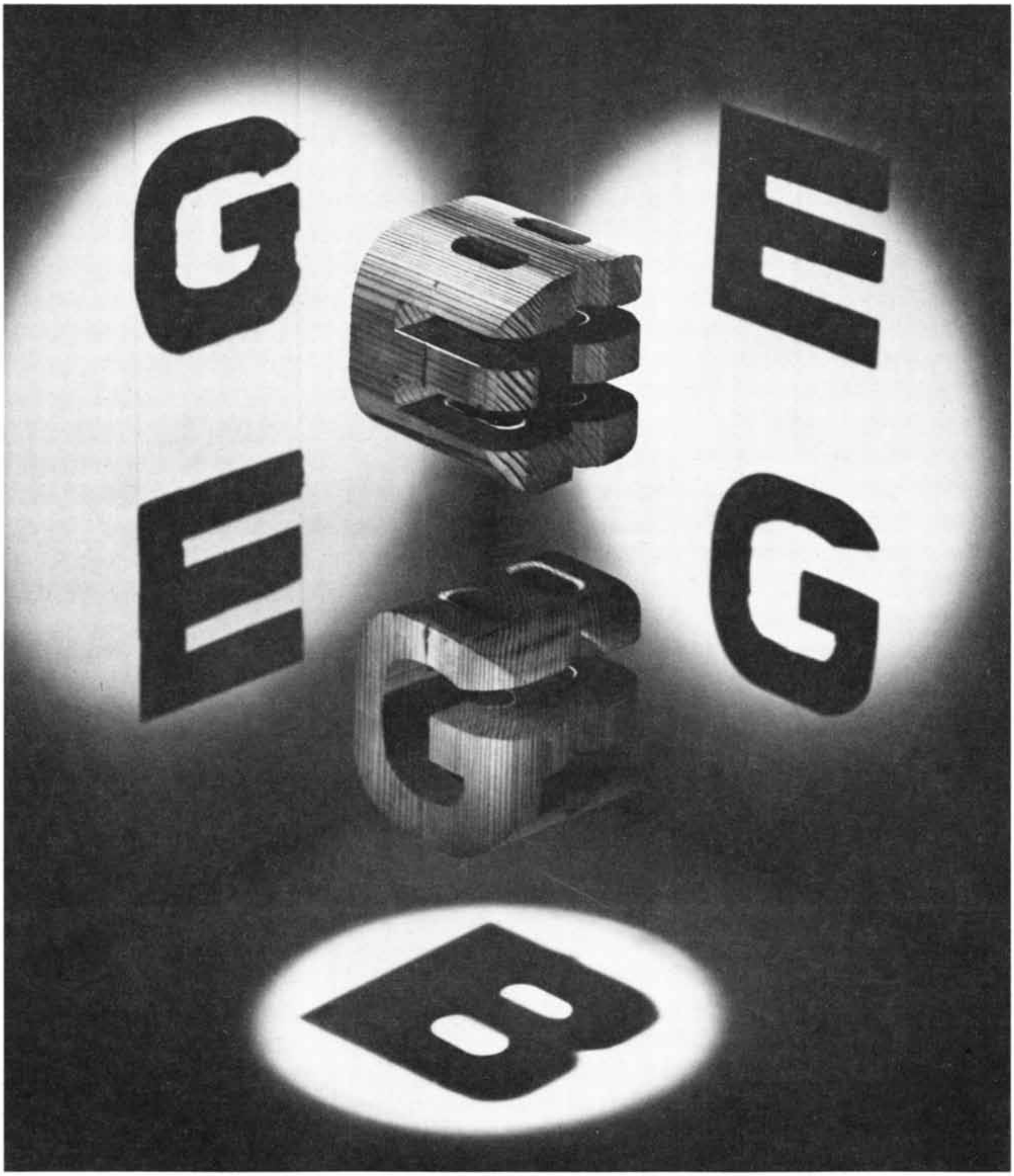


The Motor-Drive Hasselblad 500 EL/M, made in Sweden. The standard lens is a multi-coated Zeiss Planar 80mm f/2.8T.*

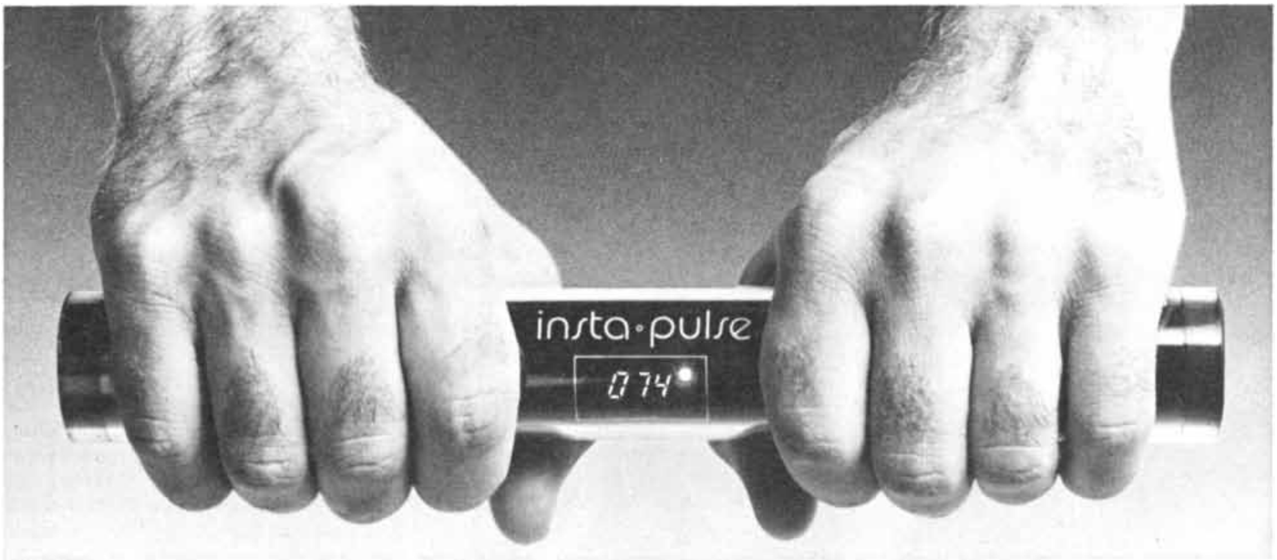
"Plenty of blank leaves, I see!" exclaims Carroll's Tortoise, glancing at Achilles' notebook. "We shall need them ALL!" The warrior shudders.

One of Hofstadter's Dialogues, "Contracrostipunctus," is an acrostic (complete with punctuation marks) asserting that if the words in it are taken backward, they provide a second-order acrostic spelling out "J. S. Bach." Another Dialogue, "Crab Canon," which is illustrated with an Escher periodic tessellation of crabs, is based on Bach's "Crab Canon" in his *Musical Offering*. As the Tortoise discusses Bach his sentences are interspersed with those of Achilles, who is discussing Escher. The Tortoise and Achilles use the same sentences in reverse order. The Crab enters briefly at the crossing point to knot the halves of their discourse together, halves that interweave in time in the same way that the positive and negative crabs of Escher's tessellation interweave in space.

The initials *A, T* and *C* (for Achilles, Tortoise and Crab) correspond to the initials of adenine, thymine and cytosine, three of the four nucleotides of DNA, the molecule with the extraordi-



G, E *and* **B** *cast as shadows by a pair of "trip-lets"*



SEE EVERY BEAT OF YOUR HEART.

Checking your heart rate is now as easy as grasping a glass of orange juice, thanks to the remarkable Insta-Pulse™ heart rate monitor from Biosig, Inc.

When held in your hands, it reads your heart rate function from the tiny electrical impulses in your finger tips. Insta-Pulse works on the same principle as a hospital electrocardiograph (EKG). Yet it weighs a mere 10 ounces, measures 11″ × 1½″ and is fully portable. It runs for up to one full year (over 100,000 readings) on a single 9-volt battery.

No waiting required.

Once Insta-Pulse is in your hands, its microcomputer chip begins to monitor every heart muscle contraction. And instantaneously translates that into heartbeats per minute—showing it with a bright, always changing LED display.

(Flash) 075 (beats per minute) . . .
(Flash) 078 (beats per minute) . . .
(Flash) 074 (beats per minute) . . . and so on.

As long as its nickel hand grips are between your fingers, pulse data continues; put Insta-Pulse down, and it turns itself off automatically.

Watch both heartbeats and heart rate patterns.

You're free to monitor your heart rate anywhere with Insta-Pulse. During a workout or after one. At home, at the gym or on the track.

Instantly, you will see how the most critical organ of your body responds to exertion. And how long it takes to return to its normal at-rest rate.

This picture of your heart recovering from activity is something you can never get with conventional pulse readings.

In seconds you can also see the effects of psychological stress, overeating, alcohol, caffeine and cigarettes. You'll learn just how valuable to your heart a short nap can be. Or check the benefits of a morning's meditation. Here, too, is an excellent tool for use in biofeedback work; with practice, you can actually develop the ability to alter your heart rate—at will.

Not a toy. Not a toy's price.

The cost of Insta-Pulse through The Sharper Image is $149.00. It's not the only heart rate monitor available to you for about that price. However, others with their cumbersome clips and wires do not provide the accuracy of Insta-Pulse, the world's first instant heart rate monitor.

And it's shock resistant as well. Insta-Pulse contains no moving parts, is 100% solid state and its circuitry is shielded from damage by a thick, clear lucite shell. A hand lanyard provides additional security. The battery is included.

Insta-Pulse comes with a ninety-day warranty from its manufacturer, and the Sharper Image's own guarantee of satisfaction: if you are not completely delighted with it, send it back within two weeks for a full refund.

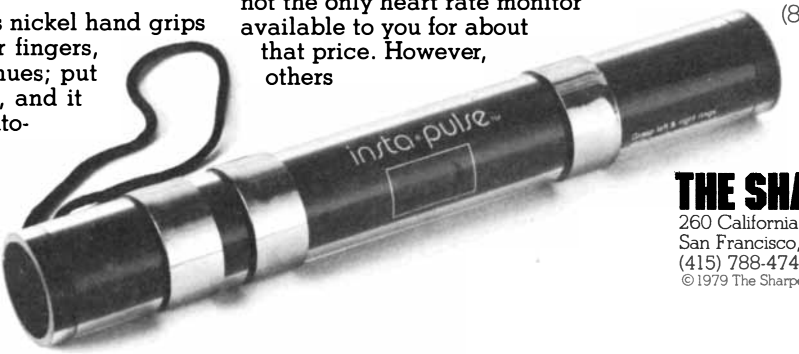
Order an Insta-Pulse now, and watch your heart function as you never could before.

ORDER TOLL-FREE

Credit card holders may use our toll-free ordering number, below. Or send check for $149.00 plus $2.50 delivery. Add $8.94 tax in California.

(800) 227-3436

In California
(800) 622-0733



THE SHARPER IMAGE™

260 California St., Dept. IP-071
San Francisco, CA 94111
(415) 788-4747

nary ability to replicate itself. Just as Achilles pairs with the Tortoise, so adenine pairs with thymine along the DNA double helix. Cytosine pairs with guanine. The fact that the initial *G* can be taken to stand for "gene" prompted Hofstadter to do a "little surgery on the Crab's speech" so that it would reflect this coincidence. The striking parallel between the tenets of mathematical logic and the "central dogma" of molecular biology is dramatized in a chart Hofstadter calls the "Central Dogmap."

The letter *G* also stands for Gödel's sentence: the sentence at the heart of his proof that asserts its own unprovability. To Hofstadter the sentence provides an example of what he calls a Strange Loop, exemplifying the self-reference that is one of the book's central themes. (A framework in which a Strange Loop can be realized is called a Tangled Hierarchy, and the letters of "sloth" turn out to stand for "Strange Loops, or Tangled Hierarchies.") Dozens of examples of Strange Loops are discussed, from Bach's endlessly rising canon (which modulates to higher and higher keys until it loops back to the original key) to the looping flow of water in Escher's *Waterfall* and the looping staircase of his *Ascending and Descending*. One of the most amusing models of *G* is a record player X that self-destructs when a record titled "I Cannot Be Played on Record Player X" is played on it.

Scott Kim's **FIGURE-FIGURE Figure**

A particularly striking example of a two-step Strange Loop is Escher's drawing of two hands, each one sketching the other. We who see the picture can escape the paradox by "jumping out of the system" to view it from a metalevel, just as we can escape the traditional paradoxes of logic by jumping into a metalanguage. We too, however, have Strange Loops, because the human mind has the ability to reflect on itself, that is, the firing of neurons creates thoughts about neurons. From a broader perspective the human brain is at a level of the universe where matter has acquired the awesome ability to contemplate itself.

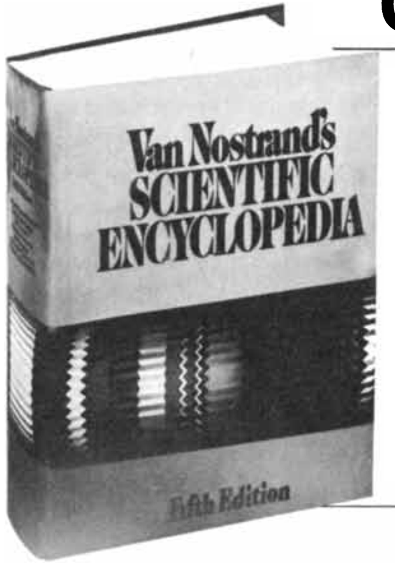
By the end of *Gödel, Escher, Bach* Hofstadter has introduced his readers to modern mathematical logic, non-Euclidean geometries, computability theory, isomorphisms, Henkin sentences (which assert their own provability), Peano postulates (the pun on "piano" is not overlooked), Feynman diagrams for particles that travel backward in time, Fermat's last theorem (with a pun on "fermata"), transfinite numbers, Goldbach's conjecture (which is cleverly linked with Bach's *Goldberg Variations*), Turing machines, computer chess, computer music, computer languages (Terry Winograd, an expert on the computer simulation of natural language, appears in one Dialogue under the anagrammed name of Dr. Tony Earrwig), molecular biology, the "mind" of an anthill called Aunt Hillary, artificial intelligence, consciousness, free will, holism v. reductionism, and a kind of sentence philosophers call a counterfactual.

Counterfactuals are statements based on hypotheses that are contrary to fact, for example, "If Lewis Carroll were alive today, he would greatly enjoy Hofstadter's book." These statements pose difficult problems in the semantics of science, and there is now a great deal of literature about them. For Hofstadter they are instances of what he calls slipping, progressing from an event to something that is almost a copy of it. The Dialogue that precedes a chapter on counterfactuals and artificial intelligence concerns a Subjunc-TV set that enables an observer to get an "instant replay" of any event in a football game

As your introduction to

Natural Science Book Club

Choose either



this $67.50 Classic for only $2.97 A saving of 95%

VAN NOSTRAND'S SCIENTIFIC ENCYCLOPEDIA.

Edited by Douglas M. Considine. Nearly 200 experts have contributed to this thoroughly revised and greatly expanded fifth edition of the most authoritative single-volume source of scientific knowledge ever assembled. Enormous 9¼" x 12" volume contains 2.2 million words, 2382 pages, 2500 photographs, drawings, and charts, and 500 tables. 7200 articles cover everything from the earth and space sciences to energy technology, from the life sciences to mathematics, from information sciences to physics and chemistry. "... an amazing book . . . for both the general and scientific reader."—*The New York Times*

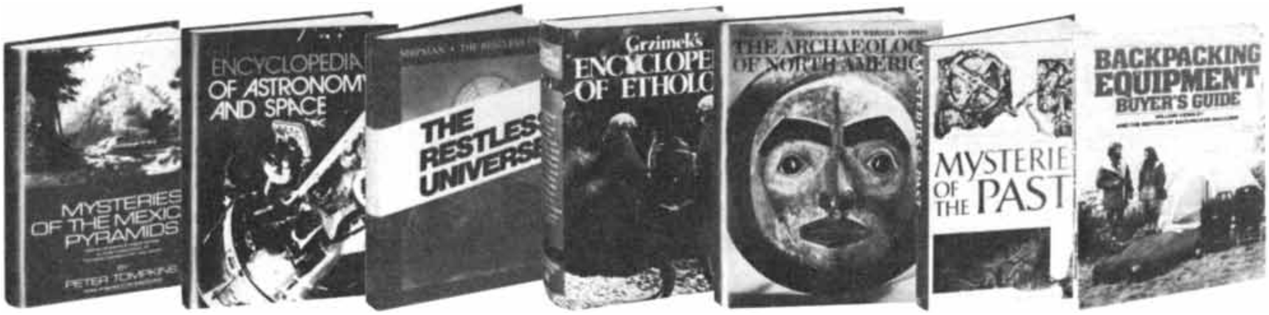
(Publishers' Prices shown)

OR

any other 3 books for only 99¢ each

if you will join now for a trial period and agree to take 3 more books—at handsome discounts—over the next 12 months

(values to $73.85)



63350. MYSTERIES OF THE MEXICAN PYRAMIDS. *Peter Tompkins.* Stunning array of 500 photos, drawings and maps illustrates the fascinating Mesoamerican pyramids which have mystified archaeologists and laymen for centuries. **$20.00**

55000. THE ILLUSTRATED ENCYCLOPEDIA OF ASTRONOMY AND SPACE. **$16.95**

73939. THE RESTLESS UNIVERSE/WHITNEY'S STAR FINDER. **$21.90**

35506. BACKPACKING EQUIPMENT BUYER'S GUIDE/WALKING SOFTLY IN THE WILDERNESS: The Sierra Club Guide to Backpacking/SIMPLE FOODS FOR THE PACK. *Softcover. The 3 count as one book.* **$19.85**

78235. SHELTER, SIGN, AND SYMBOL. Illustrated study of how homes and villages of folk cultures reflect rituals and customs—from aboriginal huts to Tibetan monasteries. **$22.50**

50800-2. GRZIMEK'S ENCYCLOPEDIA OF ETHOLOGY. Comprehensive exploration of animal behavior. 700 pages of scientific, lucid information. 774 photographs, drawings and diagrams. A landmark reference. *Counts as 2 of your 3 books.* **$39.50**

62700. MIND AND SUPERMIND. *Edited by Albert Rosenfeld.* Reports on the mechanism of memory, pacemakers for the brain, therapeutic use of mega-vitamins, and much more. **$8.95**

40695. COSMOS, EARTH AND MAN: A Short History of the Universe. *Preston Cloud.* **$14.95**

44950. THE ENCYCLOPEDIA OF NATURAL HISTORY. **$16.95**

84246. THREE ADVENTURES/LIFE AND DEATH IN A CORAL SEA/DIVING COMPANIONS. *Jacques-Yves Cousteau and Philippe Diole.* **$23.85**

63217. MURMURS OF THE EARTH. *Carl Sagan et al.* The story of the recorded messages carried by the *Voyager* space probe, man's first attempt at interstellar communication. **$15.00**

34740. ATLAS OF MAN. A gloriously illustrated, 300-page reference which covers 400 groups of peoples, tribes, and nations. 120 color photographs, 350 maps and line illustrations. **$25.00**

34210. THE ARCHAEOLOGY OF NORTH AMERICA. *Dean Snow.* Prehistoric Indian cultures, from the first crossings into this continent to the 20th century. 195 photos, charts and maps. **$18.95**

45050. THE ENDS OF THE EARTH. *Isaac Asimov.* Everything about the frozen polar wastes, from their exciting discovery to their effect on world climate. **$15.00**

62915. MONUMENTS OF CIVILIZATION: Ancient Cambodia. *Mazzeo and Antonini.* Portrays the magnificence of ancient Khmer civilization—over 100 color photographs. Oversized volume. **$25.00**

34730. THE ATLAS OF EARLY MAN. *Jacquetta Hawkes.* A panoramic overview of concurrent events in history through charts, graphs, and illustrations, from prehistory to the Roman Empire. **$20.00**

63340-2. MYSTERIES OF THE PAST. Captivating investigation of prehistory and great riddles. Outsized. Sumptuously illustrated. *Counts as 2 of your 3 books.* **$34.95**

54995. THE ILLUSTRATED ENCYCLOPEDIA OF ARCHAEOLOGY. *Glyn Daniel.* **$17.95**

82771-2. TEMPLES, TOMBS AND HIEROGLYPHS/RED LAND, BLACK LAND. An excellent introduction to the complex and fascinating subject of Egyptian history, plus an absorbing look at the customs and habits of everyday life in ancient Egypt. *Counts as 2 of your 3 books.* **$25.90**

87750. THE WORLD OF DINOSAURS. *Michael Tweedie.* **$14.95**

85095-2. UNDERWATER WILDERNESS. *Carl Roessler.* A breathtaking collection of full-color photographs, plus excellent text on life around the world's great coral reefs. *Counts as 2 of your 3 books.* **$35.00**

If the reply card has been removed, please write to
Natural Science Book Club
Dept. 4-A6E, Riverside, N.J. 08370
to obtain membership information and application.

and see how the action would have looked if certain parameters were altered, that is, if the ball were spherical, if it were raining, if the game were on the moon, if it were played in four-dimensional space and so on.

The book's discussion of artificial intelligence is also enormously stimulating. Does the human brain obey formal rules of logic? Hofstadter sees the brain as a Tangled Hierarchy: a multilevel system with an intricately interwoven and deep self-referential structure. It follows logical rules only on its molecular substrate, the "formal, hidden, hardware level" where it operates with eerie silence and efficiency. No computer, he believes, will ever do all a human brain can do until it somehow reproduces that hardware, but he has little patience with the celebrated argument of the Anglican philosopher J. R. Lucas that Gödel's work proves a human brain can think in ways that are in principle impossible for a computer.

Only a glimpse can be given here of the recreational aspects of this monstrously complicated book. In "The Magnificrab, Indeed" (a pun on Bach's *Magnificat* in D), the Dialogue that introduces a discussion of deep theorems of Alonzo Church, Turing, Alfred Tarski and others, appears a whimsical Indian mathematician named Mr. Najunamar. Najunamar has proved three theorems: he can color a map of India with no fewer than 1,729 colors; he knows that every even prime is the sum of two odd numbers, and he has established that there is no solution to $a^n + b^n = c^n$ when n is zero. All three are indeed true.

Some readers will recognize 1,729 as the number of the taxi in which G. H. Hardy rode to visit the Indian mathematician Srinivasa Ramanujan ("Najunamar" spelled backward) in a British hospital. Hardy remarked to Ramanujan that 1,729 was a rather dull number. Ramanujan rejoined instantly that on the contrary it was the smallest positive integer that is the sum of two different pairs of cubes. Hardy then asked his friend if he knew the smallest such number for fourth powers. Ramanujan did not know the number, although he guessed that it would turn out to be fairly big. Hofstadter supplies the answer: 635,318,657, or $134^4 + 133^4$ or $158^4 + 59^4$. He also wonders if his readers can find the smallest number that can be expressed as the sum of two squares in two different ways, but he hides the answer. Can you determine it before I supply it next month?

To explain the meaning of the term "formal system" Hofstadter opens his book with a simple example that uses only the symbols *M*, *I* and *U*. These symbols can be arranged in strings called theorems according to the following rules:

1. If the last letter of a theorem is *I, U* can be added to the theorem.

2. To any theorem *Mx, x* can be added. (For example, *MUM* can be transformed into *MUMUM,* and *MU* can be transformed into *MUU.*)

3. If *III* is in a theorem, it can be replaced by *U,* but the converse operation is not acceptable. (For example, *MIII* can be transformed into *MU,* and *UMIIIMU* can be transformed into *UMUMU.*)

4. If *UU* is in a theorem, it can be dropped. (For example, *UUU* can be transformed into *U,* and *MUUUIII* can be transformed into *MUIII.*)

There is only one "axiom" in the system: In forming theorems one must begin with *MI.* Every string that can be made by applying the rules, in any order, is a theorem of the system. Thus *MUIIU* is a theorem because it can be generated from *MI* in six steps. If you play with the *M, I* and *U* system, constructing theorems at random, you will soon discover that all theorems begin with *M* and that *M* can occur nowhere else.

Now for a puzzle: Is *MU* a theorem? I shall give Hofstadter's answer next month. I shall say no more about *MU* here except that it plays many other roles in the book, in particular serving as the first two letters of "Mumon," the name of a Zen monk who appears in a delightful chapter on Zen koans.

Even as simple a system as that of *M, I* and *U* enables Hofstadter to introduce a profound question. If from all the possible strings in the system we subtract all the strings that are theorems, we are left with all the strings that are not theorems. Hence the "figure" (the set of theorems) and the "ground" between the theorems (the set of nontheorems) seem to carry equivalent information. Do they really? Is the system like an Escher tessellation in which the spaces between animals of one kind are animals of another kind, so that reproducing the shapes of either set automatically defines the other? (Or so that a black zebra with white stripes is the same as a white zebra with black stripes?) In this connection Hofstadter reproduces a remarkable tessellation by Kim in which the word "FIGURE" is periodically repeated in black so that the white ground between the black letters forms the same shapes [*see illustration on page 20*]. The same concept is playfully illustrated in the Dialogue "Sonata for Unaccompanied Achilles" (modeled on Bach's sonatas for unaccompanied violin), in which we hear only Achilles' end of a telephone conversation with the Tortoise about "figure" and "ground." From Achilles' half of the conversation we can reconstruct the Tortoise's.

Other figure-ground examples are provided by the counting numbers. For example, given all the primes, we can determine all the nonprimes simply by removing the primes from the set of positive integers. Is the same true of all formal systems? Can we always take all the theorems from the set of all possible statements in the system and find that what is left—the set of nontheorems—is another complementary formal system? An unexpected discovery of modern set theory is that this is not always the case. To put it more technically, there are recursively enumerable sets that are not recursive. Thus does Hofstadter lead his readers from trivial beginnings into some of the deepest areas of modern mathematics.

The book closes with the wild Dialogue "Six-Part Ricercar," which is simultaneously patterned after Bach's six-part ricercar and the story of how Bach came to write his *Musical Offering*. (A ricercar is a complicated kind of fugue.) In this Dialogue the computer pioneers Turing and Babbage improvise at the keyboard of a flexible computer called a "smart-stupid," which can be as smart or as stupid as the programmer wants. (The computer's name is a play on "pianoforte," which means "soft-loud.") Turing produces on his computer screen a simulation of Babbage. Babbage, however, is seen looking at the screen of his own smart-stupid, on which he has conjured up a simulation of Turing. Each man insists he is real and the other is no more than a program. An effort is made to resolve the debate by playing the Turing Game, which was proposed by Turing as a way to distinguish a human being from a computer program by asking shrewd questions [see "Mathematical Games," June, 1971]. The conversation in this scene parodies the conversation Turing gives in his classic paper on the topic.

At this point Hofstadter himself walks into the scene and convinces Turing, Babbage and all the others that they are creatures of his own imagination. He, however, is as unreal as any of the other characters of the Dialogue, because he too is imagined by the author. The situation resembles a painting by René Magritte titled *The Two Mysteries*, in which a small picture of a tobacco pipe is displayed with a caption that says (to translate from the French) "This is not a pipe." Floating above the fake pepe is a presumably genuine larger pipe, but of course it too is painted on the canvas.

And how real was Magritte? How real are Hofstadter, you and I? Are we in turn no more than shadows on the rapidly flipping leaves of the book we call the universe? We are back to a Gödelian Platonism in which reality is infinitely layered. Who can say what reality really is? The book's final word, "RICERCAR," is a multilevel pun anticipated by a series of acronyms such as Turing's remark "Rigid Internal Codes Exclusively Rule Computers and Robots" as well as Bach's inscription "Regis Iustu Cantio et Reliqua Canonica Arte Resoluta" on a sheet of music he sent to Frederick the Great. Bach's six-part ricercar is from his *Musical Offering*, the story of which opens the book. In this way RICERCAR serves, in much the same way as "riverrun," the first word of *Finnegans Wake*, to twist the work into one gigantic self-referential loop.

One of the book's most amusing instances of a Strange Loop is in the Dialogue that introduces Chapter 16. The Crab speaks of browsing through a "crackpot" book on "metal-logic" titled *Copper, Silver, Gold: An Indestructible Metallic Alloy*. Hofstadter's annotated bibliography reveals that this book is by one Egbert B. Gebstadter (note the *EGB* of Egbert, the *GEB* of Gebstadter and the *EBG* initials) and was published in 1979 by Acidic Books (Hofstadter's publisher being Basic Books). Here is Hofstadter's comment: "A formidable hodge-podge, turgid and confused—yet remarkably similar to the present work. Contains some excellent examples of indirect self-reference. Of particular interest is a reference in its well-annotated bibliography to an isomorphic, but imaginary, book."

Here are the answers to last month's chess problems.

Sam Loyd's mirror-move game, in which White mates on his fourth move, has two solutions:

	White	Black
(1)	1. P-QB4	1. P-QB4
	2. Q-R4	2. Q-R4
	3. Q-B6	3. Q-B6
	4. QxB (mate)	
(2)	1. P-Q4	1. P-Q4
	2. Q-Q3	2. Q-Q3
	3. Q-KR3 (or B5)	3. Q-KR3 (or B5)
	4. QxB (mate)	

Loyd's mirror-move game in which White self-mates on his eighth move is played as follows:

White	Black
1. P-K4	1. P-K4
2. K-K2	2. K-K2
3. K-K3	3. K-K3
4. Q-B3	4. Q-B3
5. N-K2	5. N-K2
6. P-QN3	6. P-QN3
7. B-R3	7. B-R3
8. N-Q4 (check)	

Black's only possible move is PxN, which checkmates the white king.

To achieve the joke checkmate in the final chess problem given last month White simply raises his knight an inch or so above its square and shouts, "Discovered mate!"

In every age, there have always been those among us whose primary activity is inactivity.

And for those people—members of what has been variously dubbed the "idle rich" and the "leisure class"—there has always been an abundance of suitably sedentary luxury sedans.

Cars designed primarily for sitting. Cars that reach their performance peak preening in the country club driveway.

The BMW 528i, on the other hand, is a luxury sedan built to satisfy an opposite mentality. People who believe that activity—mental as well as physical—is what makes living worthwhile.

THE BMW 528i. A LUXURY SEDAN FOR SERIOUS DRIVERS.

At the Bavarian Motor Works, it is our contention that the things that seem to preoccupy the luxury car makers of the world have little to do with the one thing that makes an expensive car worth the money: extraordinary performance.

First and foremost, the engineers at BMW are racing engineers by nature and by profession.

So, while the BMW 528i provides all the creature comforts one could rationally require in any expensive automobile, it also provides a driving experience so unusual, so exhilarating it will spoil you for any other car.

Its six-cylinder engine has been called "...the most refined in-line six in the world."

Its suspension system—independent on all four wheels—provides the driver with a tactile sense of control, an uncanny feel of the road that is positively unique to BMW.

Its interior is biomechanically engineered to the nth degree.

The front seats are shaped to hold their occupants firmly in place, and so thoroughly adjustable that it is all but mathematically impossible to not find a proper and comfortable seating position.

All controls are within easy reach. The tachometer, speedometer and all ancillary instruments are large, well marked and totally visible—set in a dashboard designed in the manner of an airplane cockpit.

THE RARE OPPORTUNITY TO SHIFT FOR YOURSELF.

To some, there is simply no alternative to the singular pleasure of a manual transmission.

For the automotive purist, shifting up through the gears, smoothly and precisely, is an experience that borders on the esthetic.

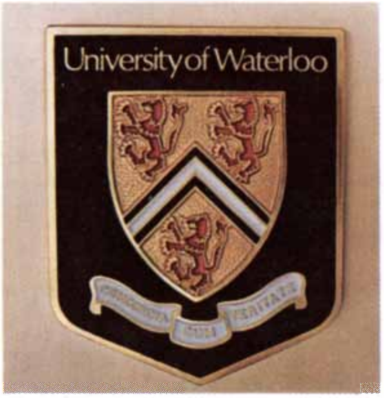
And it is particularly significant that—while a three-speed automatic transmission is available as an option—the BMW 528i offers a four-speed manual transmission as standard equipment.

If the thought of owning such a car intrigues you, call us anytime, toll-free, at 800-243-6000 (Conn. 1-800-882-6500) and we'll arrange a thorough test drive for you at your nearest BMW dealer.

THE ULTIMATE DRIVING MACHINE.
Bavarian Motor Works, Munich, Germany.

How small IBM computers students run 10,000

Although it has been in existence for less than a quarter of a century, the University of Waterloo is the third largest university in Ontario and has gained a reputation for academic innovation. About half of the 15,000 students alternate their studies at the Waterloo, Ont., campus with periods of actual work experience with employers throughout Canada. And some 80% take at least one course in computing, reflecting the University's strong commitment to the computer for instruction, problem solving and research.



With this emphasis on computing, it is not surprising that the University should be active in new ways to increase the usefulness of computing systems for students. It scored a major advance not long ago with the advent of a system for online, video-terminal program preparation by large numbers of students – a system given the acronym WIDJET (Waterloo Interactive Direct Job Entry Terminal system).

An IBM-licensed program, WIDJET is an interactive multiuser system that runs on one or more IBM Series/1 computers. At Waterloo, using three Series/1s, it provides thousands of student programmers with the advantages of ready access to the computer– without the prohibitive costs that time sharing would involve. At the same time, it represents a marked improvement, both in function and expense, over punched card methods of program preparation.

From cards to terminals

As the student computing load grew at Waterloo, increasing dissatisfaction was felt with methods which required a student submitting a program to punch a deck of cards, run the deck through a card reader and then wait at a printer for the result to be returned from the central computer – especially since the process is usually a repetitive one. Moreover, the cost of the cards and printout paper had mounted to about $21,000 a month or over $250,000 a year. "You might say a switch to video terminals was in the cards," says J. Wesley Graham, professor of computer science at Waterloo.

The low-cost interactive WIDJET system was the solution. In this system, most of the workload handled by the central computer is shifted to one or more of the Series/1 computers to which a number of terminals are linked. Programs being prepared at the terminals can be developed interactively – that is, in back-and-forth conversational mode– with the help of an editing capability resident at the Series/1 computers.

The Series/1 solution

"The Series/1 is an ideal base for such an interactive system, both in terms of its capabilities and of IBM service and reliability," says Professor Graham. "We now have 2,500 student programmers authorized to use WIDJET through as many as 200 terminals. On a normal day, we'll get 10,000 jobs submitted – up to 20,000 at a peak period. Because we've removed the constraints associated with card handling, the average submission has grown to around 200 lines of code, instead of 50. Yet with WIDJET, the six hours a week a student programmer typically puts in at a terminal nets out to only about one minute of actual processing time on our System/360 and 370 central computers. The rest is absorbed by the Series/1 and the

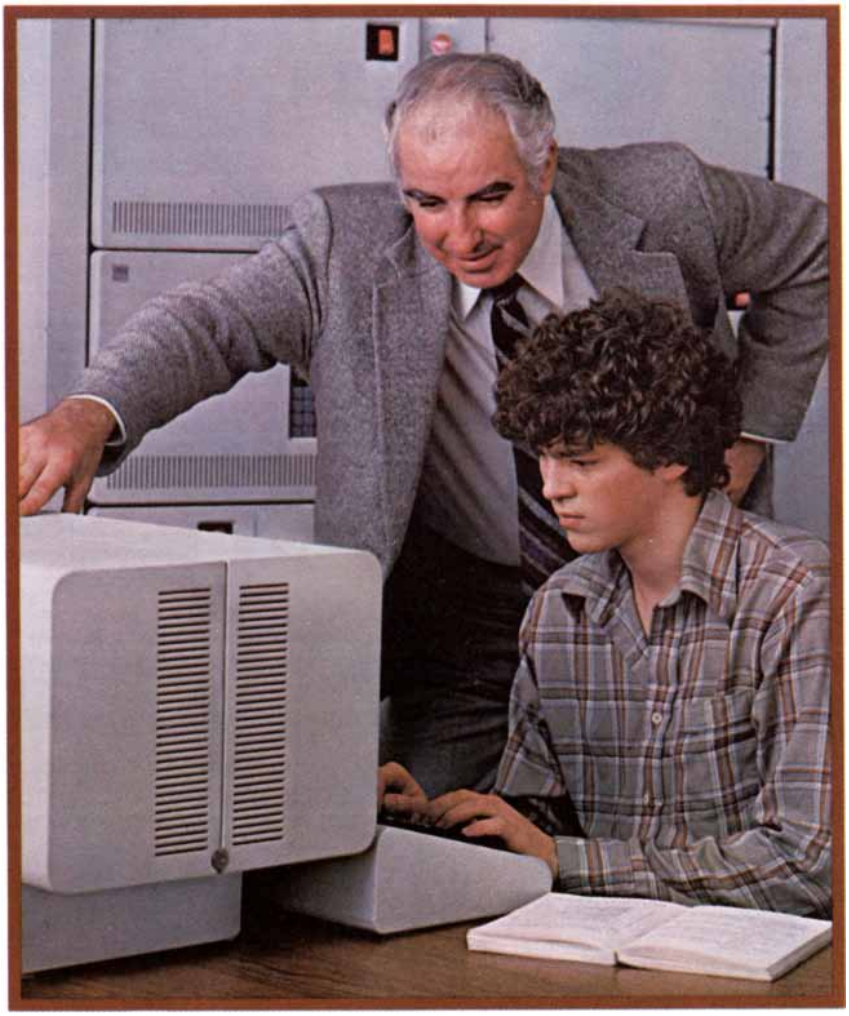
help University of Waterloo programs a day.

WIDJET system.

"And we've found that what we've saved on punched cards in a single year alone more than paid for a Series/1."

Personalized use

The Waterloo student using a WIDJET terminal keys in his or her identification plus a personal "invisible" password. Then he or she can create programs or data, edit material from a previous terminal session, or run a program. Any material the student wants saved can be placed in a personal file stored directly on the Series/1; there is no need to access the main computer when starting or ending a WIDJET session. The main computer is accessed only when a program is actually compiled or run. Since all of a student's output is identified by a personal number, no copying of another student's program is possible. As a result, students can receive a large part of their grade from an assessment of all their computer work performed during the term.



WORKING WITH WIDJET: Professor J. Wesley Graham with a University of Waterloo student at an IBM 4978 terminal.

There is also a "public library" on the Series/1 which, unlike the personal file, can be used by all students; teachers can place class announcements and program and problem assignments in this library to be copied by their students.

A standalone, too

In addition to functioning as a front-end adjunct to a central computer, the WIDJET Series/1 system can also operate in a "standalone" mode in which all jobs are submitted to the Series/1 itself. The standalone configuration can be converted to the front-end configuration by installing a multileaving remote job entry (MRJE) module. For more information, write IBM General Systems Division, College/University Programs, A06B1, P.O. Box 2150, Atlanta, GA 30301.

IBM®

BOOKS

Technology from 1900 to 1950, Galileo at work, color vision, the warm tuna

by Philip Morrison

A History of Technology, edited by Trevor I. Williams. Volume VI: The Twentieth Century, *c.* 1900 to *c.* 1950, Part I. Volume VII: The Twentieth Century, *c.* 1900 to *c.* 1950, Part II (with index to Volumes VI and VII). Clarendon Press (Volume VI $39.50, Volume VII $47.50, both volumes $82). With what pleasure and excitement the five heavy, blue-jacketed volumes edited by Charles Singer, E. J. Holmyard, A. Rupert Hall and eventually Trevor I. Williams were received by the happy reader during the mid-1950's. They were a satisfying survey of the entire history of technology, superbly illustrated, expertly prepared by a galaxy of archaeologists, historians, technologists and museum experts, mainly but not exclusively British. The work was of course imperfect (the Far East entered, so to speak, as an afterthought), but it was overall a huge success, deservedly popular worldwide, with translations into Italian and Japanese and a special large printing in the U.S. Whether it was the fire piston of Southeast Asia, the marvelous granulated goldwork of the Etruscans, the multiple grain mill the Romans built near Arles, the precision screw threads of 16th-century Nuremberg, the adventure of the early Atlantic cables or the first full-fledged cinema of the brothers Lumière in Paris as the 19th century ended, it was brought before the reader with vividness and reliable scholarship.

The 20th century was excluded. It would demand too much space, the editors argued; scholars had not yet acquired perspective, and the technical demands on the reader would be excessive. Now, a generation later, the youngest of the original editors has all the same given us a new collection: nearly 60 chapters by some 65 authors, mostly British, covering the first half of our most technical century. The work is so massive that it has had to be divided in two for reasons of bulk. (The two volumes, VI and VII, are unfortunately not easily read separately because the fine index to both is found only in Volume VII.)

Inflation and recession have taken their toll in spite of the success of the earlier work (which benefited from a generous subvention by Imperial Chemical Industries), so that this expensive continuation almost entirely lacks the drawings, wonderfully made and thoroughly documented, that enriched the first five volumes. The number of illustrations is somewhat smaller, and now nearly all of them are photographs with less meticulous documentation; good halftones are scattered through the text, whereas before they were sequestered in a special section of sized paper. Although photographs do present visual evidence with authenticity, we greatly miss the pungency, clarity and ornament offered by drawings. The technical level will be taxing for the general scientific reader only here and there (as in the detailed account of the chemical separation of plutonium), but a more generous provision of graphs, numerical tables, flow charts and the like would have much improved many of the essays.

Yet it remains true that no four or five single-topic ordinary books offer as compendious and as unified a view of the technological developments of the first half of our century, or one placed in as broad a social and historical context. Several distinguished engineers contribute chapters that have a flavor of direct participation and ripeness of judgment, although at the price of detachment, and a few essays are too brief to be more than occasional pieces. For example, the chapter on space technology covers everything from Goddard and Tsiolkovsky to the moon, at a rather popular level, in fewer than 15 pages.

Even a list of the 58 chapter titles cannot span this work. A more readable assay is provided by presenting in order some points that caught an interested eye in passage through the two thick volumes; they may evoke the entire work.

The half century was fitfully lighted by the flames of two great wars. They are well summarized, but a glaring omission is any account of the machine gun, the tank (and mechanized combat vehicles generally) and the aircraft carrier. None of these salient technological developments receives more than a passing allusion. The trendier atomic energy and nuclear weapons, on the other hand, are fully covered.

The disk memory, standard among computer peripherals for a decade or two, was once a "bootleg project" at one of IBM's laboratories, having been denied a budget.

Innovation is surveyed in economic and managerial terms as well as more directly. Technology is "one thread in the warp of an elaborately woven fabric." The famous tale of the bored relay assemblers in Western Electric's Hawthorne plant, whose work pace quickened with every change in the lighting (whether the lights were dimmed or brightened), is not forgotten. A long table lists the time between invention and application for 46 innovations. The longest intervals—50 to 79 years—were for the fluorescent lamp and the cotton picker; the one-year delay before Freon refrigerants (which had been the object of a purposive search) were put to use is the shortest interval listed.

In 1917, when a study was made that led to the slow unification of London's electricity supply, "10 different frequencies and 20 different voltages" were being supplied.

Radar and echo sounding were military developments that greatly increased the safety and productivity of commercial fisheries.

Rotary drilling, beginning in about the 1890's, has been crucial for the recovery of oil and gas; the wells have steadily gone deeper, from one mile in 1910 to four miles in 1950. Oil rigs were standing on piers offshore at Santa Barbara in the 1890's.

The pioneer large-scale open-pit mine was the huge copper pit at Bingham, Utah; the yawning cavities of this gargantuan technology have made it economic to work ores with a metal content an order of magnitude less than before.

Die-cast parts made of alloys of zinc distilled to 99.99 percent purity to reduce corrosion became staples of the automobile, from carburetors to door handles, in the late 1920's. Chromium plating came to brighten many of them.

The precipitation-hardened alloy of aluminum was found accidentally during the zeppelin epoch by Alfred Wilm in the course of studies looking toward aluminum coinage, when he saw his hardness-test results change after a weekend's delay. (Duralumin is named not from a Latin root but from the company that got the patent rights, which is at Düren in Germany.) New copper-nickel alloys lengthened the average life of steam-condenser tubes in naval vessels from between three and six months in World War I to seven years in World War II.

In 1902 Chilean guano furnished more than two-thirds of the world's nitrogen fertilizer. By the mid-1960's the use of fixed nitrogen had risen about fiftyfold, the great bulk of it synthesized by the Haber process.

Petrochemicals began in the U.S. as rivals of the coal-tar industry. Their manufacture was fed by the big flow of gasoline, which induced techniques for modifying feedstocks to improve gasoline yield and quality. In time the increase in coal prices, the spread of automobiles beyond America, the finding of natural gas and above all the development of plastic polymers based on the linear hydrocarbons made the petrochemical industry into a worldwide giant, pouring out plastics, tire synthetics and surfactants; now it is itself facing the problem of high-priced feedstock. The first plastic was made in 1846 by Christian Schönbein of the University of Basel, who isolated cellulose nitrate (suitable for "attractive vessels") from the reaction of paper with mineral acids. (The polymer story is given in some detail from its beginning, which was much before the period of the volume.)

The television-tube industry is a remarkable innovation in glassworking; it involves both centrifugal casting and shaping with the glass lathe, complete with augmented electric heating supplied to the spinning piece by currents applied through the plasma of the gas flames.

Ceramics has come into its own within this century as a body of technical and scientific knowledge, far from its ancient craft roots. Refractories, grinding wheels, electric insulators large and small, remarkable thermal insulation and even ceramic magnets are some novelties of this industry's indispensable maturity.

Transport of all kinds is well treated, aircraft in particular, both in war and in peace. Charles F. Kettering's self-starter is seen as the result of the perception that a small electric motor can be heavily overloaded for a short time with safety. The railroads reached maximum route distances all over Europe and North America a generation ago; the total mileage has fallen since then in several countries, but in Japan it has tripled since 1900. Landmark aircraft are pictured: the first practical airplane, the Wright Flyer; the first civil all-metal plane, the Junkers F13 of 1919; the first practical multiengine aircraft, the Fokker FVIIA-3m of 1925; the de Havilland Moth of 1925, the first widely used light airplane; the 1941 Sikorsky V.S. 300A, the first successful helicopter of what is now the conventional type, and the first jet fighter (flown operationally in 1944), the Messerschmitt Me 262. In 1957 the number of passengers who flew the Atlantic exceeded the number of those who went by ship.

The cruise missile, the ballistic missile and the radar- and wire-guided missile were all pioneered at Germany's wartime Peenemünde.

The elegant Gladesville Bridge carries a six-lane highway over a river near Sydney; it was built in 1964 to the design of Freyssinet with giant prestressed-concrete arches, the longest in the world. One arch rib weighs about 25,000 tons and does not depart from the true arch line by more than 10 millimeters anywhere in its 300-meter span. On the almost harborless coast of Ghana there is now a modern harbor at Tema, sheltered from the pounding waves of the deep Atlantic, a kilometer or two outside the haven, by five kilometers of rock breakwaters.

The first Otto engines were run not by gasoline but by town gas; they weighed about 200 kilograms per horsepower and were "too heavy to carry [themselves] uphill." That figure has improved about 400-fold since Victorian days for the same fundamental cycle. Automobiles are an obvious consequence; consider also how the chain saw and the bulldozer have transformed the ancient forest industries.

By 1910 or so Charles H. Norton had met Henry Ford's demand for precision production in hardened steels, the basis of the modern engine. To do so he had developed both new abrasive wheels (alumina and Carborundum, cheap and well controlled) and the key machine tool for shaping such parts as camshafts, crankshafts and cylinders in quantity: the precision grinding machine, now operated in many forms. ("Few of the millions who daily drive a car have ever heard [Norton's] name.")

Hot water was for the Victorians a "very great luxury indeed." The boiler in the kitchen range, the freestanding coal or coke boiler and the gas-fueled geyser that directly supplied the bath changed that once and for all. The geyser was invented and named in 1868; by the end of World War I the first priority in every British household was a supply of hot water.

In the comprehensive *Dictionary of Applied Physics* published in 1922 the vacuum tube went unmentioned. That could happen again a few years hence, but between the old galvanometer and today's integrated circuit came the "era of the tube."

The Zeiss Contax camera is now produced across the world by the firm of Yashica. Sic transit gloria Germaniae. Two of every three single-lens reflex cameras sold today are made in Japan.

White bread has its deficiencies, but so had the whole-meal bread of the villagers; its phytate is a contributor to trace-metal deficiencies. Long, interesting chapters cover food and water.

The narrative ends with the editor's reflections on the half century of mushrooming technology, with its two "embarrassing legacies": environmental pollution and an energy shortage. Williams recognizes the latter as the more pressing. How we came this way, rather than how we might steer now, is in fact the brightest nugget to be mined from this thick vein of ore.

GALILEO AT WORK: HIS SCIENTIFIC BIOGRAPHY, by Stillman Drake. The University of Chicago Press ($25). OPERATIONS OF THE GEOMETRIC AND MILITARY COMPASS 1606, by Galileo Galilei. Translated, with an introduction by Stillman Drake. The Burndy Library, Norwalk, Conn. ($6). "So put forward the arguments and the demonstrations, Simplicio... but not just text and bare authorities, because our discourse must relate to the sensible world and not to one on paper." Under that banner science has won many strong points since those first taken by the early heroes Galileo, Simon Stevin, Evangelista Torricelli and the rest. (There were other great names too, allies who followed another flag. Descartes, Kepler and even the incomparable Newton sought deeper causes, inferred great principles and invoked philosophical necessities, even though Newton once asserted that he did not "feign hypotheses.")

Galileo was a scientist of the sensory: the eye, the pulsebeat, the sung tune, the rolling weights and the rest. He recognized that the instruments we call the senses are only inborn biological members of a much wider set of possibilities. For example, lenses of glass can mightily aid the natural lens of protein. The matter is subtle, however, because right reason is indispensable to almost every observation.

Stillman Drake has long been the most sympathetic and among the most penetrating of contemporary Galilean scholars. Here he offers a convincing study of the scientific life of the great Tuscan, a celebration of the "narrow domain within which no appeal need be made beyond sense experiences and necessary demonstrations." Twenty-five years ago Drake published his fine translation of the *Dialogue concerning the Two Chief World Systems,* with a foreword by Albert Einstein. Einstein was pointed: "To put into sharp contrast the empirical and deductive attitude is misleading and was entirely foreign to Galileo." Old, blind but still thoughtful, Galileo himself clarified the primacy of experience: "I have learned sureness of demonstration from the innumerable advances made by pure mathematicians.... Thus far, therefore, I am Peripatetic.... And if Aristotle should see the things newly discovered in the sky, though he affirmed it to be unalterable... doubtless he... would now say the contrary. For indeed... when he says the sky is unalterable *because* no alteration has been seen there, he would now say it is alterable because alterations *are* perceived there."

To say much more on this issue of method would be to belie the book itself, which tries to avoid philosophical conjecture and the search for bookish and final influences; it puts together instead a study of improving experiment, clearer argument and growing results. Enduring meaning lies less in why than in how.

Galileo is seen first in Pisa, where he wrote against Aristotle on motion and where he was still preoccupied in the traditional way with the whys, the "search for causes as such." On to Padua, where his deft experiments continued on motion, with the inclined plane and split-second musical timing. This period ends with the "Holland spectacles," as Sarpi wrote of them, and the new heavens announced so gloriously in *The Starry Messenger*. Drake shows plainly that Galileo's moon diagrams were indeed factual. (This was brilliantly confirmed recently by E. A. Whitaker, himself a selenographer in Arizona.) The diagrams were modified only to accent the effects described in the text and were not presented as overall maps. Drake argues convincingly from manuscript sources that the Jupiter satellites too were watched with sober factuality. Galileo interpreted their motions as orbiting according to Copernican views only when those circles became the most plausible way to order the intricate dance. Not the utility of those invisible moons in the grand scheme but their daily procession was Galileo's concern.

Then to Florence, where he was always in the public eye. There was a decade of study and writing. "During these years Galileo supplied a large number of telescope lenses to various correspondents and there is little doubt that he ground these himself." Even by 1621 the scores of telescopes made in Florence by Galileo and in Venice by Sagredo and a lens grinder were hardly matched anywhere else in Europe. Galileo himself felt that the chief difficulty was in getting, outside of Florence and perhaps Venice, clear enough glass that was hard enough to take a fine polish.

The scene had long included Rome, the source of fame and of danger alike. A visit in 1616 had put Galileo under a rather friendly injunction not to make too much of the earth's motion. In 1624 he could take along a princely gift, a compound microscope (probably his own invention a few years before), and heard nothing but praise from the Pope. By 1632, with his *Dialogue concerning the Two Chief World Systems* in print, the scientist was hostage to the public man. There was a report that one Father Scheiner, on hearing a priest praise the *Dialogue* as the greatest book ever written, turned livid and shook so badly that the witness was astonished. Soon came Galileo's summons to the Inquisitor, the abjuration on his knees before the cardinals, leniency against torture and the sentence to life imprisonment, which was commuted to house arrest in Arcetri. Protestants such as Hobbes and Milton could still come to see the old man; it was harder for his old students to gain permission. The end of the high drama was at hand,

The rising young professor at Padua had sought to augment reputation and income by giving private lessons, at first on military architecture. After a few years Galileo began to improve a well-known artilleryman's instrument for his well-to-do students. Out of it he made what is properly called in English the sector, which he referred to as his *compasso geometrico et militare*. It was a brass device with legs that opened out at any set angle; each leg was marked with half a dozen graduated scales that made possible a variety of modest arithmetic calculations, without pencil and paper, by those innocent of the algorithms. The calculations included square and cube roots and determination of the fourth term, given one ratio and one other number (the rule of three).

It was a practical device that freed workmen and noble gentlemen alike from dependence on trained captains and engineers, who then were few in number. It was accurate to 1 percent or better—most effective. After a couple of years Galileo employed a live-in instrument maker to manufacture the new *compasso;* three remain of the scores that were made. Once Galileo had published his manual he also licensed the instrument's manufacture in several other places. The device entered common use all over Europe within a generation; most indications of Galileo's priority of invention were lost.

This story, told by Drake in *Scientific American* in 1976, is retold here in a translation of the very rare manual with its original figures, an inexpensive memento of the grandeur of the 17th century and a powerful temptation to people who like fine old instruments.

Galileo was science's hero in the 19th century, seen as devoted to the evidence and against speculation, a traveler with light metaphysical baggage. The reaction against this half myth was strong, and the historians have since made him a closet Platonist, a man who knew the answers before experiment, a little of a fake, certainly devout. Drake goes far to damp the swing of this particular pendulum. The experiments seem to have been real, probably even the two weights dropped from the Leaning Tower of Pisa and the swinging lamp. The historians are shown to have misunderstood the very tests they accused Galileo of neglecting.

Probably the old man really did mutter, "It moves, nevertheless"—not to the stern cardinals but to the Archbishop of Siena, who was host to the scientist after the sentencing. The unfairness of the process had shaken Galileo terribly; someone wrote that he "went through the night crying out and rambling so crazily that it was seriously considered whether his arms should be bound" lest he injure himself. His composure and perhaps his wit returned as he parted from Siena, to find a place in history he would have much enjoyed: complex, subtle and pathbreaking, no demigod but a model of human greatness.

COLOR VISION: AN HISTORICAL INTRODUCTION, by Gerald S. Wasserman. John Wiley & Sons ($18.95). Among the 700,000 buyers of this magazine perhaps two or three are unable to discriminate colors in any way; they are rod monochromats, who see only with the detectors that mediate colorless night vision in the rest of us. Their perceptions are understandable; the cause—no retinal cones—is simple and clear. The commonest form of color blindness, however, is an inability to distinguish between reds and greens. For such viewers the ability to see yellow and blue hues remains. (We know these perceptions well from detailed study of a few rare individuals in whom one eye is abnormal and the other is quite normal.) These people never report the loss of a single color sensation; for them colors always drop out in pairs. The introductory textbooks in their serried ranks all describe it differently, however; "a deduction from a theory [is] taken for a fact."

The theory, of course, is the three-color theory, based on the widely held view, known to be wrong since the first experiments of Helmholtz, that three primaries can be added to match any color. The three-color television and photographic processes are implicitly invoked. Actually those familiar systems cannot in general match vivid colors; they are clever technical compromises that can attain pleasing results, not accurate matches. Indeed, high-quality color printing actually uses four or even five colors, including black, and anyone who tries can convince himself that a brown patch, say, baffles the usual textbook model.

Of course, there is truth in the three-color model; along the way to the present fairly deep understanding we traveled that path, led by such towering names as Newton, Thomas Young, James Clerk Maxwell and Helmholtz. The whole truth is only algebraically different: to match any color one of the three primaries must sometimes be mixed with another (arbitrary) color. The true match is of two primaries to a third plus some other; there are indeed three variables, but they cannot always be additive.

This unusual text, readable and interesting throughout, takes the reader almost nonmathematically on a logical tour along the historical route to the theory of color vision. The treatment of old, overconfident partial understandings and of how they arose and declined is itself a wonderful mixing process, which makes a good match between what we all half-learned and the far subtler structures of today. Even the poet Goethe (who railed against the laboratory: "Friends, escape the dark enclosure / where they tear the light apart / and in wretched bleak exposure / twist, and cripple Nature's heart") is given his due. In fact, the centuries of color experimentation produced two classes of contrasting theories, each class with strong support from experience. One class, the component theories, appealed to three independent components in the system (as in television) and the responses of those components to stimulation. Those theories do explain much, but not all. Such clear subjective experiences as the three psychological pairs, black-white, red-green and blue-yellow, have plenty of objective consequences: striking contrast, shadow and additivity-failure effects. Again there is an algebraic relation. The three variables can be chosen from three paired opponent stimuli, which must, to be sure, include both positive and negative contributions. The three-color component theory, however, needed the same extension.

In time a more complex class of models has arisen, here called zone theories. Much more satisfactory, these models take as their premise that within the visual system there are portions—call them zones—each of which functions either in a component mode or in an opponent mode. The visual receptors may satisfy a component model, whereas the processors in the central nervous system may rely on the opponent transformations. Can we decide?

For a long time the history of color-vision study was chiefly the history of experiments with a black box: the observer and his reports. Only the invaluable natural experiments of genetics now and then made it possible to peek under the lid. For the past 50 years, however, and mainly since the 1950's, progress in neurophysiology and optical techniques has pried the box almost open. Single receptor cells and their neighbors have been studied both electrically and spectrophotometrically. There is a growing consensus from this remarkable work, usually done in the goldfish or the monkey, that we have seen correlates of both types of classical model. The retinal receptors are more complex than we might have guessed, with two-peak response, feedback and strong coupling among cells and with spatial coding for color-contrast responses. The hardware is coming clear, and the software of this marvelous computing system is beginning to be unraveled.

Wasserman, a Purdue sensory physiologist, has drawn on his special knowledge and a flair for exposition to make a modest but deep book that can clarify this tough topic for any attentive reader. From it he draws a conclusion on method: Science is a slowly self-correcting institution whose course can be true only over the long run and only if diversity of view and plurality of opportunity are encouraged. It is a world of the tentative, whose ultimate arbiter is not paper but complex and slowly won experience (even Goethe's). The reward for the scientist is the "profound pleasure of understanding that which was once not understood." The reward for society includes that understanding, and perhaps the improvement of human life.

THE PHYSIOLOGICAL ECOLOGY OF TUNAS, edited by Gary D. Sharp and Andrew E. Dizon. Academic Press, Inc., Publishers ($29.50). Tuna and their kin, the mackerel and the bonito, are fierce blue-water nomads. Their high-speed schools feed on lesser fishes and even on plankton in the open sea; they are high-performance predators, the apex (apart from mammals) of the ocean's food pyramid. A skipjack tuna can develop 40 times the power per unit of body weight released by more typical fishes of the same size. This top-of-the-line quality has caught the imagination of physiologists, and the economic value of the tuna fisheries (an annual U.S. take of $1 billion worth of fish) has provided opportunities and an incentive to understand the world of the tuna.

Once upon a time such studies were made by examining a few recently caught fish and perhaps weighing their stomach contents and sampling their blood. Now for 20 years it has been possible to study live tuna, held captive in circulating seawater tanks in the Kewalo Basin laboratory of the National Marine Fisheries Service in Honolulu. Free-swimming tuna have been tagged with tiny ultrasonic transmitters and followed across the seas; there is even tuna temperature telemetry at sea. In this fascinating, if technical, report of a workshop held in 1977 a school of experts review the state of our tuna understanding, a biological challenge only beginning to be met. The volume is in the "rapid manuscript" style (typewriter text with ragged right margins and no index), but it has a very welcome inset of eight colorplates, in addition to graphs, drawings and halftones.

Tuna are different, all right. The most striking difference was first noted in 1835: they have warmer bodies than other fishes caught in the same waters. The discoverer, John Davy, saw in tuna the "exception to the generally received rule that fishes are universally cold-blooded." It is no small effect: the temperature can rise 20 degrees Celsius or more. Tuna have either small swim bladders or none at all, and they are therefore heavier-than-water submarines, required to swim continuously to stay afloat in the open sea. They lose half of their heat, according to some measurements, by the forced ventilation of their gills, because they swim at speed with their mouth held open in order to take in oxygen more effectively than any other fishes. They digest rapidly, have large amounts of body fat near their muscles and have two kinds of muscle: red muscle for sustained aerobic effort, it appears, and white muscle for peak bursts of anaerobic power. Along with the sharks, they sport countercurrent heat exchangers in their bloodstream; a color photomicrograph of the structure is the first picture in the book.

Full understanding has not been attained. Although easy hypotheses have been plentiful, most of them have been punctured. The authors of these chapters have thrown the book at the problems, with obvious enjoyment. The chapters deal with energetics, enzyme properties, thermal biology in depth, hydrodynamic and heat-loss models of considerable sophistication, plenty of data on free and captive fish, studies and models of bluefin cruises (5,000 miles at good speed).

Do tuna need to save heat so that the muscular power plant will work better? Maybe, but some calculations show that "a Dewar flask with an internal heat source" would cook internally. Big tuna need to lose heat, not to conserve it. Yet the red muscles appear to lie below the surface only in those species found in cooler waters. What patchy feature of the distribution of the tuna's scarce nutrients (on the average only parts per billion of the waters they forage) can account for its great schools? Why, if tuna need so much energy, do they expend it fighting negative buoyancy?

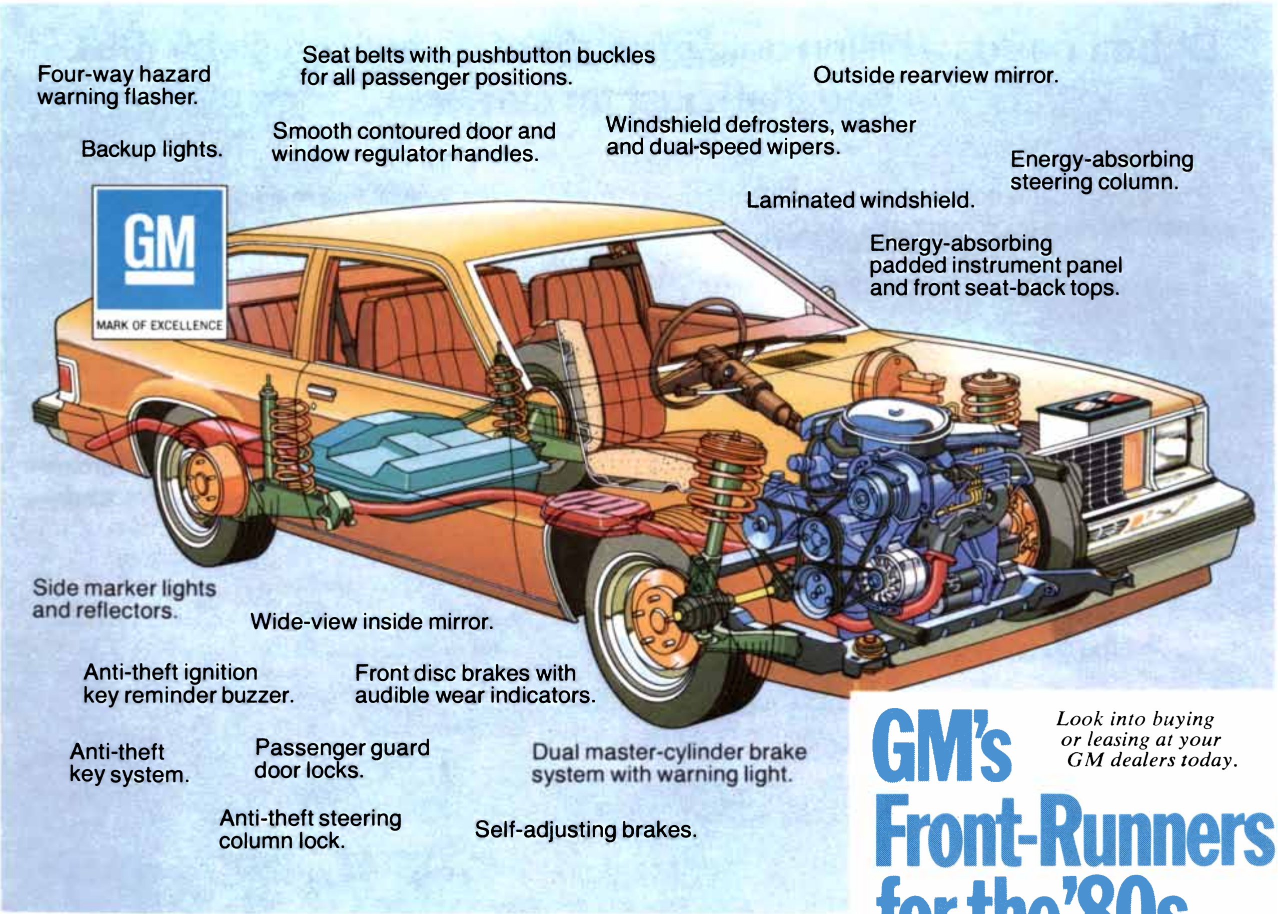
The advanced level of these discussions and the long path to knowledge ahead give an air of excitement to the technical accounts. Several authors suggest (and it seems plausible to a general reader) that the adaptations of the tuna are dynamical in nature, not to be learned from mere static measures. The fish may need to change depth rapidly in order to hunt well, making a buoyant state not so useful. They may need both to conserve heat on a sustained cruise and to lose heat fast during peak activity of modest duration. Central control over heat-exchange processes, and the consequent ability to manage temperatures in the two-muscle system, may make possible a very wide dynamic range in performance by matching a temperature rise with an increase in output—sometimes. These are not racers but versatile hunting nomads. Control may be the requisite for high performance: not only powerful dual engines but also subtle throttles, smooth transmissions and optimal heat flows.

GM's New Front-Wheel-Drive Cars

For occupant-protection and accident-avoidance, standard safety features include energy-absorbing steering column, padded instrument panel, dual master-cylinder braking system, side marker lights and more.

Every new front-wheel-drive Chevrolet Citation, Pontiac Phoenix, Oldsmobile Omega and Buick Skylark offers an impressive list of standard safety features. Among these are some developed by General Motors, such as the energy-absorbing steering column and audible wear indicators for our front disc brakes.

In addition, you'll notice certain provisions to help prevent theft. Like the safety features, we hope you'll never need them. But they're there just in case.

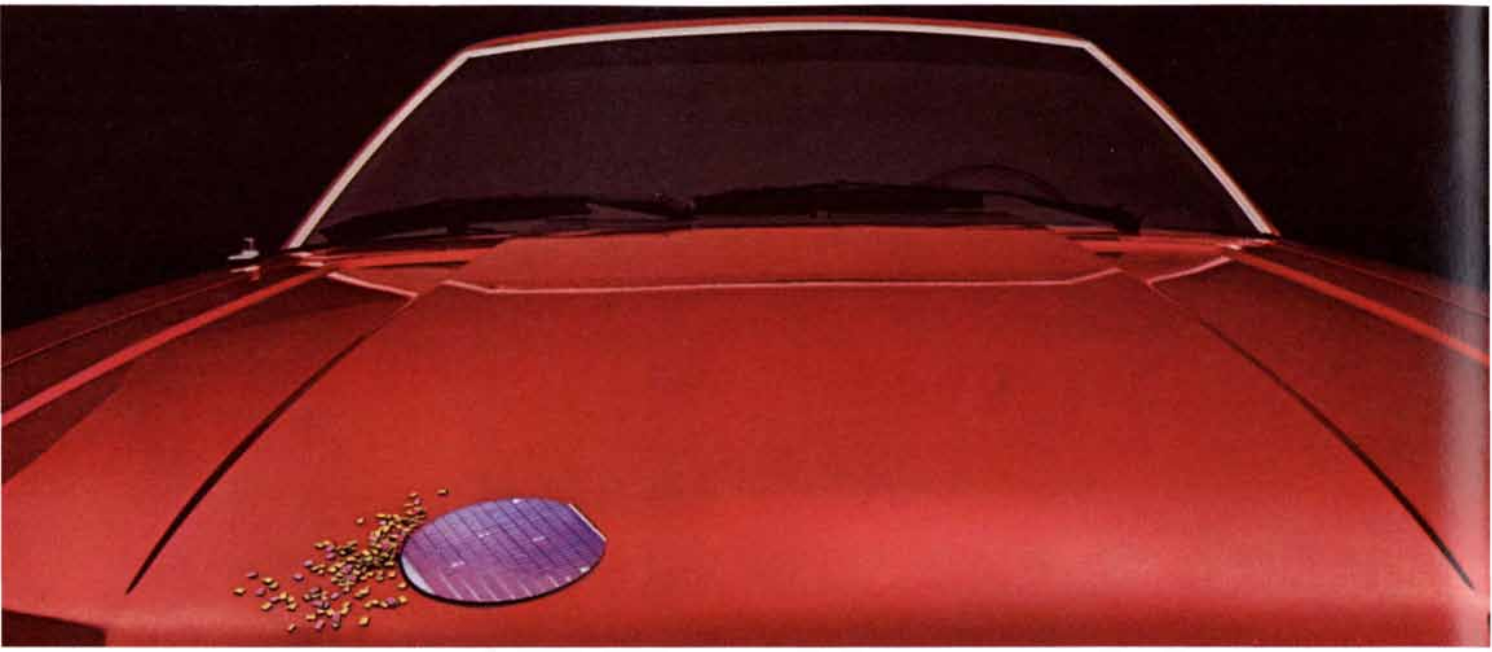


GM's Front-Runners for the '80s.

Look into buying or leasing at your GM dealers today.

Chevrolet Citation, Pontiac Phoenix, Oldsmobile Omega, Buick Skylark.

HOW THE NEW EATON PROFITS FROM CHANGE



Detroit needs a billion dollars worth of miracle chips by 1981. And that's just for starters.

The wafer on the hood holds up to 300 chips. Each chip is a microcomputer. A computer that will change the way you drive. The first major application will be the ignition system of your car.

We profit from change in truck components...

Computer chips will also direct air and fuel mixtures, govern speed, signal when anti-freeze is low, when to change oil. They'll shift gears and balance suspension when you corner. They'll increase performance and save gas.

The companies making these chips are growing fast. So the companies that make the equipment that makes the chips have to grow just as fast. And we're in the thick of that growth – through our Cutler-Hammer group, a leader in electronics and a recent Eaton acquisition.

automated storage and retrieval systems...

and hydrostatic transmissions for farm equipment.

Our base is broad and balanced. Our sales have grown to an annual rate of $3 billion. Earnings per share have increased 45% during the last five years. And with our growing involvement in advanced technology for industry, we expect to profit from change even more in the years ahead.

For the complete story write: Eaton Corporation, World Headquarters, 100 Erieview Plaza, Cleveland, Ohio 44114.

EATON

Established 1845 SCIENTIFIC AMERICAN July 1979 Volume 241 Number 1

The Prompt and Delayed Effects of Nuclear War

The prompt effects of nuclear weapons are the basis for the size of U.S. strategic forces. The delayed effects are equally great, ensuring that these forces remain a more than ample deterrent

by Kevin N. Lewis

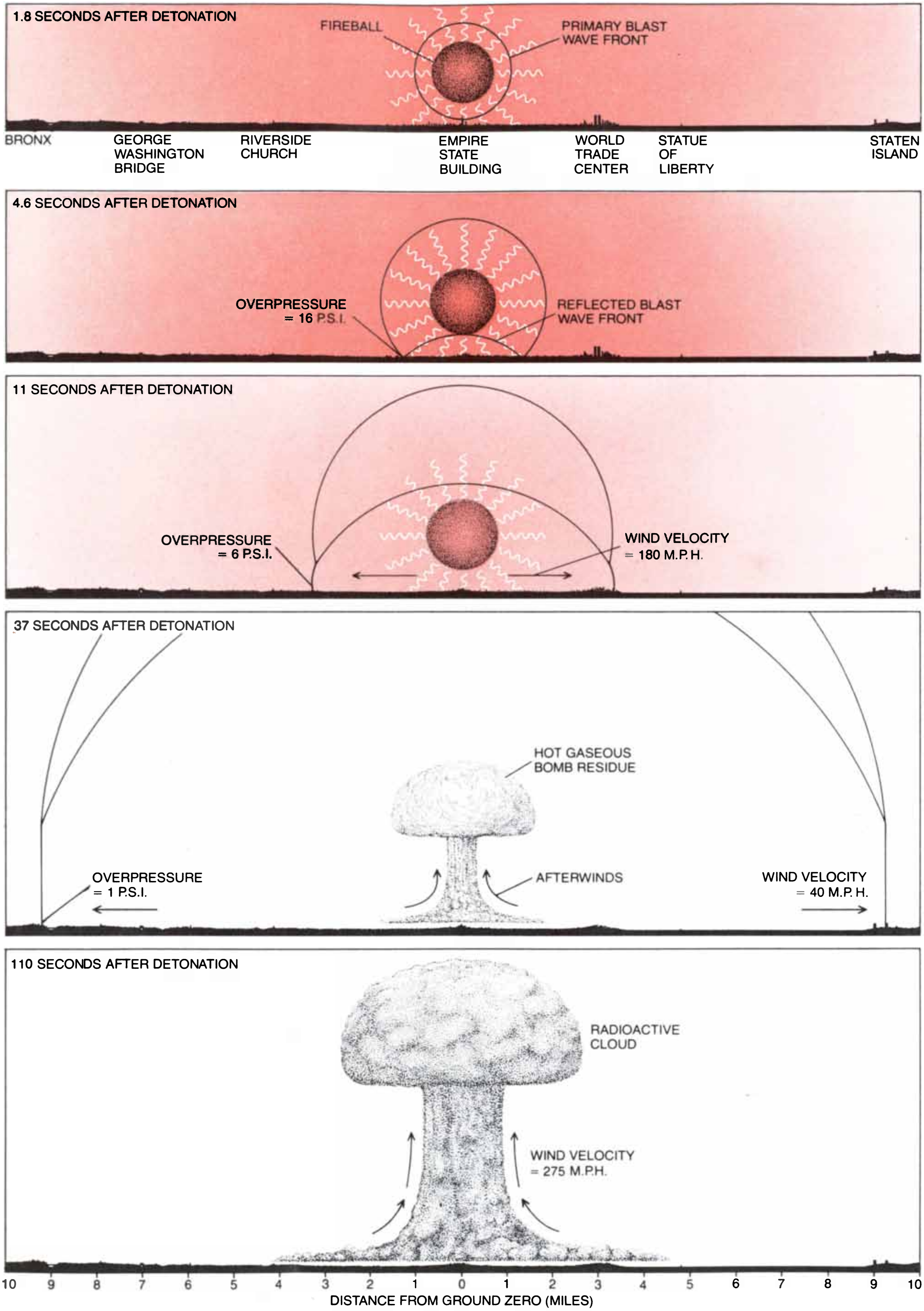
The primary purpose of this country's strategic nuclear forces is to deter the U.S.S.R. from launching an attack on the U.S. or its allies. To accomplish that mission the U.S. maintains the constant ability to inflict intolerable damage on the U.S.S.R. The long-range missile and bomber forces of the U.S. have been designed to survive even an all-out surprise attack by the U.S.S.R. in numbers sufficient to deliver a devastating retaliatory counterattack. Since the U.S.S.R. has similar forces, it is considered unlikely that either side would find it advantageous to attack the other. It is this mutual retaliatory potential, or assured-destruction capability, that is widely held to be responsible for the strategically stable military balance between the two superpowers.

Since in this view the avoidance of war depends in part on the integrity of the assured-destruction capability of the U.S., any degradation of that capability would be a grave matter. Accordingly recent assertions by some military analysts that the U.S.S.R. is actively pursuing measures to reduce the effectiveness of an American retaliatory strike have given rise to much concern. Specifically it is alleged that ambitious Russian civil-defense initiatives could create a dangerous strategic asymmetry in the absence of countervailing U.S. efforts. For example, in conjunction with a surprise "counterforce" attack on U.S. land-based missiles the U.S.S.R. could attempt to evacuate its cities, with the projected result that Russian fatalities in an all-out nuclear exchange would be substantially fewer than American ones. In such a situation the U.S. might be inhibited from further escalating hostilities, and the U.S.S.R. would then in various ways be able to impose its will. Even if all-out war were to ensue, the U.S.S.R., it is said, would be able to recover much faster than the U.S. One result of this line of reasoning has been a revival of interest in the moribund American civil-defense program; another has been the consideration of new strategic-missile targeting options designed to defeat the Russian civil-defense program.

Such hypothetical scenarios are based in part on underestimates of the damage the surviving U.S. forces could inflict on the U.S.S.R. Many estimates of this kind include only the easily calculable blast effects of nuclear weapons. They ignore the equally devastating effects of thermal radiation and ionizing radiation. When these additional effects are included in the calculations, it is clear that nuclear war remains an unmitigated mutual disaster, and that no conceivable civil-defense preparations could materially change the prospect. Therefore from an operational military point of view there is no validity to assertions that the U.S. retaliatory capability is "eroding." Moreover, it is extremely unlikely that the situation will change in the foreseeable future.

How is the damage from nuclear war estimated, and what consequences of such a war are routinely excluded from calculations of the damage? In this article I shall compare calculations frequently used to assess "adequate" levels of assured destruction with estimates of the probable wider results of a nuclear exchange between the two superpowers. The more comprehensive analysis shows that neither the U.S. nor the U.S.S.R. needs to be concerned about the integrity of its retaliatory capability. Although much of the current debate on the gravity of the Russian threat tends to ignore this fact, there can be no conceivable doubt that all-out war remains a losing proposition for both sides. Credible deterrence of course relies on many factors other than the ability to conduct a massive retaliatory attack. It is in the interest of all parties, however, that the notion of "winning" an all-out nuclear war, in the sense of one side's being able to improve its relative position at an acceptable cost, be dismissed from the strategic debate, and that the full consequences of such a calamitous event be brought to public attention.

Specific criteria of retaliatory effectiveness were first established under the direction of Secretary of Defense Robert S. McNamara in the early 1960's. Up to that time strategic military planners lacked any formal quantitative standards for determining the appropriate levels of U.S. retaliatory forces. Secretary McNamara therefore advanced the concept of assured destruction, arguing that the destruction of between 20 and 25 percent of the U.S.S.R.'s population and at least 50 percent of its industrial capacity would constitute unacceptable damage in the eyes of that country's leaders. By establishing these measures McNamara was better able to coordinate Air Force and Navy planning, to match strategic military requirements with existing force structures and to eliminate programs that were superfluous. Although the task of defining a certain level of damage had a political purpose, namely to threaten the government of the U.S.S.R. with intolerable de-


1.8 SECONDS AFTER DETONATION
FIREBALL
PRIMARY BLAST WAVE FRONT
BRONX
GEORGE WASHINGTON BRIDGE
RIVERSIDE CHURCH
EMPIRE STATE BUILDING
WORLD TRADE CENTER
STATUE OF LIBERTY
STATEN ISLAND
4.6 SECONDS AFTER DETONATION
OVERPRESSURE = 16 P.S.I.
REFLECTED BLAST WAVE FRONT
11 SECONDS AFTER DETONATION
OVERPRESSURE = 6 P.S.I.
WIND VELOCITY = 180 M.P.H.
37 SECONDS AFTER DETONATION
HOT GASEOUS BOMB RESIDUE
OVERPRESSURE = 1 P.S.I.
AFTERWINDS
WIND VELOCITY = 40 M.P.H.
110 SECONDS AFTER DETONATION
RADIOACTIVE CLOUD
WIND VELOCITY = 275 M.P.H.
10 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 10
DISTANCE FROM GROUND ZERO (MILES)

struction, the specific percentages chosen reflected the capabilities of the U.S. strategic forces programmed at that time. The required levels of destruction were also based to some extent on the characteristics of the particular target system represented by the U.S.S.R.

The population and economic resources of the U.S.S.R. are concentrated in a remarkably small number of major urban centers. About a third of the population and nearly two-thirds of the industrial capacity are concentrated in the country's 200 largest cities. Nuclear attacks on additional cities would not appreciably increase the retaliatory damage (except for the delayed effects of radioactive fallout). McNamara's criterion of assured destruction could therefore be loosely translated into the ability to destroy the 200 largest cities in the U.S.S.R.

Given this assumption, U.S. force requirements could be set by determining the number of nuclear warheads needed to destroy the social and economic targets of importance in those 200 cities. Target planning is sensitive to many operational factors, such as the composition and layout of cities, but above all it calls for predicting accurately how the local population will be affected by the lethal effects of nuclear explosions. In actual practice retaliatory damage is predicted by matching the physical properties of nuclear explosions with the relevant target characteristics on a city-by-city basis. In calculating such damage levels U.S. planners have at their disposal a large store of information on each target, sophisticated analytical techniques and an advanced data-processing capability. The results of these detailed calculations can be approximated fairly well, however, with the aid of some simple procedures.

The yield of a nuclear weapon is usually described in terms of the quantity of chemical explosive required to release an equivalent amount of energy; a nuclear weapon is said to have the power of kilotons (thousands of tons) or megatons (millions of tons) of TNT. As in a chemical explosion, the energy from a nuclear explosion is generated very quickly in a small volume. When the nuclear explosion is set off in the air, the energy released instantaneously vaporizes the components of the warhead, creating a hot, rapidly expanding fireball. The explosion gives rise to two prompt effects that in an attack on a city can be devastating. First, as the fireball expands it sends a shock wave through the surrounding medium. The shock wave, which travels away from the point of the explosion at supersonic speeds, does blast damage to structures and people. The hot fireball also radiates thermal energy, mainly photons in the visible and infrared regions of the electromagnetic spectrum, which can cause burns and ignite materials that are not protected by some kind of opaque screen. Roughly half of the weapon's energy is eventually converted into mechanical blast motions and about a third is released in the form of thermal radiation. The rest of the energy is represented by prompt nuclear radiation and delayed thermal and nuclear radiation, none of which are treated as being important in assured-destruction planning but all of which nonetheless add to the destructiveness of a nuclear attack.

The mechanical motions of a nuclear explosion are analogous to those of a tidal wave. The shock front is literally a wall of compressed air. As it passes, structures are exposed to a nearly instantaneous rise in the local atmospheric pressure, and they may be crushed. Following the shock front are strong winds analogous to the water currents that follow a moving ocean wave. The forces resulting from these winds may also lead to the collapse of structures in the target area. Depending on their shape and construction, buildings may be vulnerable either to the shock wave or to the winds that follow it or to both. The "hardness" of a target (its ability to withstand the destructive effect of the shock wave) is generally described in terms of the induced peak "overpressure" (in pounds per square inch above atmospheric pressure) at which the target is destroyed.

Thermal radiation can lead directly to flash-burn casualties and indirectly (through the ignition of nearby materials) to flame-burn casualties, superposing both effects on blast casualties. The extent of such damage depends on both the power of the radiant energy delivered (usually measured in calories per square centimeter) and the period over which the energy is delivered. Destructive blast effects decay with distance faster than thermal effects. Therefore under ideal conditions a nuclear explosion can do substantial incendiary damage well beyond the area devastated by blast. The thermal damage, however, is much influenced by external factors, including the presence of clouds or of snow cover, the relative transparency of the atmosphere and the composition of the target. Hence thermal effects are far less predictable than direct blast effects.

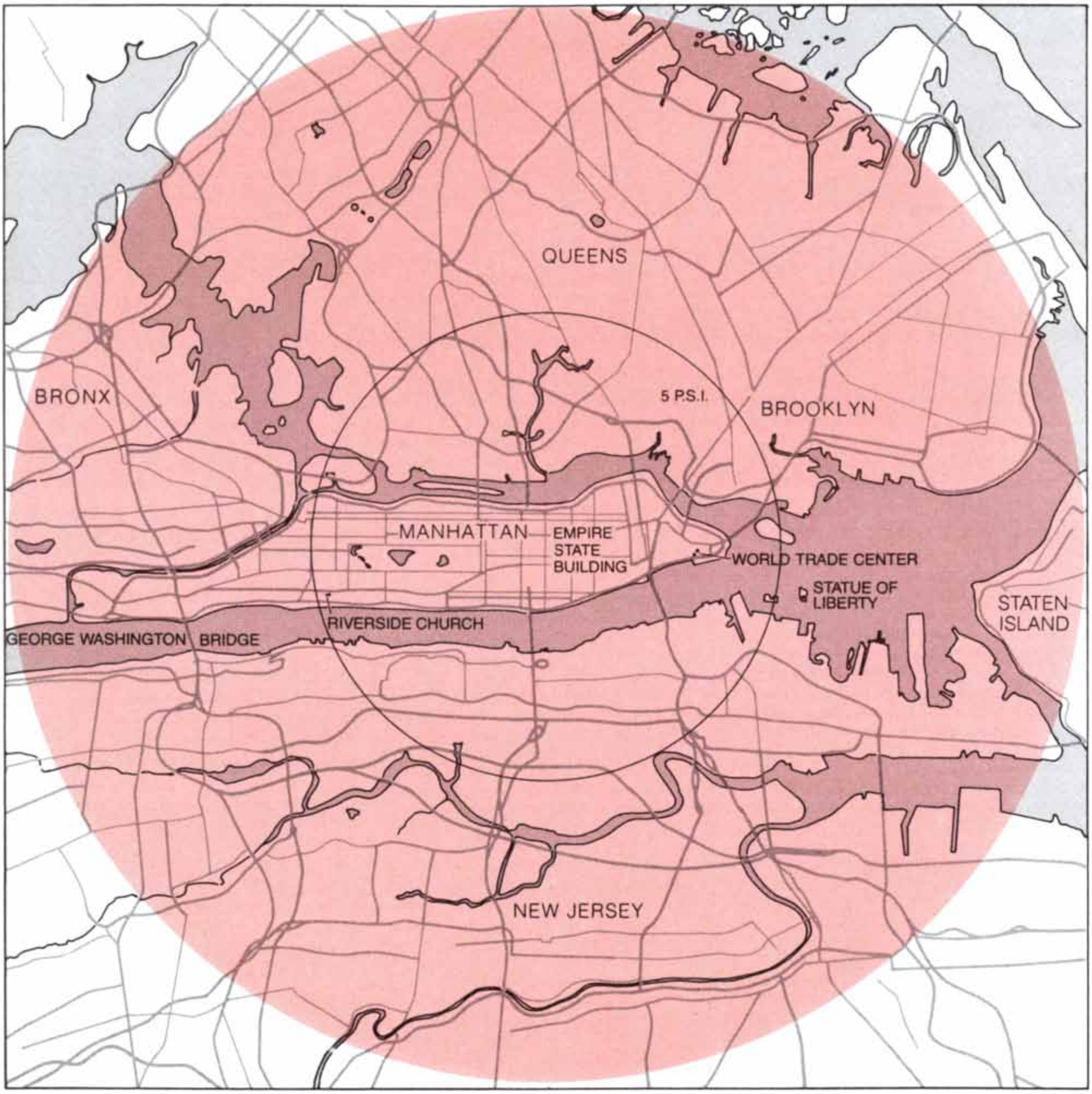
Since retaliatory forces are planned on the basis of assured damage, the consequences of an attack are typically calculated only on the basis of the more predictable blast effects. Consider the problem of allocating a suitable "package" of nuclear weapons to an urban area after a review of the targets within that city. Aim points for each weapon are selected in such a way as to ensure that the desired blast effects will cover all the targets. If the targets are close enough together, a single warhead may suffice. If the targets are dispersed or hardened, it may be preferable to allocate more than one weapon to a target area, as opposed to increasing the yield of a single weapon. This approach guards against the failure of a single large warhead, which would leave a tar-

THE PROMPT EFFECTS of the explosion of a one-megaton nuclear warhead detonated at a height of 6,500 feet over the heart of New York are depicted chronologically in the sequence of scenes on the opposite page. Immediately after such a detonation an extremely hot, luminous fireball would form. The fireball would emit intense thermal radiation (*color*), capable of causing skin burns and starting fires at a considerable distance. The explosion would also give rise to a destructive blast wave, which would move away from the fireball at supersonic speed; at 1.8 seconds after the detonation, for example, the front of the blast wave (*black circle*) would be roughly half a mile ahead of the fireball. In addition the nuclear processes responsible for the explosion would be accompanied by the emission of hard radiation, mainly gamma rays and neutrons (*wavy white lines*), which would have enough range in air to reach the ground in the midtown area. When the primary blast wave from the explosion hit the ground, another shock wave would be caused by reflection. At a certain distance from ground zero (depending on the height of the explosion and the energy yield of the weapon) the primary and reflected wave fronts would fuse near the ground to form a single reinforced Mach front; in the case of a one-megaton warhead detonated at 6,500 feet the Mach effect would begin some 4.6 seconds after detonation at a distance of 1.3 miles from ground zero. At that point the overpressure (that is, the air pressure above ambient atmospheric pressure) would be 16 pounds per square inch (p.s.i.). At 11 seconds after detonation the Mach front would have moved outward to 3.2 miles from ground zero, the overpressure at the Mach front would be 6 p.s.i. and the velocity of the wind just behind the front would be approximately 180 miles per hour; appreciable amounts of thermal radiation and nuclear radiation would continue to reach the ground. At 37 seconds after detonation the Mach front would be nearly 9.5 miles from ground zero, the overpressure at the front would be 1 p.s.i. and the wind velocity behind the front would be 40 miles per hour. (Glass would be broken at overpressures down to .5 p.s.i.) Although thermal radiation would no longer be significant, gamma rays would still reach the ground in potentially lethal amounts. The fireball would no longer be luminous, but it would still be very hot, and it would therefore rise rapidly, causing air to be drawn inward and upward, producing strong air currents called afterwinds. These winds would raise dirt and debris from the city to form the stem of what would eventually become the characteristic mushroom cloud. By 110 seconds after detonation the hot residue of the fireball, while continuing to rise, would have begun to expand and cool. As a result the vaporized fission products and other weapon residues would condense to form a cloud of radioactive particles. By this time the cloud would have risen to a height of seven miles. The maximum height attained by the cloud (after 10 minutes) would be about 14 miles. Ultimately the particles in the cloud would be dispersed by the wind, and unless there were precipitation there would probably be no early (or local) fallout of radioactive material.

get "uncovered." It also reflects the fact that few industrial and military complexes are sufficiently concentrated or have the right shape to be attacked by a single weapon of the type that currently constitutes the bulk of the U.S. strategic arsenal.

Each city has a unique set of target characteristics, but some simple rules make it possible to predict damage and fatalities. In general any structure not specifically designed for blast resistance would be destroyed if it were exposed to an overpressure of five or more pounds per square inch (p.s.i.) above the ambient atmospheric pressure of some 15 p.s.i., and those structures that would not actually collapse would typically be damaged beyond repair. Some reinforced buildings (and heavy equipment inside them) could withstand an overpressure of 40 p.s.i. or more, but if these targets were considered important, an attacker could lower the height at which his weapons were set to explode or could aim his weapons (or allocate new warheads) to achieve the desired effects. Still, as a rule of thumb an overpressure of 5 p.s.i. is considered sufficient to destroy most structures.

The human body can endure a far more intense blast than most buildings. Therefore in a nuclear attack most of



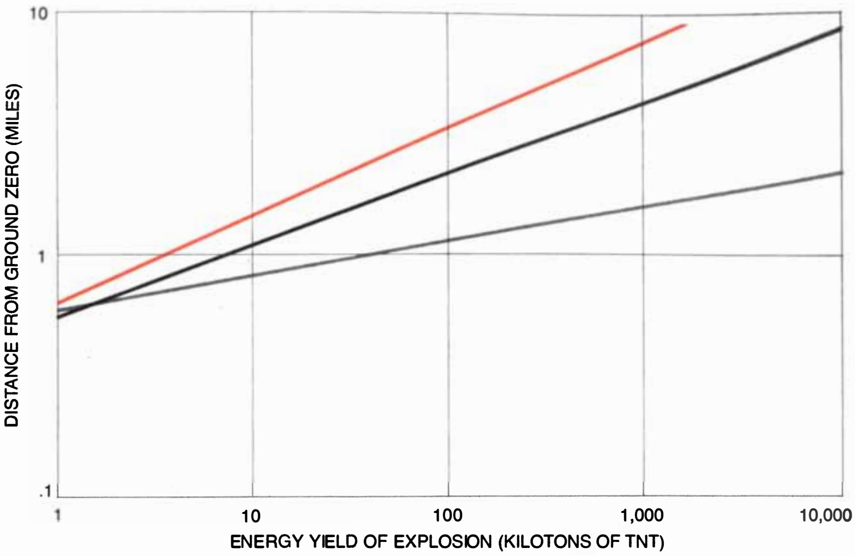
LETHAL AREA is defined by U.S. nuclear-war planners as the circular region within which the number of survivors of a nuclear explosion equals the number of fatalities outside the region. This simplifying assumption makes it possible to arrive at an estimate of the prompt fatalities resulting from a nuclear explosion by multiplying the lethal area by the population density (assuming that the population density over the entire area is uniform). As a general rule the lethal area is considered to extend roughly to the 5-p.s.i. overpressure contour, which for the one-megaton airburst represented on page 36 corresponds to a circular area with a radius of 4.3 miles (*area within black circle*). The lethal-area concept excludes several important (if less predictable) delayed effects of nuclear explosions, such as fires and radioactive fallout. On a clear day, for example, a one-megaton airburst could ignite fires as much as 10 miles away. If these fires were to consolidate into a mass fire, the entire region within that range (*colored area*) would be devastated, enlarging the lethal area fivefold.

the blast casualties would be caused by indirect effects. The bulk of the population would be at risk from being inside or near collapsing buildings, from being hit by debris thrown by the shock wave or from being hurled into an immobile surface. Thermal effects would also cause many fatalities within a certain range, regardless of external conditions. In estimating fatalities the simplifying concept of the "lethal area" is often used. Based on theoretical and empirical data developed by the Atomic Energy Commission in the 1950's, the lethal area is defined as the circular region within which the number of survivors would equal the number of fatalities outside the circle, assuming that the population density over the entire area is uniform.

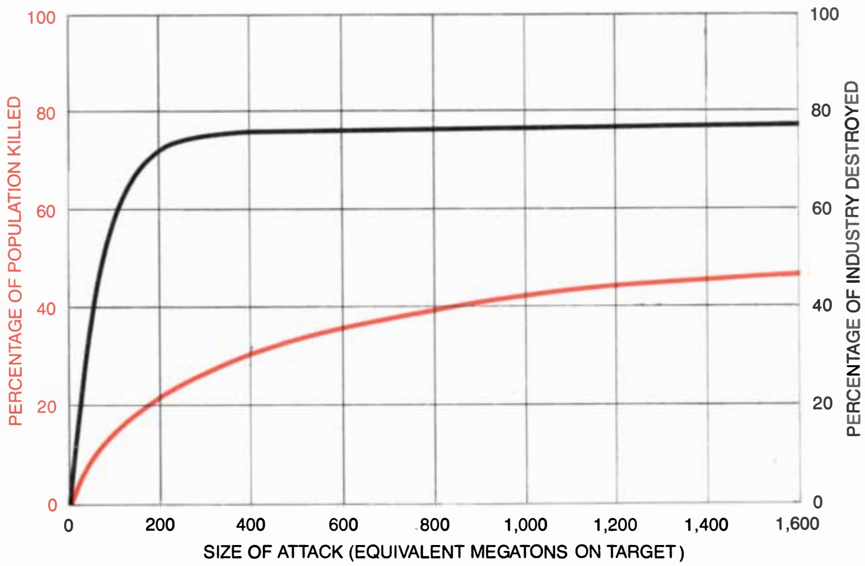
To simplify the calculations the estimated fatalities are redistributed, so that planners consider everyone within the circle to be a fatality and no one outside the circle to be a fatality. An estimate of prompt fatalities is then made by multiplying the lethal area by the population density. The experience of Hiroshima and Nagasaki and also test data indicate that for weapons in the range of 20 kilotons the lethal area extends roughly to the contour within which there is an overpressure of 5 p.s.i. Hence coverage by that overpressure is considered a satisfactory standard for calculating both the fatalities and the economic destructiveness of nuclear explosions.

Nuclear weapons will generate an overpressure of 5 p.s.i. to a distance proportional to the cube root of their yield. For this reason larger weapons are said to distribute their destructive power less efficiently than smaller ones. For example, a 100-kiloton bomb will generate an overpressure of 5 p.s.i. to a range of about two miles. Yet a warhead with 10 times the explosive yield (one megaton) will generate the same overpressure to only about twice that distance. In recognition of the inherently greater efficiency of smaller weapons, a scaled measure known as equivalent megatonnage, defined as the yield of a bomb in megatons raised to the two-thirds power, is considered a better index of countercity capability than the unadjusted yield in megatons. It was calculated by McNamara's systems-analysis staff in the 1960's that the reliable delivery of 400 equivalent megatons would kill 30 percent of the population of the U.S.S.R. and destroy 75 percent of the industrial capacity; more recently the population damage and industrial damage have been estimated to be closer to 35 and 70 percent.

In actuality these damage levels are the lowest that would result from nuclear explosions, since they are typically calculated on the basis of the predictable "prompt" effects described above. When delayed effects (fires, fallout and



TYPICAL RANGES to which three different harmful effects of nuclear weapons extend are represented here for a typical airburst as a function of the energy yield of the explosion. The colored line shows the distance to which thermal radiation can cause second-degree skin burns and ignite fires, creating the risk of a mass fire. The black line measures the radius of the 5-p.s.i. overpressure circle, within which the passage of the blast wave front, followed by 160-mile-per-hour winds, would cause massive urban destruction and a high percentage of fatalities. The gray line gives the range to which prompt nuclear radiation from the explosion would result in 100 percent fatalities. It is evident that under favorable weather conditions the destructive thermal effects of such an explosion could reach well beyond the area of major blast destruction. Prompt nuclear radiation, on the other hand, is clearly not an important damage mechanism for strategic nuclear weapons (which have explosive yields of anywhere from a few tens of kilotons to many megatons), since the areas covered by deadly radiation would also be exposed to severe blast and thermal effects. It is only at much lower yields (on the order of a kiloton or less) that prompt nuclear radiation becomes an important lethal mechanism; that relation in fact is the basic principle of the enhanced-radiation weapon, or neutron bomb.



ASSURED-DESTRUCTION CRITERION, relied on by U.S. strategic planners to determine the retaliatory potential needed by U.S. nuclear forces to deter a surprise attack by the U.S.S.R., is calibrated here in terms of the number of delivered equivalent megatons it would take to destroy key population centers and industrial targets in the U.S.S.R. (Equivalent megatons are defined as the explosive yield of a nuclear weapon raised to the 2/3 power.) Given the decreasing value of adding extra equivalent megatons to such a retaliatory attack, it is evident from these curves that the delivery on target of some 400 equivalent megatons would be more than adequate to achieve assured destruction. Population damage (*color curve*) was estimated in terms of fatalities only; industrial damage (*black curve*) was determined by calculating the "manufacturing value added" destroyed during a U.S. retaliatory attack on the U.S.S.R. (Manufacturing value added is the incremental value imparted to raw materials in any industrial process.)

so on) are introduced, the damage estimates become much higher. The delayed effects also ensure that even if the blast-damage levels cited in assured-destruction definitions were not reached, an all-out nuclear war would still result in the devastation of the combatant countries.

Prompt and delayed nuclear-weapons effects can be contrasted by considering an attack on a typical urban target, for example the greater Boston metropolitan area. The detonation of 10 one-megaton warheads, aimed at local economic and military targets, would generate an overpressure of 5 p.s.i. over more than 500 square miles. More than 1.3 million people would be killed by the prompt blast and thermal effects of the explosions, and more than 80 percent of the area's industrial capacity would be destroyed. It is likely that the secondary effects of the explosions, particularly fires and fallout, would increase these totals.

If conditions were favorable to the attack, the most devastating effect might be incendiary. Under certain weather conditions each one-megaton burst could ignite fires as much as 10 miles away. In such an attack a fire threat would presumably exist throughout much of eastern Massachusetts. Flash-induced fires would be joined by blast-triggered fires from toppled furnaces, stoves and boilers. Scattered debris and ruptured tanks and pipelines would add fuel to the fires. Firebreaks would be bridged by materials hurled by the blast. After the attack the suppression of possibly hundreds of small fires per acre would be a monumental task; water mains would be shattered and firefighting equipment and crews would be destroyed or disabled. In Hiroshima some 70 percent of the city's firefighting equipment was crushed in the collapse of firehouses and 80 percent of the firemen did not report to their posts.

Depending on weather conditions and the characteristics of the target area (particularly the density of flammable structures), the many individual fires might consolidate into one of two types of mass fire: a firestorm or a conflagration. A firestorm is driven by a strong vertical updraft of heated air, which is replaced by cool air sucked in from the periphery of the fire. A conflagration is driven in addition by a strong ground wind that was present before the attack. Whereas a firestorm continues only as long as its centripetal winds do, a conflagration can continue as long as fuel is available.

The consequence of a mass fire is total devastation within the affected area. The temperatures in a mass fire can exceed 1,000 degrees Celsius, a temperature higher than that necessary to melt glass and metal and to burn ordinarily fireproof materials. In Hiroshima an atomic bomb with a yield on the order of 15 kilotons caused a firestorm that lasted for six hours, totally destroying 4.4 square miles of the city. American cities are constructed of materials that are more fire-resistant than those in Hiroshima; on the other hand, American cities are more built up and more fuels, notably gasoline and heating oil, are available to feed fires. Most important, the yields of many modern strategic nuclear weapons exceed those exploded at Hiroshima and Nagasaki by two or more orders of magnitude. In addition much of the area under attack would be exposed to thermal radiation from more than one fireball.

Blast shelters would provide little protection against large fires. The survival of the occupants of such a shelter would depend critically on the temperature and humidity inside the shelter, and if mass fires were to start, the problem of maintaining a shelter environment in which people could survive would be aggravated beyond solution. Moreover, unless there was an independent supply of oxygen for each shelter, carbon monoxide and other toxic gases generated by

HIROSHIMA is seen from directly above in these U.S. Air Force reconnaissance photographs made before (*left*) and after (*right*) the atomic bombing of that city on August 6, 1945. The cross marks ground zero, the point on the ground directly under the explosion. The

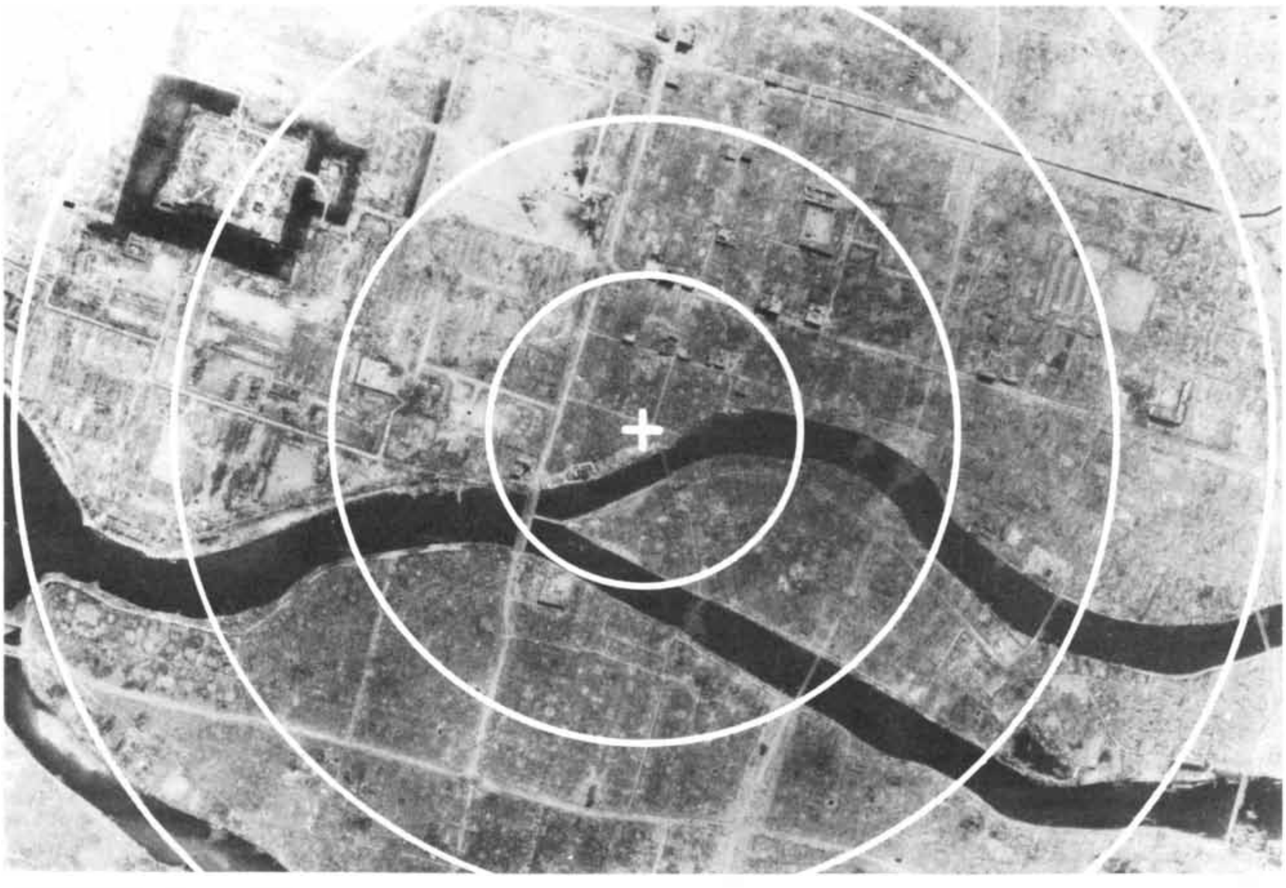
the fire could be deadly to the occupants. The heating of shelters, both by flames and by heated rubble (which could remain intolerably hot for days after the end of a fire), would jeopardize the occupants of shelters with an isolated atmosphere. In Dresden, where a firestorm ignited by chemical bombs killed more than 100,000 people in 1945, only those inhabitants who had left their shelters before the firestorm began were able to survive the twin threats of noxious gases and shelter heating.

After a nuclear attack many people would be disabled, trapped in wrecked buildings or prevented from fleeing the city because the streets were blocked by debris or fire. If mass fires were to form, which seems to be the probable result of multiple megaton bursts, the survivors among those who had escaped prompt incapacitation might be few. If mass fires were to begin in the Boston area, for example, the number of fatalities could be increased by 500,000.

Another factor not included in many assured-destruction calculations is radioactive fallout. Fallout results from the condensation of the radioactive byproducts of a nuclear explosion on materials fused by the intense heat of the fireball and (to a much smaller extent) from the conversion of nonradioactive materials into radioactive ones by the absorption of neutrons from the nuclear reactions of the explosion. If a nuclear weapon were to be exploded at or near the earth's surface, fallout would be an acute threat. Large amounts of debris would be scooped up into the rising cloud, later to fall out (or more likely be washed out) of the cloud in lethal amounts for hundreds of miles downwind. A dose of ionizing radiation measuring between 400 and 500 rems (an index of the biological effects of different types of radiation on man) delivered over a period of several days would kill half of the people who had been exposed. A dose of between 200 and 300 rems would kill somewhat fewer than 20 percent (assuming prompt medical treatment), but severe radiation-related blood symptoms, including diminished immunological response, could add extra fatalities by contributing to lethal infections. If 10 one-megaton weapons were exploded at ground level (to maximize fallout rather than blast and thermal effects), as many as a million New England residents who were not exposed to the immediate blast and thermal effects of the nuclear explosions would be subjected to dangerous levels of radiation. Even with optimistic assumptions about the availability of shelter and provisions, the fallout fatalities that would be added to the Boston-area toll could be as many as 500,000. An attack of this type might well mix airbursts and ground bursts to create maximum levels of both kinds of damage.

The number of fatalities from fire and radiation would grow steadily after such an attack, in part because medical facilities and personnel would be destroyed. Burn victims would present an exceptional medical problem, since serious burn cases require intensive and immediate treatment if they are to survive. The ability of any medical system to handle large numbers of such casualties is limited even in peacetime. The influx of some 50 survivors of the collision of two jet airliners on Tenerife in the Canary Islands a few years ago put a strain on burn centers in the U.S., which have a maximum capacity of about 130 patients. After a nuclear attack, of course, the number of burn cases would be orders of magnitude greater, and access to medical treatment would be far more difficult.

Existing medical services would be further burdened by the incidence of injuries well beyond areas of widespread mortality. The danger of injury from projected missiles (mainly shards of glass from shattered windows) would exist more than eight miles out from the

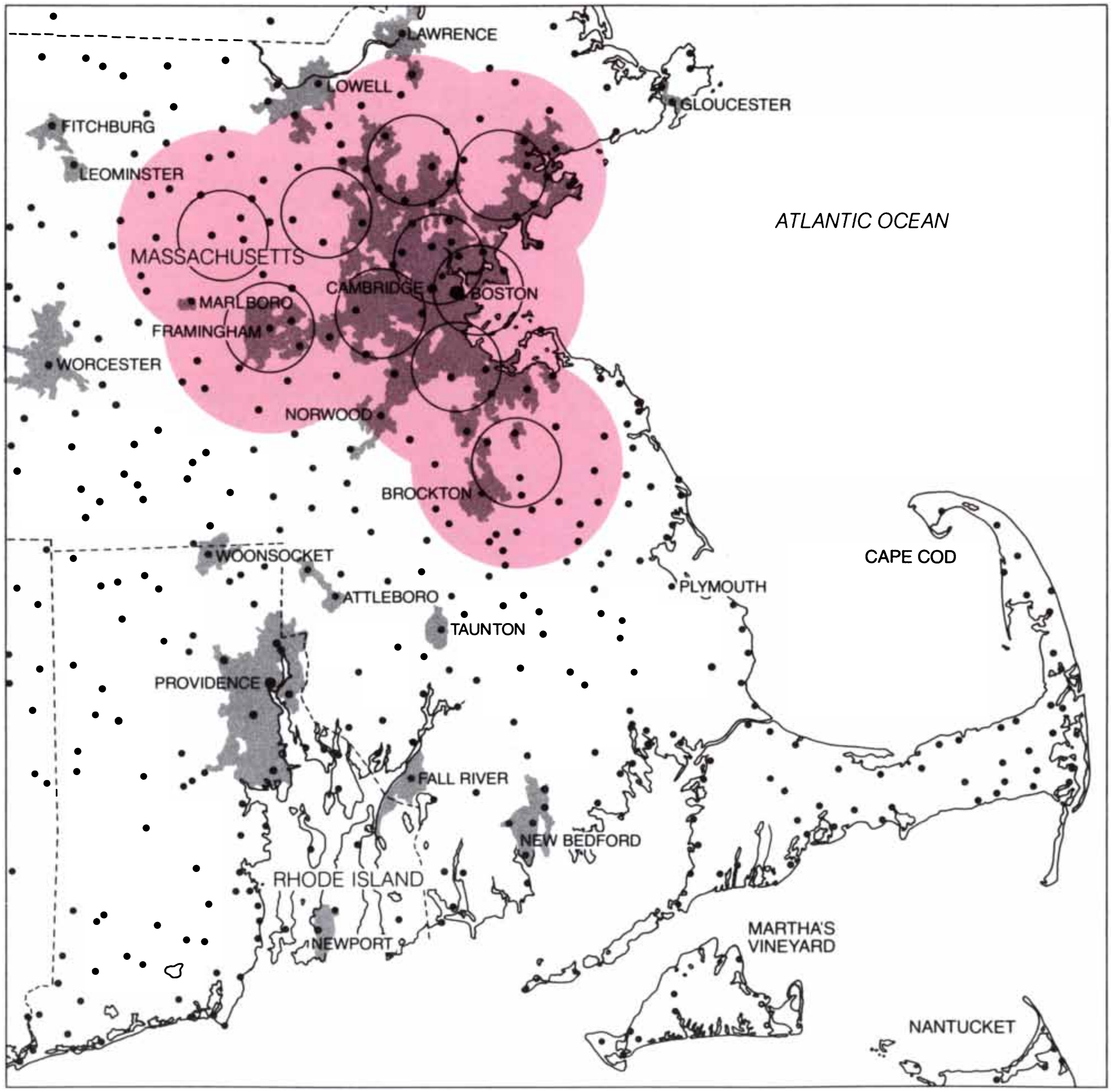


concentric circles are at 1,000-foot intervals. The firestorm following the prompt effects of the explosion lasted for about six hours and totally destroyed 4.4 square miles of the city. The explosive yield of the weapon that caused this devastation was on the order of 15 kilotons.

center of a one-megaton blast, and severe burns could be common out to nine miles, depending on weather conditions. Many victims of burns, radiation sickness and other mortal injuries who did not die immediately would require intensive (but under the circumstances unavailable) medical care. The management of less severely injured people and the very young, the very old and those with special medical needs would be complicated by the scarcity of food, shelter and medicine.

The survivors of an all-out nuclear attack would include many who would be permanently incapacitated by crippling injuries, blindness and other causes. Any medical effort would be further degraded by the destruction of public-health facilities and personnel, the proliferation of disease-causing organisms (which tend to survive high radiation levels) and other difficulties, such as the seemingly insoluble problem of disposing of the dead. The total regional casualties following an attack on the Boston area with both airburst and groundburst nuclear weapons could well exceed two million dead, with roughly the same number wounded or sick.

The assured-destruction concept also ignores certain strategic issues. It is typically argued that the U.S.S.R. is pursuing two types of program that would enable it to blunt the effectiveness of a U.S. retaliatory attack. The first program seeks to reduce the number of U.S. warheads arriving at their targets by de-

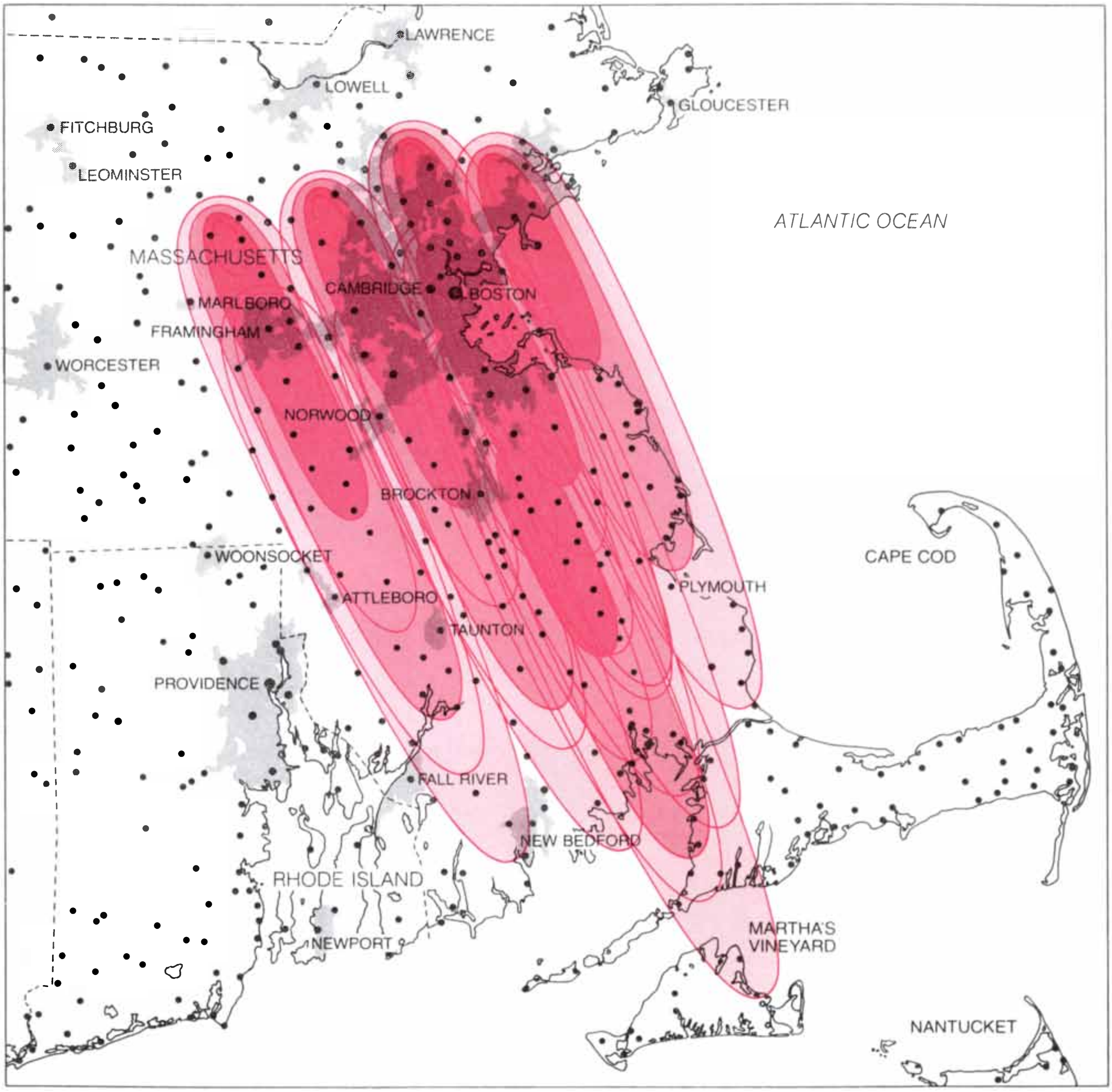


HYPOTHETICAL ATTACK on the greater Boston metropolitan area, which is outlined on these two pages, serves to contrast the prompt and delayed effects of multiple nuclear explosions. In both cases shown the attack consists of the detonation of 10 one-megaton nuclear warheads, which are aimed at local economic and military targets. In the illustration at the left it has been assumed that all the weapons have been detonated at an altitude that has been selected to maximize blast and thermal effects. Black circular outlines in the illustration correspond to regions exposed to an overpressure of at least 5 p.s.i.; each of these areas is 4.3 miles in radius. The colored areas represent the regions exposed to severe fire and burn risk on a clear day; each area in this case has a radius of 10 miles. The principal delayed effect of the attack suggested by the illustration is the risk of a regionwide firestorm or conflagration, which could add 500,000 fatalities to the assured-destruction estimate of 1.3 million killed by the prompt blast and thermal effects of the explosions. In the illustration

stroying U.S. strategic forces in a surprise attack and by intercepting as many surviving warheads as possible before they reach their targets in the U.S.S.R. The second program seeks to minimize the damage done by arriving U.S. weapons by evacuating urban residents and by dispersing and hardening industrial sites. Because the Russians could thereby "deny" important urban and economic targets to U.S. attack it is maintained that U.S. forces would fail to satisfy the assured-destruction damage levels, and that the U.S.S.R. would retain the industrial base, personnel and administration necessary for a rapid postwar recovery.

In spite of the alleged success of the U.S.S.R. in these endeavors, neither strategy could effectively reduce the devastation of an all-out nuclear war. Furthermore, neither effort would appreciably enhance the potential of the U.S.S.R. for recovery. On the contrary, such schemes only appear to reduce U.S. retaliatory capabilities in the perspective of the narrow and arbitrary definition of assured destruction discussed above.

Such analyses ignore the fact that even under the worst circumstances the U.S. would be able to mount a more than adequate retaliatory attack. Any Russian plan to degrade the U.S. assured-destruction capability would face the formidable task of reducing U.S. forces substantially below the level of 400 deliverable equivalent megatons. (Actually the task might be even more



at the right it has been assumed that all 10 of the warheads have been detonated at ground level in order to maximize the effects of radioactive fallout. (Typical January wind patterns have been assumed in drawing the contours.) The dark-colored areas are those that are covered by an amount of radiation that would be fatal to at least 80 percent of the exposed population. The medium-colored areas are those in which at least 50 percent of the exposed population would die of radiation sickness. The light-colored areas are the probable extent of the region in which clinical radiation symptoms would be evident in much of the exposed population, resulting in perhaps 20 percent fatalities. (Presumably the survivors would also be subjected to the effects of additional long-term radioactive fallout from attacks on neighboring regions.) The total number of casualties in the Boston region following an attack that made use of a suitable combination of airburst and ground-burst nuclear weapons could well exceed two million dead, with approximately the same number wounded or sick.

difficult, because a well-planned American attack of even 200 equivalent megatons could still promptly kill a fifth of the U.S.S.R.'s population and destroy more than two-thirds of its industry, thereby satisfying the requirements of assured destruction.)

It is extremely unlikely that a preemptive Russian first strike could achieve this goal. For one thing, 400 equivalent megatons is only a fraction of the current U.S. nuclear arsenal. More than half of the current U.S. arsenal of more than 6,000 equivalent megatons is carried by missile-launching submarines on station, by bombers on alert at Strategic Air Command bases and by silo-based missiles, all of which are capable of going into action within a few minutes of a Presidential order. The rest of the U.S. strategic forces consist mainly of bombers not on alert and submarines in port for maintenance. If a Russian surprise attack were to destroy many land-based U.S. missiles in their silos and all the nonalert bombers and submarines, more than 2,000 equivalent megatons would remain available for retaliatory action. Even if an unexpectedly large number of U.S. weapons were to malfunction or to be destroyed in flight, more than 1,500 equivalent megatons could still be delivered with high confidence. These figures assume "worst case" conditions from a U.S. perspective: if some warning were available prior to such a Russian attack, extra bombers and submarines could be alerted and the number of deliverable equivalent megatons would more than double.

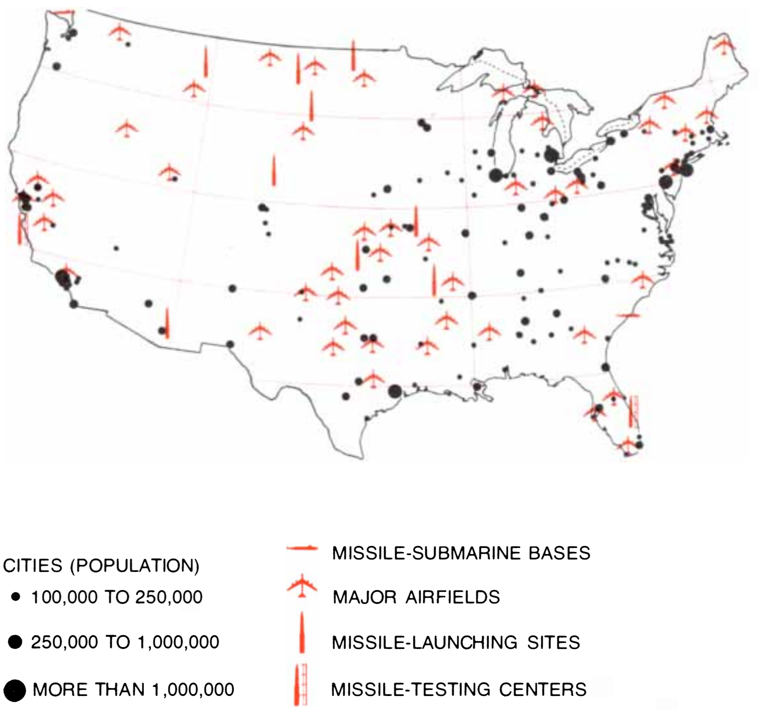
Because of the availability of what are sometimes described as "overkill" forces any effort to reduce the numbers or effectiveness of the arriving U.S. warheads is bound to fail. For example, Moscow is protected by an anti-ballistic-missile (ABM) system that is limited by treaty to 100 missile launchers. (Currently only 64 missiles are deployed in that system.) In the event of a missile attack those missiles could destroy a certain fraction of the incoming missiles. U.S. planners could easily compensate for this potential attrition by several strategies, one of which would be to allocate extra warheads to the "Moscow package" based on generous theoretical assumptions about the effectiveness of the Moscow ABM system.

The assured-destruction criterion implicitly assumes what may be the least probable scenario for a general nuclear war. It is extremely unlikely that an all-out war would begin with a massive surprise attack by either side on the other side's cities. Such a war would more probably follow an escalating crisis that might begin with limited nuclear strikes on military targets. These alternative scenarios imply that city populations would have ample warning of a possible or probable nuclear attack, with the result that evacuation and other tactics for reducing damage could be pursued. If this were to be the situation, the character of the city population would clearly have changed by the time of an attack. Since assured destruction is calculated according to peacetime population densities and the calculations rely on certain assumptions about the disposition of city dwellers and workers on a day-to-day basis, the programmed fatality levels may not be reached under realistic circumstances.

The assured-destruction criterion also assumes that a general nuclear war would consist of a single massive "countervalue" strike: a strike against both military and economic targets. Countervalue strikes might well, however, remain at a relatively low level of intensity for some time. A limited countervalue exchange might consist of attacks on industrial locations away from large cities in order to discourage escalation to attacks with the largest possible number of fatalities. (The Russians in particular have built some key installations in remote areas, where they could be attacked with a relatively low level of fatalities.) Strikes against cities might be preceded by a warning or an ultimatum, which would clearly encourage evacuation. A general war might even begin with a slow campaign of "city-trading." In other words, virtually any change in assumptions radically alters the context of the assured-destruction scenario and casts doubt on the accuracy of fatality estimates.

In fact, except in the circumstances of certain specialized scenarios, there will always be an opportunity for a country to evacuate its urban centers to some extent, regardless of the degree to which the country has prepared for such evacuation. (For example, on September 1, 2 and 3, 1939, the British government evacuated some 1.5 million women and children from Britain's major cities, and in the same three days an additional two million people moved out on their own initiative.) The wide availability of private-automobile transport in the U.S. probably more than compensates for any current Russian evacuation plans and training.

Nevertheless, allegations of evacuation planning in the U.S.S.R. have inspired much concern in the U.S. According to a recent report of the Central Intelligence Agency, if the Russians were to have at least one week to thoroughly evacuate their cities and shelter refugees against radioactive fallout, war-related casualties could be reduced to the "low tens of millions, about half of which would be fatalities." Some analysts have gone so far as to term these fatality levels "acceptable" in view of the fact that the U.S.S.R. suffered 20 million dead in World War II. Even if evacuation could reduce the number of prompt fatalities, however, the degree of damage the U.S. could inflict on the unprotectable economic resources of the U.S.S.R. would be so great that the U.S.S.R. would be eliminated as a major industrial power.



RELATIVE CONCENTRATIONS of potential population targets and military targets in the U.S. and the U.S.S.R. are suggested by these two maps. The black dots indicate the location of the largest cities in each country. The colored symbols designate strategic-weapons instal-

One purpose of any campaign of strategic bombardment is to reduce an enemy's potential for supporting armed forces in the field. In World War II factories, transportation systems and power plants were attacked. One goal of such bombing campaigns was to destroy industries on which other economic sectors relied, depriving those sectors of essential inputs and leading to an expanding industrial incapacitation. The bombing of Germany failed to have this effect, in part because of limitations on the size of chemical-explosive payloads. Attacks on a given target system had to be spread out over many raids, and so the surviving facilities could be "jury-rigged" to compensate for the damage done to certain parts of the industrial network. Civilians left homeless in attacks could be housed in nearby towns that had not been damaged. Even after the atomic bombing of Hiroshima and Nagasaki enough aid was available in surrounding communities to significantly aid the survivors.

The deployment of large numbers of nuclear weapons has radically changed the context of strategic bombing. The forces currently deployed by the U.S. and the U.S.S.R. are able to destroy the entire industrial structure of any nation. Moreover, this damage can be done all at once, so that little assistance would be available for those targets that had come under attack.

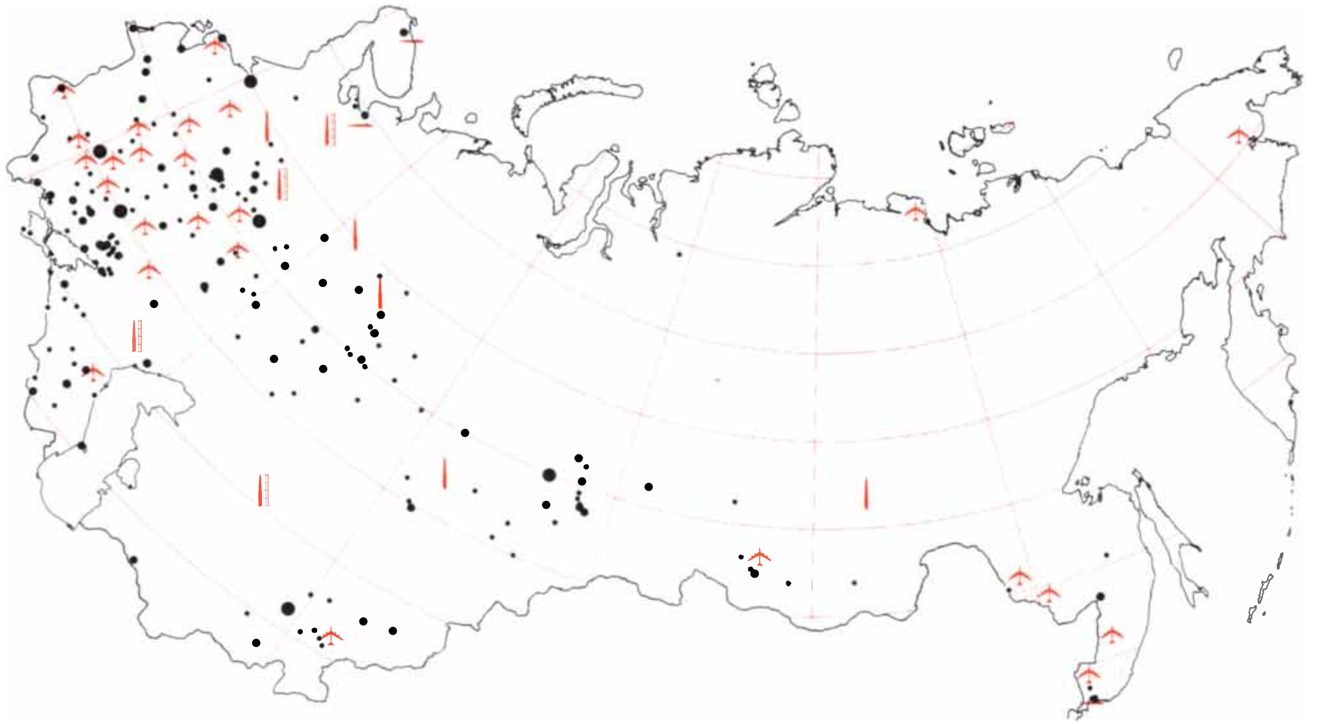
In both the U.S. and the U.S.S.R. a limited number of facilities comprise the bulk of the productive capacity in many major industries. The centrally planned economy of the U.S.S.R. in particular has many vulnerable bottlenecks and choke points. Hence the destruction of a single target or very few targets could disrupt production in many other industries. Because of this concentration 100 equivalent megatons, corresponding to the payload of the missiles carried by five or six Poseidon submarines, would be sufficient to destroy crucial industries without which the Russian economy could not sustain itself.

For example, a study conducted recently by the Office of Technology Assessment of Congress showed that a U.S. attack on petroleum refineries in the U.S.S.R. could, with only some 40 low-yield nuclear warheads, destroy about three-fourths of the U.S.S.R.'s entire refining capacity. Comparatively few warheads could also destroy the transportation, energy, maintenance and management resources needed for any postwar economic recovery. The Russian energy system is particularly vulnerable to attack and is crucial for recovery. For instance, nearly all the intercity freight in the U.S.S.R. is shipped over electrified rail networks, whereas much of it in the U.S. goes by truck.

These kinds of figures should not be taken as evidence that the U.S. economy is somehow less vulnerable than that of the U.S.S.R. The Russians have more than enough warheads to cover similar U.S. targets. Rather, it is instructive to remember, as the CIA report noted, that "the coordination of requirements with available supplies and transportation is a complex problem for Soviet planners even in peacetime, let alone following a large-scale nuclear attack on the U.S.S.R."

Even if the Russian evacuation plans were successful, they would only defer, not prevent, the impact of the war on civilians. A nation's fixed medical, technical and educational base would, after all, be destroyed in a nuclear war. Recovery stockpiles and facilities could also be targeted. If some food, pharmaceuticals, clothing, equipment and spare parts did survive, there would be neither the administrative structure to allocate the goods nor the transport to ship them where they were needed. The destruction of refineries and electric-power stations could interdict resupply, and shortages could develop quickly. Perishable goods, including many foods and drugs, would be lost if electric power were cut off. The devastation of housing would make summer life difficult and winter existence intolerable. This would be particularly true in the U.S.S.R., where outside cities there are few alternative forms of shelter such as hotels. In short, civil defense might protect some people, but it could not prevent the widespread destruction of property essential to the support of life. The economic interdependence of an industrialized nation is a vulnerability that cannot be defended.

A nation's administrative and social structure would also be disrupted by nu-



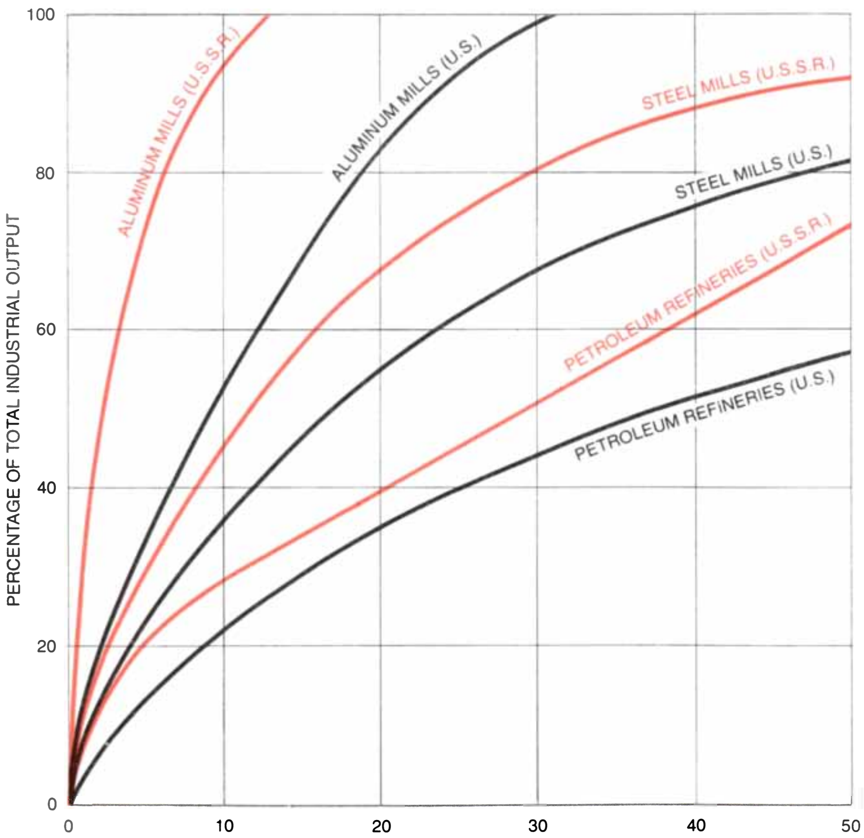
lations: major airfields, missile-submarine bases, land-based-missile launching sites and missile-testing sites (*see key at bottom left*). In addition to the installations shown here the U.S. has a variety of strategic forces stationed elsewhere in the world (mainly on the island of Guam and in Alaska). In general suitable targets for nuclear attack are more concentrated in the U.S.S.R. than they are in the U.S.

clear attack to the point that a political system might be shattered beyond reconstitution. Although special bunkers are being constructed to protect the bureaucratic and internal-security apparatus of the Soviet government in the event of war, the U.S. does not lack the means to attack those shelters.

The delayed effects of a nuclear war between the U.S. and the U.S.S.R. would propagate far beyond the borders of the antagonists and their allies. Worldwide effects would result mainly from the fact that the stem and cloud of most nuclear explosions would penetrate the stratosphere and deposit several kinds of radioactive material in it. Unlike the lower part of the atmosphere, the stratosphere lacks the moisture and shear motions needed to quickly sort out particulate and gaseous matter. Since such materials would remain in the stratosphere for a long time, their effects would be diluted. One consequence of this long residence time, however, would be wide dispersion. Thus although stratospheric effects would be less intense than lower-atmospheric ones, they would last longer and be more widespread.

A report issued by the National Academy of Sciences in 1975 listed three effects of nuclear war that might have adverse worldwide impacts. First, stratospheric ozone might be depleted, because nitrogen oxides made from atmospheric nitrogen and oxygen by the heat of nuclear explosions would be injected into the stratosphere, where they would aid in the conversion of ozone into molecular oxygen. Second, the deposition of large amounts of dust in the upper atmosphere could alter the amount of solar radiation arriving at the earth's surface. Third, hazardous radioactive isotopes could be dispersed through the stratosphere, falling out slowly on a worldwide scale.

Stratospheric ozone plays an important role in life on the earth by screening out harmful ultraviolet radiation. The NAS report estimated that a 10,000-megaton nuclear war could destroy half of the ozone in the Northern Hemisphere and about 30 percent of the ozone in the Southern Hemisphere. As opponents of the supersonic transport aircraft and fluorocarbon spray-can propellants have contended, the depletion of the ozone layer could lead to a variety of medical and environmental problems. Higher cancer rates and harmful effects on plants, including crop plants, could result. The destruction of stratospheric ozone on this scale could upset the thermal structure of the upper atmosphere and lead to worldwide temperature changes. After such a war ozone levels might not return to normal levels for many years.



KEY INDUSTRIAL TARGETS are also more concentrated in the U.S.S.R. than they are in the U.S., as these three pairs of curves demonstrate. As a consequence fewer nuclear warheads would be needed to cripple the U.S.S.R.'s production of such vital materials as steel, petroleum and nonferrous metals. In addition the economies of both countries are characterized by crucial bottlenecks. For example, only one plant at Pavlodar in the U.S.S.R. does work essential to 65 percent of the aluminum industry of the country. By the same token close to 80 percent of the iron ore shipped in the U.S. travels through one set of locks at Sault Sainte Marie.

A single one-megaton surface burst would also project thousands of tons of fine dust into the stratosphere. The dust could absorb, reflect and scatter radiation arriving from the sun or reflected from the earth, and there have been suggestions that this effect could lead to a change in the weather at the earth's surface. According to the NAS study, however, a 10,000-megaton war would inject no more dust into the stratosphere than was thrown up by the explosion of the volcano Krakatoa in 1883. By extrapolating from such volcanic events the NAS report concluded that only a slight change in surface weather conditions might result.

Radioactive isotopes would be distributed worldwide by stratospheric transport processes. Since these isotopes would have a relatively long residence time in the stratosphere, many of the dangerous short-lived ones would decay before they could reach the ground. Nevertheless, some hazardous isotopes, such as strontium 90, cesium 137, iodine 131 and carbon 14, would persist and might enter the food chains of the biosphere. The NAS report did not suggest that this fallout would have the kind of worldwide lethal consequences for human life that are depicted in novels such as Nevil Shute's *On the Beach*. Regional concentrations of fallout in the combatant nations (and neighboring nations) could nonetheless present an acute radiation hazard to many evacuees and rural residents who might not have been directly imperiled during an attack on cities. Less intense "hot spots" could appear at greater distances, with adverse biological consequences. Few parts of the attacked country would escape the threat of fallout, since a thorough attack would cover economic and military targets nationwide, leaving most areas contaminated.

Atmospheric phenomena are complex, and it is not clear how a 10,000-megaton nuclear war might influence climate. Although the NAS study estimated that the effects of ozone depletion and dust loading probably would not have an irreversible impact on global weather patterns, the report did indicate that changes of a much more serious nature could not be excluded. The possibility of synergistic actions among these various effects cannot be ignored. For example, it has been noted that a global cooling of only one degree C. could eliminate all wheat growing in Canada.

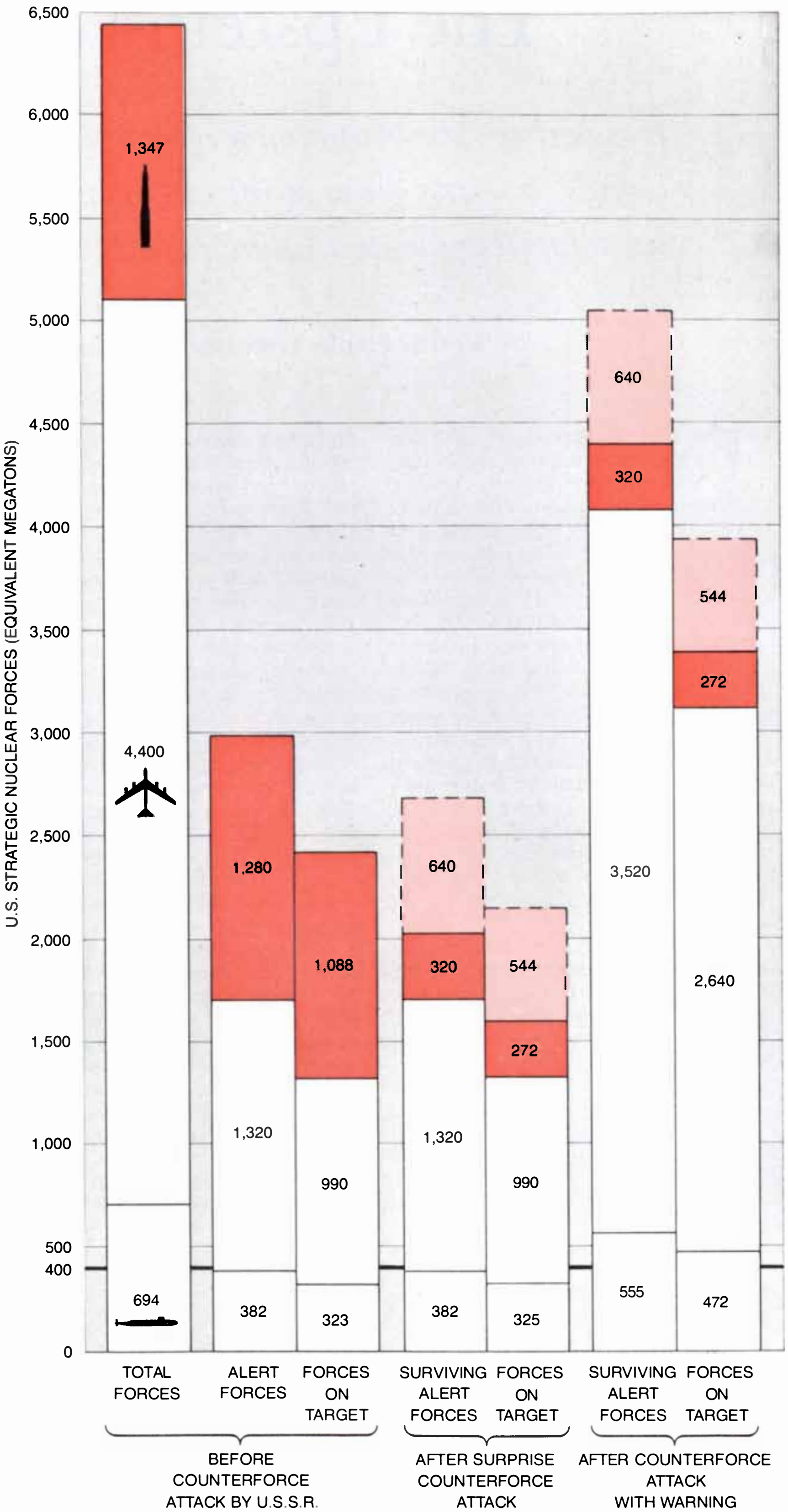
Direr possibilities include the expansion or melting of polar ice.

The NAS report did not examine possible changes in continental weather resulting from effects such as fires. A single 10-megaton airburst could ignite a forest fire covering thousands of square miles. The burning of the broad grasslands and forests of the U.S. and the U.S.S.R. could defoliate the natural ground cover, thereby changing the reflectivity of the earth's surface and giving rise to weather changes. Particulate combustion products thrown into the atmosphere by forest fires would absorb and reflect solar radiation, and they would also act as nucleation centers for the formation of water droplets and ice crystals, thereby increasing the cloud cover and altering the distribution of precipitation. Such local effects could exacerbate the worldwide phenomena cited above.

Finally, just as the various components of a national economy are interlocked, so nations themselves are interdependent. The destruction of the economies of the major powers by a nuclear war would be a massive blow to the economies of nations dependent on those powers for the exchange of commodities and technology. The less developed countries in particular would suffer, since at this stage of their development they need to import technology from more developed countries.

In sum, the cumulative effects of an all-out nuclear war would be so catastrophic that they render any notion of "victory" meaningless. The formal methodologies of the assured-destruction scenarios do not reveal the full extent of these effects. Moreover, arguments that throw doubt on the sufficiency of the deterrent capability of the U.S. exclude some of the most profound and long-lasting of these effects. When the delayed effects of all-out war are taken into consideration, it should become clear that no countermeasure would significantly lessen the degree of devastation that would surely occur. Even if a highly efficient program for the evacuation of cities could substantially reduce prompt fatalities, it could not prevent the delayed social consequences of industrial and economic devastation. The magnitude of either the prompt disaster or the delayed one would be so great that neither disaster could ever be considered tolerable.

There are many steps that could be taken by both sides to diminish the likelihood of an all-out nuclear war. Many of them are now the subject of strong disagreement. One step in the right direction would be to reframe the currently misleading concept of assured destruction in more realistic terms to reflect the full extent of the catastrophe that would be represented by a nuclear war.



U.S. STRATEGIC ARSENAL would retain far more deliverable nuclear weapons than would be necessary to accomplish the assured-destruction mission, even after an all-out surprise attack by the U.S.S.R. on strategic military targets in the U.S. If there were any warning available before such an attack, the number of U.S. nuclear weapons that could be delivered on targets in the U.S.S.R. would increase considerably. The heavy black line across the bottom of the chart indicates the 400 equivalent megatons thought to be sufficient to kill 35 percent of the people in the U.S.S.R. Strategic forces above this level are referred to as overkill capability.

The Epstein-Barr Virus

It is perhaps the commonest of all the viruses that infect human beings. It is the cause of infectious mononucleosis, and there is substantial evidence linking it with two types of human cancer

by Werner Henle, Gertrude Henle and Evelyne T. Lennette

The ability of certain viruses to induce tumors in animals under natural or experimental conditions has been known for more than 70 years. As early as 1908 V. Ellerman and O. Bang in Denmark demonstrated that leukemia could be transmitted by a virus in fowl, and in 1911 Peyton Rous of the Rockefeller Institute for Medical Research showed that a sarcoma could be similarly transmitted in chickens. Since that time many examples of tumor-causing viruses have been found in the mouse, the frog, the rabbit, the guinea pig, the cow, the cat, the dog and even some nonhuman primates. Such findings led to an intensive search for similar agents in man, based on the reasonable assumption that human beings would not be exempt from natural scourges that plague their fellow animals.

To date the prime candidate for a virus involved in cancers of man is the Epstein-Barr virus. The fascinating story of this virus began in the late 1950's, when Denis Burkitt, a British surgeon working at Makerere University in Uganda, saw his first case of an unusual tumor that would later bear his name. The patient was a young boy suffering from a massive tumor of the jaw that proved to be a lymphoma: a malignant growth of the lymphoid tissues. The tumor was remarkably fast-growing, doubling in size every 24 to 48 hours. It caused grotesque disfigurement and finally death by metastasis to other parts of the body.

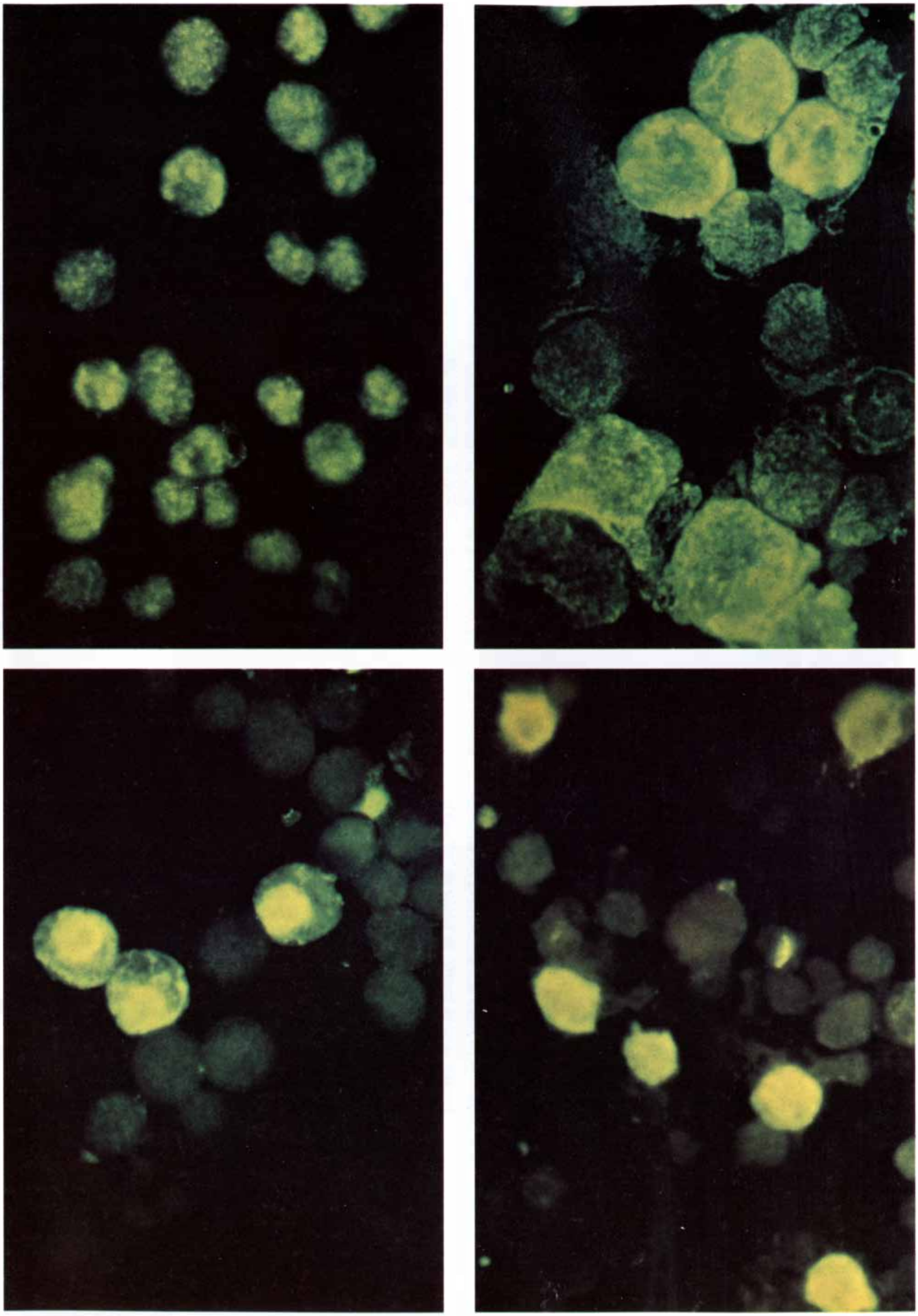
Burkitt soon had ample opportunity to observe the disease further, because a surprisingly large number of young patients were brought to him. The lymphoma of the jaw occurred most frequently in children six to eight years old; it was rare in children older than 14 and unknown in those less than a year old. It was not uncommon to find clusters of patients of both sexes in neighboring families within a short period of time. This distribution of the tumor was reminiscent of viral infections of childhood, such as poliomyelitis before the introduction of vaccines against it. Burkitt proceeded single-handedly to gather epidemiological evidence about the lymphoma, not only in Uganda but also throughout Africa. On his now historic Land Rover safaris, ranging over thousands of miles, he collected data on the patients, their residence, their family and their other diseases. From this evidence he was able to piece together an impressively complete picture of this previously undescribed cancer.

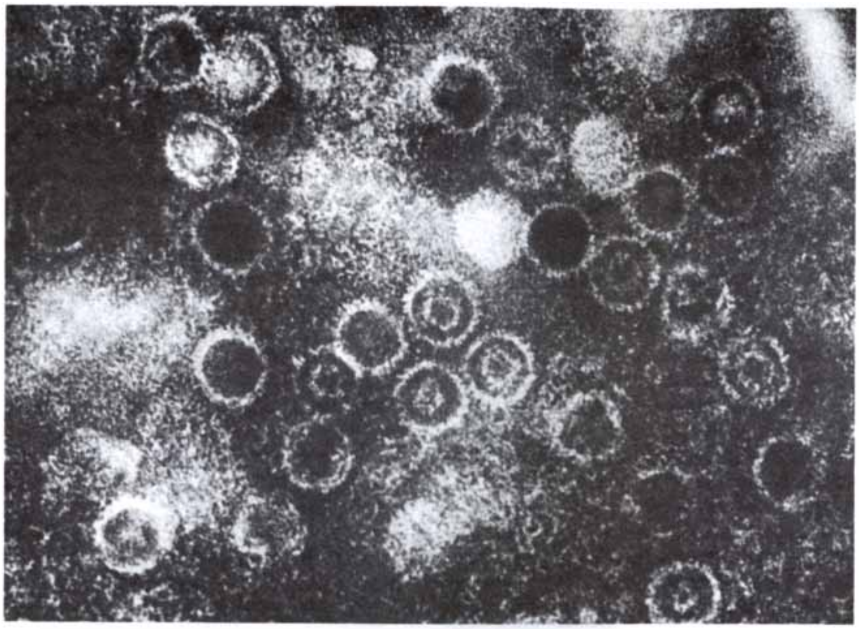
Burkitt's lymphoma was and still is one of the commonest childhood cancers in Africa, with a rate of one case per 10,000 children per year. It is seen only in areas that have an elevation of less than 1,800 meters, an annual rainfall of more than 60 centimeters and an average temperature of more than 16 degrees Celsius. These areas are coextensive with those where mosquito-transmitted infections such as yellow fever and malaria are endemic. Burkitt also found evidence that the lymphoma could spread from one area to adjacent areas and that it was influenced by seasonal variations. These findings prompted him to suggest in 1958 that the cause of the tumor was a mosquito-transmitted infectious agent, or at least that such an agent was a precipitating factor in the disease.

Later, thanks again to Burkitt and the physicians working with him, it became possible to treat the lethal lymphoma. Even in early trials the tumor responded dramatically to injections of cyclophosphamide, melting away within a few days after the initiation of treatment. Even more effective drugs with fewer side effects have been developed since then. In many cases, however, the tumor recurs eventually and may ultimately prove resistant to chemotherapy, resulting in death. Depending on the initial extent and site of the tumor, between 30 and 60 percent of the patients survive for five years or more without evidence of disease.

Burkitt's lymphoma became the subject of research in many laboratories. In 1964 M. Anthony Epstein, B. G. Achong and Y. M. Barr at Middlesex Hospital in London examined cultured cells from biopsy specimens of Burkitt's lymphoma in the electron microscope. Inside the cells they observed particles of the virus that was later to bear the name Epstein-Barr virus. The virus was classified on the basis of its structure as belonging to the herpes group of viruses, which have double-strand DNA as their store of genetic information. At the time three human viruses of this type were known: herpes simplex virus, which causes fever blisters; cytomegalovirus, which induces serious congenital malformations and a variety of syndromes in older patients, and varicella-zoster, which causes chickenpox and shingles. The discovery of a herpes virus in the

VIRAL ANTIGENS are foreign proteins that induce specific antibody responses in animals and man. Antigens of the Epstein-Barr virus can be detected within infected human lymphocytes (white blood cells) by sensitive fluorescent-antibody techniques, as is shown in the fluorescence micrographs on the opposite page. The discovery of virus-associated proteins in two human tumors has provided strong circumstantial evidence for a link between the Epstein-Barr virus and cancer in man. The majority of virus-infected lymphocytes do not yield progeny virus particles but do express a viral antigen in the cell nucleus called the Epstein-Barr nuclear antigen (EBNA). When fluorescent antibodies combine with this antigen, they cause the cell nucleus to glow brightly but the cytoplasm to remain dark (*top left*). Before the cycle of viral replication begins in a productively infected cell the early antigens appear. The first antigen, the diffuse component (*D*), is found in both the nucleus and the cytoplasm (*top right*). The second, the restricted component (*R*), is confined to a large mass in the cytoplasm (*bottom left*). Following the replication of the viral DNA the late antigens appear; these antigens are protein constituents of the virus itself. One of the late antigens is the viral-capsid antigen (VCA), which is present in both the nucleus and the cytoplasm (*bottom right*). Each of the three diseases associated with the Epstein-Barr virus gives rise to a distinct pattern of antibodies to viral antigens.

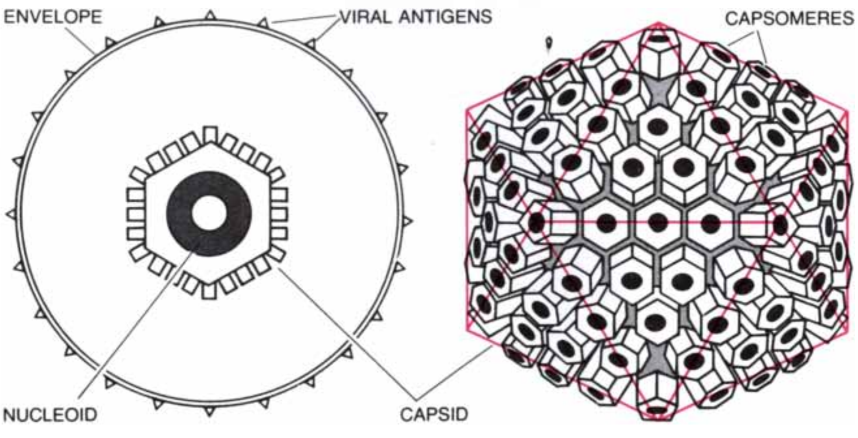




EPSTEIN-BARR VIRUS PARTICLES are enlarged some 96,000 diameters in this electron micrograph made by M. Anthony Epstein of the University of Bristol Medical School. The capsids, or protein shells, of the virus particles stand out against the darkly stained background. Some of them have been penetrated by the negative-contrast material, revealing that they are either empty or contain viral DNA. Unstained virus particles look white. The virus has been isolated from two human tumors: Burkitt's lymphoma and nasopharyngeal carcinoma.

Burkitt's lymphoma cells did not arouse great interest, however, because at the time no herpes virus was suspected of causing cancer in animals. Hence the virus particles in the tumor cells were assumed to be merely casual passengers.

As Epstein was attempting without success to enlist the help of British virologists in identifying the virus, we were trying to obtain specimens from Burkitt's lymphoma patients in Africa for our laboratory at the Children's Hospital of Philadelphia. A fortunate series of events put Epstein in touch with us, and within days a set of his cell cultures was crossing the Atlantic to our laboratory. Soon after the arrival of the cultures we were able to demonstrate that the Epstein-Barr virus was a new type of herpes virus. We came to this conclusion when we failed to transmit the virus to various standard cell cultures, chick embryos and experimental animals that were susceptible to infection by the three known human herpes viruses.



STRUCTURE of the Epstein-Barr virus is diagrammed. It belongs to the herpes group of viruses, which store their genetic information in the form of double-strand DNA. The infectious virus particle consists of three components: a doughnut-shaped central core, or nucleoid, which contains the viral DNA in condensed form; a capsid, which is icosahedral in shape and made up of hollow, tubular protein subunits called capsomeres, and a protective envelope derived from either the nuclear membrane or the outer membrane of the host cell. The envelope incorporates a number of viral proteins (antigens) that were manufactured and inserted into the host-cell membrane before the assembly of complete virus particles began. During infection the viral envelope fuses with the outer membrane of a susceptible host cell, and the nucleocapsid (the capsid and the viral DNA within) enters cell. Viral DNA then either persists in a repressed state or it diverts enzymatic machinery of host to manufacture of new virus particles.

Convinced that we were dealing with a previously unknown virus, we began to investigate the link between the Epstein-Barr virus and Burkitt's lymphoma. Our main approach was to look for antibodies to the virus in patients' serum. Once someone has been exposed to a virus he almost invariably manufactures specific antibodies to it that circulate in his bloodstream. The presence of circulating antibodies to a given virus in a patient is therefore a clear sign of past exposure to that virus.

To determine whether the Epstein-Barr virus played a role in the genesis of Burkitt's lymphoma we sought to answer the following questions: Do all patients with Burkitt's lymphoma have antibodies to the Epstein-Barr virus? Are the antibodies to the virus limited to patients with the disease or are they also found in healthy individuals or in patients with other diseases? If the antibodies are not limited to Burkitt's lymphoma, do patients with the lymphoma show a significantly higher titer (concentration in the blood serum) or a greater variety of antibodies to the Epstein-Barr virus than other people?

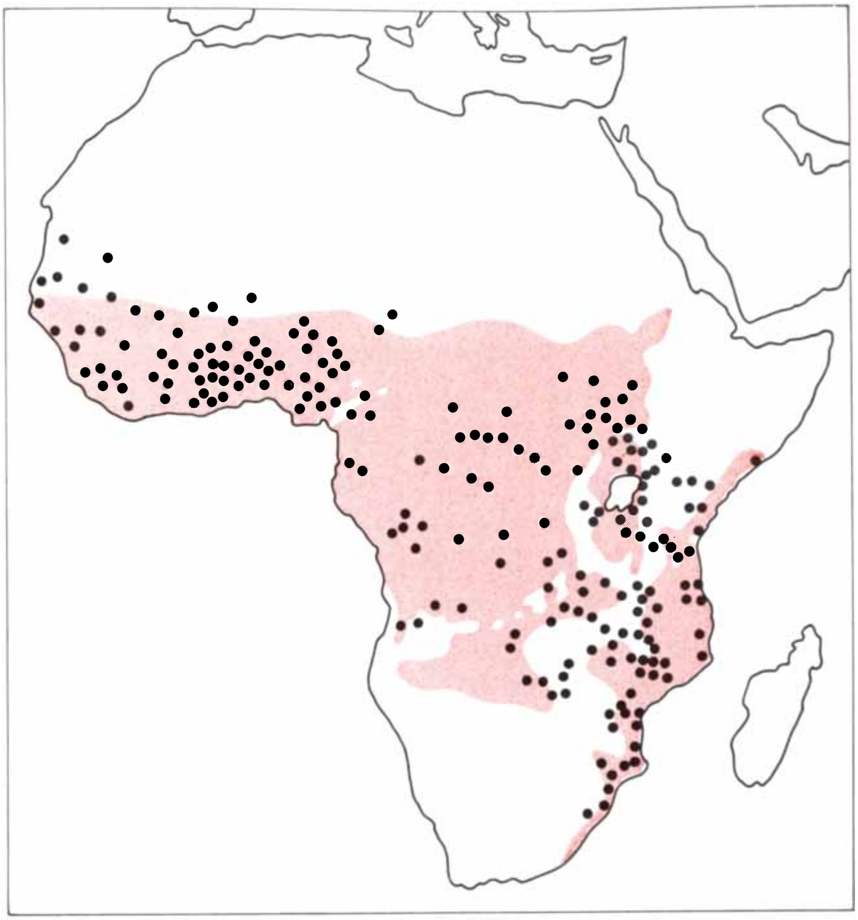
To detect the patients' antibodies to the Epstein-Barr virus we employed the sensitive technique known as indirect immunofluorescence. In this technique cultured cells containing the unknown virus are fixed on a glass slide. Serum from a patient suspected of harboring antibodies to the virus is then put on the slide, so that the patient's antibodies can bind to the viral antigens (foreign proteins) either inside the cell or on the cell surface. The cells are now washed, and a second set of antibodies, which react specifically with all human antibodies and are conjugated with the molecules of a fluorescent dye, are incubated with the cells to which the patient's antibodies have already bound. When the cells are illuminated with ultraviolet radiation in a microscope, the patient's antibodies are revealed as fluorescent areas within the virus-infected cells. The fluorescence is bright because several fluorescent antibodies can bind to one of the patient's antibodies.

With this simple but elegant assay we were able to confirm that the Epstein-Barr virus is indeed a novel virus, because antibodies to each of the known human herpes viruses did not react with the Burkitt's lymphoma cells. We then obtained serum from a Nigerian child with Burkitt's lymphoma who had been flown to Bethesda, Md., for treatment at the Clinical Center of the National Institutes of Health. In the indirect-immunofluorescence test the child's serum caused a small proportion of the Burkitt's lymphoma cells to fluoresce brightly. When individual fluorescent cells were examined in the electron microscope, they showed the presence of numerous particles of the Epstein-Barr virus; the nonfluorescent cells contained no virus particles.

To our surprise antibodies to the Epstein-Barr virus were found not only in children with Burkitt's lymphoma but also in the serum of nearly all healthy African children tested. Even more unexpected was the finding that antibodies to the virus were present in children from many parts of the world, which implied that the Epstein-Barr virus has a worldwide distribution and that almost no one escapes infection. It soon became evident that the age at which antibodies to the virus are acquired depends on the prevailing living conditions. Under the conditions of crowding and poor hygiene prevalent in the developing countries practically all children manufacture antibodies to the virus by the age of three or four. In the more advanced countries the acquisition of antibodies to the virus is often delayed until adolescence or adulthood.

Measurement of the antibody titer to the Epstein-Barr virus in children with Burkitt's lymphoma revealed that the patients on the average had titers from eight to 10 times higher than those of healthy children. The patients with the lymphoma were therefore reacting to the virus in an abnormal fashion. We found, however, that patients with some other diseases also had elevated titers of antibodies to the Epstein-Barr virus. Foremost among them were patients with nasopharyngeal carcinoma, who had antibody titers that were 10 times higher than those of healthy individuals or of patients with other tumors of the head and neck. The connection between this carcinoma and the Epstein-Barr virus was initially made by Lloyd J. Old and his colleagues at the Memorial Sloan-Kettering Cancer Center, who observed that serum from patients who have nasopharyngeal carcinoma forms precipitates with extracts of cultured Burkitt's lymphoma cells. This finding implied that the Epstein-Barr virus is intimately associated with both diseases.

Nasopharyngeal carcinoma is a malignant tumor that arises in the epithelial lining of the space behind the nose. It afflicts mostly adults and occurs at high frequency (10 per 100,000 people per year) in southern China and at somewhat lower frequencies in Tunisia and East Africa. American Eskimos living in Alaska also suffer from the disease, which is rare elsewhere in the world. The tumor usually takes one of two lethal courses: it may grow through the bones of the skull to invade the brain, or carcinoma cells may spread from the primary tumor through the lymph vessels, forming metastases in the lymph nodes of the neck and then in the liver, the lungs, other organs and the bones.



DISTRIBUTION OF BURKITT'S LYMPHOMA in Africa reveals that clusters of cases (*black dots*) are found in areas where malaria is endemic because of climatic factors (*light color*). Burkitt's lymphoma is a tumor of the jaw and other tissues in African children that is closely associated with the Epstein-Barr virus. This evidence suggests that malaria may be a secondary factor in genesis of the lymphoma. Data are from work of Denis Burkitt and others.

The ubiquitousness of antibodies to the Epstein-Barr virus in healthy populations led us to suspect that the virus was primarily the cause of a common, self-limited illness in addition to being apparently associated with the two malignancies. By a remarkable coincidence the identity of this disease emerged through an accidental discovery made in our laboratory. Late in 1967 one of our young technicians developed classical symptoms of infectious mononucleosis: sore throat, fever and enlarged lymph nodes in the neck. Before her illness she had shown no antibodies to the Epstein-Barr virus, but with the onset of the disease antibodies appeared in her serum. This fortuitous event gave the first clue that the Epstein-Barr virus might be the long-sought cause of infectious mononucleosis, which was first described by T. P. Sprunt and F. A. Evans of the Johns Hopkins University School of Medicine in 1920.

In its classical form infectious mononucleosis causes fever, tonsillitis, swollen lymph nodes and an enlarged liver and spleen. In spite of its name it does not seem to be highly infectious; no major epidemics have been reported. The majority of the patients are young adults, and uncomplicated cases run their course in from one week to four weeks with complete recovery. Fewer than 1 percent of the patients have complications (such as encephalitis, pneumonia and kidney failure), and fatal cases are very rare. Some clinicians, however, have described mononucleosis as a self-limited leukemia. Patients with the disease show a marked increase in the number of white blood cells, which are often morphologically abnormal. The proliferating white cells include the B lymphocytes (which originate in the bone marrow and manufacture antibodies) and the T lymphocytes (which mature in the thymus and participate in cellular immune reactions). In patients with mononucleosis the abnormal proliferation of lymphocytes soon subsides; in leukemia the cells continue to multiply out of control.

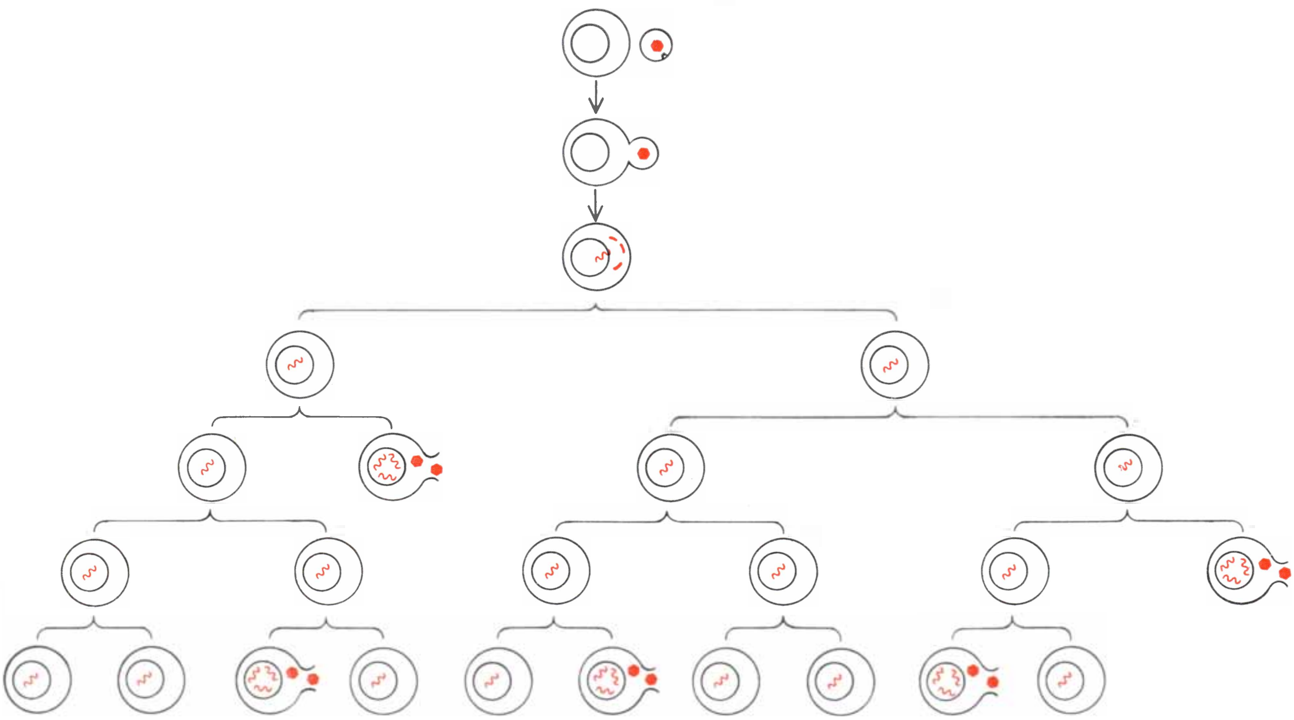
To confirm a causal relation between the Epstein-Barr virus and mononucleosis we needed additional samples of serum from patients with the disease. We remembered that James C. Niederman and Robert W. McCollum of the Yale University School of Medicine had been engaged since 1958 in a prospective study of mononucleosis in college students, collecting serum from hundreds of students before (at entry into college), during and after the disease. Their valuable collections were generously made available to us for testing. We found that antibodies to the Epstein-Barr virus were always absent before the illness but were present during and after the illness. These findings led us to conclude unequivocally that the Epstein-Barr virus is responsible for at least one disease: infectious mononucleosis. Subsequent epidemiological studies have indicated that people who have had infectious mononucleosis do not have a higher probability of contracting cancer later in life. This finding is consistent with the fact that virtually everyone becomes infected with the Epstein-Barr virus, including those who never develop mononucleosis.

Much progress has been made in elucidating how the Epstein-Barr virus causes mononucleosis. It is now known that the route of infection in nonimmune individuals is through the mouth and the respiratory tract, either by the use of contaminated eating or drinking utensils, intimate oral contact (hence the familiar label "kissing disease") or the inhalation of virus particles. The initial target cells for the virus are possibly the epithelial cells of the mouth and the throat, but the virus sooner or later infects B lymphocytes, which disseminate it throughout the lymphatic system. (For some reason the Epstein-Barr virus is highly selective for B lymphocytes and does not infect T lymphocytes.) The first signs of illness usually develop from four to seven weeks after infection; during the acute phase of the illness virus particles can be detected in the saliva of almost all patients.

In the course of its evolution the Epstein-Barr virus has established a nearly perfect symbiosis with its human host, ensuring the survival of both. Although infection with the virus leaves a permanent immunity, the virus persists in a latent state in the B lymphocytes of the lymphatic system and is spontaneously activated in a few cells. Infectious progeny virus particles are therefore released periodically into the saliva, ensuring the spread of the virus to nonimmune individuals.

Under primitive conditions such as those prevailing in prehistoric times and in many parts of the world today primary infections with the Epstein-Barr virus take place at an early age, when they do not cause significant illness, perhaps because of the immaturity of the target cells or of the immune system. Under the improved hygienic conditions and living standards of the developed world primary infections with the virus are often delayed until late childhood and adolescence, when they are more likely to cause the symptoms of mononucleosis. For example, in college populations in the U.S. between 35 and 80 percent of the students at entry have antibodies to the Epstein-Barr virus, and about 10 to 15 percent of those without antibodies become infected every year. About two-thirds of these belatedly infected individuals experience significant illness. This explains why mononucleosis is essentially a disease of the economically advanced countries.

The Epstein-Barr virus resembles some of the other herpes viruses in that it can go "underground," persisting in a silent, or repressed, state in dividing human lymphoid cells without inducing the manufacture of progeny virus and the death of the host cell. Cultured human B lymphocytes that are latently infected with the virus may be classified in two groups. "Producer" cell lines are those in which the viral genome (the complete set of viral genes) may become



SPONTANEOUS INDUCTION of replication of the Epstein-Barr virus may take place in the descendants of a single cell from a "producer" culture of Burkitt's lymphoma cells. Many cells containing the viral genome (the complete set of viral genes) persist indefinitely, dividing every 18 to 24 hours. In up to 10 percent of the cells, however, the repressed viral genome is activated, leading to the manufacture of virus particles. Liberation of progeny virus is inevitably accompanied by the death and lysis (dissolution) of the host cells.

activated spontaneously in a small proportion of the cells (up to 10 percent), giving rise to viral antigens and occasionally to infectious virus particles. "Nonproducer" cell lines are those in which the viral genome is never spontaneously activated. Even when the viral genome of a nonproducer cell is experimentally activated, the process is usually abortive: viral antigens are manufactured but virus particles are not.

The presence of viral genomes in nonproducer cells was first demonstrated in 1970 by Harald zur Hausen and Heinrich Schulte-Holthausen of the University of Erlangen. They showed that strands of viral DNA extracted from infectious virus particles paired up with complementary strands of DNA extracted from nonproducer cells. The latently infected cells were found to contain as many as 50 complete copies of the viral genome.

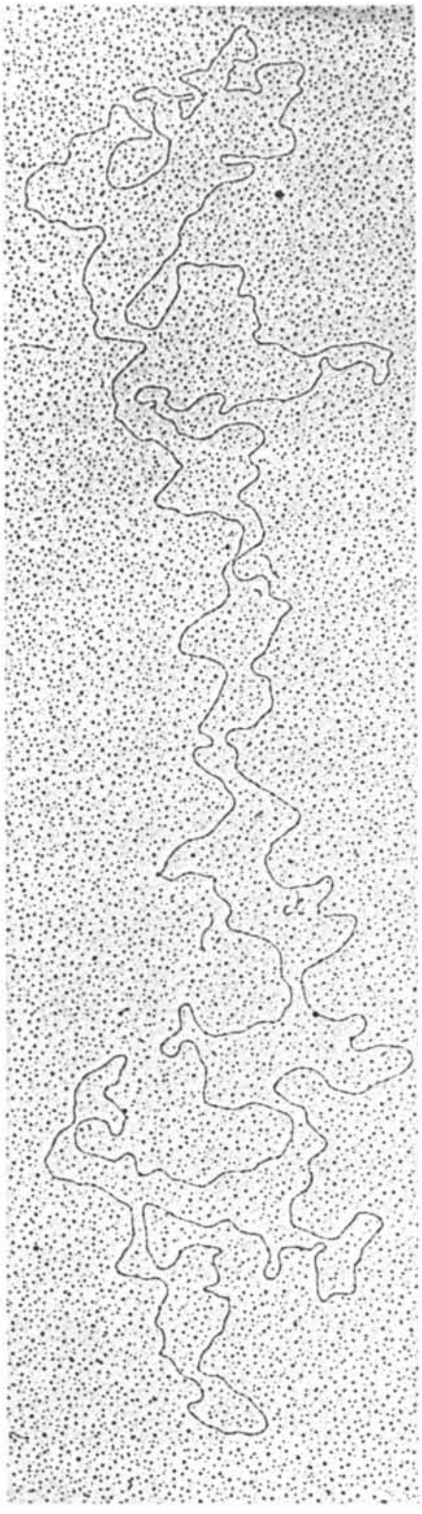
Alice Adams and Tomas Lindahl in George Klein's laboratory at the Karolinska Institute in Stockholm discovered that only part of the viral DNA could be separated from the host-cell DNA by spinning it in a centrifuge. The separable fraction consists of circular plasmids of viral DNA, which are the same length as the DNA strands packed into infectious virus particles. The nonseparable fraction represents copies of the viral DNA that have been linearly inserted into the DNA of the host cell, where they are replicated along with the rest of the host chromosomes.

In an occasional cell of a producer cell line the repression of the viral DNA is reversed, leading to the active manufacture of virus particles and the death of the cell. The virus particles shed from these producer cells into the culture medium are of two types: transforming and lytic. The lytic virus induces the manufacture of progeny virus particles in newly infected cells. The transforming virus, which is released by the majority of producer cells, causes B lymphocytes to multiply indefinitely when they are grown in tissue culture. (Uninfected B lymphocytes are unable to grow and divide in culture and usually die off within two or three weeks.) If the transforming virus is inactivated by heating or treatment with antibodies or is removed from the culture medium by passing it through an extremely fine filter, the B lymphocytes will not be transformed. These experiments have clearly established that the transforming agent is the virus itself.

The Epstein-Barr virus also transforms lymphocytes in the living human being: B lymphocytes obtained from the blood and lymph nodes of virus-infected individuals grow readily in cell culture but similar cells from uninfected individuals do not. Giovanni Rocchi and his colleagues at the University of Rome have shown that in healthy people with latent Epstein-Barr viral infections one lymphocyte in 10 million is transformed; in patients with infectious mononucleosis one lymphocyte in 5,000 is transformed. These results suggest that the Epstein-Barr virus is potentially tumor-inducing because it endows B lymphocytes with the capacity for unrestrained growth and proliferation, a notable property of cancer cells. The transformation of lymphocytes appears to be induced by copies of the viral DNA that have been linearly inserted into the host chromosome. Elliott Kieff of the University of Chicago School of Medicine is currently analyzing the viral DNA in detail in an effort to identify the specific DNA segment that codes for the transforming activity of the virus.

Although most of the viral DNA in nonproducer cells is repressed, as is indicated by the absence of viral replication, a small fraction of the viral DNA is expressed in the form of new proteins, both within the nucleus of the cell and at the cell membrane. These proteins are not structural proteins of the transforming virus but are coded by it. Since the virus-associated proteins differ from the normal cell proteins, they induce the manufacture of antibodies. One of these antigens is the Epstein-Barr nuclear antigen (EBNA), which is situated in the nucleus of the cell in association with chromatin: the material that incorporates the DNA. It was discovered by Beverly Reedman and Klein, who showed further that EBNA is expressed by all cells carrying the viral genome, including the cells of Burkitt's lymphoma and nasopharyngeal carcinoma. In nonproducer cells the presence of EBNA in the nucleus is the only indication that the cells have been transformed by the Epstein-Barr virus.

By manipulating the growth condition of latently infected lymphocytes in cell culture, or by treating them with certain drugs, nonproducer cells can be induced to enter the lytic cycle, which culminates in the death and lysis (dissolution) of the cell. The lytic cycle involves the manufacture of several additional antigens and, if it progresses far enough, the liberation of progeny virus particles. Soon after the induction of the lytic cycle and before any replication of viral DNA can be detected, three virus-specific antigens appear in the cells. These antigens are termed the early antigen complex. The first antigen, the diffuse component (*D*), appears initially in the cell nucleus and spreads throughout the cytoplasm. The second antigen, the restricted component (*R*), is localized in a large mass in the cytoplasm adjacent to the cell nucleus. The third antigen, the early membrane antigen, is inserted into the membrane of the host cell. In



DNA MOLECULE carrying the genetic information of the Epstein-Barr virus is present in multiple copies in both latently and productively infected human lymphocytes. Most of the DNA molecules are in a circular form, as is shown here, and replicate independently of the host-cell chromosomes. The rest of the viral-DNA molecules are linearly integrated into the DNA of the host cell and are copied along with it. The DNA inside the virus particle is in linear form. Micrograph was made by Gunnar Bjursell of University of Aarhus.

many cells the infectious process stops at this stage.

In other cells, however, the lytic cycle continues with the replication of the viral DNA. Following this step the late viral antigens appear: proteins that are structural components of the capsid, or protein shell, of the virus particle. The late antigens include the viral-capsid antigen (VCA) and the late membrane antigens, which are inserted into the cell membrane. Progeny virus particles are assembled in the cell nucleus, after which some of them pass into the cytoplasm of the cell and are extruded through the cell membrane. In the process the viral capsids are coated with a protective envelope derived from the membranes of the host cell. The liberation of the progeny virus particles—indeed, the mere synthesis of the early antigens—leads inevitably to the death of the cell.

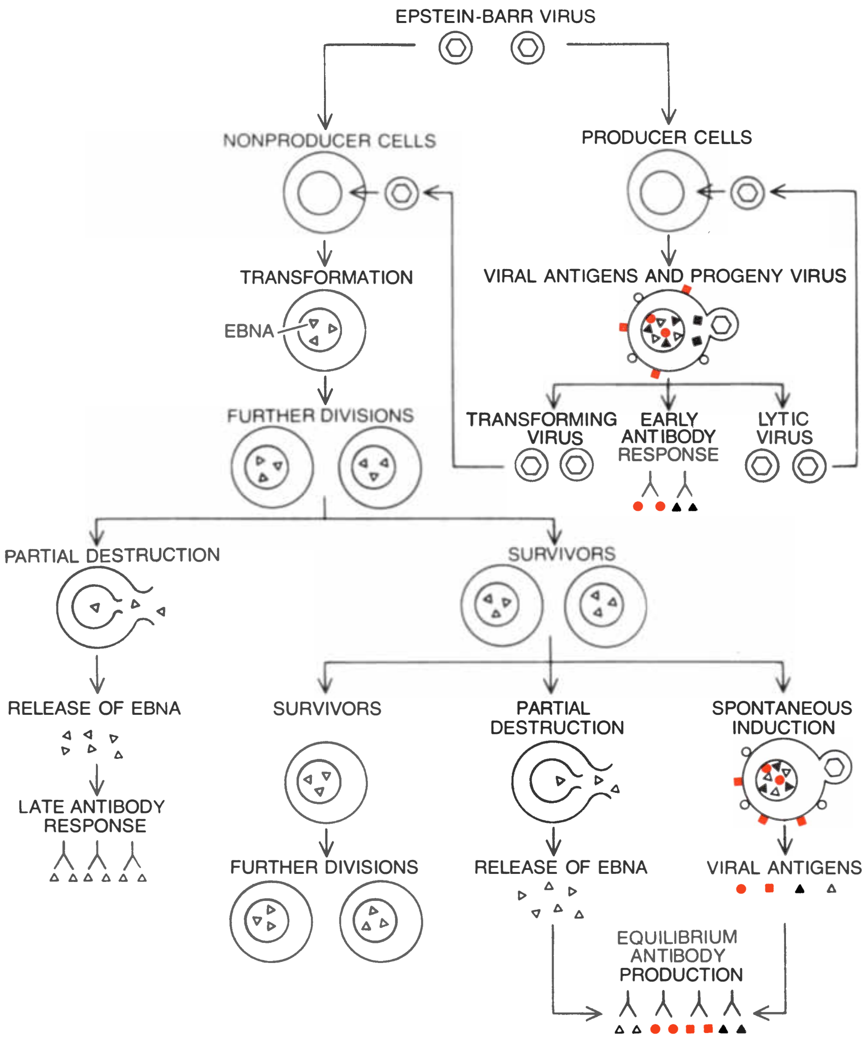
The sequence of events following infection with the Epstein-Barr virus in man is much less clear, because the incubation period between exposure and the appearance of clinical symptoms of mononucleosis is between 30 and 50 days. One way of following the course of the infection is to monitor the patient's immune response. In primary viral infection the first antibody to appear after infection is immunoglobulin M (IgM), which is specific for VCA. During the four-to-seven-week incubation period the production of IgM stops and is replaced by the production of immunoglobulin G (IgG). By the time the patient seeks medical attention the IgM and IgG antibodies to VCA have usually reached their peak levels.

More than 80 percent of patients with mononucleosis also develop transient IgG antibodies to the *D* component of the early antigen complex by the time of the onset of illness. In contrast, antibodies to EBNA develop late in the course of the disease and then persist for life.

Patients with infectious mononucleosis also develop heterophil antibodies, so named because they are able to cross-react with unrelated antigens from different animal species. For example, human heterophil antibodies will agglutinate (clump) red blood cells from sheep or horses. Heterophil antibodies were discovered in 1932 by John R. Paul and W. W. Bunnell of the Yale School of Medicine and were found to be highly specific for mononucleosis. Detection of these antibodies in the patient's serum is still the diagnostic method of choice for mononucleosis because of the simplicity of the assay. As far as can be determined, however, the antigen responsible for the heterophil antibody is not directly related to the Epstein-Barr virus, and not every patient develops these antibodies.

As we have indicated, children in developing countries are exposed to the Epstein-Barr virus very early in life. Recent studies of infants in Ghana, conducted by us in collaboration with Robert J. Biggar of the National Institutes of Health, have shown that maternal antibodies to the virus pass through the placenta from the mother to the fetus and protect the newborn infant from infection for up to six months after birth. Within the next 12 months, however, 80 percent of the infants studied developed their own antibodies to the virus without manifesting any clinical signs of illness. The antibody response of infants to the Epstein-Barr virus is comparable to the responses seen in adult cases of mononucleosis, with two striking exceptions: heterophil antibodies are either absent or reach only barely significant levels, and transient IgG antibodies to the early antigen complex are directed against the *R* component rather than the *D* component. (Antibodies to the *R* component are seen characteristically in patients with Burkitt's lymphoma.) The significance of the different antibody profiles of infants and adults with Epstein-Barr viral infections is not understood.

Whereas the evidence linking the Epstein-Barr virus to mononucleosis is unequivocal, the involvement of the virus in the two cancers is less clear-cut. So far the case for a causal relation between the Epstein-Barr virus and



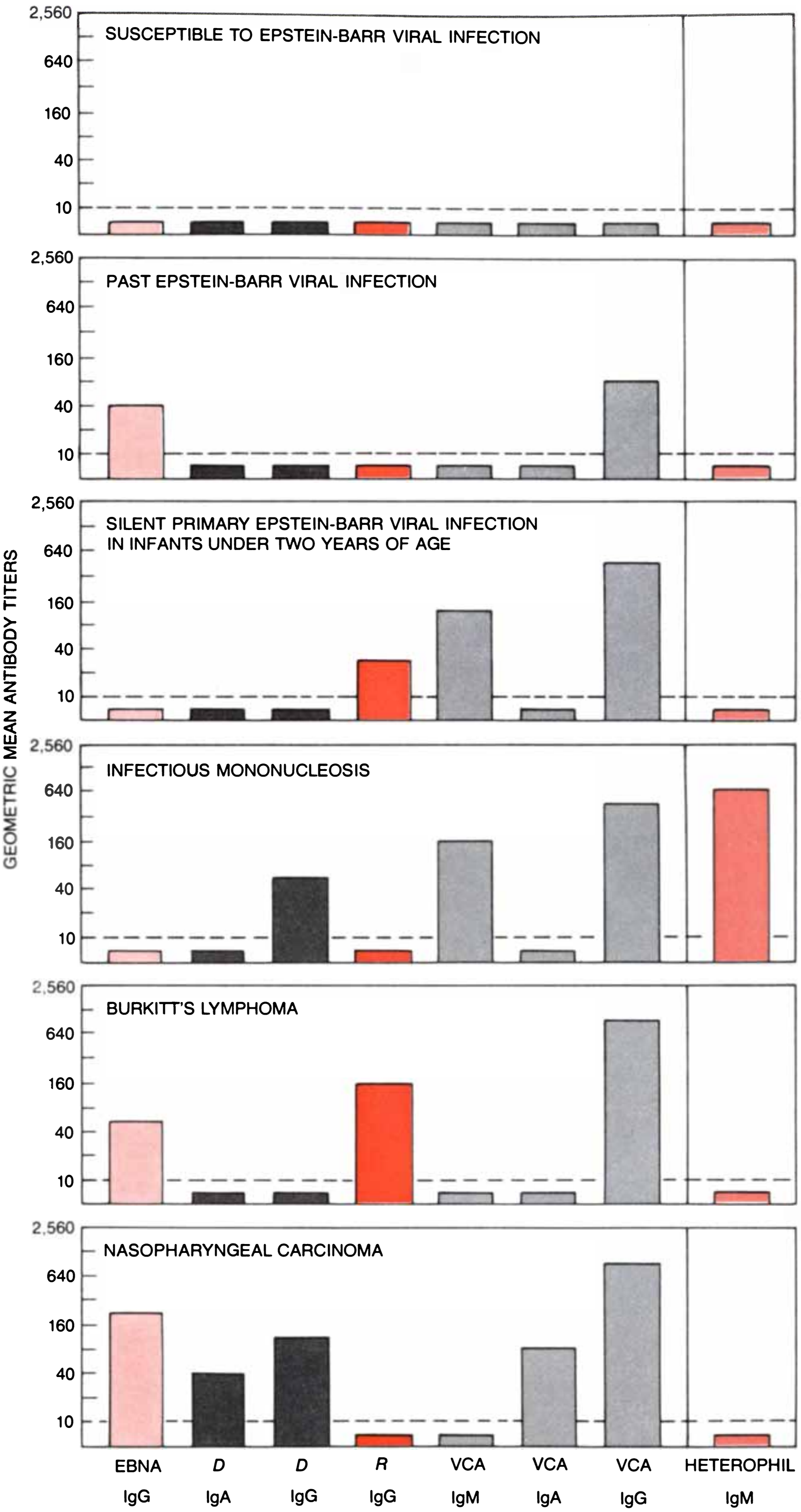
SEQUENCE OF EVENTS during infectious mononucleosis in man is summarized in this flow chart. Many B lymphocytes become lytically infected by the Epstein-Barr virus as the infection spreads; they manufacture viral antigens and virus particles before their inevitable death and lysis. The viral antigens induce an early antibody response. The progeny virus particles are mostly of the transforming type; B lymphocytes infected by them remain alive and divide indefinitely. The transformed cells harbor copies of the viral DNA and express both the Epstein-Barr nuclear antigen and an antigen on the cell surface that becomes the target of activated T lymphocytes. The ensuing destruction of transformed B lymphocytes by the T lymphocytes leads to some of the clinical signs of mononucleosis; it also releases enough EBNA to induce a late antibody response to this antigen. The transformed cells are never completely eliminated. Some of the survivors are spontaneously induced to manufacture virus particles. Others are continually weeded out, liberating antigens and inducing an equilibrium level of antibody.

Burkitt's lymphoma rests on the following evidence: (1) the detection of viral DNA or viral antigens in the tumor, (2) the transformation of cultured B lymphocytes by the virus, (3) the induction of lymphomas by the virus in nonhuman primates, (4) the demonstration of higher antibody titers to the virus in lymphoma patients compared with the titers in a control population and (5) the correlation between antibody patterns and the prognosis of the tumor. All five of these criteria have provided considerable support for a causal relation between the Epstein-Barr virus and Burkitt's lymphoma, although important questions remain to be answered.

The DNA of the Epstein-Barr virus has been found in the great majority (98 percent) of biopsies from Burkitt's lymphomas in Africa but not in any other type of lymphoma. Each tumor cell contains multiple copies of the viral genome. Lytically infected cells that make the early antigens, VCA or progeny virus particles are not observed in fresh tumor biopsies but appear in one or two days when the cells are placed in culture. It is likely that lytically infected cells arise in small numbers within the tumor but are rapidly destroyed by the immunological defenses of the patient as soon as they express the early membrane antigen on the cell surface. Indeed, as early as 1966 Klein and his colleagues had detected an antigen on the surface of Burkitt's lymphoma cells and found that it was often coated with the patient's own antibodies. The Epstein-Barr virus is therefore intimately associated with the Burkitt's lymphoma cells.

Cultures of B lymphocytes transformed by the Epstein-Barr virus differ, however, in important respects from cultures of Burkitt's lymphoma cells, which are also derived from B lymphocytes. Whereas the virus-transformed lymphocyte cultures arise from a number of different cells, the Burkitt's lymphoma cultures are monoclonal: each tumor arises from a single malignantly transformed cell. In addition the great majority of Burkitt's lymphoma cells show a chromosomal abnormality that is not seen in transformed lymphocytes or in the lymphocytes of patients with mononucleosis. This abnormality, which was first observed by G. Manolov, Y. Manolova and their colleagues in A. Levan's laboratory at the University of Lund, involves the translocation of a small segment from one of the No. 8 chromosomes to the long arm of one of the No. 14 chromosomes, and of a small piece from the No. 14 chromosome to the No. 8 chromosome. Similar but not identical abnormalities of chromosome No. 14 have been detected in other lymphomas.

The origin and function of the chromosomal aberration in Burkitt's lymphoma is not known, but it appears to



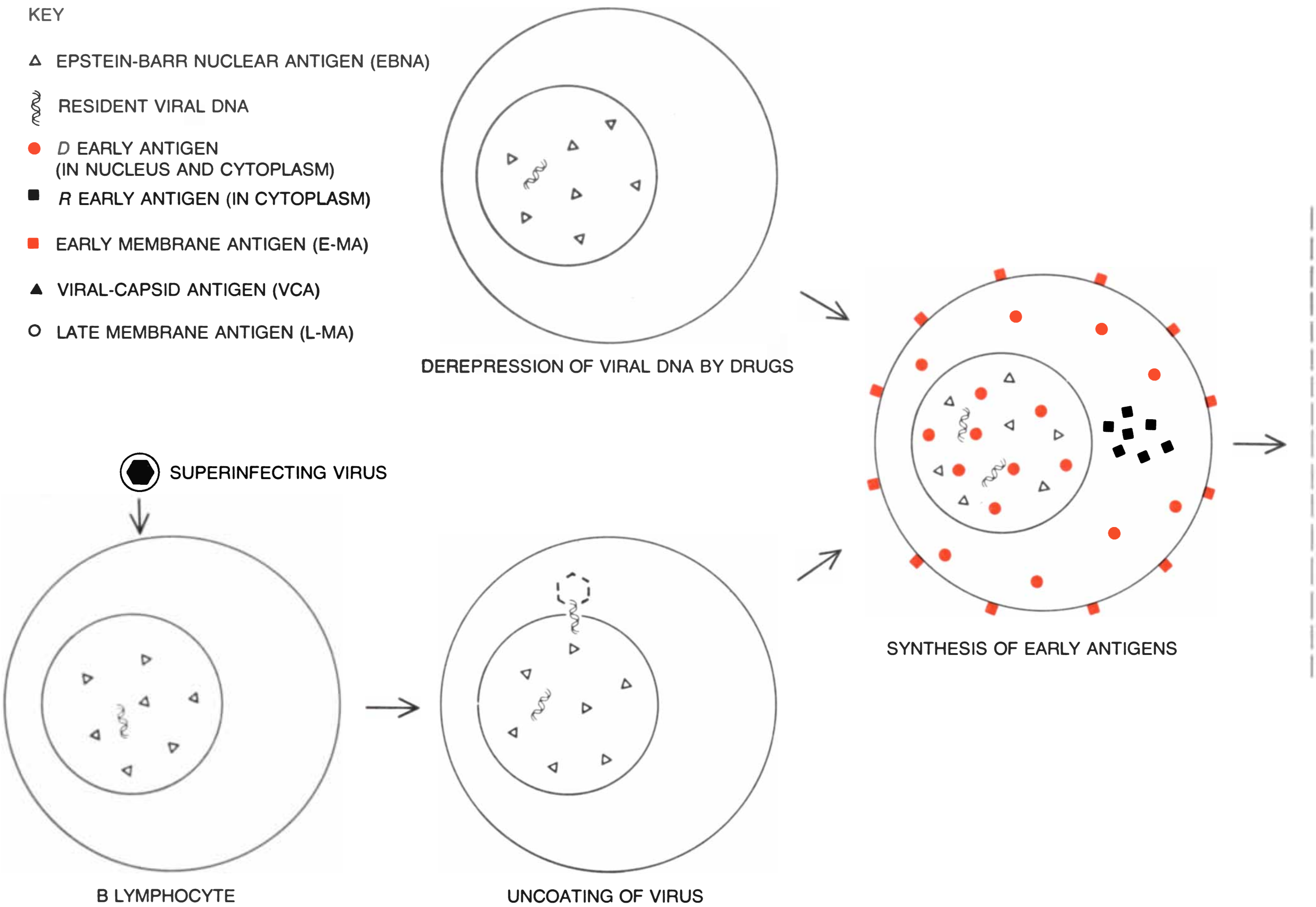
PATTERNS OF ANTIBODIES manufactured in response to the Epstein-Barr virus differ significantly in the various diseases associated with the virus. Characteristic patterns have been identified for silent primary infections in infants less than two years old, for infectious mononucleosis in young adults, for Burkitt's lymphoma and for nasopharyngeal carcinoma. Patients may manufacture antibodies of three different immunoglobulin classes (IgM, IgA and IgG), which are directed against these virus-associated antigens: the Epstein-Barr nuclear antigen, the viral-capsid antigen and the diffuse and limited components of the early antigen complex. Antibodies to the heterophil antigen are present only in patients with mononucleosis.

confer a biological advantage on the affected cells. Whatever causes the aberration might be an important predisposing factor or cofactor in the genesis of Burkitt's lymphoma, since lymphocytes derived from patients with mononucleosis do not have the abnormal chromosomes. Here again, however, the situation is not clear-cut: the translocations are observed in most but not all Burkitt's lymphomas associated with the Epstein-Barr virus, and they are found also in Burkitt's lymphoma cells not associated with the virus. It seems likely that the transformation of lymphocytes by the Epstein-Barr virus is only a first step in the process of malignant transformation and that other events are required to make the cells fully malignant.

Experiments on subhuman primates have supported the association of the Epstein-Barr virus with Burkitt's lymphoma. Thomas Shope, D. C. DeChairo and I. George Miller of the Yale School of Medicine injected transforming virus into marmosets and other New World monkeys and observed a variety of responses. Some of the animals manufactured antibodies to the virus without manifesting clinical illness, others showed a transient proliferation of lymphocytes and the rest developed fatal lymphomas in from six to 14 weeks. The lymphoma cells expressed EBNA, and the animals manufactured antibodies to VCA and the early antigen complex. The lymphomas induced in monkeys by the Epstein-Barr virus do not, however, conform in every respect to Burkitt's lymphomas in man. They are diffuse rather than solid and develop in weeks rather than the years that often elapse between primary viral infection in African children and the appearance of Burkitt's lymphoma. These experiments have nonetheless clearly demonstrated that the Epstein-Barr virus can induce malignant lymphomas.

Patients with Burkitt's lymphoma generally have high antibody titers to VCA and the *R* component of the early antigen complex; antibodies to the *D* component may develop in the final stage of the disease. Long-term follow-up studies of individual patients have revealed that those patients who respond initially to therapy but who maintain high or increasing antibody titers to VCA and the *R* antigen generally face repeated and ultimately fatal recurrences of the tumor, even after apparently disease-free periods of from three to six years. In contrast, patients who show a steady decline in antibody titers to the *R* antigen after surgery or chemotherapy have a good chance of long-term survival without evidence of disease. Antibodies to the Epstein-Barr virus are therefore good indicators of the prognosis in Burkitt's lymphoma.

Although practically all African children have been silently infected with the Epstein-Barr virus by the age of three, the peak incidence of Burkitt's lymphoma falls between the ages of six and eight. To investigate this puzzling discrepancy a prospective study of the lymphoma was organized in 1972 under the auspices of the International Agency for Research on Cancer in the West Nile District of Uganda, a region where the disease is endemic. Blood samples were taken from more than 45,000 children aged from one to eight, and the serums were stored for future reference. Between 1972 and 1977, 14 of these chil-



LYTIC CYCLE of replication of the Epstein-Barr virus can be induced in a cultured Burkitt's lymphoma cell by superinfection with the virus or by derepression of the resident viral genome with antimetabolite drugs. Once the synthesis of the early antigen complex

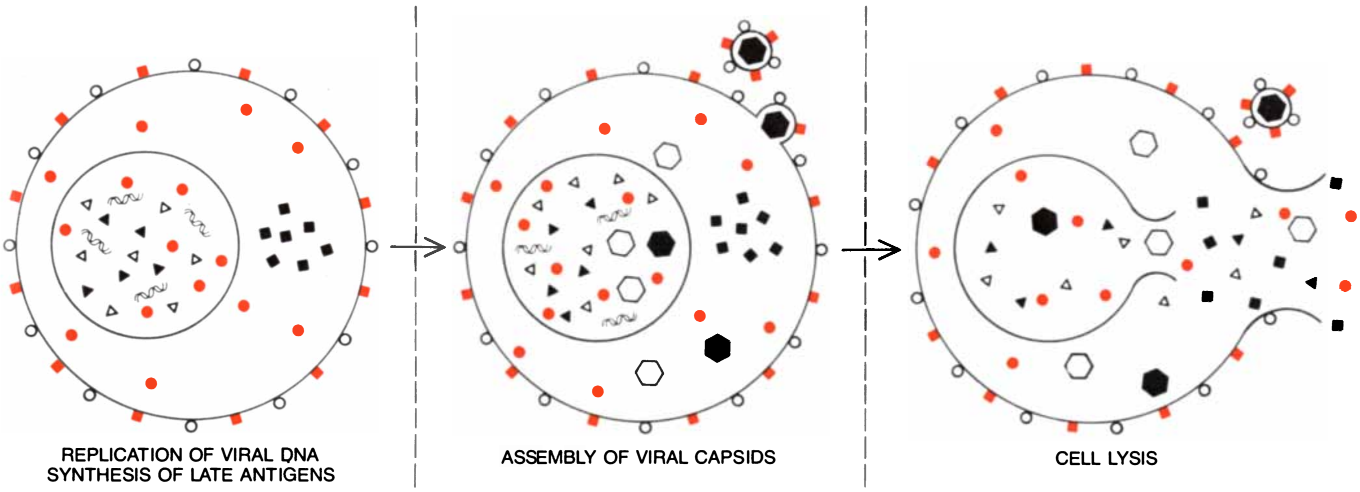
dren developed Burkitt's lymphoma. The presickness serums of the children were then tested for antibodies to the Epstein-Barr virus and compared with the serums from age- and sex-matched healthy children from neighboring families and numerous other controls.

It was found that the sick children had acquired antibodies to the virus from seven to 54 months before the tumor was detected and that their antibody titer to VCA exceeded the titer of nearly all the healthy control children. It was therefore not surprising that the appearance of the tumor was not accompanied by a significant rise in VCA antibodies over the presickness levels. Ten of the 14 patients, however, acquired new antibodies to the *R* antigen. These results ruled out the possibility that Burkitt's lymphoma is a rapid consequence of a rare, delayed primary infection with the Epstein-Barr virus in African children. We concluded that a heavy, persistent viral carrier state, evident from high antibody levels, predisposes children to Burkitt's lymphoma but that additional factors are required for the development of the tumor.

Malaria, which is highly endemic in regions where Burkitt's lymphoma is prevalent, could be one such factor. Heavy infection with malaria parasites might predispose a child to Burkitt's lymphoma because malaria is known to suppress the immune system and to enhance the proliferation of lymphocytes. An epidemiological study of the effect of malaria-control programs on the incidence of Burkitt's lymphoma is now in progress and should shed some light on the apparent link between the two diseases. Another important factor in the genesis of the lymphoma might be the age at which primary infections with the Epstein-Barr virus occur. Infections early in infancy might transform immature B lymphocytes, enabling them to persist in the body and eventually give rise to lymphoma cells. Further study of these early infections might therefore prove rewarding.

Although Burkitt's lymphoma is endemic in several regions of both East and West Africa, it also arises sporadically in other parts of the world, including the U.S. About 80 percent of the American cases, however, are not associated with the Epstein-Barr virus: no viral-DNA-positive or EBNA-positive cells are detected in the tumors, and the patients have antibody profiles comparable to those of normal children. The remaining 20 percent of the American cases of Burkitt's lymphoma are associated with the Epstein-Barr virus and are indistinguishable from their African counterparts, except that in the U.S. the age range of the patients is shifted toward adolescence and young adulthood. Recently a cluster of four young adults with virus-associated Burkitt's lymphoma was observed in rural Pennsylvania in one year. The patients lived within 30 miles of one another, and two of them, who were related by marriage, had shared the same household for several years.

The cases of Burkitt's lymphoma that are not associated with the Epstein-Barr virus (not only the 80 percent of American cases but also about 2 percent of the African ones) clearly present a dilemma: either the lymphoma associated with the virus is a disease different from the lymphoma that is not associated



begins, the production of all cellular proteins stops and the cell will inevitably die, even though infectious process may be arrested at various stages (*indicated by broken lines*). Most cells express early antigens, fewer express late antigens and very few liberate virus particles.

with the virus (although the tumors are morphologically indistinguishable) or Burkitt's lymphoma has at least two causes, one of them being the Epstein-Barr virus. Whatever the answer, the secondary factor thought to be essential for the development of the tumor cannot be malaria in the American cases of virus-associated Burkitt's lymphoma. No other secondary factors have been implicated, however, nor has a different virus been found in Burkitt's lymphomas that are not associated with the Epstein-Barr virus.

The case for an association between the Epstein-Barr virus and nasopharyngeal carcinoma is based on the same criteria as those for an association between the virus and Burkitt's lymphoma. All biopsies of the carcinoma obtained anywhere in the world have so far revealed the presence of the viral DNA, which is lacking in all other carcinomas of the head and the neck.

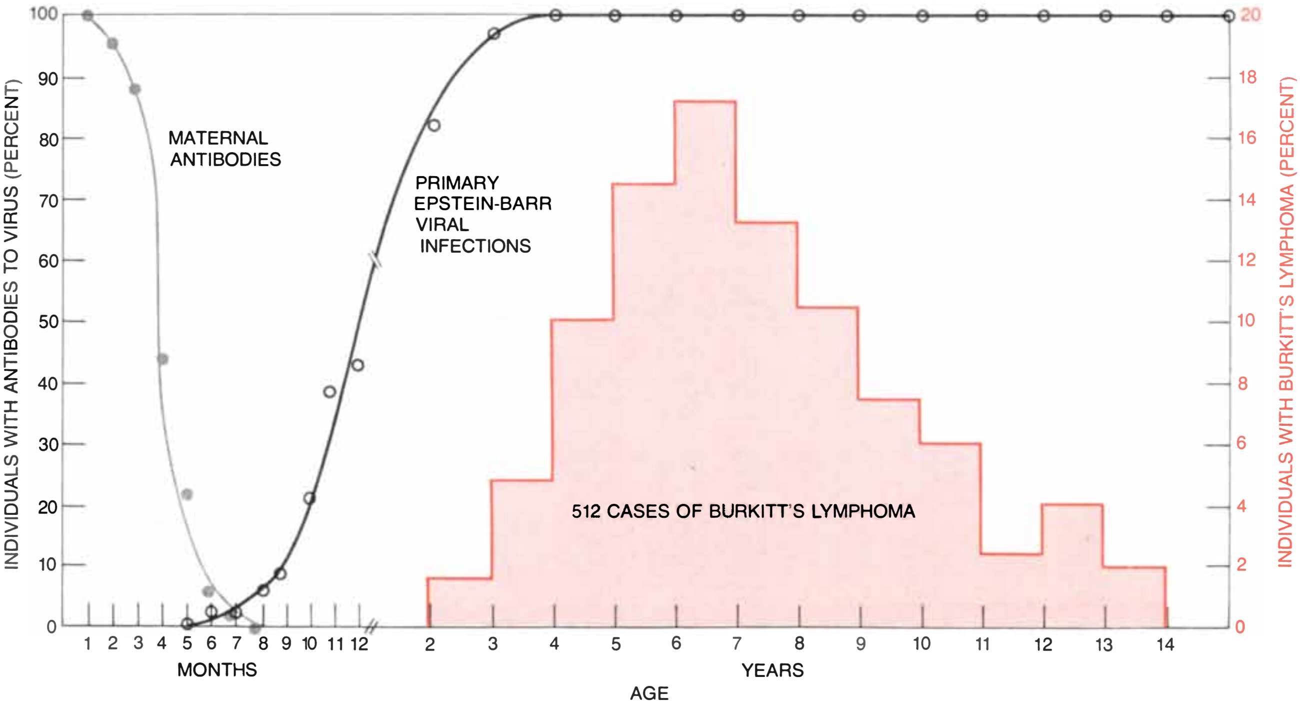
A number of experiments have demonstrated that the viral DNA is in the carcinoma cells and not in the lymphocytes that infiltrate the tumor. For example, imprints of biopsy fragments on glass slides deposit carcinoma cells, which can then be shown to contain EBNA. Lytically infected cells have not been detected in the tumors, presumably because such cells are rapidly destroyed by the patient's immune defenses. When the carcinoma cells are placed in culture, however, the latent viral genome can be activated with drugs, giving rise to the manufacture of the *D* antigen, VCA and even infectious progeny virus particles that are capable of transforming B lymphocytes in culture. Nevertheless, no cultured normal epithelial cells have been transformed by the Epstein-Barr virus, nor have carcinomas been induced by injecting the virus into nonhuman primates. Presumably the reason for these failures is that inappropriate routes were used for administering the virus or that unknown cofactors are required.

Patients with nasopharyngeal carcinoma exhibit a characteristic antibody response to the Epstein-Barr virus. Unlike patients with Burkitt's lymphoma, they manufacture a high titer of antibody to the *D* component of the early antigen complex but not to the *R* component. Another unique feature of the carcinoma is the production of antibodies of the immunoglobulin class A (IgA). Antibodies of this type to VCA and the *D* antigen are frequently present at high levels.

In collaborative studies with John C. Ho of Queen Elizabeth Hospital in Hong Kong we have demonstrated that both IgG and IgA antibody titers to VCA and the *D* antigen in patients with nasopharyngeal carcinoma rise as the tumor increases in size and then decline after successful treatment. In some of the patients there was an initial decline in the titers over a period of from two to four years, after which the trend reversed and the titers rose again. These patients had a subsequent recurrence of the tumor or suffered metastases. Monitoring antibody titers to the Epstein-Barr virus in patients with nasopharyngeal carcinoma therefore provides a way of assessing the effectiveness of therapy and of predicting imminent tumor recurrences.

These observations suggest that the Epstein-Barr virus plays a causal role in nasopharyngeal carcinoma, but as with Burkitt's lymphoma a number of puzzling questions remain. Although primary infections with the Epstein-Barr virus also occur early in life in regions where the carcinoma is endemic, the tumor develops in China only in adults, with a peak incidence at about 40 years of age. In Tunisia the age distribution is bimodal, with about 20 percent of the cases developing before the age of 20. The long interval between the primary viral infection and the development of the tumor again suggests the participation of cofactors. Among the Chinese a genetic predisposition for the tumor seems evident from the occurrence of multiple cases in some families through three or four generations. Chinese patients with the carcinoma also show a certain histocompatibility antigen (HLA) at a high frequency, but this antigen is not observed in the Tunisian patients. Another contributing factor may be the ingestion or inhalation of carcinogenic substances such as nitrosamines, which are present in smoked fish and the smoke of wood fires and incense.

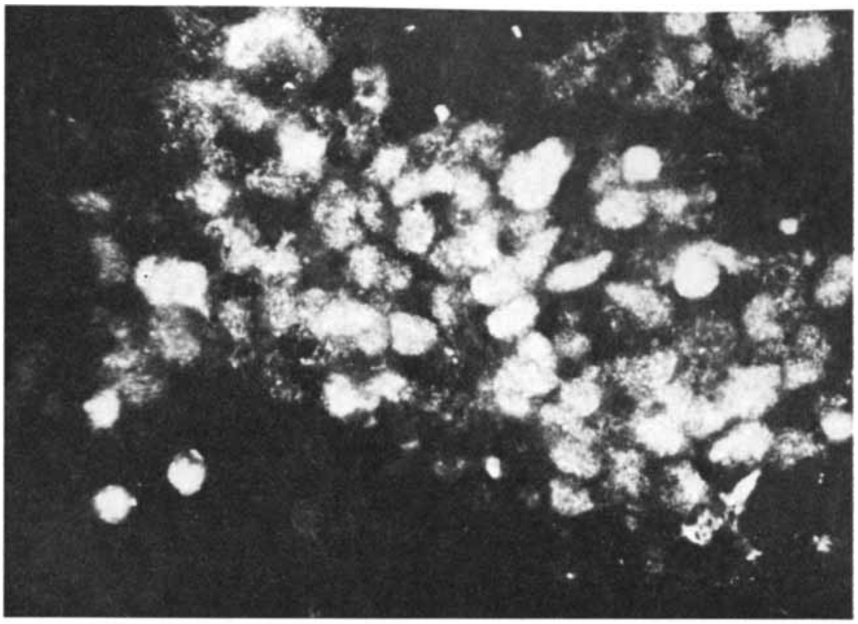
In summary, the Epstein-Barr virus is

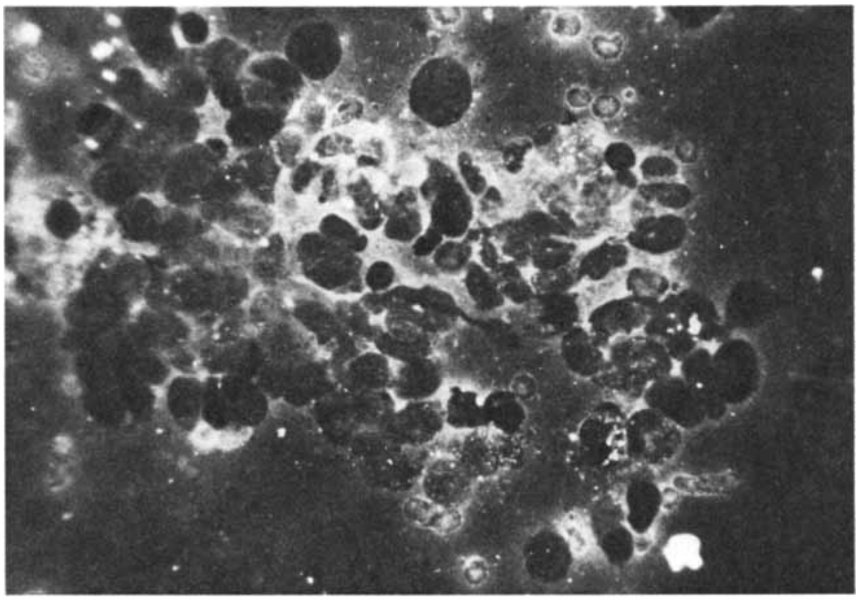


PRIMARY INFECTIONS with the Epstein-Barr virus in African children develop soon after the decline of maternal antibodies to the virus, which were transmitted to the fetus across the placenta. In this graph the time of primary viral infections is shown in relation to the decline of maternal antibodies and the age of onset of 512 cases of Burkitt's lymphoma. Note that the peak incidence of the tumor is reached several years after the primary viral infection, suggesting that the genesis of the tumor involves factors in addition to the virus.

one of the most widely disseminated human viruses. It is definitely the cause of infectious mononucleosis, and all the available evidence points to its involvement in the genesis of Burkitt's lymphoma and nasopharyngeal carcinoma. Both of these tumors arise from the uncontrolled division of a single malignantly transformed cell, meaning that the Epstein-Barr viral genome must have been present in that first cell. It seems improbable that the true cause of the tumors, whatever it is, would for no specific reason always single out cells carrying the viral genome. Such cells, after all, represent only a minute fraction of the total B-lymphocyte population from which Burkitt's lymphoma arises. A similar consideration might hold for the precursors of the carcinoma cells. It is therefore unlikely that the Epstein-Barr virus is merely a harmless passenger in these tumors. If it is not the primary factor, it is at least a priming one.

It seems clear that several events are required to make the virus-transformed cells fully malignant. These events could be the action of carcinogens of a chemical, physical or biological nature. Severe malnutrition, the concurrent action of a disease such as malaria, a decline in the immunological defenses of the host or an imbalance in any one of a large number of body functions may be further contributing factors.





PRESENCE OF A VIRAL ANTIGEN associated with the Epstein-Barr virus in a biopsy specimen of nasopharyngeal carcinoma supports the link between the virus and the genesis of this cancer in man. Imprints were made of the biopsy specimen on glass slides. The imprints were stained by fluorescent antibodies to the Epstein-Barr nuclear antigen (*top*) and serum lacking antibodies to the antigen (*bottom*). The nuclei of the carcinoma cells glow brilliantly in the presence of the antibody but are dark in its absence. Technique was developed by authors in collaboration with Dolly Huang and John C. Ho of Queen Elizabeth Hospital in Hong Kong.

A vaccine against the Epstein-Barr virus, in addition to being desirable for the prevention of mononucleosis, could provide further support for a causal role of the virus in Burkitt's lymphoma and nasopharyngeal carcinoma if it would prevent their occurrence in areas where those cancers are endemic. The development of such a vaccine, however, is beset with obstacles that are possibly insurmountable. For one thing, the virus cannot as yet be produced in large quantities and can so far be obtained only from cells tainted with malignant traits. Moreover, killed-virus vaccines made from herpes viruses have generally proved to be unsatisfactory or to provide only transient protection. Live-virus vaccines would be possible if the virus could be attenuated (weakened), but such attenuation might increase its cancer-causing potential in unpredictable ways.

There is a further problem. In regions where Burkitt's lymphoma or nasopharyngeal carcinoma are endemic the vaccine would have to be administered during the short and uncertain period after the maternal antibodies to the Epstein-Barr virus (which are transmitted to the fetus across the placenta) have disappeared from the infant's blood and before the infant becomes naturally infected. The maternal antibodies would prevent the vaccine from inducing the manufacture of new antibodies by the infant's own immune system, and natural infection with the virus may occur soon after the maternal antibodies have disappeared. This time element presents a logistical problem that would be difficult to solve under the field conditions prevailing in the Burkitt's lymphoma belt of Africa. Nevertheless, efforts should be made to overcome these technical obstacles, if for no reason other than the prevention of infectious mononucleosis.

This account serves to illustrate the difficulties inherent in providing a final proof for the involvement of viruses in the causation of human cancers. In spite of the fact that all the indirect criteria for linking a virus to a human cancer have been satisfied for the Epstein-Barr virus, uncertainty remains about the exact role of the virus in the two malignancies. Indeed, a final proof to everyone's satisfaction may never be attained. For those willing to accept the indirect evidence, however, the Epstein-Barr virus is the foremost candidate for being the first known human cancer virus.

The Central Parsec of the Galaxy

Infrared and radio observations indicate that the center of our galaxy harbors an ultracompact object, possibly a massive black hole, embedded in a dense, swirling mass of stars, gas and dust

by Thomas R. Geballe

In the early 1930's Karl Jansky, a young physicist employed by the Bell Telephone Laboratories in New Jersey, built the first radio telescope to assist him in identifying sources of interference that could hamper transatlantic radio-telephone communication. One major source of static Jansky discovered was a "steady hiss," as he described it, that came from the sky and was strongest in a direction far to the south of his instrument. Over the course of time he observed that the source of this interference followed the motion of the fixed stars and hence originated outside the solar system. The direction of peak intensity, toward the constellation Sagittarius, coincided well with what was already considered by astronomers to be the location of the center of our galaxy.

Jansky's observations were not only the first radio-astronomical observations but also the first detection at any wavelength of the nucleus of the galaxy, which is shrouded by clouds of interstellar dust. Later, with the benefit of the advances in electronics during World War II, radio astronomy earned recognition as a full-fledged scientific enterprise. In the 1960's and 1970's infrared technology, which had lagged behind its radio counterpart, began to provide sensitive detectors and spectrometers that could be used in conjunction with large optical telescopes. Because of their ability to "penetrate" the intervening clouds of dust, radio and infrared astronomy have contributed most of the knowledge that has been gained about the central region of the galaxy. In the past few years continued advances in both fields have enabled astronomers to begin to explore in detail the innermost and very densest region.

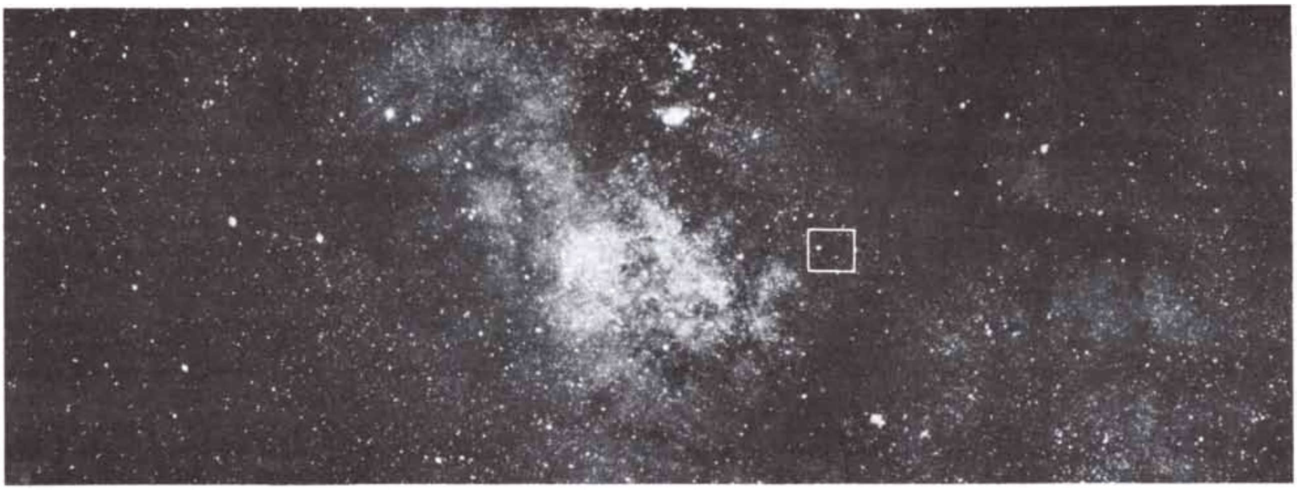
Galactic nuclei are dense in mass solely as a consequence of gravitation. The galaxies in the universe are thought to have formed from density nonuniformities in the gas ejected by the "big bang" in which the universe was created. In time the nonuniformities contracted gravitationally to form galaxies, and within them smaller clouds collapsed by gravitation to form stars and other objects. All the objects in a galaxy are also subject to a net gravitational pull toward its nucleus. Although the rotation of the galaxy slows the infall of matter, and momentary outbursts (for example supernovas) may eject matter from the center, frictional forces inevitably cause the center to acquire a mass density much higher than that of the outer regions of the galaxy. The ultraluminous objects known as quasars, which can be observed at vast distances, may be an extreme result of the process. The center of even a normal-looking spiral galaxy such as our own, however, might be expected to harbor almost the entire galactic astronomical zoo: main-sequence stars, red giants, white dwarfs, gaseous nebulas, dust clouds, neutron stars and perhaps even a black hole—all packed together in a tight swarm.

Galactic nuclei are the probable source of the enormous expulsions of material that have been observed in some distant galaxies. It appears that the energy necessary for such expulsions can have come only from dense galactic nuclei. Our galaxy also shows evidence of past explosive events, albeit relatively mild ones, in the large clouds of gas observed by radio astronomers to be moving outward from the nucleus.

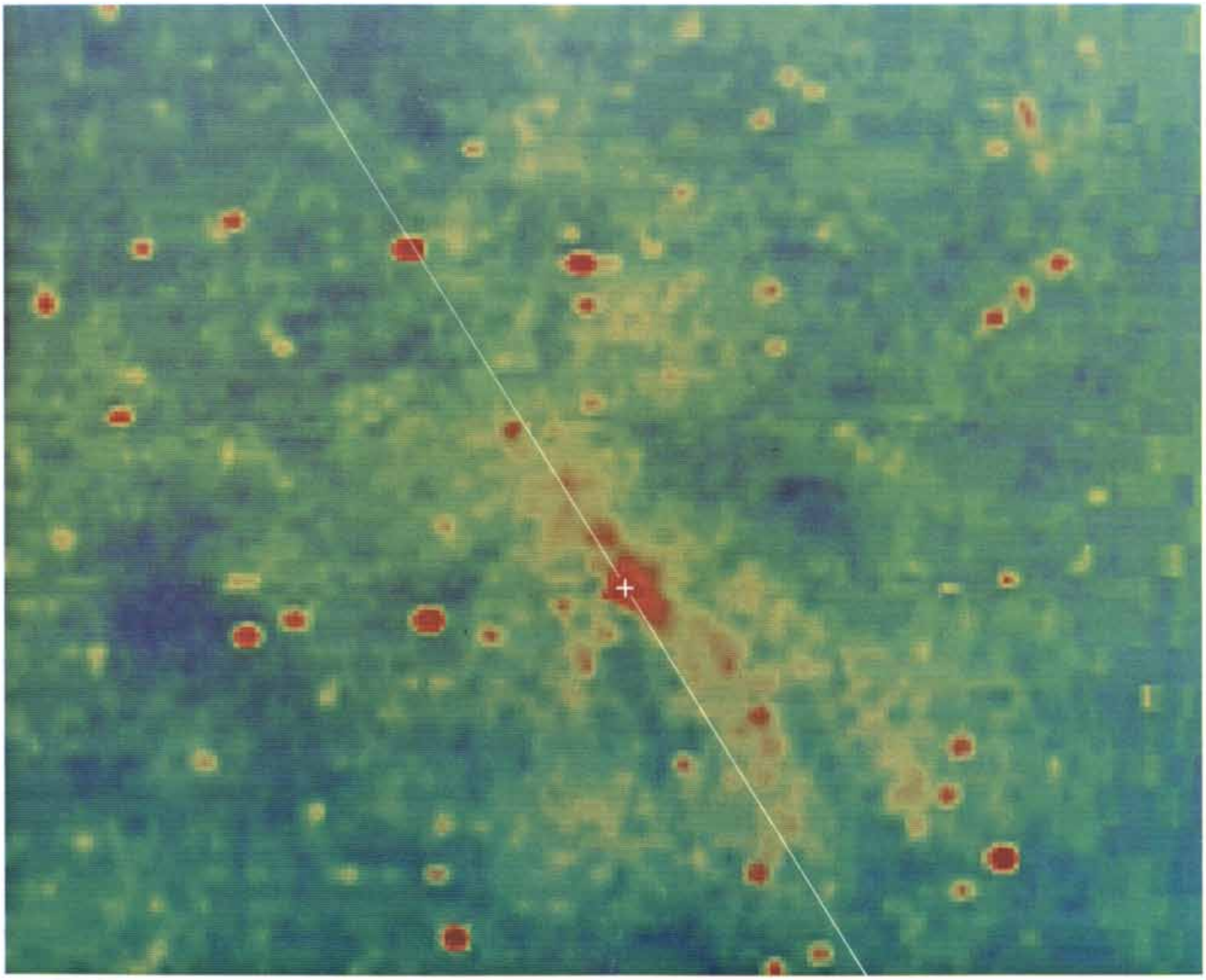
Perhaps the most exciting discoveries currently being made about the nucleus of our galaxy concern its central parsec, a volume only 3.26 light-years in diameter. At the distance of the center one parsec subtends an angle of 20 seconds of arc, which is equal to the angle subtended on the earth by a medium-size crater on the moon. Within that volume, which would fit comfortably between the sun and its nearest stellar neighbor, there are millions of stars and a variety of other objects whose nature astronomers are just beginning to understand. There is also evidence that an amount of material a few million times more massive than the sun is concentrated at the very heart of the central parsec, perhaps in the form of a black hole. Although the black hole cannot be observed directly, its existence is suggested by the behavior of the surrounding material.

I have chosen the unit of a parsec for this account not because it has any inherent virtue over some other unit, such as the light-year, but simply because the parsec is the astronomer's habitual unit of distance. One parsec is the distance from the earth to an object whose position as observed from the earth seems to shift back and forth from its average position by an angle of one arc-second (1/3,600 degree) as a result of the parallax as the earth travels around the sun. The distance to the center of the galaxy is about 10,000 parsecs, or about 20 times greater than the distance to the well-known Great Nebula in Orion. On the other hand, the center of the galaxy is about 70 times closer to the solar system than the nucleus of any other galaxy is. Therefore one should be able to perceive the details of the nucleus of our galaxy 70 times more clearly than those of the nucleus of any other galaxy. One problem, of course, is that when a telescope is pointed toward the center of our galaxy, it does not distinguish among objects lying in the foreground, in the nucleus or in the background. There are various stratagems, however, for making such distinctions and for deciding which features are actually at the galactic center.

The largest telescopes and the most sensitive photographic emulsions are able to record only the merest trace of visible light emitted from the millions of stars packed within the nucleus of our galaxy. Because of intervening clouds of dust only one photon of visible light in about 10^{11} photons survives the journey of 30,000 years from the galactic center to the earth. Even the rare surviving photons are swamped by the exposure-limiting background radiation consisting of visible photons scattered in the earth's atmosphere. In the infrared region of the spectrum, however, a much larger fraction of the photons complete their long journey. In the radio region photons from the galactic nucleus are only rarely scattered and absorbed. At



CENTER OF THE GALAXY lies hidden in the southern Milky Way in the direction of the constellation Sagittarius. The distance from the solar system to the center of the galaxy is about 10,000 parsecs, or slightly more than 30,000 light-years. The light from the millions of stars in the nucleus of the galaxy is completely obscured by interstellar dust. Infrared and radio emission from the galactic nucleus, however, is detectable. The rectangle, which is about twice the diameter of the full moon, outlines area that is shown in the infrared map below.

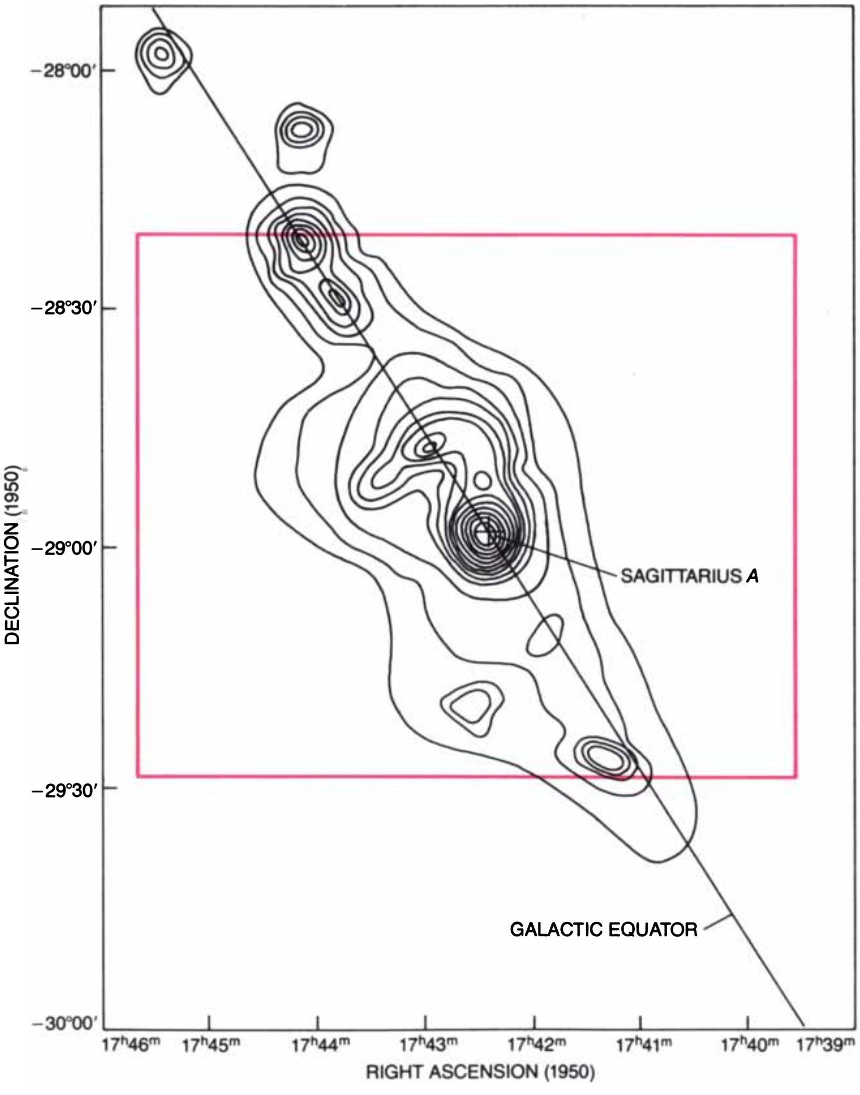


INFRARED VIEW OF THE GALACTIC CENTER was made at a wavelength of 2.2 microns (four times the wavelength of green light) by Eric E. Becklin and Gerry Neugebauer of the California Institute of Technology and the Hale Observatories. Their observing instrument was the one-meter telescope at the Las Campañas Observatory in Chile. In this computer-generated color-coded picture the intensity of the radiation increases from blue to red. The cross marks the center of the galaxy. It can be seen that the infrared radiation tends to concentrate along the galactic equator, indicated by the thin white line. At 2.2 microns cool red stars are the principal source of radiation detected from the galactic center. Individual bright spots represent red-giant stars in the foreground. Patchiness is caused by interstellar dust.

the other end of the spectrum, in the X-ray region, photons also travel unimpeded, but they do not survive passage through the earth's atmosphere and therefore can be recorded only by instruments carried to high altitude. Our atmosphere, however, is transparent at most radio and some infrared wavelengths. As a result most of what has been learned of the central region of the galaxy has been provided by radio telescopes and by sensitive infrared detectors placed at the foci of reflecting telescopes.

If our galaxy were free of dust, one could use a large optical telescope to record the entire central parsec of the galaxy on a single photographic plate with a resolution of about a second of arc. With a telescope in an artificial satellite the resolution could be improved by a factor of 10 or 20. That improvement might be just sufficient to resolve some 100,000 stellar objects within a diameter of 20 arc-seconds, corresponding to the central parsec of the galaxy. If the central parsec actually contains a few million stellar objects, as now seems probable, they would merge into bright clumps, and perhaps into a single brilliant mass, even if they were photographed from an artificial satellite at high resolution.

One can obtain an impression of what a dust-free view of the galactic nucleus might look like from photographs taken of the nucleus of the Great Nebula in Andromeda, the nearest spiral galaxy similar to our own. Because of the way the two galaxies are oriented in space the nucleus of the Andromeda galaxy can be photographed along a line of sight that is largely free of dust, which is concentrated in the central planes of the two systems.



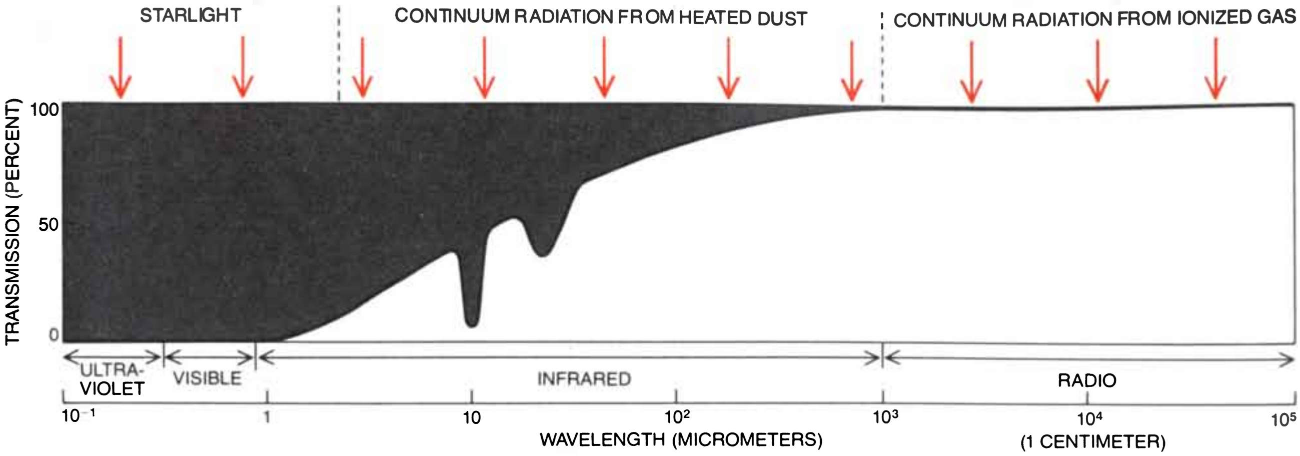
RADIO MAP OF THE GALACTIC CENTER was made at a wavelength of 3.75 centimeters by Dennis Downes and Arthur E. Maxwell of Harvard University and M. L. Meeks of the Massachusetts Institute of Technology with the 36.6-meter radio telescope at Westford, Mass. The principal source of the radio emission is ionized gas. The rectangle encloses the region represented in the picture made at infrared wavelengths on the preceding page. Ionized gas and stars are concentrated along galactic equator. They are brightest at the radio source Sagittarius *A*.

The fact that the center of our galaxy is invisible at visible wavelengths has undoubtedly been a stimulus to efforts to observe it at other wavelengths. In order to obtain a "picture" at infrared and radio wavelengths it is necessary to scan successive strips of the sky (by moving the telescope, for example) while recording point by point the number of photons striking the detector. In this way one can build up a television-like image with a resolution that is simply the area of the sky the detector sees at any instant. The limit to the resolution is established by the effective aperture of the telescope: the higher the ratio of aperture to wavelength, the higher the maximum achievable resolution. In most cases data acquired in this way are converted into contour maps on which the lines connect points of equal brightness in the same way that lines on a topographic map connect points of equal elevation. Recently with computer assistance it has become fairly easy to convert such data into color-coded two-dimensional pictures.

Before proceeding to the central parsec of the galaxy I shall describe briefly what has been learned from radio and infrared studies of the extended nuclear region in which it lies. Radio contour maps of the region have been made at many wavelengths. All the maps reveal a distribution of radiation that is elongated along the galactic equator. The position of the strongest peak, known to radio astronomers as Sagittarius *A*, is also common to all the maps, and it has long been suspected of coinciding with the galactic center. Sagittarius *A* must have contributed to the steady hiss detected by Jansky, although his instrument lacked the resolution to distinguish it from other peaks nearby.

The sources of emission that appear on radio maps of the galactic center are the clouds of hot interstellar gas known as H II regions. These regions, which are common in our galaxy, consist of ionized atoms, that is, atoms that have been stripped of one or more of their electrons. The atoms, mostly those of hydrogen, are ionized by the ultraviolet radiation from nearby stars (or other sources) that are much hotter than the sun. The positively charged atomic fragments and the negatively charged free electrons in an H II region emit radio waves over a broad range of wavelengths when they are accelerated, for example during near-collisions with other charged particles. In contrast, electrically neutral gas, such as the H I regions (which consist largely of hydrogen in atomic form) or molecular clouds (which consist largely of hydrogen in molecular form), emits little radio energy except at a few discrete wavelengths.

A different picture of the extended nuclear region emerges from observations



RADIATION SOURCES in the galactic nucleus include stars, heated dust and the clouds of ionized gas known as H II regions. The term H II describes hydrogen atoms that have been dissociated into protons and electrons by the ultraviolet radiation from nearby hot stars. The electrons and protons emit radiation over a broad band of wavelengths (the continuum radiation) as they are accelerated by electrical forces, primarily those they exert on one another. The curve shows how much continuum radiation survives passage through the interstellar dust to reach instruments on the earth. Before radio and infrared astronomy the galactic nucleus was inaccessible to observation.

made in the infrared at a wavelength of 2.2 microns, roughly four times longer than the visible wavelengths. Whereas the radio emission comes from hot interstellar gas, the 2.2-micron radiation is predominantly starlight. Most of the 2.2-micron radiation comes from red giants, which are large, luminous, highly evolved stars with surface temperatures as low as 2,000 degrees Kelvin, compared with 6,000 degrees K. for the sun. Although at most only a few percent of the stars in the galactic center are likely to be red giants, these objects should be fairly evenly distributed among the younger and brighter stars, and therefore in principle they should provide a good indication of the distribution of all stars in their neighborhood.

A color-coded picture of the extended central region of the galaxy at 2.2 microns appears at the bottom of page 61. Some of the patchy appearance of the picture is accounted for by variations in the amount of interstellar dust between the solar system and different portions of the galactic center. The brightest pointlike objects in the picture are probably red giants and are undoubtedly foreground objects. This inference is supported by their fairly uniform distribution across the picture.

Even with the foregoing exceptions it is clear that the 2.2-micron emission, like the radio emission, is concentrated along the galactic equator. The direction of maximum brightness at 2.2 microns coincides with a bright radio peak within Sagittarius *A* known as Sagittarius *A* West. Maps at longer infrared wavelengths, those emitted by the interstellar dust itself, also have their brightest peaks near Sagittarius *A* West. Hence the concentrations of stars, hot gas and cooler dust are all highest in the same direction. There can be little doubt that in this direction lies the actual center of the galaxy.

From the 2.2-micron observations, which were made by Eric E. Becklin and Gerry Neugebauer of the California Institute of Technology and the Hale Observatories, one can estimate the mass of stars at the center of the galaxy. Becklin and Neugebauer found that the strength and distribution of the 2.2-micron emission from the center of the Andromeda galaxy, designated M31, is comparable to the emission from the center of our galaxy. Since the nucleus of M31 can be clearly photographed at visible wavelengths, one can estimate with fair accuracy the number of stars that must be represented. By assuming that the ratio between infrared radiation and visible radiation is the same for the center of our galaxy as for M31, Robert H. Sanders and Thomas Lowinger of Columbia University and J. H. Oort of the University of Leiden calculated the distribution of stellar mass in the center of our galaxy. They estimate that a sphere with a diameter of one parsec centered on the bright central peak of the 2.2-micron emission contains a mass in stars equivalent to about two million suns.

The most densely packed globular cluster of stars outside the galactic nucleus has about 10,000 stars within its central parsec. Within one parsec of the sun there are no other stars. Assuming that each star in the central parsec is similar to our sun, the above estimate that the average distance between stars in the central parsec is about one light-week, or roughly 100 times the distance from the sun to the planet Pluto. If the earth were within the central parsec, the millions of nearby stars would bathe the planet with an amount of light equal to that of several hundred full moons. Whether life could exist within the radiation-filled central parsec is another question.

In observing a region as small and as distant as the central parsec of the galaxy it is desirable to work with instruments that have the highest possible angular resolution. The largest radio telescopes operating at short wavelengths are limited to a resolution of about one minute of arc, a resolution low enough to blur completely any spatial structure in the central parsec. By using two or more radio telescopes separated by a large distance and employing the technique of interferometry, however, the resolution can be increased substantially. Unfortunately the southerly location of the galactic center has made it almost inaccessible to interferometry with the large radio telescopes that until recently existed only at high northern latitudes. The situation will improve quite soon, however, with the completion next year of the Very Large Array radio facility near Socorro, N.M. This instrument will finally make it possible to explore the central parsec with an angular resolution comparable to that now attainable at infrared wavelengths.

The high resolution made possible at radio wavelengths by interferometry has already been applied in a limited way to the study of Sagittarius *A* West. In 1974 Bruce Balick, then of the Lick Observatory, and Robert L. Brown of the National Radio Astronomy Observatory in Green Bank, W.Va., discovered by radio interferometry that Sagittarius *A* West has near its center a radio-emitting object that is both exceedingly small and exceedingly bright. This compact radio source has since been shown to have a central core whose diameter is no larger than .001 arc-second, or barely 10 times the distance from the earth to the sun (assuming that the object is indeed at the center of the galaxy).

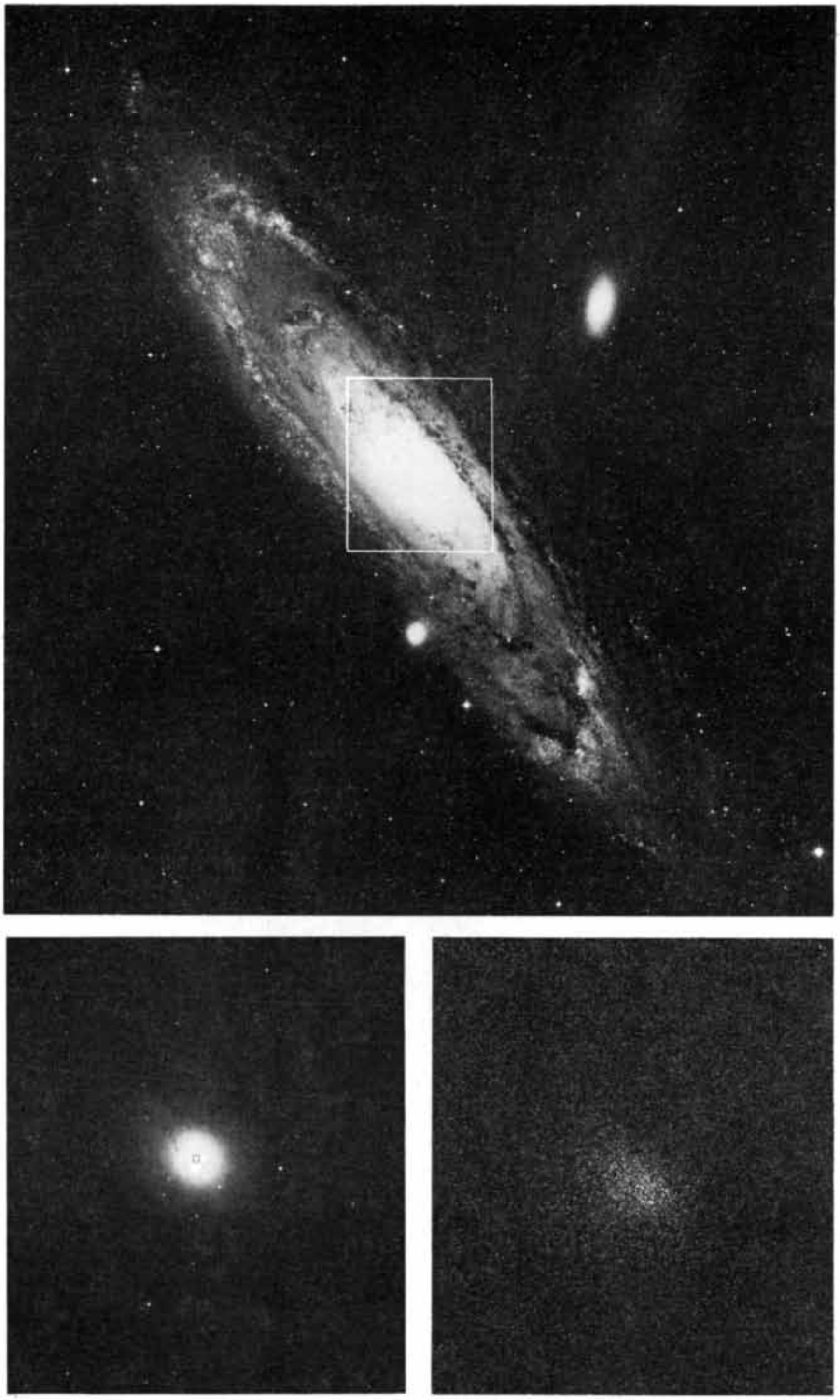
Although the compact radio object bears some resemblance to certain kinds of radio-emitting stars, it also meets some of the qualifications for the way matter should appear as it is drawn into a black hole. It is thought that a black hole may be created when a massive star that has exhausted its nuclear fuel can no longer generate the internal pressure needed to counterbalance the inward pull of gravity. Under such circumstances the interior of the star will collapse until it is so small and dense that known physical laws cannot describe it. The gravitational force near such a collapsed object is so great that there is a region surrounding it from which essentially nothing, not even light, can escape; the region is a black hole.

Material in the strong gravitational field just outside a black hole will be accelerated to high velocities and heated by collisions to very high temperatures. As a result a black hole should be surrounded by a small but intense disklike source of radio radiation, looking rather like the compact radio source. At the galactic center, where gas and dust are abundant, one would expect a black hole to swallow up matter at a comparatively high rate, growing rapidly in mass and in gravitational influence. The possibility that Sagittarius *A* West harbors a massive black hole at the center of the galaxy has undoubtedly entered the thoughts of many astronomers. I shall examine recent evidence for this possibility as we proceed.

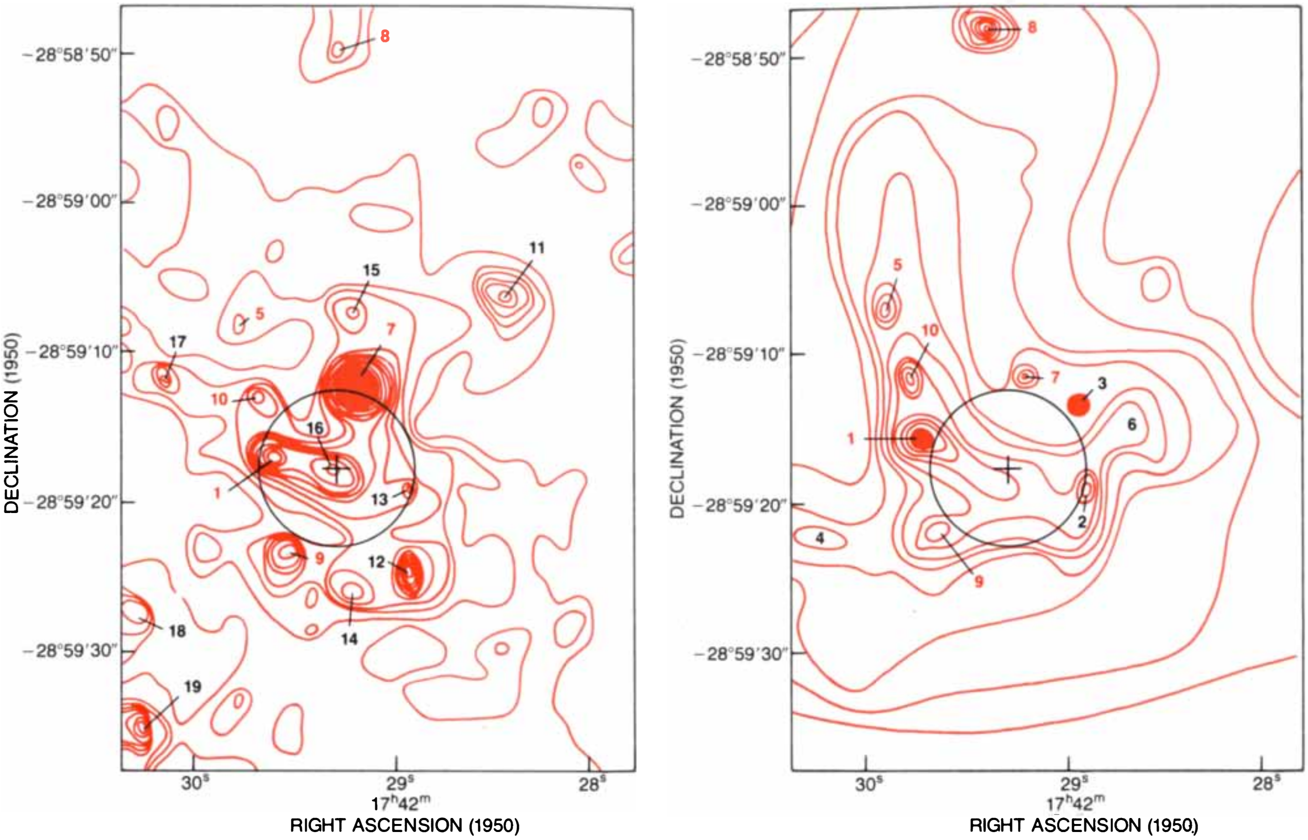
The detailed examination of the central few parsecs of the galaxy has been much easier at infrared wavelengths than in the radio region. The study has been carried out at several infrared wavelengths over the past half-dozen years, principally by G. H. Rieke and Frank J. Low of the University of Arizona and by Becklin and Neugebauer. At each wavelength they have found a bright concentration of objects within the diameter of one parsec coincident with Sagittarius *A* West. At different wavelengths, however, the relative luminosity and distribution of the objects are strikingly different.

The observations at 2.2 microns and 10 microns perhaps best demonstrate the differences [*see illustration on opposite page*]. At each wavelength many distinct objects can be observed, but only six of them show up at both wavelengths. In other words, objects prominent at one wavelength are not prominent at the other. For example, the object known as IRS 1 (infrared source 1), which is the brightest object at 10 microns, is only moderately bright at 2.2 microns. IRS 7, by far the brightest object at 2.2 microns, is barely detectable at 10 microns. It would therefore seem that the objects being observed at different wavelengths belong to different classes.

What kinds of object are they? All are brighter in the infrared than the great majority of stars would appear to be at a distance of 10,000 parsecs. Since, as I have mentioned, red giants are particu-



NUCLEUS OF A NEARBY SPIRAL GALAXY, the Great Nebula in Andromeda, also known as Messier 31 (M31), can be photographed at visible wavelengths even though it is 70 times farther away than the nucleus of our own galaxy because there is little intervening dust to block the passage of light. In M31, as in our spiral system, dust is concentrated along the plane of the galactic disk. The two galaxies are oriented in space in such a way that the nucleus of M31 is seen unobstructed. In the photograph at the top, which shows the entire galaxy, the central region is heavily overexposed. The region within the rectangle is shown in the photograph at the bottom left, a one-minute exposure with the 200-inch Hale telescope on Palomar Mountain. Although most of the stars that nearly fill the area within the rectangle in the top picture are now invisible, the myriads of stars tightly packed within the central 100 parsecs are still overexposed. The photograph at the bottom right shows the central 20 parsecs of the nucleus, the area enclosed within the rectangle of the second picture. The fuzzy region in the middle of the third picture, roughly five parsecs across, represents the light from millions of stars. The white dots elsewhere are not stars but grains of photographic emulsion, which are much enlarged. The photograph was made with a 20-centimeter balloon-borne telescope by E. S. Light, R. E. Danielson and Martin Schwarzschild of Princeton University. Infrared observations suggest that stars in the nucleus of M31 and in the nucleus of our galaxy are similarly distributed.



INFRARED MAPS of the galactic center plotted at two different wavelengths, 2.2 microns (*left*) and 10.6 microns (*right*), bring out different features. The circle on each map represents what is thought to be the central parsec of the galaxy. Only six of the most prominent infrared sources appear on both maps: 1, 5, 7, 8, 9 and 10. Some of the sources that are bright at 2.2 microns are almost certainly red giants, stars that are both cool and extremely luminous. IRS (for infrared source) 7, IRS 11 and IRS 12 are red giants. Except for IRS 3 and IRS 7 most of the sources on the 10.6-micron map are thought to be warm dust clouds associated with H II regions. The cross on both maps coincides with the bright radio-emitting object within Sagittarius *A* West, which conceivably surrounds a black hole. The 2.2-micron map was made by Becklin and Neugebauer with the 200-inch telescope. The 10.6-micron map is a composite of several maps obtained with various telescopes by G. H. Rieke of the University of Arizona and C. M. Telesco and D. A. Harper of the University of Chicago.

larly bright at 2.2 microns, it is natural to ask whether any of the bright 2.2-micron sources at the galactic center are red giants. There is a way to answer the question. In the atmospheres of the coolest stars molecules are able to form. One species that forms in abundance is carbon monoxide, which absorbs radiation beginning at 2.3 microns and extending to longer wavelengths. The absorption occurs most strongly in luminous red giants. In 1975 observations made by groups at the University of Arizona and at Cal Tech revealed absorption dips at 2.3 microns in the spectrum of three of the galactic-center sources (IRS 7, IRS 11 and IRS 12), showing that each source consists of at least one luminous red giant. On the other hand, one of the sources that is bright at 2.2 microns, IRS 16, shows little or no carbon monoxide absorption. The identification of IRS 16 is still open to question. Its position seems to coincide with the compact radio source in Sagittarius *A* West, so that it is of particular interest.

A second object that does not show carbon monoxide absorption, IRS 1, belongs to the class of objects that are much brighter at 10 microns than at 2.2. Very little of the 10-micron emission from objects of this class can be direct stellar radiation because even the coolest stars are much fainter at 10 microns than at 2.2. Infrared astronomers have now found many sources scattered around the galaxy that are unusually bright at 10 microns. In almost every case the source of the radiation has been identified as warm dust. The dust particles absorb visible and ultraviolet radiation from nearby stars or other luminous objects and reradiate the absorbed energy at infrared wavelengths. Something similar probably happens in the galactic nucleus, and one can assume that the peaks in the map at 10 microns represent dust clouds in or near the central parsec.

The interstellar gas in our galaxy is about 100 times more abundant by mass than the interstellar dust, and its presence is always implied by the presence of dust. It therefore seems reasonable that the radio-emitting gas of Sagittarius *A* West is gas associated with dust clouds and that it is heated by the same source or sources warming them. If this is the case, one would expect high-resolution radio maps of Sagittarius *A* West to reveal structure similar to that seen in the maps at 10 microns. There were indications in 1974 from radio-interferometry studies that Sagittarius *A* West is indeed a collection of more compact clouds of ionized gas. At the time, however, it was difficult to determine whether or not any of the emission peaks in the infrared and radio maps were spatially associated.

The story up to this point approximately covers the state of knowledge about the galactic nucleus in 1975. It was known that within the central parsec there was a massive concentration of stars whose visible light is hidden by dust. Within the central few parsecs there had been found infrared-emitting sources of several types. Of the sources that were bright at 2.2 microns some had been shown to be cool and luminous red giants; most of the other sources might also harbor cool stars but they had not yet been identified. The majority of the objects that were bright at 10 microns resembled compact H II regions, but although it was known that ionized gas was present at the galactic center, it was not yet known whether the gas was physically associated with any of these infrared sources. An intriguing source of radio emission, which might be a supermassive object at the very center of the galaxy, had been found to coincide with a peculiar bright source that emits strongly at 2.2 microns. Finally, nothing whatever was known about the motions of all these objects; the material in the central parsec might be collapsing, expanding, rotating or moving randomly.

The stars, gas and dust at the galactic center emit radiation whose variation in intensity with respect to wavelength is fairly smooth; the radiation is therefore called continuum radiation. Another potentially rich source of information is line radiation: the radiation emitted and absorbed at specific and characteristic frequencies by atoms and molecules. The spectrum of an astronomical object usually consists of both a smooth continuum and lines. Each line is associated with a specific energy level of an atom or a molecule and has an intensity, a central wavelength and a width.

The intensity of the line is a measure of the temperature and the abundance of the emitting substance. If the source is moving toward the observer, the line is shifted to a shorter wavelength; if the source is receding, the line is shifted to a longer wavelength. Such Doppler shifts inform the astronomer about radial velocity: the velocity of a source along the line of sight. Velocities toward the observer ("blue shifts") are termed negative; those away from the observer ("red shifts") are termed positive. When the source atoms or molecules are moving at varied speeds, the shifts vary. The line is "smeared out" and is said to have a width. Therefore the width of a line can reveal the range of radial velocities in the source.

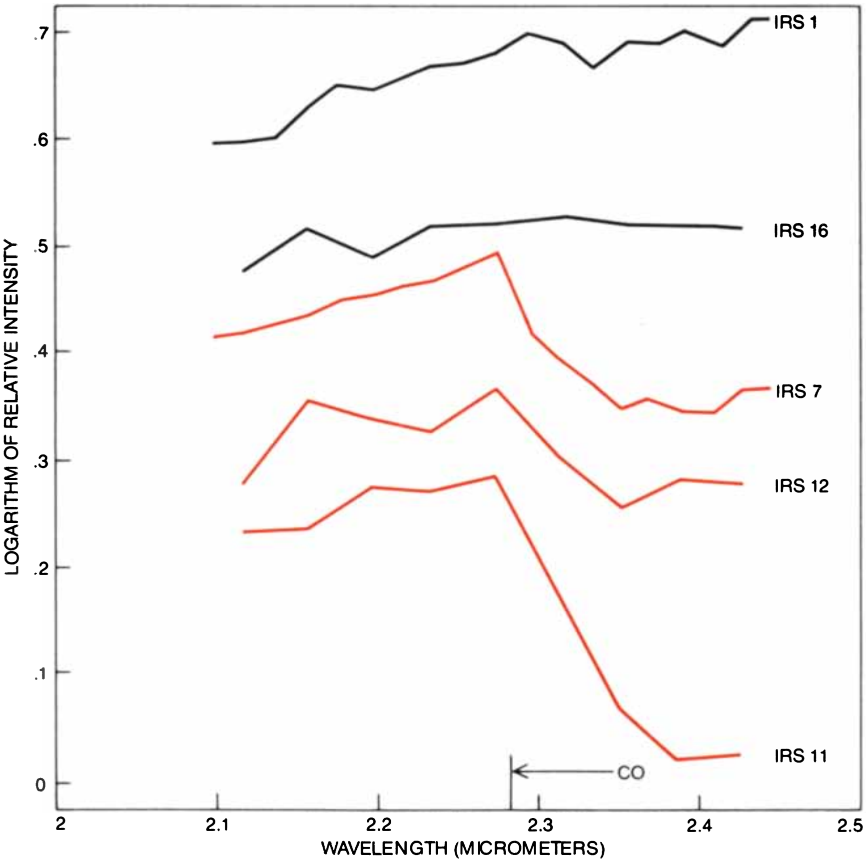
In 1973 David Aitken, Barbara Jones and James Penman of University College London succeeded in detecting for the first time an infrared emission line originating from the galactic center. The line's wavelength near 12.8 microns identified its source as being singly ionized neon, or Ne II: neon missing one of its 10 electrons. The same neon line had earlier been found to be one of the brightest infrared emission lines in several ionized regions elsewhere in the galaxy. These regions also reveal themselves by continuum radiation in the radio region of the spectrum.

Using the known intensity of radio radiation from Sagittarius *A* West and assuming that the abundance of neon in the ionized gas of Sagittarius *A* West is normal, Aitken, Jones and Penman calculated the expected intensity of the neon line. The observed intensity of the line was slightly greater than their predicted intensity, which implied that the gas at the galactic center is richer in neon than gas observed elsewhere. The prediction was close enough, however, to imply that the same ionized gas produces both the infrared neon line and the radio continuum emission. Hence the London work showed that the 12.8-micron line of Ne II, representing only a tiny fraction (approximately .01 percent) of the ionized gas, could serve as a tracer in studying the motion and distribution of the ionized gas of Sagittarius *A* West, which was putatively at the very center of the galaxy.

The London group worked with an infrared instrument that admitted a circular beam whose diameter corresponded roughly to one parsec at the galactic center. The resolution of their spectrometer was too low to determine the width of the neon line or whether it exhibited a Doppler shift. In the department of physics at the University of California at Berkeley, where I was then a graduate student, a succession of students working with Charles H. Townes had been developing infrared spectrometers capable of high spectral resolution. In 1975 Eric Wollman, John Lacy and I, together with Townes and David Rank of the University of California at Santa Cruz, measured the shape of the Ne II line with the 1.5-meter telescope of the Cerro Tololo Inter-American Observatory in Chile (where Sagittarius passes directly overhead). What we found made it clear that Sagittarius *A* West is not a typical galactic H II region. The neon line had a width of several hundred kilometers per second, whereas for an H II region velocities in the range of 30 kilometers per second are normal.

At about the same time radio observations of emission lines generated by the recombination of hydrogen nuclei and electrons made by Thomas Pauls, Peter G. Mezger and Edward Churchwell of the Max Planck Institute for Radio Astronomy in Bonn also showed that the ionized gas in Sagittarius *A* West was in rapid motion. Whether the motion was organized or turbulent and how the neon and other ionized gas was distributed within the central parsec had not yet been determined.

The answers to these and other questions called for observations of the 12.8-micron line on a finer spatial scale. At the Lick Observatory in 1975 and 1976 we observed several positions within the central parsec and found that both the line strength and the radial velocity varied with position. In 1977 Lacy completed a new and more sensitive spectrome-



INFRARED SPECTRA OF FIVE SOURCES close to the galactic center show that three, IRS 7, IRS 11 and IRS 12, carry the "signature" of red giants: a drop in intensity at wavelengths longer than 2.3 microns. Such stars have an atmosphere rich in carbon monoxide, which strongly absorbs radiation at 2.3 microns and beyond. IRS 1, an H II region, and IRS 16, in Sagittarius *A* West, do not exhibit such absorption. The spectra were obtained with the 200-inch telescope by Becklin, Neugebauer, Steven V. W. Beckwith, Keith Matthews and C. G. Wynn-Williams.

ter that made it possible to observe the Ne II line in even greater detail. In 1977 and 1978 Lacy, Townes, Fred Baas (visiting from Leiden) and I, working with this instrument at the new 2.5-meter Du Pont telescope at the Las Campañas Observatory in Chile, got complete coverage of the Ne II line over the central parsec and beyond it. Working with aperture diameters corresponding to as little as .2 parsec at the galactic center, we obtained spectra on a rectangular grid of positions and made contour maps of the line intensity at several different velocities.

Our data show that there is great diversity in both the intensity of the neon line and the velocity of the neon observed in the direction of the central parsec. In a few directions the intensity of the line exhibits peaks at more than one velocity. Sagittarius *A* West clearly cannot be considered a uniform region of ionized gas but rather must be composed of a number of independent smaller sources moving with respect to one another. As I have mentioned, much of the infrared continuum radiation at 2.2 and 10 microns also comes from groups of small sources. It was important to determine whether the distribution of ionized neon was related to either the 2.2- or the 10-micron maps.

In order to investigate this question we noted the positions of maximum intensity of the Ne II line at specific velocities. When we plotted the positions of the velocity components on the infrared-continuum maps, we found that the maximum intensities coincided with many of the sources that were bright at 10 microns. For example, IRS 1 is at the same position as a peak in the neon line at zero velocity. Near IRS 2 there is little zero-velocity ionized neon, but there is a sharp peak in the ionized neon approaching the observer at about 280 kilometers per second, a velocity not seen elsewhere in the central parsec. At least five other infrared sources, IRS 4, IRS 5, IRS 6, IRS 9 and IRS 10, coincide with localized velocity components of the Ne II line.

The spatial coincidences of hot ionized gas and warm dust are convincing evidence that most of the bright sources on the 10-micron continuum map are compact H II regions within Sagittarius *A* West. From the intensity of the neon line we can estimate that the ionized material in each of these regions amounts roughly to one solar mass. The mass of the associated dust, which is responsible for the 10-micron continuum emission, is much less than that. The diameter of a compact H II region is typically a quarter of a parsec. Since there are about two million stars in the central parsec, it is likely that each compact H II region harbors as many as 100,000 stars.

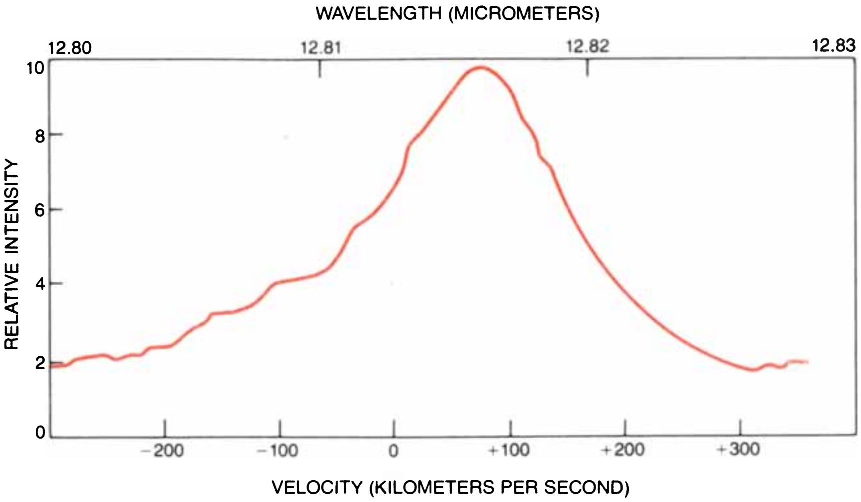
Only two of the sources on the 10-micron map are definitely not H II regions. One of them, IRS 7, is already known to be a red giant. The other, IRS 3, is more mysterious. IRS 3 cannot be a red giant because it is not bright at the shorter infrared wavelengths. Nor can it be an H II region because there is a distinct drop in the intensity of the Ne II line in the vicinity of the source. This suggests that IRS 3 obscures part of the ionized gas. Perhaps IRS 3 is a cool and opaque cloud of dust and gas, either recently ejected by a star or collapsing gravitationally to become a star, in front of the central parsec.

Each of the compact H II regions in the central parsec is moving at a velocity whose radial component is given by the Doppler shift of its associated neon line. Most of the H II regions are well separated in space or in velocity and therefore must be independent objects. Nevertheless, study of all the neon spectra indicates that there is a collective motion of the central parsec. This is seen most easily by noting that a line running roughly north-south through the middle of the region separates the ionized neon that is predominantly approaching from the ionized neon that is predominantly receding. The simplest interpretation of this division of velocities is rotation; the dividing line would then correspond to the axis of rotation.

Although evidence for rotation has been found so far only in the ionized gas, it seems reasonable that the rest of the material in the central parsec (including the stars) is moving in much the same way as the ionized gas. On the basis of the Doppler shifts of the neon line observed a few tenths of a parsec from the apparent axis one can estimate that the time required for gas, dust and stars inside the central parsec to make one trip around the galactic center is roughly 10,000 years. (The sun takes about 200 million years.) The orientation of the axis is peculiar: it is almost perpendicular to the axis of rotation of the main galactic disk. This suggests that the nuclear region is a separate structure, distinct from its surroundings. It is not yet known how far out from the center the peculiar axis of rotation extends.

The motions of the ionized gas in the central parsec, whether or not they are part of a collective rotation, are influenced by the gravitational force of the mass within the same volume and can be utilized to calculate that mass. Since the rest of the galaxy is distributed more or less symmetrically around the center, with most of its mass a great distance from the center, it exerts almost no net gravitational force on the central parsec. The total mass of the central parsec, as estimated from the neon-line velocities, is between five and eight million times the mass of the sun. Of that total it appears from study of the 2.2-micron observations that only two million solar masses are in the form of stars of all types. The total mass of dust and gas within the region is in comparison negligible: probably no more than about 10 solar masses.

It is therefore fascinating to speculate whether the difference of from three million to six million solar masses between the apparent stellar mass and the



EMISSION OF SINGLY IONIZED NEON, Ne II, present in gas clouds in the central parsec of the galaxy, was studied by the author and his colleagues with an infrared spectrometer attached to the 1.5-meter telescope at the Cerro Tololo Inter-American Observatory in Chile. Examination of the Ne II emission line near 12.8 microns reveals that the line has been substantially broadened by motion of the gas clouds in which the neon is a minor constituent. Motion toward the observer (designated as negative) shifts the line to a higher frequency, or shorter wavelength, whereas motion away from the observer (designated as positive) has the opposite effect. The observed velocities of ionized neon range from roughly −200 to +200 kilometers per second along the line of sight, providing evidence of strong gravitational forces near the center of the galaxy. The intensity of continuum radiation near 12.8 microns from the warm dust in the nucleus is given by the level of the radiation on each side of the broadened neon line.

total mass is significant. Does it represent missing mass still to be accounted for? Or are the estimates inaccurate? One observation may be pertinent. Measurements of the infrared neon line and of the radio lines generated by the recombination of hydrogen nuclei and electrons show that within the central few parsecs the gas velocities tend to increase toward the center. It has been pointed out by Luis Rodriguez and Eric J. Chaisson of the Center for Astrophysics of the Harvard College Observatory and the Smithsonian Astrophysical Observatory, and by others as well, that such an increase could not be due to the gravitational force of the central stellar cluster unless the cluster had an unusually dense core or possibly included a compact, massive and presumably nonstellar central object.

In considering theoretical models to account for the observations of the central parsec let us assume (at least initially) that the nucleus is not being observed at a particular epoch. The observations have shown that the central parsec contains an extremely dense cluster of stars, together with a sizable amount of dust and ionized gas, which are distributed nonuniformly and are in rapid motion. A satisfactory model should be able to explain not only the presence of such features but also their continued presence.

Because of the gravitational attraction toward the center of the galaxy it is not surprising to find the center populated by a dense cluster of stars. Occasional newcomers to the cluster are probably ordinary stars that had previously circled the galactic nucleus at some greater distance and that have lost angular momentum, perhaps as a result of near-collisions with other stars. Conversely, through the same process stars within the nucleus may occasionally acquire enough velocity to escape from it. Hence the stellar population of the central parsec probably remains close to a steady state or at most changes very slowly.

In contrast, dense concentrations of dust and ionized gas are normally short-lived. In the outer parts of the galaxy there are two classes of such objects: the H II regions found in the vicinity of newly formed hot, luminous stars, and the planetary nebulas, which are also H II regions but whose ionization is due to certain old stars that are usually hotter but less luminous. In fact, ultraviolet radiation from roughly 100 stars of the type that ionize planetary nebulas would be needed to account for the amount of ionized gas observed in the central parsec, but as we have seen only about 10 ionized clouds are observed there. An identification of the clouds as planetary nebulas can be questioned for this reason. On the other hand, only about 10 highly luminous, newly formed stars of the type that ionize normal H II regions would be needed to do the job. Other typical sources of dust and gas, such as the "wind" of particles emitted by stars or the infall of matter from surrounding regions, would not give rise to the highly nonuniform distribution of interstellar material inferred from the infrared-continuum and neon-line observations.

In a region as densely populated with stars as the central parsec is, however, an additional mechanism for producing dense clouds must be considered: stellar collisions. A disruptive collision, one in which two stars actually make contact, will tear material out of one or both of them, creating a turbulent and initially dense cloud that disperses into the more rarefied surroundings. Calculations by Townes, Lacy and David Hollenbach indicate that within the central parsec collisions in which at least one participant is a red giant are an important source of interstellar material. The clouds that would be created by such collisions have sizes, velocities, internal motions and lifetimes that are consistent with the properties of the clouds actually observed. Moreover, the estimated collision rate may be adequate to yield clouds in the appropriate number.

The next question is: What is the source of the ionizing radiation needed to convert the clouds in the central parsec into H II regions? I have mentioned hot, luminous stars and planetary nebulas as possible ionizing agents. Hot, luminous stars remain that way for only a few million years, and so they cannot have traveled far from where they are now observed, which is in the galactic nucleus itself. Star formation, however, is thought to be a delicately balanced process involving the contraction of large clouds of gas and dust over a period of roughly a million years. In the turbulent environment of the galactic nucleus, where orbital periods may be as short as 10,000 years, it may not be possible for large volumes of gas and dust to contract into stars before they are thoroughly disrupted.

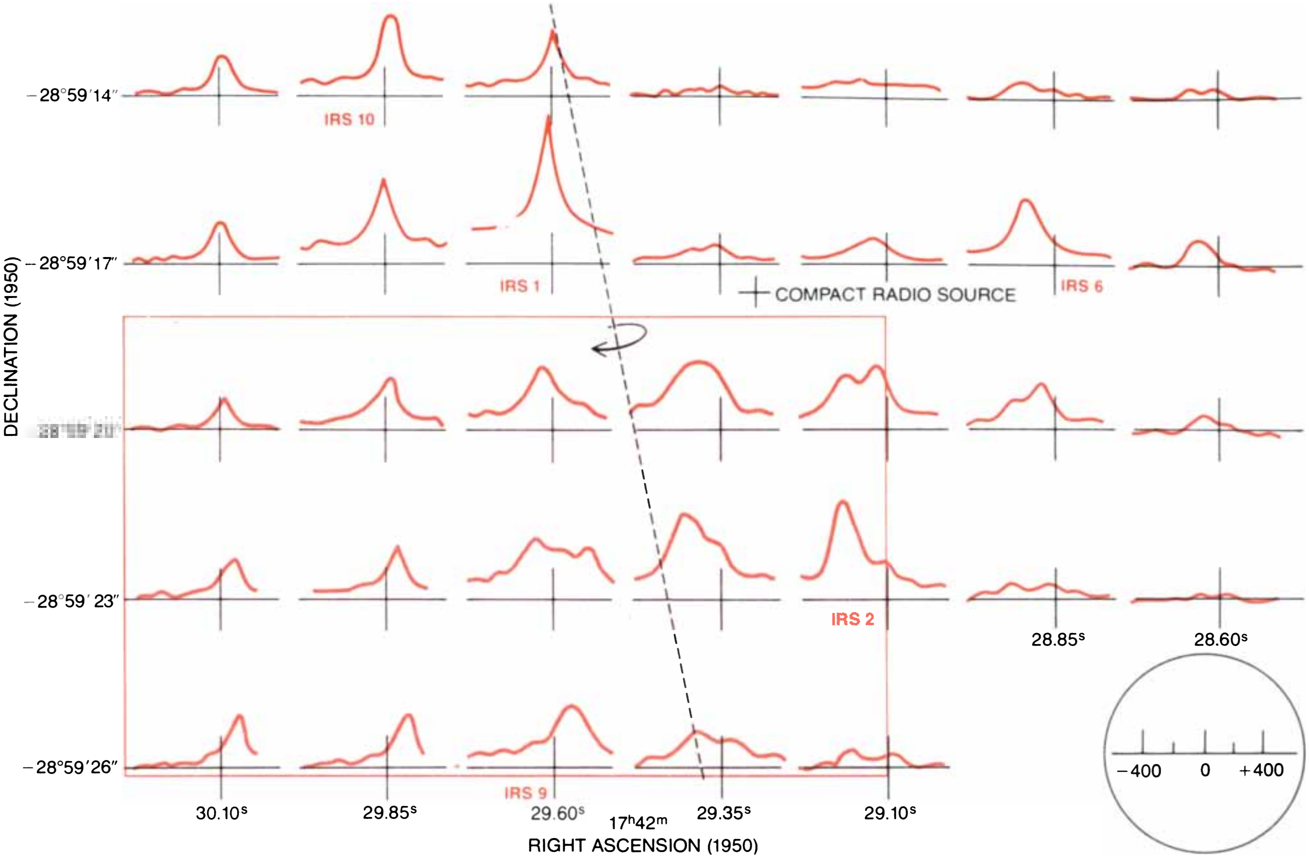
The most important constraint on the type of ionization source in the central parsec appears to be the uniformly low level of ionization in the region. Typical H II regions elsewhere in the galaxy usually contain, in addition to singly ionized neon, atomic species that are multiply ionized, that is, lacking two or more electrons. In order to create the more highly ionized states very energetic ultraviolet photons from extremely hot stars are needed.

In the course of our observations of Ne II we also searched intensively within the central parsec for spectral lines of doubly ionized argon (Ar III) and triply ionized sulfur (S IV) at their known infrared wavelengths. These lines are often quite strong in H II regions outside the galactic nucleus. The results at various positions in the central parsec were negative, with the exception of a very weak Ar III line detected in the direction of IRS 1. Recently a strong infrared line of Ar II (singly ionized argon) from the central parsec was detected by Steven P. Willner, Ray Russell and Richard Puetter of the University of California at San Diego, Thomas Soifer of Cal Tech and Paul Harvey of the University of Arizona. Their result implies a moderate overabundance of argon in the central parsec and, when it is combined with our result, makes it clear that in the central parsec the ionization state is unusually low.

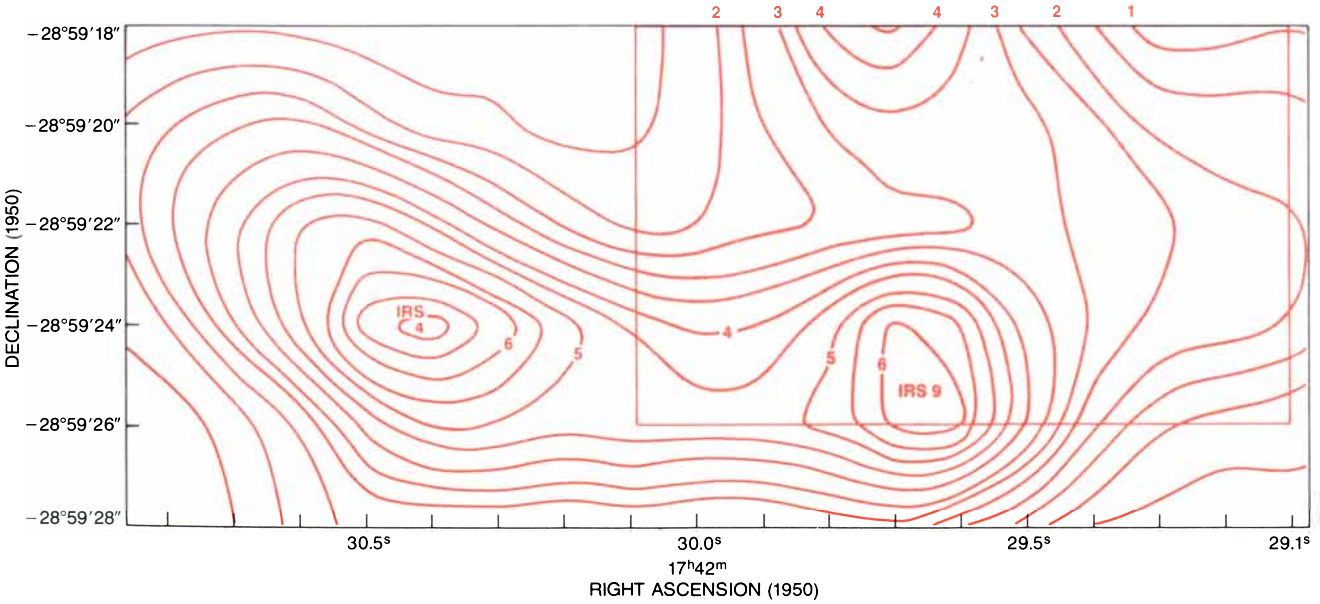
The central stars of planetary nebulas, which are usually very hot, would seem to be ruled out as the source of ionizing radiation because most of them give rise to high ionization states. The situation is also peculiar, however, if several hot, luminous, recently formed stars are the principal source. Unless the most massive and hottest stars are somehow inhibited from forming at the galactic center, or unless those that formed most recently have cooled significantly, one would expect the ionizing stars to exhibit a wide range of temperatures. Such a temperature range would lead to variations in the ionization state of the gas throughout the central parsec. The absence of Ar III and S IV throughout the entire central parsec seems to imply either that all the ionizing stars are nearly identical or that there is one extremely luminous but relatively low-temperature ionizing source in the central parsec. An object of the latter type could not be an ordinary star.

Such considerations lead to a rather exotic model of the central parsec: a model dominated by a massive black hole at the center. Material being swallowed up by a black hole is thought to be first stored outside it in a rotating disk, which under the right conditions is expected to radiate strongly in the ultraviolet. The uniformity of the ionization state of the central parsec might be explained by the presence of such an object. A natural candidate for the black hole is the compact radio object, with a diameter of less than .001 arc-second, in Sagittarius *A* West.

The expected properties of massive black holes have been calculated by many theoretical astrophysicists. The spectrum and the intensity of the ionizing radiation from the disk depend only on the mass and the accretion rate of the black hole. Townes, Lacy and Hollenbach have shown that quantitative agreement with the observed luminosity and ionization state of the H II region in the central parsec of the galaxy would be provided by a black hole that had a mass of a few million suns and accreted matter at the moderate rate of about 10^{-5} solar mass per year. The necessary mass of accreting material could be sup-



DETAILED STUDY OF THE NEON LINE illustrates the complexity of motion within the central parsec of the galaxy. The grid of spectra is arranged according to position in the sky and includes slightly more than the central parsec. In the lower right-hand corner a circle gives the beam size of the spectrometer; a scale gives the velocity of gas motion in kilometers per second. For each spectrum the position of origin coincides with zero line-of-sight velocity, with zero intensity of emission and with the position to which the telescope was pointed. Several of the localized velocity components of the neon line coincide with sources that are bright at 10 microns (IRS 1, IRS 2, IRS 6, IRS 9 and IRS 10), indicating that the ionized gas is physically associated with warm dust clouds. The sloping broken line separates regions of ionized neon that are predominantly approaching (*right*) and receding (*left*). Such a well-defined separation suggests that material in the central parsec is rotating in the sense that is given by the arrow. The region within the rectangle reappears in the illustration below.



INTENSITY OF THE NEON LINE in the southeastern part of the central parsec demonstrates the association of the ionized neon with two specific sources, IRS 4 and IRS 9, whose continuum radiation originates in warm dust clouds. The contour lines show the intensity of neon with velocities of about +125 kilometers per second. The peaks coincide with the two infrared sources. Spectra made at the positions of the sources show that IRS 4 is actually receding at 110 kilometers per second and IRS 9 at 140 kilometers per second. Observations for illustrations on this page were made by the author and his colleagues with 2.5-meter telescope at Las Campañas Observatory.

plied many times over by infall, stellar winds, planetary nebulas and the debris of stellar collisions within the central parsec. Even allowing for some of the material's being blown away by the intense radiation pressure from the accretion disk, from perhaps two million stars and from other luminous objects in the central parsec, there would seem to be ample material to feed the black hole at the required rate.

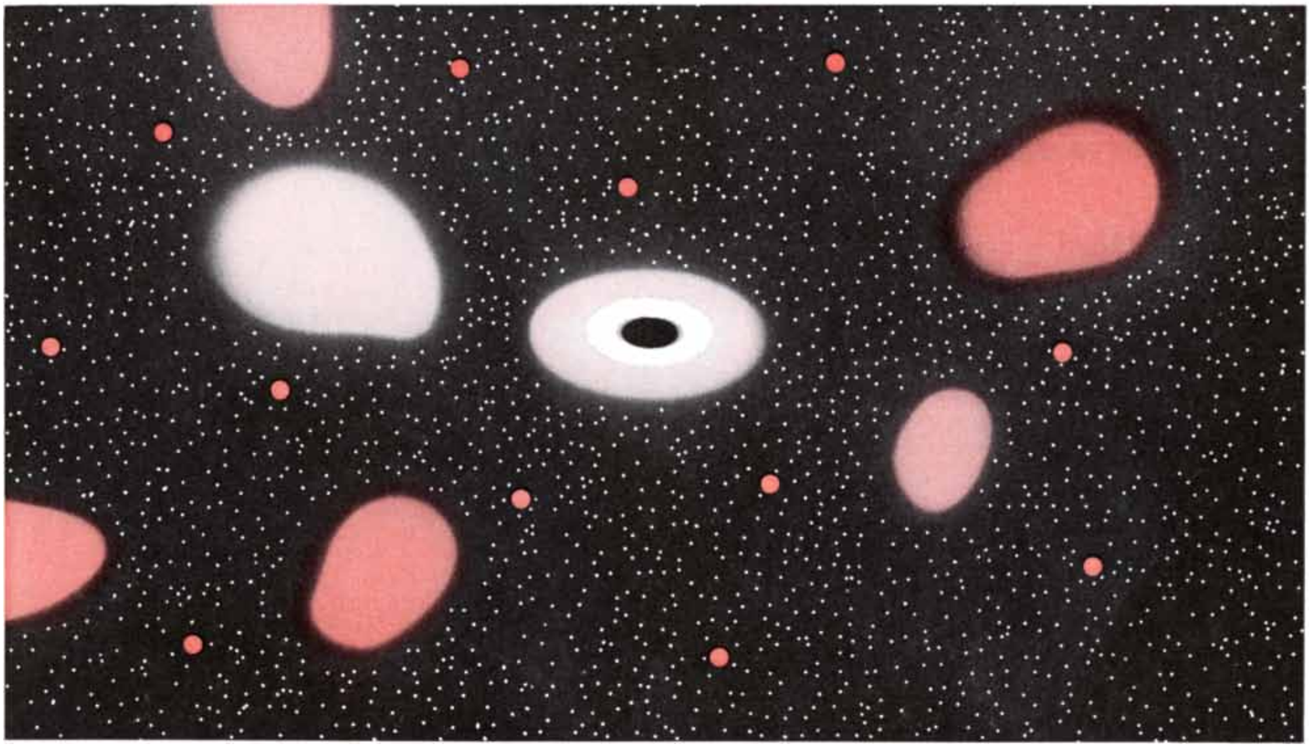
It is perhaps significant that the intensity of infrared radiation emitted by the peculiar object IRS 16, in which the compact radio source appears to be located, is roughly the same as the intensity that would be emitted in the vicinity of a black hole accreting matter at the above rate. Finally, the existence of a black hole with a mass of a few million suns is consistent with the difference between the estimated total mass in the central parsec and the mass of the stars in the same volume. A black hole of that mass fits easily into a volume the size of the compact radio source.

The evidence for the presence of a massive black hole appears reasonable, but it is indirect. More direct tests might come from future radio and infrared observations of the small region around the compact radio sources. For example, the intense gravitational field in the vicinity of such a massive black hole should give spectral lines originating in the source a Doppler width of 1,000 kilometers per second or more. When our infrared spectrometer with a beam diameter of .2 parsec was centered on the compact radio source, a possible location of the black hole, the neon line did not appear to be that wide. A weak and broad component that could not have been detected by our instrument may nonetheless have been present. Significant changes in the accretion rate of the black hole, which may occur on a time scale of only a few years, should eventually give rise to observable changes in the ionization state of the central parsec. Hence continued monitoring of all the infrared emission lines is important.

One might well ask again: Are we really observing the central parsec of the galaxy? In my view the answer is a strong yes. First of all, we are certainly pointing our telescopes in the direction of the central parsec; that is indicated by all the infrared and radio observations, perhaps most clearly by the 2.2-micron observations of the stellar distribution. Second, the rich array of infrared sources observed in this direction has few if any counterparts elsewhere in the galaxy. The high positive and negative velocities obtained from the neon line are unique, and the enormous mass densities implied by the 2.2-micron and neon-line measurements have not been found elsewhere.

It is possible that the central parsec is the source of phenomena that are observed far beyond the galactic nucleus. Sudden large increases in the accretion rate of the proposed black hole, supernovas or bursts of star formation near the galactic center may be the cause of the outward-moving molecular clouds that have been observed by radio astronomers. Such outbursts would eject material from the galactic nucleus in the form of high-velocity clouds. The material eventually would collide with and accelerate the observed molecular clouds, which in turn could be the debris of earlier explosions.

Although the nucleus of our own galaxy is a powerful energy source, it is quite weak compared with many other galactic nuclei. Perhaps that is one reason we are here and able to study the galactic nucleus. At present it may be going through a quiet phase between more explosive epochs. The continued study of the galactic nucleus will lead to a better understanding of it and perhaps also of physical events in the nuclei of more active galaxies.



SCHEMATIC VIEW OF GALACTIC CORE, roughly three parsecs across, postulates the existence of a massive black hole (*shown much larger than scale*) at the very center. The luminous color patches represent clouds of interstellar gas and dust, which are created at the rate of one every 1,000 years or so by stellar collisions within the dense central cluster of stars. Some of the gas and dust, gravitationally pulled toward the black hole, forms a hot, rotating accretion disk before it is swallowed and disappears. Intense ultraviolet radiation from the accretion disk ionizes the gas in the surrounding clouds, which then emit the infrared lines and the radio continuum radiation that have been detected from the direction of the central parsec. Dust in the clouds is also heated by radiation from the accretion disk and nearby stars and emits infrared continuum radiation. Much of the short-wavelength infrared radiation is emitted by a few red giants (*colored dots*) within the dense central star cluster. The white dots in the drawing are normal stars, of which there are about two million.

The new hypersensitization

Despite all that has been learned in recent decades by radio astronomy and x-ray astronomy from above the atmosphere, nobody is advocating that the optics of the great telescopes in the world's observatories be salvaged to make beer bottles.

Time on those magnificent instruments gets ever harder to earn.

At the Royal Photographic Society's tercentenary celebration of the Greenwich Observatory, R. L. Jenkins *(left below)* of the Kodak Research Laboratories in Harrow, England, brought glad tidings of more time to parcel out to the deserving.

Latter-day three-hour exposures put a pinch on telescope time that was never felt when photons from afar had merely to land on the Astronomer Royal's personal retina.

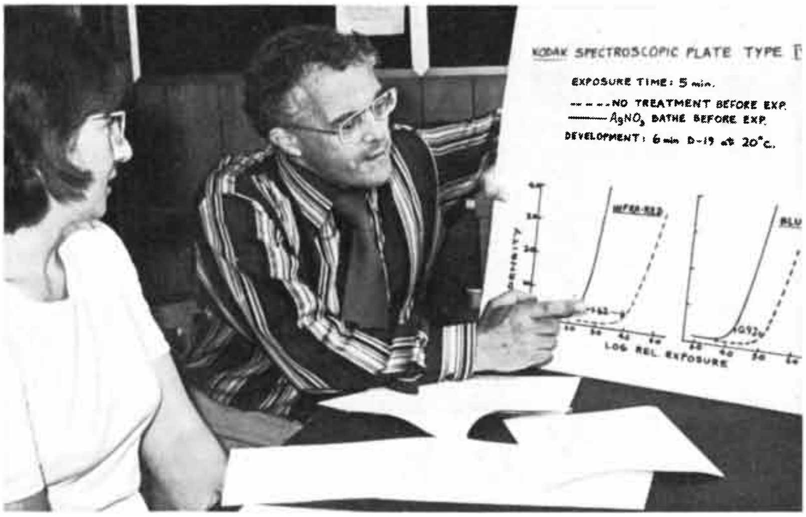
Jenkins' research partner G. C. Farnell shows below that bathing a plate in silver ions just before exposure greatly cuts the time needed to pick up all that's going to be picked up before sky light takes over and obscures the view. The speed enhancement is greatest for the plates we make for infrared astronomy.

Though the treatment must be done very carefully for good results, a dip before exposure into nothing more advanced technologically than .001N aqueous silver nitrate solution markedly increases speed, particularly towards long exposure times. In other words, low-intensity reciprocity failure is greatly reduced.

There is no reason why plates couldn't have been preexposure dipped into $AgNO_3$ before Rosemary Jenkins and Geoff Farnell were born. Astronomers are doing it now. Hindsight! It's so clear and penetrating!

If this is important to you and news to you, and you have to know more, write Kodak, Dept. 55-W, Rochester, N.Y. 14650; ask for Publication P7-670.

Kodak



Creationism Evolves

The creationist assault against the concept of evolution through natural selection, and indeed against the fact of evolution, is being pressed on a wide front and with increasing vigor. The mode of attack has changed. Now the effort is neither to ban the teaching of evolution nor to require the teaching of creation according to Genesis; both approaches have run into trouble with the U.S. Constitution. The effort instead is to obtain state legislation or administrative regulations requiring that courses and textbooks dealing with the origin and development of life and of man present creationism as a "scientific" alternative to evolution. Creationism denies the geological evidence for a world billions of years old, the fossil evidence for the gradual appearance of new species and the biological evidence for the evolution of man from lower primates. The current creationist effort is exemplified by recent events in Iowa and California.

State legislation actually forbidding the teaching of evolution in public schools was finally invalidated on Constitutional grounds only in 1968, when an Arkansas "monkey law" was declared unconstitutional by the U.S. Supreme Court. (The 1925 conviction of John T. Scopes in Tennessee's famous trial had been reversed on technical grounds; the Tennessee law was not repealed until 1967.) Ironically, it was at just about the same time that creationist sentiment was building up anew, for various reasons including the growth of fundamentalist Christian sects, concerns about the erosion of established social values and doubts about the virtues of science and technology (see "The Science-Textbook Controversies," by Dorothy Nelkin; SCIENTIFIC AMERICAN, April, 1976).

A new drive against evolution was launched, led not by theologians but by creationist engineers and scientists, members of the Creation Research Society, whose proclaimed goal was "the realignment of science based on theistic creation concepts" and who were "committed to full belief in the Biblical record of creation and early history." Their first success was the inclusion of language in the "science framework," or guidelines, published by the California State Board of Education in 1970, to the effect that "several theories," including "creation in scientific terms," are necessary to explain the origin of life. There was strong opposition to this "equal time" concept by science teachers, and the language was eliminated in favor of Darwinian concepts in a revision of the framework in 1974 and when a new framework was promulgated in 1978. Now the creationists have filed for an injunction against the Board of Education, seeking withdrawal of the 1978 framework on the ground that it does not contain the creationist language and therefore violates the rights of creationist students and their families.

The plaintiffs are individual creationists and the Creation-Science Research Center, a foundation based in San Diego that handles lobbying and similar public-affairs activities. The headquarters of the creationist movement appears to be the Institute for Creation Research, a division of Christian Heritage College in San Diego. The college, which was established in 1970 under the sponsorship of the Scott Memorial Baptist Church, teaches its students "the foundational importance of special creation in every subject." The college, the institute and the creationist movement as a whole do not seem to be formally supported by any major religious organization; many fundamentalist Protestant groups that might be expected to support a literal interpretation of Scripture tend to shy away from activities that seem to threaten the separation of church and state. The institute nonetheless seems able to mobilize creationists to lobby for legislation and to influence state and local boards of education, many of which have adopted policies at least allowing individual schools or teachers to select creationist instructional materials.

After the creationists' initial California success they opened drives in several states to get the equal-time principle mandated by law. In 1973 legislation was approved in Tennessee requiring that any public-school textbook discussing an opinion or a concept about the "origin or creation of man" must label it "a theory," not "scientific fact," and must give equal attention to "other theories, including, but not limited to, the Genesis account in the Bible." The Bible was declared to be a "reference work" rather than a textbook to exempt it from the "theory" disclaimer, and "all occult or satanical beliefs of human origin" were excluded from the equal-time provision.

Acting on behalf of the National Association of Biology Teachers, Frederic S. Le Clercq, a Knoxville attorney, moved against the law in the Federal courts, and in 1975 a U.S. Court of Appeals declared the statute "unconstitutional on its face." The court held that exempting the biblical account from the disclaimer requirement provided "a clearly defined preferential position for the Biblical version of creation as opposed to any account of the development of man based on scientific research and reasoning" in violation of the First Amendment. Moreover, the exclusion of "occult" or "satanical" beliefs would necessarily lead to government "entanglement" with religion: "the God of some men has frequently been regarded as the Devil incarnate by men of other religious persuasions."

This past winter and spring bills to require equal time for creationist views were introduced in several state legislatures, but with a difference: the language was modified in an apparent attempt to avoid First Amendment problems. The major effort was made in Iowa, where a bill introduced into the state senate required simply that "whenever the origin of human kind or the origin of the earth is taught in the educational program of the public schools of this state, the concept of creation as supported by scientific evidence shall be included." Support for the bill came primarily from creationist students at Iowa State University, where a number of faculty members, including the dean of the College of Engineering, are creationists. The bill was reported out of committee, but when it was pointed out that enforcing such a statute would cost money and that the bill should be referred to the appropriations committee, support waned and the bill did not come to a vote. It may be revived next year.

Coal: Prospects and Problems

The amount of coal mined and burned directly as fuel in the U.S. could double or even triple by the year 2000 without requiring either major technological innovations or a substantial relaxation of current Federal regulatory standards, according to a recent report issued by the Congressional Office of Technology Assessment. Nevertheless, the report warns, the ultimate environmental and social costs of such a development are uncertain and could turn out to be prohibitive.

The 411-page OTA report, *The Direct Use of Coal: Prospects and Problems of Production and Combustion,* was prepared at the request of the House Committee on Science and Technology as one in a series of studies intended to assist Congress in the development of a national energy policy. In this instance the OTA staff was given the assignment of analyzing the benefits and risks of a projected massive shift to coal and away from other fuels over the next two decades.

Although the researchers who worked on the OTA Coal Project were not able to identify "any significant violations of existing environmental standards that inevitably would result from a substantial rise in coal use," they point out in

Using space technology to save lives on Earth: The Perkin-Elmer RMS III

Left on Mars by Viking I and II in 1976, Perkin-Elmer gas analyzers sampled the Martian soil for signs of life. In Skylab, when the astronauts exercised, the same basic instrument analyzed their respiratory condition. In every U.S. nuclear submarine, a similar Perkin-Elmer device stands watch over the purity of the air breathed by the crew.

In 1972, Perkin-Elmer put the technology developed for such space and defense applications to work in the health care field with the introduction of our first Medical Gas Analyzer. Since then, hundreds of

these units have been used for pulmonary testing, physiological and anesthesia research, stress testing and other related applications

The Medical Gas Analyzer, an exceptionally rugged and maintenance-free Perkin-Elmer mass spectrometer, is the heart of our new Respiratory Monitoring System (RMS III). By measuring and monitoring a patient's inspired and expired gases, the system provides the physician with information for improved respiratory and anesthetic management.

Up to 16 patients can be sequentially monitored with one RMS III. A unique rotary valve selects the patient line to be sampled and a small amount of gas is drawn into the mass spectrometer and analyzed. Computation and readout take place at a central monitor, located at the nurses' station, with remote data displays available for each operating room or bedside.

RMS III gives simultaneous inspired and expired values for up to six different gases, and compares them with pre-set levels entered via the system's keyboard. Audio and visual alarms are activated if any of these gases exceeds prescribed limits.

Each patient's breathing can be viewed as a moving wave form on the monitor screen. Eight- or twenty-four-hour trends of various physiological parameters as well as gas concentrations can also be displayed on the screen for each patient.

In many cases, developing problems can be anticipated before they would otherwise be apparent.

Perkin-Elmer's Series Sixteen: Newest wave of powerful small computers for industry

From vans bouncing over desert tracks to the clanging production lines of heavy machinery factories, Perkin-Elmer 16-bit small computers have proven themselves a rugged, versatile breed.

Now we've brought out a new line of these popular computers, called the Series Sixteen. With Series Sixteen and new software—including FORTRAN—system builders can combine the most advanced computer control with economy and ruggedness to handle many jobs in manufacturing, automation, industrial design and data collection.

Drawing on 12 years' experience with industrial computers, Perkin-Elmer designed the new Series Sixteen for optimum price/performance, and packaged it for flexibility—it's a family of processors that answers a wide range of needs. Series Sixteen incorporates features that previously were extras, features like hardware multiply divide, power fail, single board memory expansion and expansion power.

Designed to maximize printed circuit modularity, all units can be readily customized, or upgraded, or interfaced with larger computers.

That Series Sixteen's versatility and reliability provide an excellent solution for industrial applications is shown by some of the jobs earlier Perkin-Elmer small computers are doing all over the world.

They travel the wilderness in vans, collecting and taping seismic data from geological test blasting over possible oil formations.

They prepare and distribute the numerical control tapes which automate the fabrication of expensive, high-tolerance machine parts

One of the big U.S. auto makers uses our small computers for carburetor testing and communicating results to a 32-bit computer, while a European car manufacturer uses them for control of robots in assembly-line automation.

And Perkin-Elmer 16-bit computers monitor the flow of petroleum through pipeline networks to and from tank farms on five continents.

With hardware proven for dependability and low cost, the new Series Sixteen stands out for state-of-the-art technology as well: high density LSI circuitry, fast Metal Oxide Semiconductor memory, dual bus architecture for high-speed and low-speed devices, no fewer than 255 vectored interrupts and battery backup for memory retention during power failures.

Software includes a comprehensive resource-sharing operating system, language processors for CAL Macro, BASIC and, especially, an enhanced scientific FORTRAN IV.

The new Perkin-Elmer LC-75: Keystone unit in a sophisticated system for chemical analysis

In determining the chemicals present in unknown substances, our new liquid chromatography Methods Processor routinely performs feats of quantitative and qualitative analysis once regarded as impractical, or even impossible. Use of the system—built around our powerful new LC-75 spectrophotometric detector with Autocontrol—is adding rapidly to basic chemical knowledge. It is also satisfying more practical needs in such applications as analysis of plastics and food, crime lab work, pharmaceutical development, dye and cosmetics manufacture, and monitoring environmental purity, to name just a few.

Liquid chromatography separates the molecular species in any substance which can be dissolved in an appropriate solvent. In essence, a sample is pumped under high pressure, at controlled temperatures, down a tube filled with small packed particles (such as silica gel). The molecules with most bonding affinity for the packing material are retained longest in the tube and emerge from it last. By flushing the packed particles with increasingly strong solvent and examining by spectrophotometry the effluents and the order in which they emerge, a chemist can identify ingredients and calculate how much there is of each in the sample

Obviously, the spectrophotometer is a key element in the process And fittingly, the LC-75 with complete electronic digital control (Autocontrol) is the most revolutionary module in the system. It's the first spectrophotometric detector designed specifically for LC. It can shine through a dissolved sample an analytical wavelength precisely variable from 190 to 600 nanometers. Trace analysis is excellent—even with tiny samples. It takes the operator through even the most sophisticated chromatographic peak analysis step by step. And quantitative capability is superior—providing the widest range of usable sample concentrations plus reliably repeatable results. Autocontrol unlocks the tremendous spectroscopic power of the LC-75.

Combine the LC-75 with the four other modules in the Methods Processor system (Series 3B pump, Model 420 Automatic Sampler/Injector, LC-100 column air-bath oven, and SIGMA 10 Chromatography Data Station) and you get infinite options for performing exactly the kind of analysis you need.

For more information on these products, please write: Corporate Communications, Perkin-Elmer Corporation, Main Ave., Norwalk, CT 06856.

PERKIN-ELMER

Responsive Technology

their report that "some adverse environmental impacts are likely to occur with increased coal mining and combustion because of conscious economic trade-offs, lack of efficient control devices or insufficient enforcement of existing standards." Among the negative effects for which permanent controls are not always available they cite the problems of acid drainage and ground-level subsidence long after the abandonment of underground coal mines. In addition, they write, some emissions from coal burning "may prove to be inadequately regulated because of scientific uncertainty concerning the cause and significance of their effects." Assuming a high projected rate of expansion in domestic coal use, "emissions of sulfur dioxide, nitrogen dioxide, carbon dioxide... and fine particulates will all increase significantly by the end of the century under... present or pending regulations."

The main concerns arising from such projected increases in coal-related air pollution focus on whether or not they will "cause or aggravate three potential environmental problems whose magnitude and importance cannot yet be evaluated adequately." The three problems in question are (1) the apparent buildup of carbon dioxide in the atmosphere resulting in part from fossil-fuel combustion, a trend that may lead in turn to significant changes in global climate; (2) the possibly serious public-health consequences of long-term exposure to low levels of air pollutants; (3) the risk of further damage to aquatic and terrestrial ecosystems caused by "acid rain" resulting from increased coal combustion. Additional evidence tending to confirm current concerns about these problems "could force a reappraisal of our environmental control strategy or even our commitment to increased coal development."

Under the heading of occupational health and safety the OTA report estimates that a tripling of coal production by the end of the century would result in approximately 370 fatalities and 42,000 disabling injuries per year among coal workers (almost three times the current rates), not counting the "thousands of workers" who would be disabled by various respiratory diseases (known collectively as black-lung disease). Furthermore, the report adds, "important questions have been raised concerning the effectiveness of enforcement of the existing dust standard and the adequacy of the research on which it is based."

In short, the OTA concludes, although there are no obvious technological, social or environmental obstacles to prevent the U.S. from relying on coal as its "most productive energy source well into the 21st century," the future of direct coal combustion remains "a mixture of promise and risk." Future legislative activity bearing on coal, the Congressional advisory office predicts, "will

The *Newest* Magazine Of Today's Astronomy Enthusiast

STAR & SKY

Become a Charter Subscriber

(See the special Charter Subscription Offer below)

YOU GET INFORMATION...
probing articles by astronomers and astronomy journalists, accompanied by photographs of all the wonders of deep space.

Thought-provoking articles by such noted authorities as Patrick Moore, John Gribbin, Stewart Sharpless and Wallace Tucker...by astronomy writers like Robert Sheaffer, Julius Benton, Ray Moses...by amateur correspondents Jay Gunter, Joe Rao and Ben Mayer...

YOU GET NEWS...
in-depth reports on the whole range of exciting discoveries taking place in such diverse fields as:

- black holes
- "ring" galaxies
- space exploration
- spectroscopy
- quasars
- pulsars
- cosmology
- planetology
- the search for extraterrestrial intelligence

Monthly departments like ***Astronomy Update,*** which keep you in touch with the moment-to-moment, fast-breaking developments in astronomy and the space sciences...and ***Event Horizon,*** which keeps you current with the activities of amateur astronomers.

YOU GET FACTS...
you'll read about ● ***The Solar Connection*** — the link between sunspot activity and the Earth's changing climate. ● ***The Gas Clouds of Orion*** — birthplace of stars. ● ***The Beta Lyrae Star System***—a unique binary system. ● ***Telescopes for Amateur Astronomers*** — a guide to purchasing the right scope for your individual requirements. ● ***The Death of Massive Stars*** —a study of stellar evolution. ● ***The Vela-X Pulsar*** — a detailed look at this supernova remnant. ● ***Observing Saturn: Visual Photometry and Colorimetry*** — an examination of the variation in intensities of the belts and zones of Saturn... and much, much more...

YOU GET TIPS ON ASTROPHOTOGRAPHY...
every month you'll get into such departments as the ***Astrophoto Gallery,*** which showcases — in brilliant color and black-and-white — several pages of an individual amateur's sky photographs. Prize-winners like John Sanford, Walter Hamler, David Healy and Rick Dilsizian...and showings of the work of proficient newcomers.

In conjunction with this is the new addition, ***Astrophotographer's Notebook*** — an ongoing series of lessons on film, lenses, locales, techniques and more, by veteran astrophotographer John Sanford.

YOU GET UNIQUE DEPARTMENTS...
"Who's Who" in Astronomy is our monthly column about the prominent figures in the astronomical world and an in-depth examination of the work they are involved in. Articles on such people as Arno Penzias and Robert Wilson in *Listening for the Echo of the Big Bang*...Raymond Davis who is *Puzzling Over Missing Neutrinos*...Beatrice Tinsley and her contribution to cosmology in *Brightness Counts*...

Asteroid of the Month is a column which features a minor planet for each month of the year, complete with tracking charts to help you find them.

Lee's Forum is an innovative, informative question-and-answer page. Associate Editor and experienced amateur, Lee Najman, fields questions about all aspects of observing, and promotes a dialogue — of a "hands-on" nature — in which observers relate their problems and share their successes.

YOU GET TIPS ON OBSERVING...
Deep Sky Vistas is a monthly guide to the various deep sky phenomena. Amateur and writer, Ron Morales describes in detail how to hunt for and find galaxies, open clusters, diffuse and planetary nebulae, variable stars, and other interesting objects.

Cosmic Frontiers explores the different attractions in, and aspects of the current evening sky with details on the backgrounds of the phenomena being observed.

Skywatch gives you planet information for each month, moon data, accounts of variable stars, i.e. which ones will be visible in that month and their brightness...what meteor showers will occur...special events like occultations...

You also get a variety of easy-to-read charts, maps and timetables to help you follow the ever-changing spectacle of the nighttime sky...to let you know what and when to observe.

From planetary apparitions to periodic comets...meteor showers to moonrise... solar eclipses to supernovas...WE COVER IT ALL!

Be an active observer or an "armchair" astronomer, but whatever your level of interest, ***let STAR & SKY turn you on to the Universe!***

So don't delay, the sooner you return your Charter Subscription Order Form, the faster you enter the dramatic world of ***dark skies...***

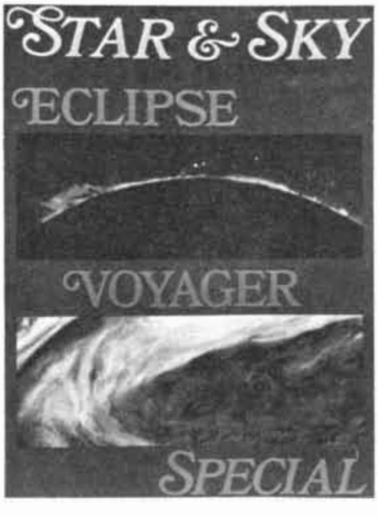
GUARANTEE: If you become dissatisfied with ***STAR & SKY,*** tell us why, and get a refund on all unmailed issues.

Free...

Get your Eclipse/Voyager Special Issue (a $2.00 value) *FREE*...simply *enclose payment with your order.*

A fabulous volume, overloaded with color pictures, recording every thrilling aspect of these two historic *Super-Events*...pre-totality atmospheric changes, coronas, prominences, "diamond rings," Bailly's beads...by such noted astrophotographers as Robert T. Little, Dr. Jay Pasachoff, John Sanford, and Ben Mayer — who covered the event at 32,000 feet above the earth in a jet!

In the same volume is complete coverage of the Voyager 1 Mission...close-up photos of Jupiter's surface, moons, and satellites — all in big, beautiful, brilliant color with accompanying text by astronomy writer, Terence Dickinson.



STAR & SKY, P.O. Box 324, Southport, CT 06490

Turn me on to the Universe!

☐ **I am enclosing payment with my order, so send me my FREE *Eclipse/Voyager Special Issue...***

☐ $27 for two years (24 issues)

☐ $15 for one year (12 issues)

☐ Please bill me.

Canada & Mexico: add $3 per year
All other foreign: add $6 per year

Name____________________

Address____________________

State__________ Zip__________

"le"

This 2-letter French word may well be the most successful (if not the only) 2-letter marketing strategy in the annals of advertising.

This simple, yet eloquent "le" has put Renault's "Le Car" happily in the U.S. car-buying public's consciousness. It positions; it defines; and for the media-inundated consumer, provides an easy, memorable handle. Usually it takes many more letters and many more words to do all this.

Happily our little "le" has put across the fact that we are <u>the</u> car company of France. You can't blame us for sighing *"Vive le 'le'!"*

RENAULT

fall into three categories: further amelioration of negative impacts; promotion of coal to reduce the use of other fuels; and oversight to ensure that existing legislation is being appropriately implemented."

The Chemistry of Acupuncture

Acupuncture, the ancient Chinese art of relieving pain by inserting fine needles at specific points and depths on the surface of the body, has attracted much attention outside China in recent years. Some non-Chinese physicians have applied the technique and have maintained that they get similar results; others have speculated on how such effects could be explained, if not by hypnotic suggestion, then by known principles of nerve physiology. Now a Chinese investigator, Chang Hsiang-tung of the Shanghai Institute of Physiology, surveys acupuncture in the light of the newest findings in the physiology of pain.

Writing in *Chinese Medical Journal,* Chang describes how in China until quite recently the pain-suppressing effects of acupuncture were explained only according to the metaphysical tenets of traditional Chinese medicine. This view holds that pain and disease result from a disruption or blocking in the circulation of vital energy through 12 hypothetical meridians (channels) in the body, each of which is associated with a particular organ system. By inserting acupuncture needles at certain points in the meridians (which are not correlated with specific anatomical structures) the blocked channels can be cleared and the healthy state of energy flow restored.

Attempts by both non-Chinese and Chinese investigators to relate the empirical practice of acupuncture to modern physiology were initially based on the observation that many acupuncture points lie close to clusters of nerve endings, suggesting that the flow of energy through the hypothetical meridians might be analogous to the functioning of the nervous system and the various chemical systems that mediate its regulation and control. The obvious question then arose of how needles inserted into the body at specific points could activate the nervous system to yield analgesia in distant parts of the body. Also to be explained was the fact that the analgesia takes about 20 minutes to develop and persists for some time after the needles have been withdrawn.

It has long been known that there are two major classes of nerve fibers that carry pain information from the peripheral pain receptors to the spinal cord: the low-threshold *A-delta* fibers, which give rise to sharp, mild pain, and the high-threshold *C* fibers, which give rise to severe, burning pain. Chang suggests that mild stimulation of the acupuncture needles (traditionally accomplished by twirling the needles) selectively activates the low-threshold fibers in the deep tissues underlying the acupuncture points. The resulting impulses are transmitted to relays in the spinal cord and the brain, where they interfere with the transmission of messages from the high-threshold fibers, so that severe pain is kept from reaching conscious awareness. This "gate" hypothesis is supported by a number of experimental findings. For example, if a local anesthetic is injected into the tissues under an acupuncture point, it completely abolishes both the characteristic mild discomfort caused by the needle and the remote analgesic effect.

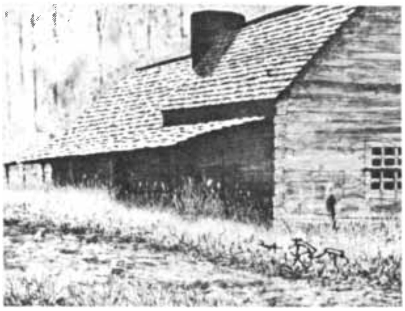
The specificity of acupuncture points, writes Chang, seems to lie in the segmental organization of the spinal cord: each segment of the cord innervates a particular region of the trunk and the limbs. The acupuncture points on the arms are usually stimulated to relieve pain in the head and the upper body, and those on the legs are stimulated to relieve pain in the lower body. This observation suggests that the low-threshold pain impulses generated by acupuncture needling travel to the spinal cord and block the transmission of severe-pain messages entering the same segment of the cord or adjacent segments.

Chang points out that although the specific nerve-cell interactions that give rise to acupuncture analgesia are not well understood, there has been much interest in the involvement of certain neurotransmitters: the chemicals that carry excitatory and inhibitory messages between nerve cells at relay centers in the spinal cord and the brain. Experiments on monkeys and human beings have demonstrated that the analgesic effect of acupuncture can be appreciably reduced by the administration of the drug naloxone, which blocks the action of morphine. This finding has been interpreted to mean that acupuncture triggers the release in the central nervous system of morphinelike peptide molecules: the short chains of amino acids called enkephalins and the longer chains called endorphins. Indeed, enkephalin and another peptide, substance P, have been implicated in pain perception, substance P with the transmission of pain-related impulses and enkephalin with their suppression.

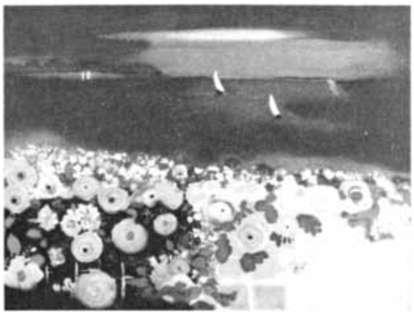
The rapid and transient action of a neurotransmitter such as enkephalin is hard to reconcile, however, with the slow induction and lingering action of acupuncture. It is therefore unlikely that enkephalin alone is responsible for the analgesic effect. Indeed, B. Sjölund and his colleagues in Sweden have shown that during acupuncture there is an increase in the concentration in the spinal fluid of a class of endorphins designated Fraction I, which are distinct from enkephalin. The rise in Fraction I endorphins in the spinal fluid is local-

What your home could have in common with the Met, the Tate, and the Louvre

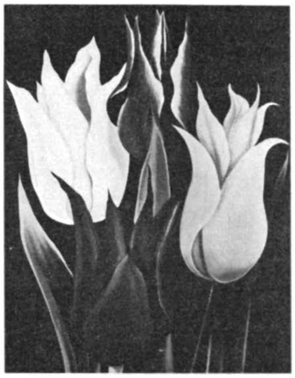
Gilles Figuier's *L'Etang de Sologne,* original lithograph. Signed limited edition of 175.



Don Whitlock's *Homestead,* original lithograph. Signed limited edition of 195.

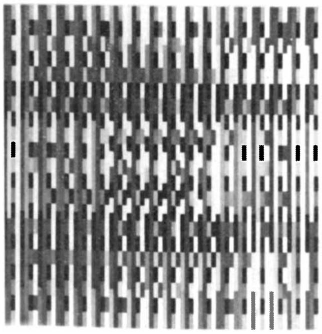


Marcelle Stoianovich's *St-Jean-Cap-Ferrat,* original lithograph. Signed limited edition of 225.



Lowell Nesbitt's *Four Tulips,* original serigraph. Signed limited edition of 175.

Bernard Charoy's *Valerie,* original lithograph. Signed limited edition of 300.



Yaacov Agam's *Evening,* original serigraph. Signed limited edition of 270.

In your own home, you can exhibit original work by artists who are represented in the world's great museums and galleries.

Artists like Calder. Chagall. Dali. Miro. Picasso.

At the Original Print Collectors Group, we offer moderately priced limited edition prints by these and other well-known artists.

Our offerings include signed original etchings, engravings, lithographs, woodcuts, and silkscreen prints.

Each print we offer is signed and numbered by the artist. Custom-framed. Accompanied by a certificate of authenticity and a full money-back guarantee.

And each print has been chosen not only for its beauty but for its investment value.

(Works by the artists we offer tend to appreciate, sometimes as much as 30% in a single year. Perhaps that's why *The Wall Street Journal, Business Week, Money,* AP and UPI have quoted us as an authority on investing in art.)

For more information, send in the coupon. There's no obligation.

You'll find out about the special pleasure of owning original art, instead of just visiting it.

Original print collectors group, Ltd.
120 East 56th Street, Dept. SA-20, New York, N.Y. 10022

☐ PLEASE SEND ME YOUR FREE COLOR BROCHURE, LATEST NEWSLETTER AND OTHER INFORMATION.

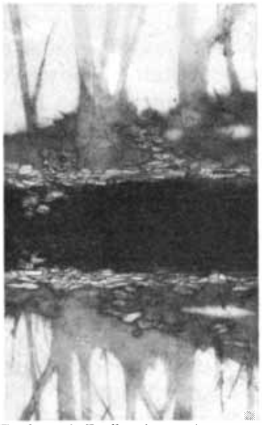
Name ______________________

Address ______________________

City ______________ State ______________ Zip ______________

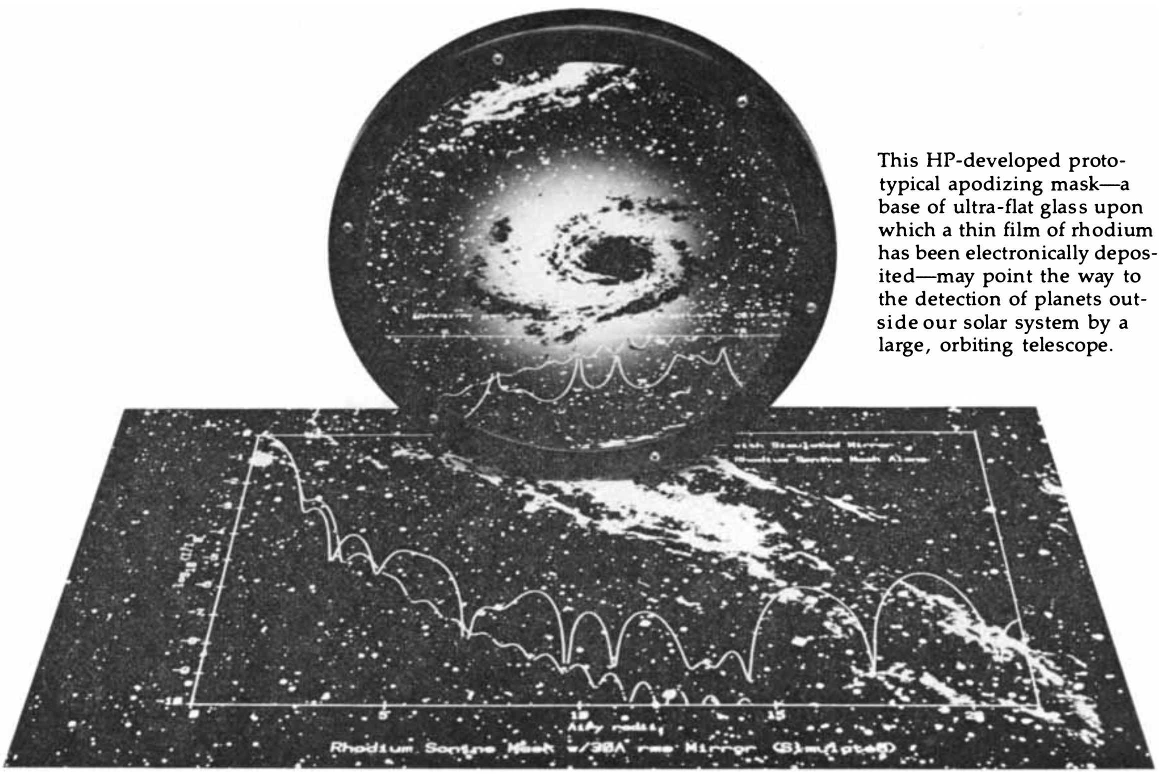
Rolf Rafflewski's *Ile St. Louis,* original lithograph. Signed limited edition of 225.

André Brasilier's *Pastorale,* original lithograph. Signed limited edition of 150.



Rodman's *Reflection—Autumn,* original collograph. Signed limited edition of 150.

HP measurement and computer advances



This HP-developed prototypical apodizing mask—a base of ultra-flat glass upon which a thin film of rhodium has been electronically deposited—may point the way to the detection of planets outside our solar system by a large, orbiting telescope.

There's more to telescopy than meets the eye.

No one has ever seen the planet of another star. The small angle of separation and enormous brightness ratio makes such a planet virtually disappear in the glare of its star. In a NASA-sponsored feasibility study, HP's Physical Electronics Lab has used its resident optical expertise to devise solutions to the problem of detecting planets outside our solar system.

Do other stars have planetary systems like our own? The answer would be of great interest to astronomy. Planetary systems are thought to be generated along with stars during their contractive phase, and are therefore possibly quite common. But so far, evidence for their existence is indirect and questionable, predicted entirely on the slight wobble detected in the motion of some nearby stars.

Two HP people, Vice-President of R&D Barney Oliver and physicist Larry Hubby, have been doing optical and astronomical calculations to determine whether it would be possible to see planets of other stars by putting a large telescope into orbit around the earth. The actual telescope would presumably be a self-contained space vehicle, similar to the Large Space Telescope to be launched by the Space Shuttle in the early 1980's.

While an orbiting telescope could indeed give astronomers their first view of the heavens unobstructed by the atmospheric scattering that degrades the resolving power of earthbound telescopes, other problems would still exist:

Diffraction of starlight occurs when it strikes the sharply defined rim of a telescope aperture. This creates a series of subsidiary "ripples" that extend to infinity, and which can totally obscure the far fainter images of a star's planets, as the ripples from a stone cast into a pond would obscure our perception of grains of sand also striking the water's undulant surface.

The solution proposed by Oliver and Hubby is *apodization*—literally, "cutting off the feet"—a term coined by P. Jacquinot to signify reducing or eliminating the ripples that flank strong lines in a spectrograph. An apodization mask, rather like a specialized camera filter but which gradually reduces the light transmission of the aperture from the center toward the rim, will theoretically reduce diffraction ripples by softening the hard edge that generates them. Through extensive analysis (using an HP desktop computer to perform all calculations) Oliver and Hubby established the feasibility of such a mask, as well as its required characteristics.

With an apodized telescope in orbit, astronomers should be able to examine some 275 stars within the range of 33 light years, with excellent chances of detecting large extrasolar planets if any, in fact, exist.

extend your possibilities.

HP's first UV-VIS spectrophotometer lights the way to faster, nondestructive, chemical identification.

Electronics simplifies the optical system, reduces its moving parts to one in the new HP 8450, a powerful, computer-controlled, ultraviolet-through-visible-light spectrophotometer that generates full spectrum and multicomponent analysis in seconds.

Spectrophotometers identify unknown chemical compounds by measuring their ability to absorb light. Conventional instruments do this by the time-consuming method of sending light through a prism mechanically rotated so that one wavelength at a time passes through the sample to a detector. The first spectrophotometer to be offered by HP has a revolutionary optical system. It passes light first through the sample and then onto a holographic grating, which disperses it across two parallel arrays of light-sensitive diode detectors—one array measuring visible, the other ultraviolet, light. Resulting spectra are shown one second later on the instrument's built-in video display.

The reversed optics technique coupled with the high speed data handling capability of the HP 8450 enables chemists to make full-spectrum, multicomponent analyses of mixtures containing up to seven components, at an unprecedented 360 samples per hour!

The built-in, 44k-word microcomputer permits the user to program operating parameters and measure virtually any sample with minimal sample preparation and minimal training. Communication is through a friendly, interactive keyboard, sealed to prevent damage from chemical spills.

After an analytical run, the output from the HP 8450 may be directed to an optional plotter, printer/plotter, or other HP-IB (Hewlett-Packard Interface Bus) compatible device. For use with other data systems, RS232C serial output is also included.

Prices start at $20,000*.

HEWLETT hp PACKARD

1505 Page Mill Road, Palo Alto, California 94304

For assistance call: East (301) 258-2000, West (213) 877-1282,
Midwest (312) 255-9800, South (404) 955-1500, Canada (416) 678-9430

Mail to: Hewlett-Packard, 1505 Page Mill Road, Palo Alto, CA 94304
Please send me further information on

( ) HP 8450 UV-VIS spectrophotometer

Name ______________________
Company ______________________
Address ______________________
City ______________ State ________ Zip ________

*Domestic U.S. prices only.

00946

ized mainly at the level of the spinal segments receiving input from the needled acupuncture point. It therefore seems that acupuncture triggers the liberation in the spinal cord (and possibly in the brain as well) of long-lasting morphine-like peptides that suppress the transmission of pain impulses. These peptides do not appear to be released into the bloodstream, however, because elevated levels of endorphins in the blood after acupuncture have not been detected.

Chang cites the work of investigators in China indicating that acupuncture causes an increase in the amount of the neurotransmitter serotonin in the brain. Selective destruction of the serotonin-containing cells in the brain stem of experimental animals reduces the analgesic effect of acupuncture, whereas electrical stimulation of these cells enhances the effect. How serotonin might interact with the endorphins, however, is not yet understood.

Such inquiries into the physiology of acupuncture, Chang writes, have already resulted in important modifications of the technique. Many Chinese physicians now choose acupuncture points on the basis of the segmental-innervation theory rather than the traditional meridian theory. A large number of acupuncture points have been shown to be redundant or superfluous, and they have been omitted without sacrificing the analgesic effect. The resulting simplification of the technique has saved patients unnecessary discomfort from the repeated insertion of needles. Electrical rather than mechanical stimulation of the needles is also gaining ground, since it reduces the danger of contamination of the operation field by the acupuncturist and minimizes the tissue damage caused by needling. Finally, drugs that enhance the effectiveness of acupuncture (including many drugs that by themselves are not analgesics) are coming into clinical use. Some Chinese physicians are also experimenting with "acupressure," the application of finger pressure to acupuncture points, which in some cases has been shown to be as effective as needling in relieving pain.

Chang observes that since acupuncture does not entail the hazardous and unpleasant side effects of anesthetic agents and opiate drugs its potential value in analgesia and anesthesia is clear. He concludes: "This age-old Chinese healing art is now already placed on a solid scientific basis and is indeed rejuvenated and revitalized. There is no doubt that acupuncture will become a new addition to the world medical armamentarium for combating pain."

Holes in History

Black holes, those putative ghosts of collapsed massive stars from which no matter, light or signal of any kind can escape, would seem to be a brainchild of modern cosmology. The fact is that the concept goes back to the 1770's, writes Simon Schaffer of the University of Cambridge in a recent issue of *Journal of the History of Astronomy.*

John Michell, an 18th-century astronomer who was the rector of Thornhill in Yorkshire, discussed the attraction of light by gravity in a 1784 issue of *Philosophical Transactions of the Royal Society of London.* In an article titled "On the Means of discovering the Distance, Magnitude, &c. of the Fixed Stars, in consequence of the Diminution of the velocity of their Light, in case such a Diminution should be found to take place in any of them, and such Data should be procured from Observations, as would be farther necessary for that Purpose," Michell argued that if light is subject to gravity, the velocity of the light emitted by a star would be diminished because of the star's gravitational pull. Hence the mass of a star could be determined by measuring the retardation of the starlight. Michell calculated that the escape velocity from the sun is 497 times smaller than the speed of light. The escape velocity is proportional to the radius, and so he maintained that any star with the sun's density but with a radius 497 times larger would trap all the light it emitted.

In 1796 the French mathematician and physicist Pierre Simon de Laplace argued in the same vein, although apparently he was not aware of Michell's work. In his *Exposition du système du monde* Laplace concluded that any star with the density of the sun and a radius 250 times larger should be invisible because the escape velocity at its surface would be greater than the speed of light. Michell's and Laplace's discussions of black holes were of course based on Newtonian mechanics, whereas current discussion is based on the general theory of relativity, the special theory of relativity and quantum mechanics. Therein lies a crucial difference between the two discussions: Michell and Laplace merely conjectured, without observational or theoretical support, that light might have particle properties and so might be influenced by gravity. Today it is known with certainty that this is the case. The laws of modern physics demand that light have particle properties, and experiments have revealed what they are. It is also known today that the operative mechanism in a black hole would not be the retardation of the velocity of light but the containment of the particles of light, or photons, by the curvature of space-time.

Piltdown Won't Down

It has been 26 years since Piltdown man, at that time one of the main fossils bearing on human evolution, was shown to be a fraud. In 1953 scholars at the British Museum (Natural Histo-

An example of permutational art in which the photographer "enters" the environment at a chosen point and follows a geometric path around that point, stopping to record the scene with an SX-70 camera en route. The resulting sixteen prints are scrutinized on the spot and later recombined to form the finished work.

Today's Bunker Ramo is giving your broker a lot of the information he's giving you.

On the floor of the New York Stock Exchange and in brokerage houses coast to coast, many thousands of people are looking to Bunker Ramo for the very latest information every trading day.

What they're looking at is Bunker Ramo's Market Decision System 7—the nationwide network of terminals, computers and over 100,000 miles of high-speed data circuits that provides the securities industry with accurate up-to-the-second data on stocks, bonds, options and commodities.

In addition to such current data as open, high, low, net change, and volume, Market Decision System 7 also displays statistical data and market trends. Among many other features, it can supply a standard stock watch on up to 72 issues per terminal, a limit watch on up to 12 issues per terminal, ticker displays, statistical data bases, and Dow Jones News/Recall. For those who make investment decisions, it is a vital, intelligent system.

Today, Bunker Ramo is a major producer of the data terminals and minicomputers that, together, comprise the information systems widely used in banking and insurance, as well as the securities industry. In addition, through our well-known Amphenol® connectors, we have a leading position in the aerospace and aircraft industries, a strong foothold in telecommunications, and a growing role in fiber optics.

Most important, we are a company with a reservoir of high technology, a solid base of innovative products, and a strong management team.

Which is why today's Bunker Ramo is a company that's on the move. For our latest financial reports, write Bunker Ramo Corporation, Dept. FR, 900 Commerce Drive, Oak Brook, Illinois 60521.

BUNKER
RAMO

ry), where "*Eoanthropus dawsoni*" was housed, announced that the purported "dawn man" was a contrived combination of genuinely prehistoric human cranial fragments and a modern ape's lower jaw with doctored molar teeth. Ever since then a dual question has gone unanswered: Who perpetrated the hoax and why?

In his detailed 1955 account of the initial acceptance and eventual reexamination of Piltdown man, Joseph S. Weiner, a member of the British Museum's group of investigators, cited incidents in the career of Charles Dawson, a solicitor and semiprofessional fossil hunter, suggesting that Dawson was not above doctoring the fossils he collected, for example chemically staining them to give them a greater patina of antiquity. It was Dawson who first turned up the five human cranial fragments of "*Eoanthropus*" in a Sussex gravel pit and later unearthed the ape jaw. Chemical analyses in 1953 showed that all these objects had been artificially stained.

Still, few scholars were willing to credit Dawson with the sophisticated knowledge required, for example, to grind the ape's molar teeth into an approximation of a human pattern of wear before planting the jawbone for later discovery. Who else, then, was learned enough and had access to anatomical material (including not only the ape's jaw but also appropriately ancient remains of other animals found in association with the cranial bones) to stage such a successful hoax? Notable among those who were taken in was the keeper of the geology department of the museum, Sir Arthur Smith Woodward, who had enthusiastically joined Dawson in searching for more remains in the gravels of Piltdown Common.

Last year a 94-year-old British geologist, James A. Douglas, dictated a rambling "deathbed statement" to a neighbor on the Isle of Wight. At age 53 Douglas had succeeded to the professorship of geology at Oxford following the death of William J. Sollas, aged 87; Sollas (1849–1936) had been professor for some 40 years. In his statement Douglas declared that Sollas had profoundly disliked Smith Woodward (who had dismissed a line of fossil research pursued by Sollas as being trivial) and considered him to be uninformed in general and ignorant in particular about fossil man, Sollas' own specialty. Douglas suggested that Sollas had supplied Dawson with the doctored ape jaw. Sollas supposedly expected that the deception would soon be discovered, to Smith Woodward's embarrassment. Then some 40 years passed and both men were dead before the hoax became known.

When Douglas' charge was made public two of the British Museum workers who had exposed the hoax, Weiner and Kenneth P. Oakley, came to Sollas' defense. The only concrete item of evidence against Sollas that was cited by Douglas, coming under the heading of opportunity, was his possession of a chemical (potassium bichromate) similar to the one Weiner and Oakley had detected as being a staining agent. Pointing out that Dawson was known to be well supplied with bichromate of his own, Weiner and Oakley further dismissed Sollas' supposed motive for the hoax as the idiosyncratic view of his accuser. As the matter now stands, Sollas appears to be acquitted of the charge, and both the motive and the perpetrator of the Piltdown hoax remain unknown.

The Day the Earth Screamed

Many people who have mused on volcanoes have wondered what would happen if a well were drilled in exactly the right hot spot. Well, it happened. Borehole 4 (one of five) was drilled 1,138 meters to tap a hot-water aquifer at the Námafjall geothermal field in Iceland in 1968. From then until September 8, 1977, the borehole duly produced steam, as it was meant to do. At 11:45 P.M. on that day it suddenly shot a thin column of molten rock 15 to 25 meters into the air. The column gave off quantities of sparks and cinders. This phase of the eruption, which was accompanied by a steady roar, lasted for about a minute. For the next 10 to 20 minutes the new volcano was quiescent. The final phase, which lasted for about a minute, consisted of a series of shots of glowing scoria (rough, vesicular lava). The borehole then returned to the production of steam.

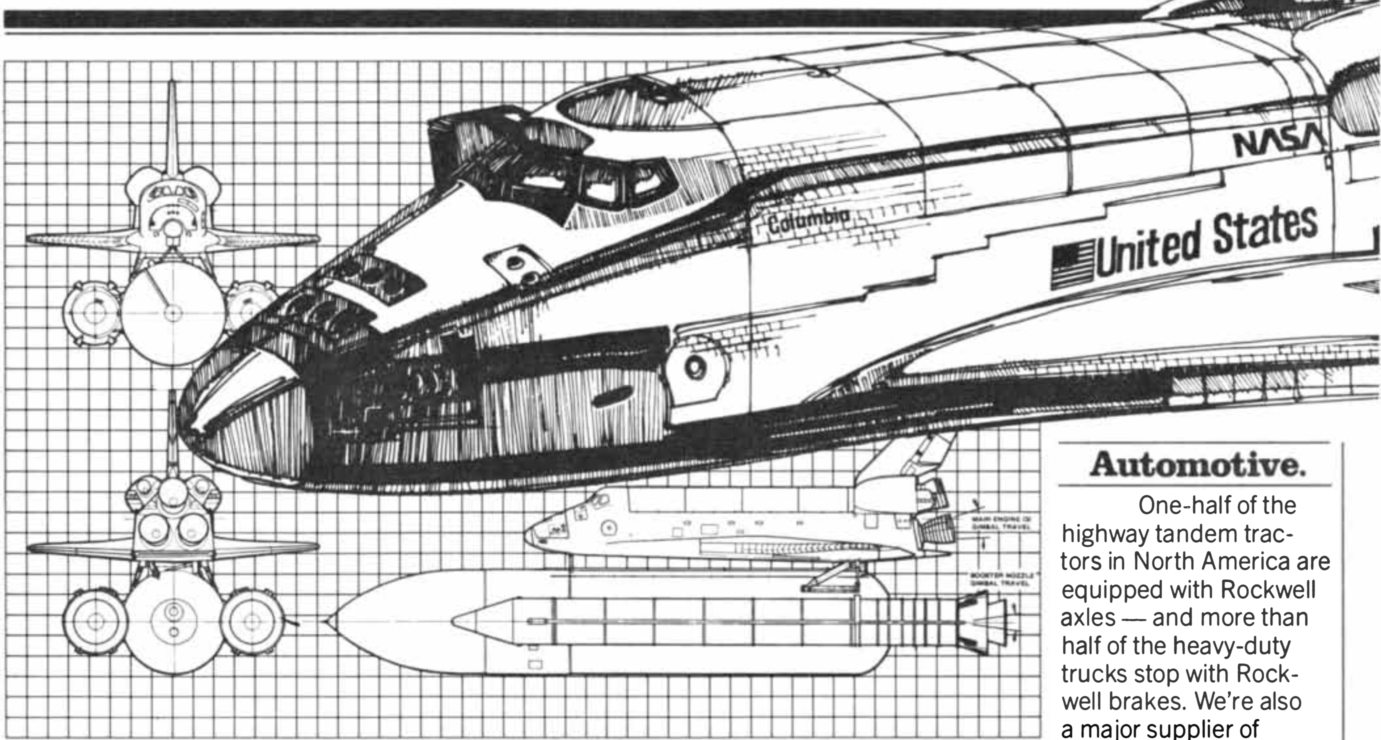
The event is described in *Nature* by Gudrún Larsen, Karl Grönvold and Sigurdur Thorarinsson of the University of Iceland. They point out that the field is about nine kilometers south of the Krafla caldera (volcanic crater) and that magma is fed continuously into magma reservoirs below the caldera at a depth of about three kilometers. The geothermal-energy field lies on the Krafla fault swarm, the scene of periodic rifting events in which segments of the swarm from five to 15 kilometers long "are activated with movements on faults, earthquakes and the formation of new steam fields." The five boreholes of the geothermal field were deliberately placed to intersect faults at depths favorable to tapping these sources of energy.

The authors conclude that the eruption in the borehole "was part of a rifting event that took place during that day" and that it provides evidence for a large-scale horizontal movement of basaltic magma from the caldera through the fault swarm. The magma was apparently injected into the borehole through a steeply dipping fault that intersects the hole above the aquifer. The volume of the eruption is estimated at 26 cubic meters, the weight at 3,100 kilograms (three metric tons).

Rockwell is more than the builder of America's Space Shuttle.

Much more.

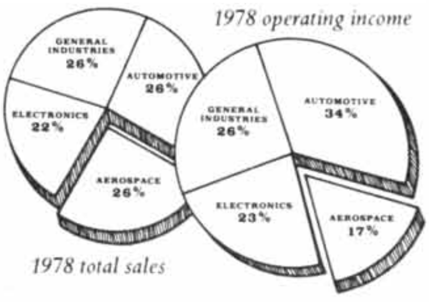
Designed for 55 separate firings.



Automotive.

One-half of the highway tandem tractors in North America are equipped with Rockwell axles — and more than half of the heavy-duty trucks stop with Rockwell brakes. We're also a major supplier of drivelines, steel and styled aluminum wheels, mechanical devices, castings, stainless steel wheel covers, reinforced plastic and other components for trucks, trailers, buses, vans and passenger cars.

If you think of Rockwell International as a company that makes spaceships, you have good reason. We are prime contractor to NASA for Space Shuttle orbiters and their main engines, and for integration of the entire Shuttle system and selected payloads. The Rockwell-built orbiter is the world's first reusable spaceship, capable of at least one hundred missions into space and back. It will launch like a rocket, haul like a truck, work like a space station, then return to Earth and land like a glider.

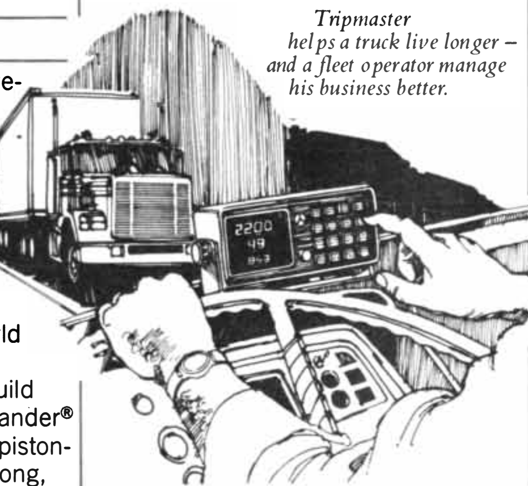


But aerospace is only part of the Rockwell story. Rockwell International is a major multi-industry company, applying advanced technology to a wide range of products — in aerospace, automotive, electronics and general industries. Following are some examples of our balanced diversification.

Aerospace.

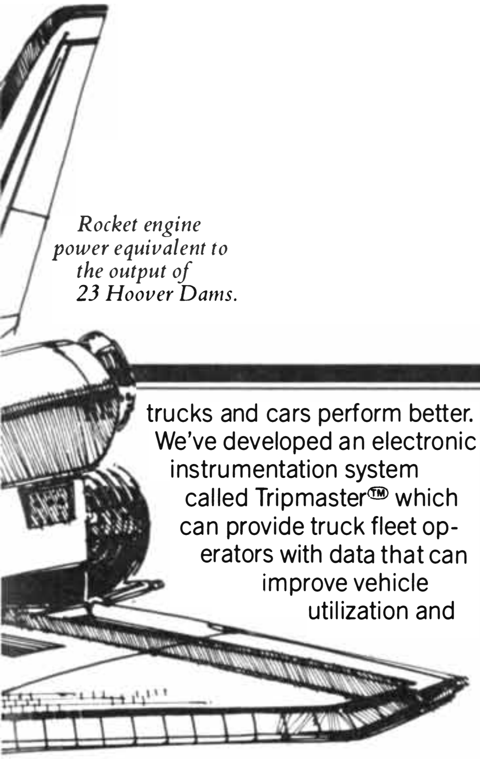
Rockwell aerospace involvement extends well beyond the Space Shuttle. Our rocket engines have been used to launch over two-thirds of all U.S. manned space flights and satellites, and we're building the Navstar satellites that are bringing totally new standards of accuracy to world navigation.

For general aviation, we build Sabreliner® business jets, Commander® 690B propjets, and Commander piston-powered aircraft. And we have a long, proud history as a designer and builder of U.S. military aircraft.

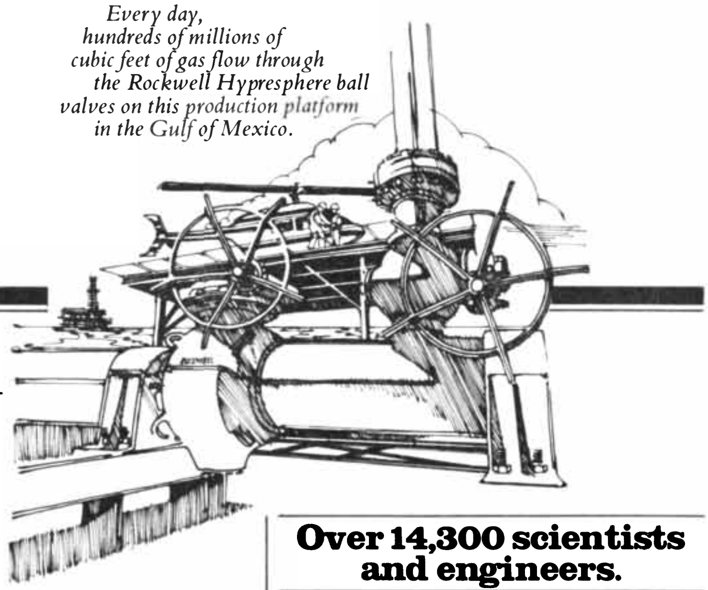


Tripmaster helps a truck live longer – and a fleet operator manage his business better.

In this age of rising costs, Rockwell is creating products that can help

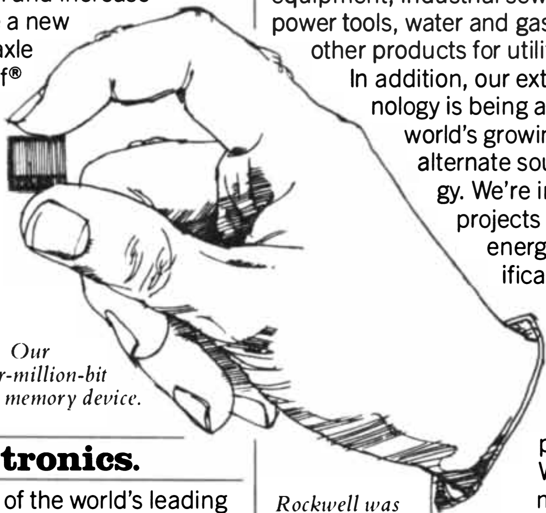


Rocket engine power equivalent to the output of 23 Hoover Dams.



Every day, hundreds of millions of cubic feet of gas flow through the Rockwell Hypresphere ball valves on this production platform in the Gulf of Mexico.

trucks and cars perform better. We've developed an electronic instrumentation system called Tripmaster™ which can provide truck fleet operators with data that can improve vehicle utilization and operating efficiency. Some of our components that can help cut a trucker's fuel consumption and increase payloads include a new aluminum front axle beam, Taper-Leaf® springs and Stopmaster® brakes. Our weight-saving products for cars include reinforced plastic components and aluminum wheels.



Our quarter-million-bit bubble memory device.

Electronics.

We're one of the world's leading suppliers of avionics — communications, navigation and flight control equipment for air transport, general aviation and government aircraft. We also make microelectronic systems and devices and guidance and control systems. And we manufacture and install telecommunications systems for businesses and governments worldwide.

One of Rockwell's latest electronic achievements is the production of bubble domain memories. We developed and market a device measuring about two-fifths of an inch square that can store up to 256,000 bits of information. It has no moving parts and retains its data without use of power.

Mechanical mass storage computer memories can now be replaced by smaller, faster, more reliable, more energy-efficient systems.

General Industries.

Most of America's major daily newspapers are printed on Rockwell-Goss presses. We also make textile equipment, industrial sewing machines, power tools, water and gas meters and other products for utilities.

In addition, our extensive technology is being applied to the world's growing need for alternate sources of energy. We're involved in projects for nuclear energy, coal gasification, flue gas desulfurization, and solar, wind and geothermal power. We also manufacture valves — plug, pipeline ball, gate, globe, angle and check valves for America's power needs. Our Flow Control Division is one of the world's leading suppliers of high-technology valving for U.S. and international energy markets. Rockwell valves are used in oil and gas production, transmission and processing; natural gas distribution pipelines; coal slurry pipelines; synthetic fuel processing plants; electric power plants; shipboard power plants; and in many other energy-related markets.

Rockwell was the first company in the world to produce a 256,000-bit memory for commercial applications.

Over 14,300 scientists and engineers.

Of our 114,000 employees, one in eight is either a scientist or an engineer. They constitute about one percent of America's total scientific-engineering community. This technological base positions us for leadership in each of our product areas. It also makes our corporate slogan, ". . . where science gets down to business," a fact.

Business is good.

Earnings per share

Rockwell International's total sales for 1978 were over $5.6 billion. Fiscal 1978 was the most profitable year in Rockwell's history. Net income, up 23% over fiscal 1977, was $176.6 million, generating a record $5.02 per share.

For more of the Rockwell story, please write us:
Rockwell International,
Department 815S,
600 Grant Street,
Pittsburgh, PA 15219.

Rockwell International

...where science gets down to business

The Origins of the First Powered, Man-carrying Airplane

The Wright brothers' "Flyer" of 1903 was not just a lucky effort by two bicycle mechanics from Dayton but the outcome of an intensive program of research, engineering and testing

by F. E. C. Culick

In a series of flights on December 17, 1903, Wilbur and Orville Wright became the first to pilot a powered airplane. As is well known, they had designed and built the craft themselves. To this day, however, the notion persists that the Wright brothers were essentially bicycle mechanics who more or less stumbled on their successful design. Even among aeronautical engineers ignorance of what the Wrights accomplished is widespread. In tracing the steps by which the brothers arrived at their success, and in building a model of their 1903 "Flyer" for tests in a wind tunnel, I have come to recognize how remarkable their achievements in research, engineering and testing were. Their work deserves to be set in a richer historical context.

On the day of the first flights the brothers took turns as pilot. Wilbur was the pilot on the fourth, last and longest flight of the day, staying aloft for 59 seconds. Nearly four years passed before anyone else was able to fly for as long as a minute, and even then the machine was not fully controllable. By that time the Wrights had developed a maneuverable airplane capable of flying for more than an hour.

As late as 1908, when the Wrights first flew publicly, no one else yet understood the need for lateral control, much less the function of the vertical tail. Therefore no one else could execute proper turns. No one else knew how to make propellers correctly. Above all, no one else had pursued a comparable program: doing the necessary research, constructing his own aircraft and doing his own flying, so that he understood the entire problem. The Wrights were able to outstrip all the other people who were trying to fly because they recognized the problems that had to be faced and solved.

The brothers began their work in 1899, when Wilbur was 32 and Orville was 28. Wilbur, who was the informal leader in much of the work, realized they would have to follow a systematic program of research and engineering. He summarized his view in a lecture he gave in September, 1901, at a meeting of the Western Society of Engineers: "The problems of land and water travel were solved in the 19th century because it was possible to begin with small achievements and gradually work up to our present success. The flying problem was left over to the 20th century, because in this case the art must be highly developed before any flight of any considerable duration at all can be obtained." Four years later the Wrights would have a practical airplane.

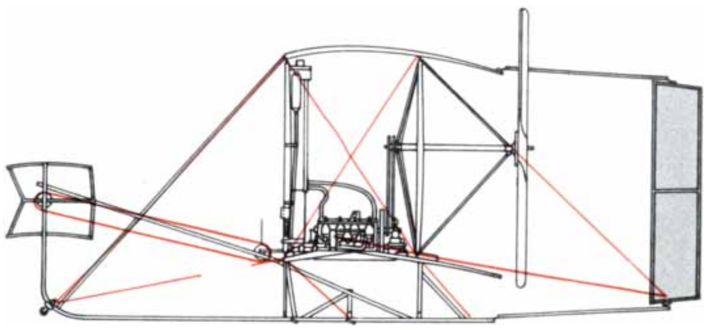
Their invention of the powered, man-carrying airplane happened partly as a natural development from the work of others. To an extraordinary degree, however, the Wrights' own contributions were crucial. Their technical achievements were stunning, particularly when one considers where the effort to achieve human flight stood in 1899.

Four Pioneers

The people trying to build flying machines had followed several strategies. Observations of the flight of birds motivated the efforts, and so it was natural that the earliest ideas always entailed the use of flapping wings. Sir George Cayley (in 1799) was the first to understand that manned flight would be more easily achieved if the means of generating lift were separated from the means of propulsion.

Cayley is recognized as the originator of the airplane. His designs and conclusions were based on careful observations and experiments. He conceived the configuration now viewed as being conventional: a body or fuselage supporting one main wing, with horizontal and vertical tails positioned aft. He sucessfully flew gliders, one of them large enough to carry a boy on at least one occasion. Cayley himself never flew. Among his many other notable contributions were demonstrations that curved surfaces are better than flat ones for providing lift; the idea that lateral balance or stability can be gained by giving the wing a dihedral angle, that is, making it with raised tips, and the concept of employing a horizontal tail to achieve longitudinal stability.

What Cayley began to comprehend were the intimate connections between the geometry of an aircraft, the forces



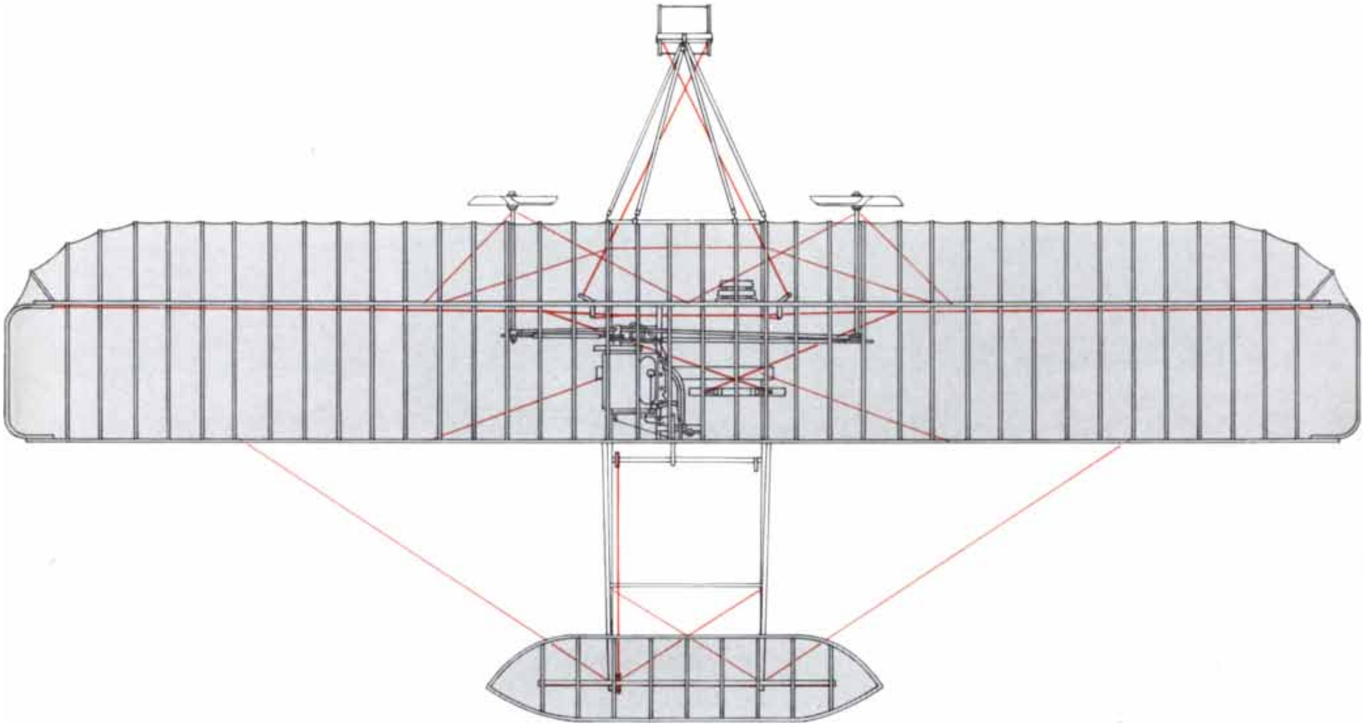
WRIGHTS' FLYER OF 1903, portrayed at the left in a view from the left side and at the right in views from above and in front, was the first powered, man-carrying airplane to fly successfully. Its most distinctive feature was the forward horizontal control surface. The twin vertical

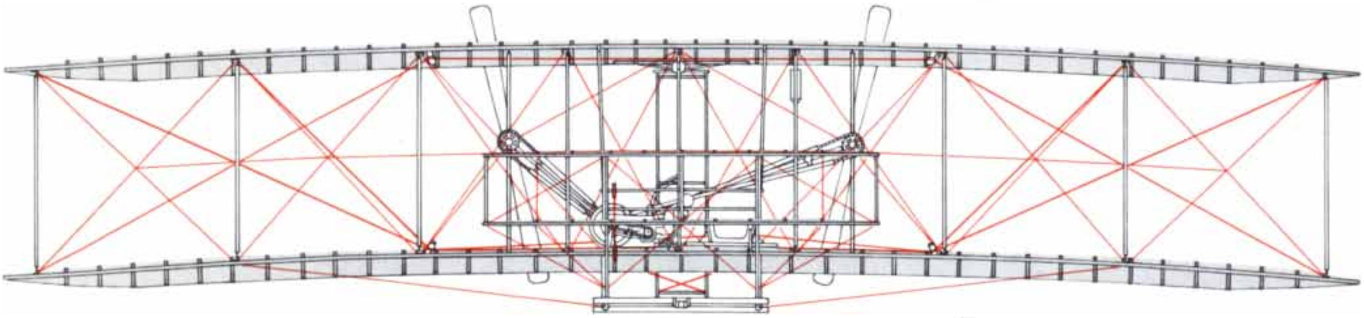
acting on it and its stability in flight. An airplane has a vertical plane of symmetry passing through its longitudinal axis. Motions that do not deflect the craft in directions out of that plane are called longitudinal motions. The commonest motion is pitching, in which the nose moves up or down. Sideslip, roll about the longitudinal axis and yaw about the vertical axis are collectively called lateral motions.

The tendency of any object in steady motion to return to its initial state following a disturbance is called stability. There are two kinds of stability, longitudinal and lateral, corresponding to the two classes of motions of an airplane. In a modern aircraft longitudinal stability is provided mainly by the smaller horizontal surface, which is usually placed at the rear as Cayley advocated. Lateral stability is the effect of many causes, including the dihedral angle of the wings and the vertical tail. The designs of early flying machines often included a vertical tail for steering, by analogy with the rudder of a ship. The true function of the vertical tail as a contribution to stability was not apparent until the work of the Wrights.

The giant of early French aeronautics was Alphonse Pénaud. He was the first to use wound strip rubber to power a propeller-driven model airplane. Independently of Cayley he conceived of the horizontal tail as a means of achieving longitudinal stability. Moreover, he understood how the tail worked and gave the first explanation of its function in producing stable flight.

Otto Lilienthal, a mechanical engineer working in Germany, made major contributions that directly influenced the Wrights. In 1889 he published a book, *The Flight of Birds as the Basis for the Art of Aviation,* that contained the results of extensive experiments he had carried out with the help of his brother. That work provided the first useful data on the lift and drag of curved airfoils. Lilienthal also experimented in gliding, beginning in 1891, and became the first successful hang-glider pilot. He tried both monoplane and multiplane gliders





tail could not be moved except when the pilot warped the wings to make a turn. Flying prone, he achieved warping by means of wires attached to a hip cradle. Wing warping, which was an invention of the Wrights, fulfilled the function now uniformly done with ailerons. Wilbur and Orville Wright built the entire airplane, including the gasoline engine and propellers, which they had to design from scratch.

and was killed while testing a monoplane.

In the U.S. the most serious student of aeronautics during this period was Octave Chanute, an eminent civil engineer. With his writings and his own experiments he brought Lilienthal's work to the U.S. Because of his age (he was 68 in 1900) he did no flying himself. He had several protégés who tested his gliders and their own designs.

Chanute designed a biplane glider that represents the beginnings of modern aircraft. (The design has served recently as the basis for several hang gliders.) Chanute adopted Pénaud's aft tail for longitudinal stability, and for better lift he designed a cambered wing section similar to one tested by Lilienthal. Less obvious is Chanute's most important contribution: his choice of structure for the biplane. It is the Pratt truss, which had been patented in 1844 as a design for railroad bridges. The two wings were connected by solid vertical struts, which carried compressive loads. Tensile loads were transmitted by crossed diagonal wires joining the struts in both the lateral and the fore-and-aft plane. The result was a rigid, lightweight structure that became the basis for all externally braced biplanes.

Lilienthal and Chanute were trying to solve the problem of longitudinal stability. Although they were able to design their gliders to generate sufficient lift, they had great difficulty maintaining balance in flight. They understood that achieving balance meant making the center of lift coincide with the center of gravity. The difficulty arises because the center of lift moves if the attitude of the aircraft is disturbed, as by a gust. In a hang glider the pilot can restore balance by shifting his position in order to move his center of gravity and change the attitude of the aircraft. If this maneuver is accomplished correctly and consistently, the combination of the glider and the pilot forms a stable system.

When an aft horizontal tail is properly installed, it will (as Pénaud understood) act to restore balance. The reason is that the effective center of lift of the wing and tail together remains fixed even if the attitude of the airplane is disturbed. The airplane is inherently stable, and the pilot need not shift his position. A model airplane is stable for this reason. A full-size aircraft requires more control, and so the tail is made movable and the pilot can change its lift at will. With a movable tail an unstable airplane can be flown by a sufficiently skillful pilot.

The Wright brothers were introduced to flying when as young boys they were given a small toy helicopter similar to one invented by Pénaud. News of Lilienthal's death in 1896 aroused their interest in the problem of flying. In May, 1899, Wilbur wrote to the secretary of the Smithsonian Institution asking for a list of the available literature on flight. Among the items he decided to buy was Chanute's book *Progress in Flying Machines,* written in 1894. After studying the book Wilbur wrote to Chanute, initiating an exchange of letters that continued for nearly 10 years. For most of that period Chanute was a source of both information and encouragement to the Wrights. The letters and the diaries of the Wrights constitute a detailed chronicle from which one can reconstruct the events leading to powered flight.

The Wrights Begin

By the time the brothers began experimenting seriously they had acquainted themselves thoroughly with the aeronautical literature. They understood the prevailing views that longitudinal stability required a Pénaud tail, that lateral stability should be achieved by having a dihedral angle in the wings and that steering could be accomplished with a vertical tail or rudder. Yet they eventually rejected all three notions, producing designs that were quite different from the basic one conceived by Cayley and generally accepted by their contempo-

EARLY GLIDER was designed by Otto Lilienthal, who is seen here flying it over a brickyard on the outskirts of Berlin in 1894. Lilienthal was the first pilot of what would now be called a hang glider. He was killed in 1896 when one of his gliders lost lift and went into a sudden dive. The irregular horizontal line passing near the heads of the spectators in this photograph is a crack in the original glass negative.

raries. Their first powered airplane, the 1903 Flyer, had a forward horizontal surface, the tips of the wings drooped with negative dihedral and the vertical tail moved only when the wings were flexed to turn the aircraft. The reasons for the configuration adopted by the Wright brothers tell the story of their research and engineering.

Probably the single most important motivation in the Wrights' program was that they themselves were resolved to learn to fly. That determination influenced virtually all their decisions. Since they were to be pilots, they had to give careful consideration to stability, control and safety. They were not inclined to build a machine, hope it would work and hire somebody else to do the testing. Moreover, they were clearly not going to be satisfied just with gliding straight ahead.

Lilienthal's death and a similar accident in England convinced Wilbur at the outset that to be safe an airplane should have the horizontal stabilizing surface in front of the main wing (a system that came to be called the canard configuration). No one at the time understood stalling (the loss of lift), and the Wrights had no scientific basis for determining the relative merits of aft and forward surfaces. While they were experimenting, the theoretical conditions for stability were being developed by others, but not until after 1903 was it widely known that an airplane can be made longitudinally stable with either an aft or a forward horizontal surface. The forward surface, which was the most unusual feature of the early Wright aircraft, had for the brothers the advantage that the pilot could see it, thereby gaining a measure of safety by being able to respond quickly if something went wrong with it in flight.

Unlike his predecessors, Wilbur did not assume that the problem of lateral stability had been solved by Cayley's invention of dihedral. "My observations of the flight of buzzards," Wilbur wrote to Chanute, "lead me to believe that they regain their lateral balance, when partly overturned by a gust of wind, by a torsion of the tips of the wings." Having made this fundamental observation, Wilbur then had to solve the problem of inventing a comparable method of control for a man-carrying aircraft.

He eventually realized the idea as wing warping, which was the predecessor of the ailerons that fulfill the function in modern aircraft. With wing warping the Wrights were able not only to stabilize the airplane in a straight path but also to turn it. And that is how they advanced far ahead of their competitors. By employing a movable horizontal surface and wing warping the Wright brothers fused the tasks of stability and control, thereby making their most important contribution to the development of the modern airplane.

HAVE SAID ABOUT THE AUDI 5000.
ING ABOUT THE AUDI 4000.

and dealer delivery charges additional. Metallic paint, additional.

For some months now, there have been reports from the Continent about an exciting new car — "the Audi 5000's scrappy kid brother"— designed by the same team of engineers that created the 5000.

In the February issue of Road & Track there appeared an advance road-test report on this new car, written by R & T's "Letter from Europe" columnist (whose credentials as a driver include the winning of the LeMans 24-hour classic some years ago).

Here is what this gentleman had to say about the new Audi 4000—or the Audi 80, as it is called in Europe—after "a quick run through Germany, Switzerland, Italy and France and over several passes in the Alps, as well as over some 250 miles of motorways where the Audi cruised at around 90-100 mph..."

"...a brilliant piece of engineering..."

"...when I tried the car on the 'colonial' tracks of a rival manufacturer's proving ground ... there wasn't the slightest squeak or rattle. The light weight is obtained by proper stressing of the various components of the shell and systematical weight saving where it can be done without harm to comfort, quality and strength. This is a game at which Ferdinand Piëch, who leads the Audi developmental team, was unbeatable when he designed racing cars from 1965 to 1972, and he certainly hasn't lost his touch."

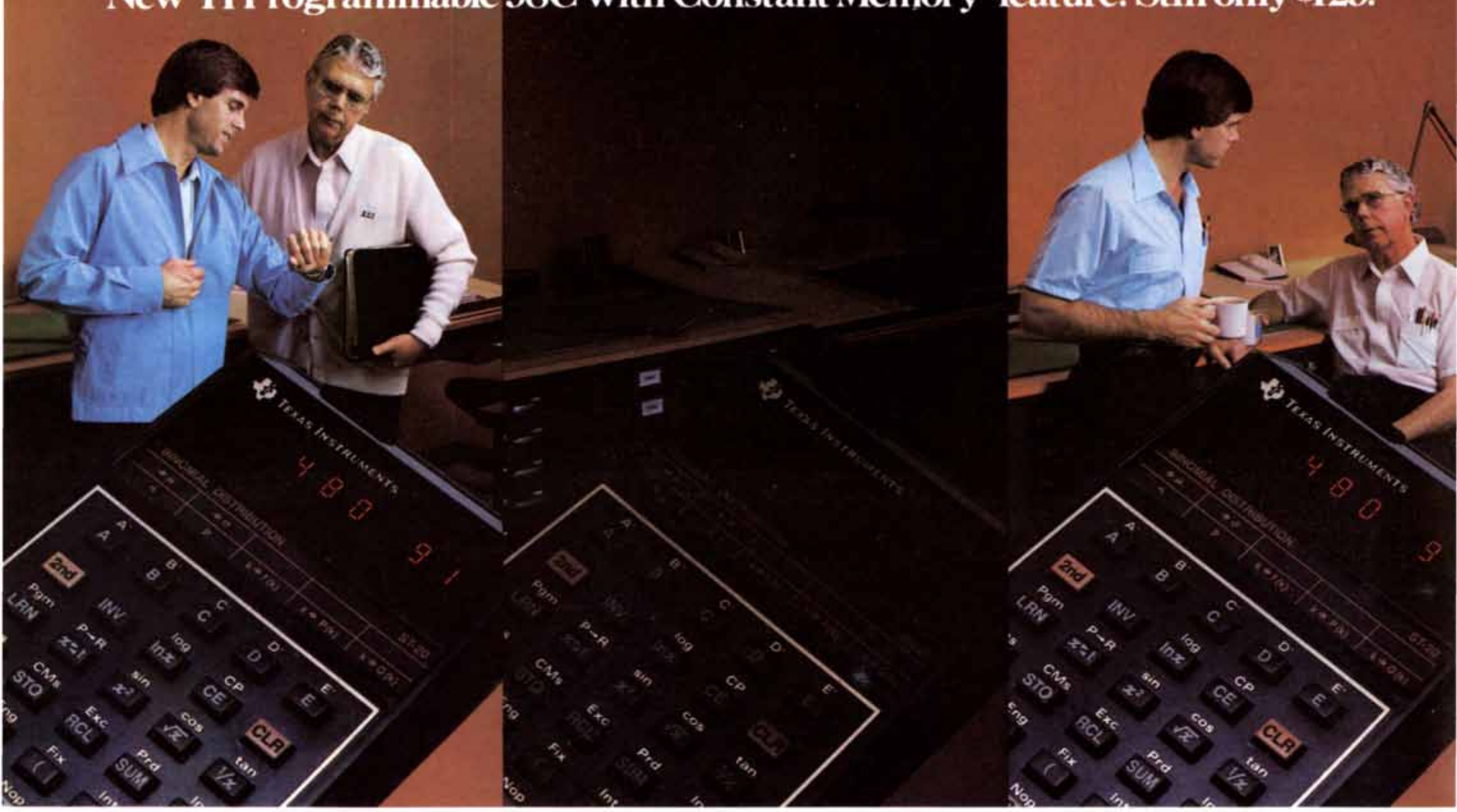
"...by far the best car announced in the past year. In fact it's the best full-size European car under two liters."

(—March "Letter from Europe" column, R & T.)

Test drive the new Audi 4000, and find out why the Europeans are so impressed with it. For nearest Porsche+Audi dealer, call this toll-free number: (800) 447-4700. (In Illinois, (800) 322-4400.) He can also supply you with information on leasing and overseas delivery.

PORSCHE+AUDI

Nothing Even Comes Close



If you enter a 480-step program today, the new TI Programmable 58C will remember it tomorrow...even if you turn it off overnight!

We haven't changed the price, but we've made the TI-58 even more versatile by adding a Constant Memory feature that retains data or program information even when the calculator is turned off. You can turn off your TI Programmable 58C right in the middle of the most complex problems, for hours or even days, and pick up right where you left off when you turn it back on.

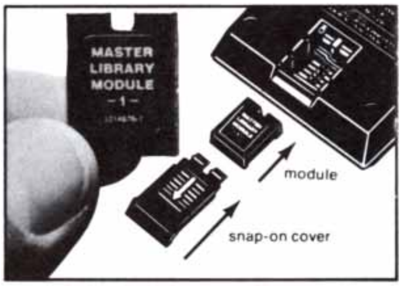
The TI Programmable 58C, with plug-in application libraries and more than 170 functions and operations in scientific, engineering, and statistical fields ... now an even greater value with Constant Memory feature.

More easy-to-use software.

Take advantage of the versatility of the new TI Programmable 58C and new software. Twelve optional *Solid State Software*™ libraries are now available. Each plug-in library module contains dozens of prewritten programs. Solid-state libraries are available in a variety of fields including engineering, science, statistics, and business. From $35.*

The *Professional Program Exchange* lets you exchange programs with members. Over 1,500 programs are already available.

Specialty Pakettes are conveniently formatted program listings available in 16 subject areas. Just follow the program listings and tackle a wide variety of problems without the need to do any conventional programming. With the TI Programmable 58C's Constant Memory feature, you won't lose your program if you have to turn the calculator off. $10* per subject.



In addition, *Custom Software Modules* can be developed for specialized applications and a unique *Professional Productivity Program* is available for corporate use.

Printing and plotting capability.

Add the PC-100C printer/plotter to your TI Programmable 58C and you can plot, prompt, and print inputs and outputs automatically. $200.*

Official Computer and Calculator Supplier of the 1980 Olympic Winter Games

Free literature.

We want to tell you more about our products and services. For additional information write: Texas Instruments, Programmable Calculators, P.O. Box 53, Lubbock, Texas 79408.

*U.S. suggested retail price.

Texas Instruments technology –bringing affordable electronics to your fingertips.

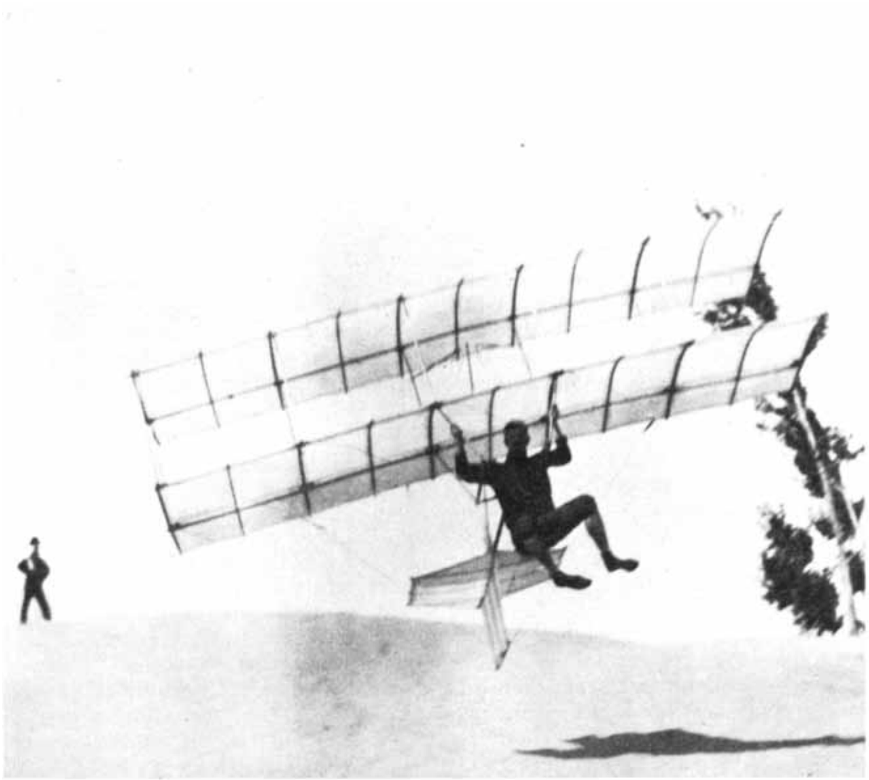
TEXAS INSTRUMENTS
INCORPORATED

45645A

Several months after Wilbur had corresponded with the Smithsonian Institution, he built a biplane kite with a five-foot wingspan. Kites as they are now known (particularly box kites) had been developed quite recently. Wilbur followed the latest construction practices, but the basic structural design of his kite was the Pratt truss, which he had heard about from Chanute. He removed the fore-and-aft bracing wires in order to have the wings free to slide with respect to each other, thereby producing the wing warping. The only rigid parts of the structure were the connections between the control surface and the uprights to which it was attached. The four control cords were arranged so that the operator could exercise both longitudinal and lateral control of the kite, independently or simultaneously.

Flight Testing

Wilbur flew this first test vehicle one afternoon in August, 1899, when Orville was away on a camping trip. Although such a kite is somewhat sensitive to fly, Wilbur succeeded, convincing himself that his ideas for longitudinal and lateral control were sound. He found also that the kite would fly with the horizontal tail either forward or aft. This remarkable kite was the first flying device that could be controlled both laterally and longitudinally.



BIPLANE GLIDER was designed in 1896 by Octave Chanute, a civil engineer who was the most serious student of aeronautics in the U.S. before the Wrights began their work. They carried on a long correspondence with him, and he was on hand for many of their early flights. The pilot for this flight in 1897 in Indiana was A. M. Herring, one of Chanute's associates.

The design of the 1899 kite was the basis for a scale-up version that was small enough to be flown as a kite but large enough to support a man. It had a wingspan of 17 feet and a wing chord (width) of five feet; its weight was 50 pounds empty, 190 pounds with a pilot, who flew prone. With the pilot aboard the wing loading was 1.15 pounds per square foot. (A high-performance sailplane today might have a wing loading of five pounds per square foot.)

Wilbur intended to fly this glider-kite of 1900 initially from a tower to which it would be attached by ropes. If steady winds were available, he and his brother could gain hours of experience in operating the controls without the risk of flying free. They estimated that in five years of work Lilienthal had logged only about five hours in the air, which was hardly enough time in which to learn how to fly.

Chanute suggested a number of places where the Wrights would be able to find steady winds at the time of year when they could experiment. (Because of constraints imposed by their bicycle business they could leave their home in Dayton, Ohio, only between September and January.) After verifying the probable wind conditions with the Weather Bureau they decided to go to Kitty Hawk on the coast of North Carolina. Kitty Hawk was a small, isolated settlement that could be reached only by boat after a day's train ride from Ohio. The brothers not only had to transport their aircraft and all their equipment from the mainland but also had to obtain materials for building their tower.

The Wrights spent less than a month near Kitty Hawk in 1900, experimenting with their glider between October 5 and 18. Most of the time they operated the glider as a kite, first from a 10-foot tower but usually from the ground with hand-held ropes. The total time in free flight was about two minutes, of which roughly 10 seconds was spent with a man on board. Only Wilbur flew. The controls worked well when the glider was flown as a kite, but in the brothers' limited amount of piloted-flight time they did not master the adaptation of the controls to direct operation.

It is apparent from the diaries that the first experiments were often frustrating. Orville noted: "We tried it with tail in front, behind, and every other way. When we got through, Wil was so mixed up he couldn't even theorize. It has been with considerable effort that I have succeeded in keeping him in the flying business at all." The flights were short and there were many hard landings. Fortunately the structure was simple and easily repaired.

With spring scales attached to the restraining ropes the brothers made measurements of lift and drag. They determined the speed of the wind with an anemometer. These were the first quantitative data for the performance of any full-scale lifting structure.

Although the Wrights' flying program in 1900 was short, it yielded fundamental results and served to "confirm the correctness of our original opinions." The brothers proved their techniques of longitudinal and lateral control; they believed even more strongly that they had to practice flying, and from their measurements they learned that flying in a prone position greatly reduced the drag on the machine.

Results of Testing

Not all their conclusions were so positive. It was at this time that Wilbur decided not to make the wings with a dihedral angle. Although the configuration aided stability in still air, the slightly raised tips of the wings tended to accentuate the effect of gusts from the side. Since the Wrights had lateral control through wing warping, they did not need the stabilizing effect of dihedral.

By far the most disappointing results of the experiments in 1900 were the values for lift and drag that the brothers found. The drag of the structural frame when no lift was being generated was much less than they had expected. Yet for a given angle between the wing and the wind the total drag was greater and the lift smaller than they had anticipated. (The Wrights were working from data on a flat plate perpendicular to an airstream reported by John Smeaton in 1759 and on measurements for cambered airfoils made by Lilienthal.) The brothers observed in their flight tests that the glider operated at an angle of incidence much larger than the one they had expected. Lift increases with angle of incidence and must equal the weight of the glider for steady flight. At a given angle of attack the Wrights' wings produced less lift than the brothers had predicted on the basis of Lilienthal's data.

That was a surprising result, but the Wrights believed their own data and considered three possible explanations: (1) the fabric covering was not sufficiently airtight, (2) the camber of their airfoil (1 : 22) was too shallow for Lilienthal's data to be applicable or (3) Lilienthal's data were wrong. They favored the second reason and designed their next glider, which was to be flown in the fall of 1901, with a more highly cambered airfoil.

With the interpretation of their results in 1900 the Wrights stopped depending on the work of others. From then on their progress was to be entirely the result of their own discoveries. They began work on their 1901 glider during the winter. To get more lift they increased the wingspan to 22 feet. With a chord of seven feet the total covered wing area was 290 square feet; the stabilizer area was 18 square feet. This was by far the largest glider anyone had tried to fly. The weight of the structure had become 98 pounds and the wing loading, with a pilot, was .78 pound per square foot. The only other major changes were the increase of the camber from 1 : 22 to 1 : 12 and modifications of the controls so that the pilot could more easily operate both the stabilizer and the wing warping. Although the glider did not look much different from the 1900 model, it was in some respects the most important of the Wrights' research devices. Indeed, the year 1901 was the watershed for almost all their later accomplishments.

Expanded Testing

Now the test program was expanding. The brothers built a wooden shed at Kill Devil Hills, four miles from Kitty Hawk, where large sand dunes afforded good launching platforms. Because of business pressures they conducted their flying experiments earlier than they had originally planned, from July 10 to August 20.

According to their diaries, Wilbur did all the flying in 1901. On the first day of testing, July 17, he made 17 glides without having tested the machine as a kite. These first flights were nearly disastrous, mainly because of the problem of longitudinal control. Wilbur found that full deflections of the forward horizontal control (the canard) were often required and that the flight path of the glider was severely undulating. Twice the angle of incidence was so high that the machine lost forward motion and stalled. Each time the glider simply mushed slowly to the ground rather than diving as Lilienthal's had. These two recoveries from the conditions Wilbur had feared satisfied the brothers of the worth of the canard configuration. From then until 1910 they had no interest in trying the aft tail.

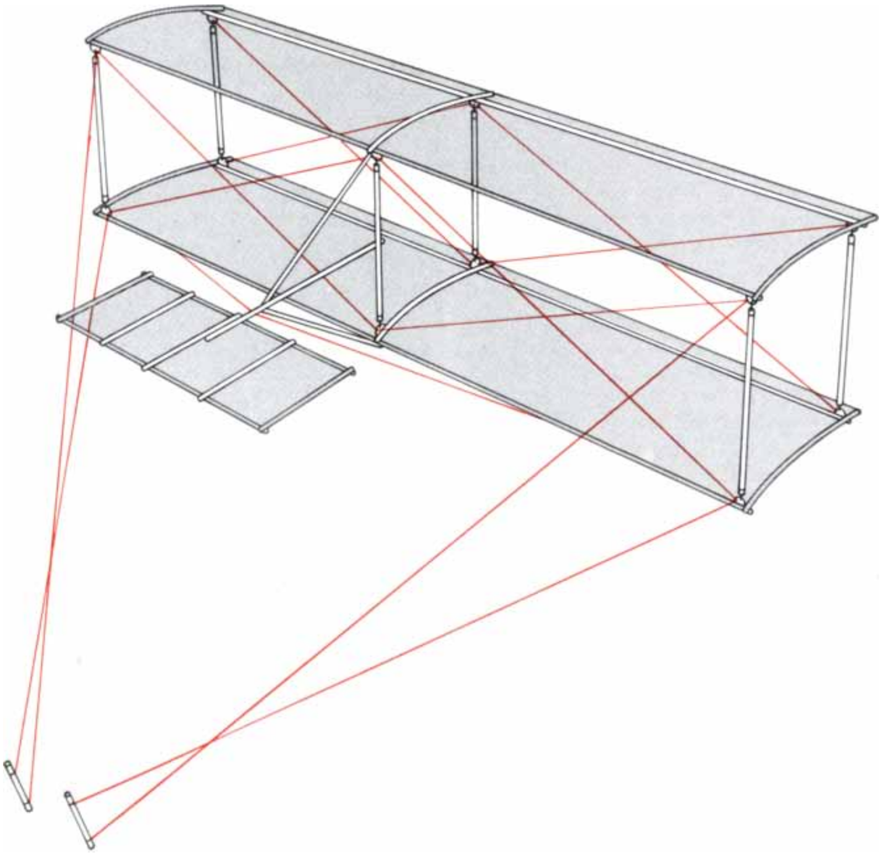
Wilbur was faced with the problem of determining why the glider was behaving so erratically. No noticeable improvement followed a reduction of the area of the control surface from 18 square feet to 10. The brothers then flew the glider as a kite and in one afternoon found that the center of lift was traveling backward as the angle of attack was decreased.

It was known at the time that for a flat plate the center of lift progresses continuously from the center, when the plate is perpendicular to the airstream, to the leading edge, when the angle of incidence is zero. In testing a full-scale wing the Wrights discovered a fundamental property of airfoils: camber causes the motion of the center of lift to reverse as the angle of incidence is changed.

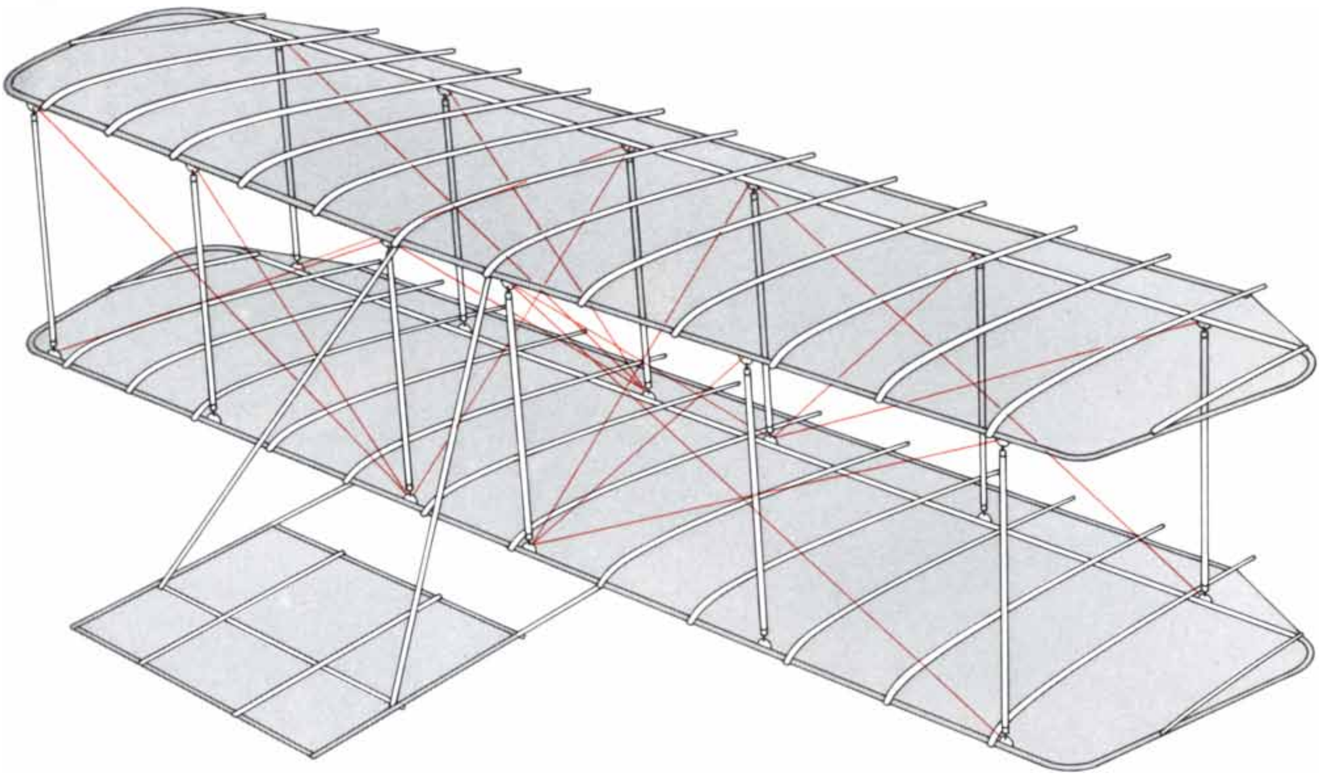
The reversal of the motion of the center of pressure with angle of incidence was the cause of the difficulties Wilbur had in controlling the glider. The solution lay in reducing the camber of the wings. The brothers went back to the camber of the 1900 glider by installing king posts on the lower wing and rigging truss wires to pull the middle sections of the ribs downward on both wings. The results were immediate and satisfying. Wilbur made some 30 glides after the modification; the longest one lasted 17.5 seconds and the greatest distance was 390 feet.

During the last week of testing Wilbur made a fundamental observation about the behavior of an aircraft when it is turned. He recorded in his diary that the "upturned wing seems to fall behind, but at first rises." On August 22 he wrote to Chanute: "The last week was without very great results though we proved that our machine does not turn (i.e., circle) toward the lowest wing under all circumstances, a very unlooked for result and one which completely upsets our theories as to the causes which produce the turning to right or left."

Turning an airplane is done not by steering with a rudder but by rolling about the longitudinal axis. To turn to the right, for example, the right wing is lowered and the left one is raised. The direction of the total lift is thus tilted to the right of the vertical and part of the lift acts to accelerate the airplane to the right. It was the Wrights who discovered



FIRST AIRCRAFT built by the Wright brothers was this biplane kite of 1899. It had a wingspan of five feet and embodied as its basic structure the Pratt truss, which had been designed for railroad bridges and adapted for aircraft by Chanute. The kite also embodied Wilbur Wright's concept of wing warping. The four cords gave the operator lateral and longitudinal control.



GLIDER-KITE OF 1900 was the first Wright aircraft large enough to support a man. It was a scale-up version of the 1899 kite, having a wingspan of 17 feet and a chord (wing width) of five feet. The Wrights tested it, mainly as a kite, at Kitty Hawk, N.C., in October of 1900. Only Wilbur flew as the pilot, for a total of about 10 seconds. Although the controls worked well when the Wrights flew the aircraft as a kite, Wilbur found in his brief attempt at piloting that he could not master the adaptation of the controls to direct operation.

the correct method of turning, a discovery made possible by their invention of wing warping for lateral control.

While Wilbur was practicing turns he discovered what is now called adverse yaw. Since initially the lift on the raised wing is greater than that on the lowered one, so is the drag. The different drag forces tend to turn the airplane in the direction opposite to the one intended when the rolling motion is initiated. Wilbur's observations led to the installation of the vertical tail on the 1902 glider.

Because Wilbur was simultaneously a theorist, an inventor, a builder and a test pilot he had in three weeks of flying been able to make two crucial discoveries, one being the influence of camber on the motion of the center of lift and the other being adverse yaw. His responses essentially completed the configuration of the aircraft.

The Wind Tunnel

The experiments of 1901 convinced the Wrights that Lilienthal's data for curved airfoils were wrong. More careful measurements had confirmed the result found in 1900 that the predicted lift was too large. Beginning in the fall of 1901 the brothers carried out a lengthy series of tests in a wind tunnel.

Theirs was not the first wind tunnel or even the first in the U.S. They probably got the idea from articles in *Aeronautical Annual,* which was published by James Means in Boston in the years 1895–97. The contribution of the Wrights was characteristically to the point. They recorded accurate data for many airfoils and wing shapes. They did not make absolute measurements but rather compared the forces acting on an airfoil with those acting on a flat plate oriented perpendicular to the airstream. As a result of their tests they had by midwinter all the data they would need to design their aircraft for the next decade.

The Wrights also decided on the basis of their own data that Lilienthal's measurements were not as inaccurate as they had initially believed. For their predictions of lift on their gliders they had needed a constant (related to the drag on a flat plate), which had been inferred from the data published by Smeaton. By comparing the wind-tunnel findings with the measurements of lift on their gliders the Wrights determined that the value of the constant was wrong. The error had caused them to overestimate the lift by 40 percent.

On August 28, 1902, the Wrights arrived again at Kill Devil Hills with another glider. This model had a span of 32 feet and a chord of five feet, so that its total area and wing loading were about the same as those of the 1901 glider. From the wind-tunnel tests the brothers had learned that a long, narrow wing, which has a higher aspect ratio, is more efficient than one with the same area but a shorter span. Their choice of cross section for the airfoil was also based on the wind-tunnel tests.

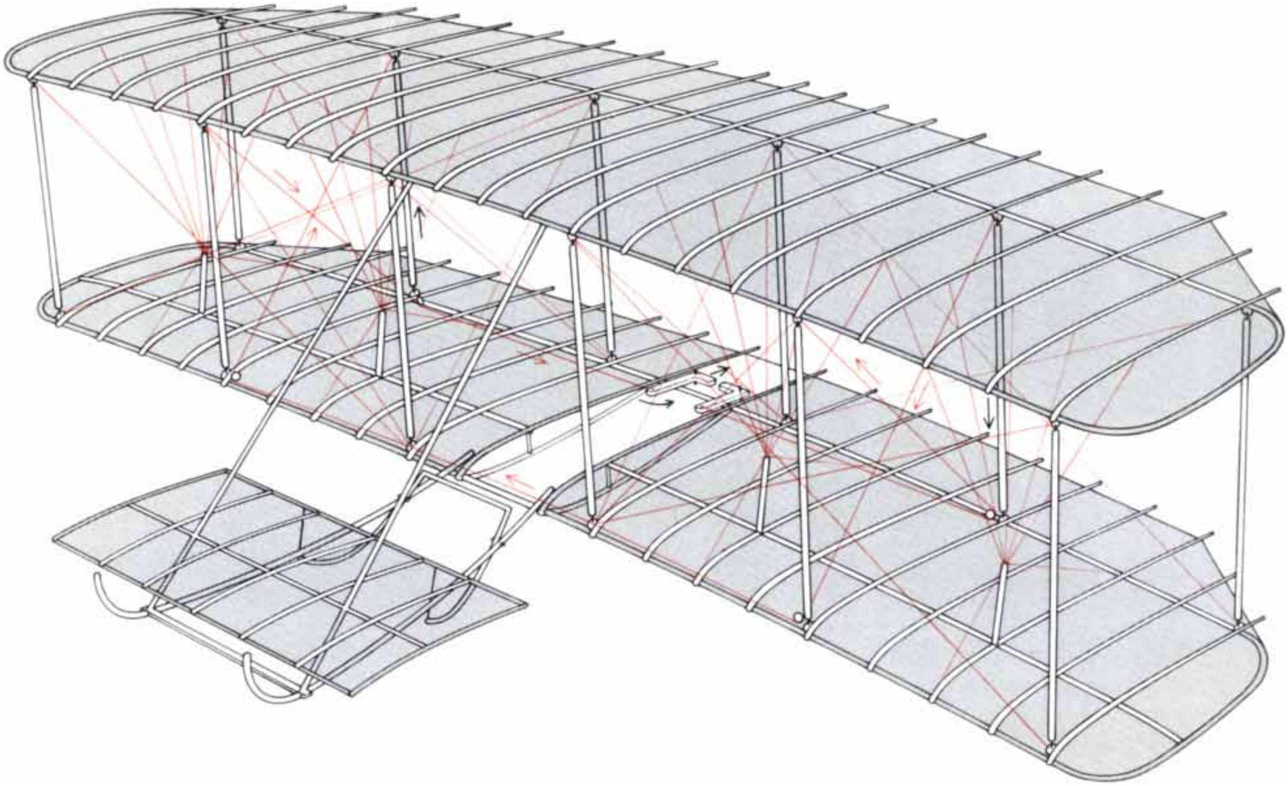
The 1902 machine was the first to include the hip cradle the brothers devised to enable the pilot in a prone position to operate the warping wires by shifting his hips laterally. The deflection of the horizontal surface was changed by operating a lever with the left hand.

The most obvious and important change of configuration was the addition of a double vertical tail. It had a total area of 12 square feet. This surface was rigidly mounted in the 1902 glider and was never intended for steering. The Wrights had figured out correctly that the vertical tail was needed to counteract the adverse yaw.

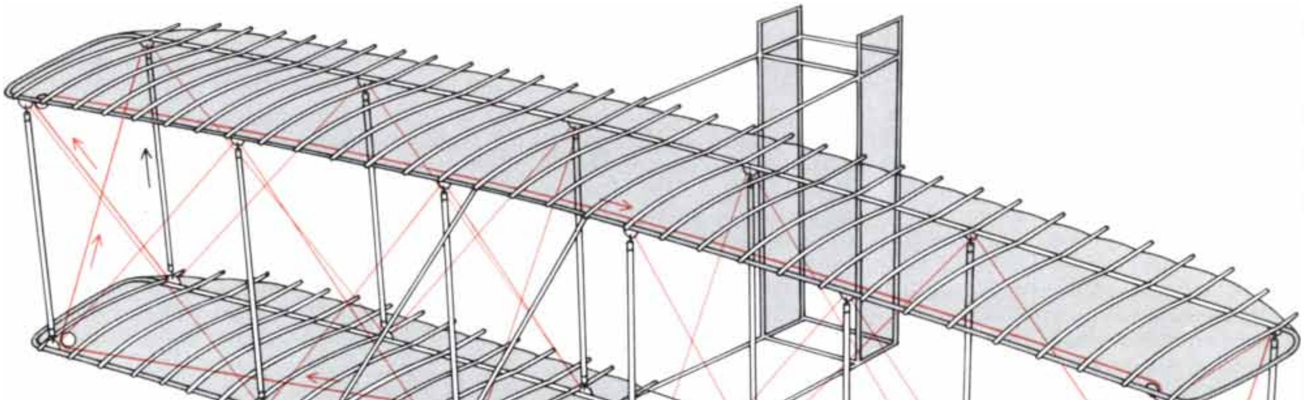
Tests of the glider as a kite showed the brothers they had a better airplane. They found that with the new wing they had a much-improved lift-drag ratio and that the difficulties with longitudinal control were reduced. Moreover, the fixed vertical tail did act to reduce the yawing tendency in a turn.

Initial glide tests showed an excessive response to side gusts, and so the wings were set with negative dihedral. Further testing showed that the 1902 glider gave the pilot even more trouble with lateral control than the 1901 machine had. When the glider was hit by a side gust, it had an uncontrollable tendency to oscillate in yaw.

Much more serious problems arose in several of the turns. Both brothers expe-



GLIDER OF 1901 was by far the largest glider anyone had tried to fly. It had a wingspan of 22 feet and a chord of seven feet; the weight without pilot was 98 pounds. In tests at Kitty Hawk, with Wilbur as the pilot, the glider at first followed a strongly undulating flight path. The Wrights modified the aircraft to reduce the camber they had added to the wings since 1900 and then found that it flew much better.



GLIDER OF 1902 provided the Wright brothers with most of their flying experience. It was the first of their aircraft to have the double vertical tail, which was added to compensate for the tendency of the craft to slew in a turn, and to include the hip cradle. The craft had a wingspan of 32 feet and a chord of five feet. The Wrights had learned from wind-tunnel tests that a long, narrow wing is more efficient than a shorter one with the same area. Another feature of the 1902 glider is that the wings are set in negative dihedral, that is, they droop.

rienced the beginning of a spin in the direction of a turn. The stall-spin sequence in a turn has been estimated to be the cause of loss of control in 70 percent of all flying accidents. The Wright brothers were the first to discover it. They began a turn, say to the right, by warping the trailing edges down on the left wings and up on the right. The force acting on the vertical tail also tended to turn the airplane to the right, counteracting the adverse yaw. If the turn became too steep, the warping had to be reversed. This maneuver gave rise to adverse yaw acting in the same direction as the turning effect of the fixed vertical tail. For a short time the aircraft was therefore turned even more to the right: the beginning of a spin. It could also happen that excessive downward warping of the wings inside the turn would cause them to stall and drop. Several times the brothers found that this series of events ended when the wing that was inside on the turn dug into the sand and the airplane pivoted to a stop.

It was Orville's idea to correct the behavior by replacing the double fixed tail with a single movable one. Wilbur suggested that its operation should be tied to the warping wires; it therefore could not be operated alone. Nevertheless, a reversal of the warping also reversed the yawing tendency produced by the movable tail, and the net result was a significant reduction in the tendency to spin.

These various adjustments completed the configuration and the control system the brothers employed in the Flyer of 1903. Not until the end of 1904 did they disconnect the vertical tail and finally achieve independent control of pitch, roll and yaw. Virtually all modern aircraft depend on the same system of controls.

Providing Power

The prospects for constructing a flying machine appeared to be much more favorable after the invention of the internal-combustion engine late in the 19th century. Several people before the Wrights had worked on the problem, and their contemporary Samuel P. Langley, secretary of the Smithsonian, had reached the point late in 1903 of testing a powered, man-carrying airplane. In two widely publicized tests, in which the craft was launched by catapult from a houseboat on the Potomac, his plane failed to fly and dived into the river. The progress of the Wrights was not publicly known, but they were feeling the competition.

Orville and Wilbur had begun work on their Flyer in the fall of 1902. Their plan was to install an engine and propellers on an improved version of the 1902 glider. They assumed that they could adapt an automobile engine and that the marine technical literature would provide the basis for designing efficient propellers.

After trying unsuccessfully to obtain a suitable engine from manufacturers they decided to design and build their own. This was not a particularly difficult decision for them, because they had already built a gasoline engine to power machinery in their bicycle shop. By March, 1903, they had what they sought: an adequate engine that developed 12 horsepower and weighed 15 pounds per horsepower. They had not tried to create an exceptional engine, and their result was far inferior to the engine built for Langley by his engineering assistant Charles Manly.

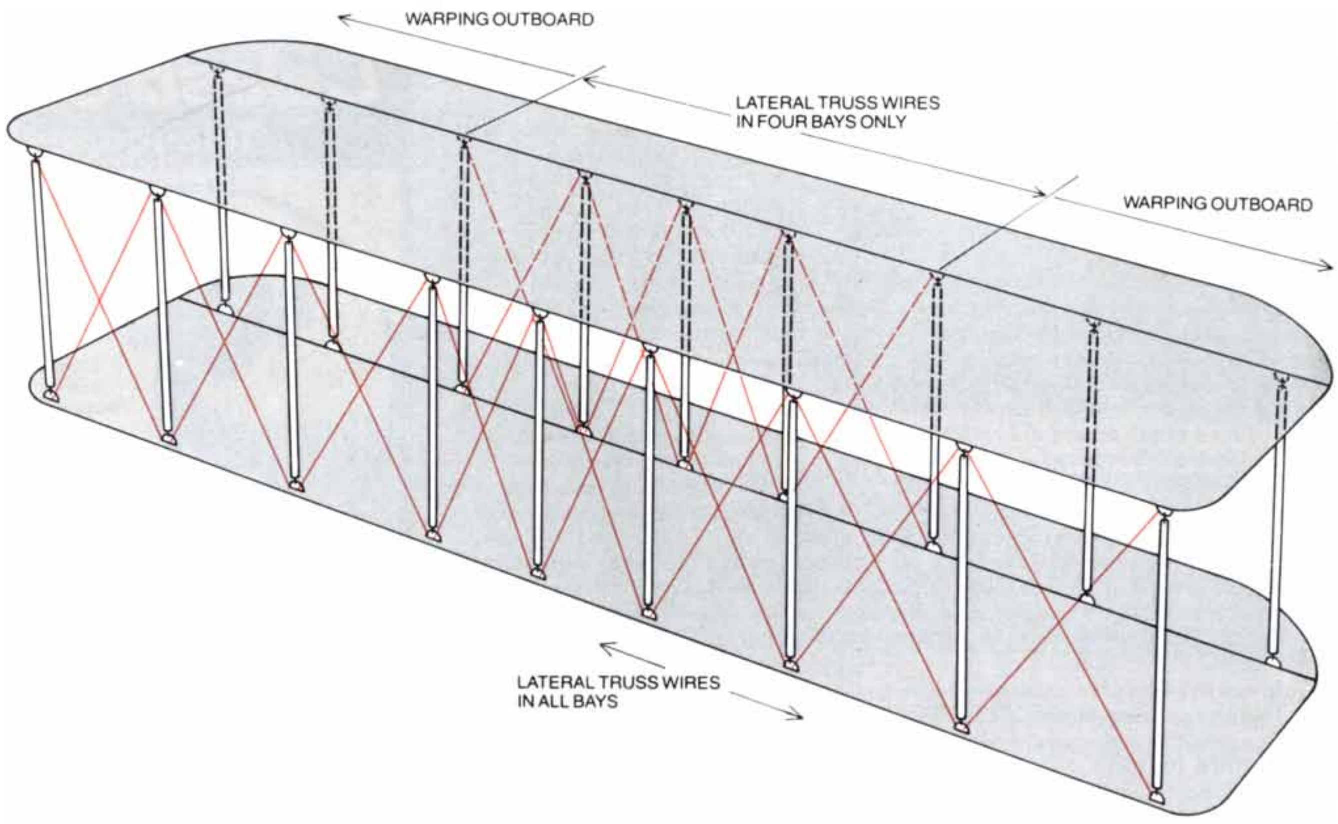
Then the brothers were surprised to find that the available propeller theory provided only one useful result: for a given output of power the best efficiency is obtained by passing the largest possible amount of air through the propeller. (This principle accounted for the rather large diameter of their two propellers: 8.5 feet.) The most significant consequence of their discovery was that they were forced to work out their own methods for designing propellers.

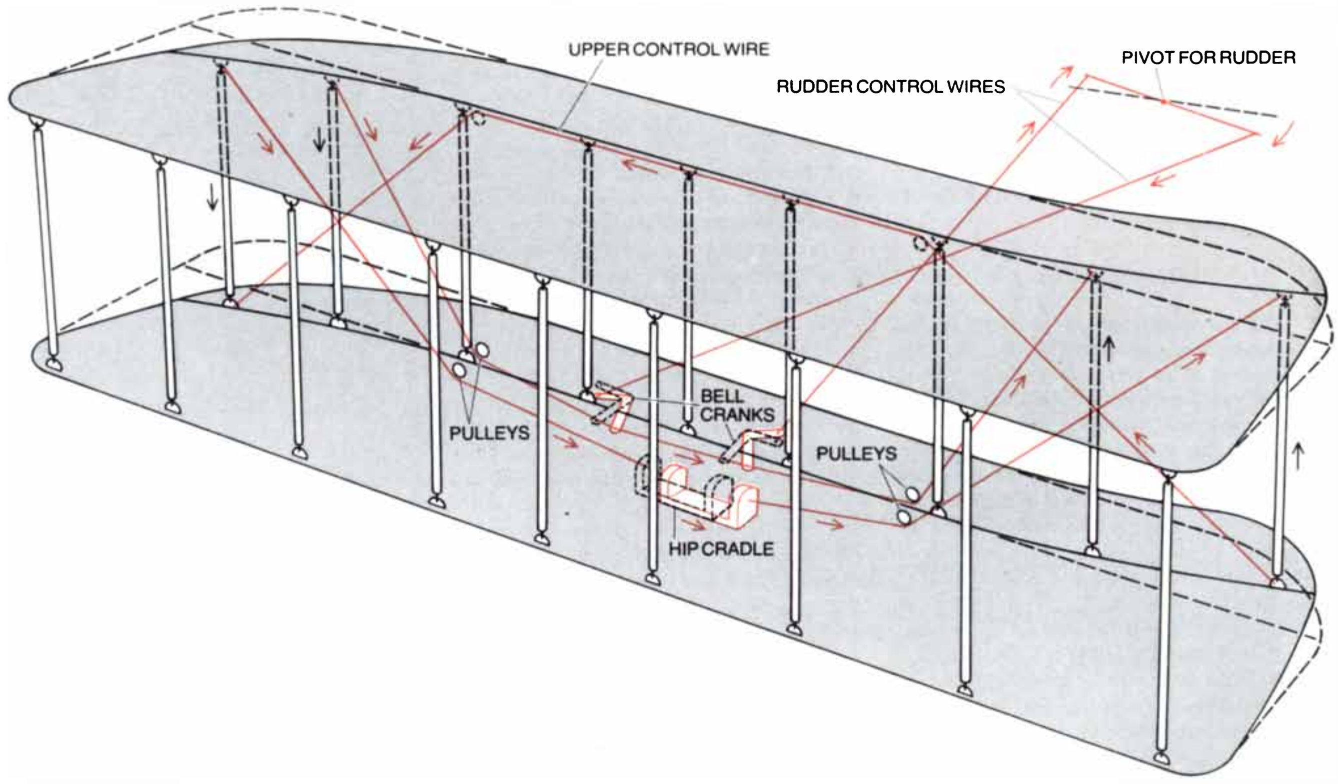
No theory existed for the detailed design of any kind of propeller. The Wrights employed their knowledge of airfoils to develop what later became known as blade-element theory, based on the idea that a small segment of a propeller blade is treated as a section of wing. Since the propeller is rotating and the airplane is moving, each such little wing is exposed to a different velocity and therefore generates different lift and drag forces. The thrust of the propeller and the power absorbed by it are found by summing the lift and drag contributions from all such blade elements.

The Wrights' analysis showed them how to design propellers with an efficiency of 70 percent or better at a time when the propellers of other investigators had an efficiency of less than 50 percent. With the brothers' superior propellers they were able to power the 1903 Flyer with their relatively modest engine. Their work on propeller design was their greatest analytical accomplishment, but for them its only value lay in the practical consequences. They never published their ideas and analyses, and modern propeller theory is traceable to other sources.

The First Flight

The engine and the propellers were installed in an aircraft that had the same configuration as the 1902 glider but was larger, more rugged and improved in construction. Anyone examining the aircraft or the engineering drawings must be impressed by the Wrights' mastery of techniques based on wood, wire





WING-WARPING SYSTEM devised by the Wright brothers as a means of executing turns is depicted. At the top is a front view of the wings; each of the eight bays between struts is fitted with truss wires, whereas only four of the bays at the rear have them. Thus the outer edges of the rear wings could be moved (*bottom*) by the wires attached to the hip cradle. The same system of wires also actuated the cranks that moved the vertical tail. In order to make a turn, say to the right, the pilot would warp the right wings up and the left wings down.

and fabric. The Flyer had a wingspan of 40 feet, a chord of 6.5 feet and a total area of 510 square feet. The weight with a pilot was 750 pounds, giving a wing loading of 1.5 pounds per square foot.

Wilbur and Orville arrived at Kill Devil Hills on September 25, 1903, this time for their longest stay, namely three months. They rebuilt their 1902 building and constructed a new one (44 by 16 feet) for the Flyer. When the weather was favorable, they sharpened their flying skills with the 1902 glider. They spent the rest of the time assembling the new airplane and carrying out preliminary tests of it.

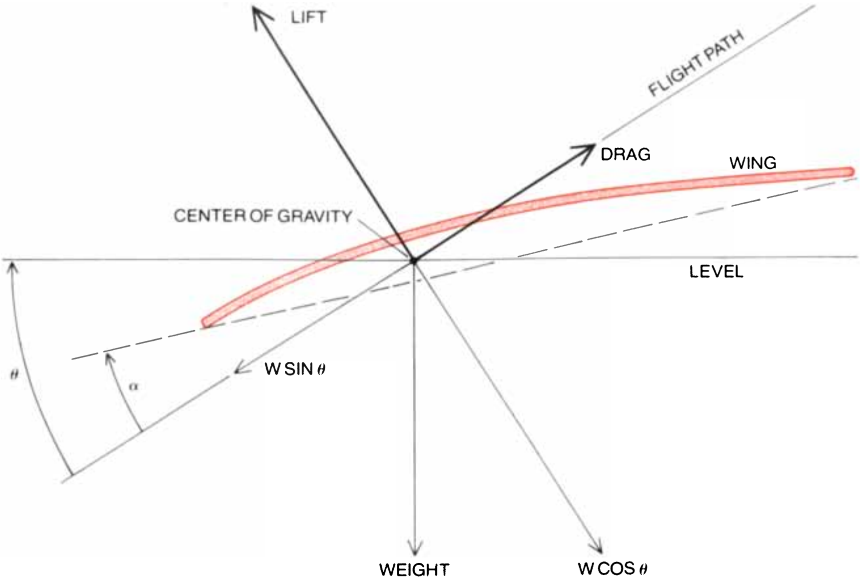
In October the brothers learned of Langley's first unsuccessful test. The news may have been encouraging, but their own chance of success was marginal and they knew it. Their airplane was 25 pounds heavier than they had expected. The machine had not been assembled before the trip to Kitty Hawk, so that the engine and the propellers had not yet been operated on it. In the first test the engine ran roughly, and then two propeller-shaft attachments failed. They were sent to Dayton for repairs on November 5.

Until the shafts were returned on November 20 the brothers could do little. Time was running short, inasmuch as the Wrights knew Langley was preparing for another test. Weather conditions were becoming a problem, with the temperature occasionally dropping below freezing at night, but by November 23 enough tests had been run to measure the total thrust: 132 pounds. Because the brothers had estimated the drag to be 95 pounds they were now confident their machine would take off and fly.

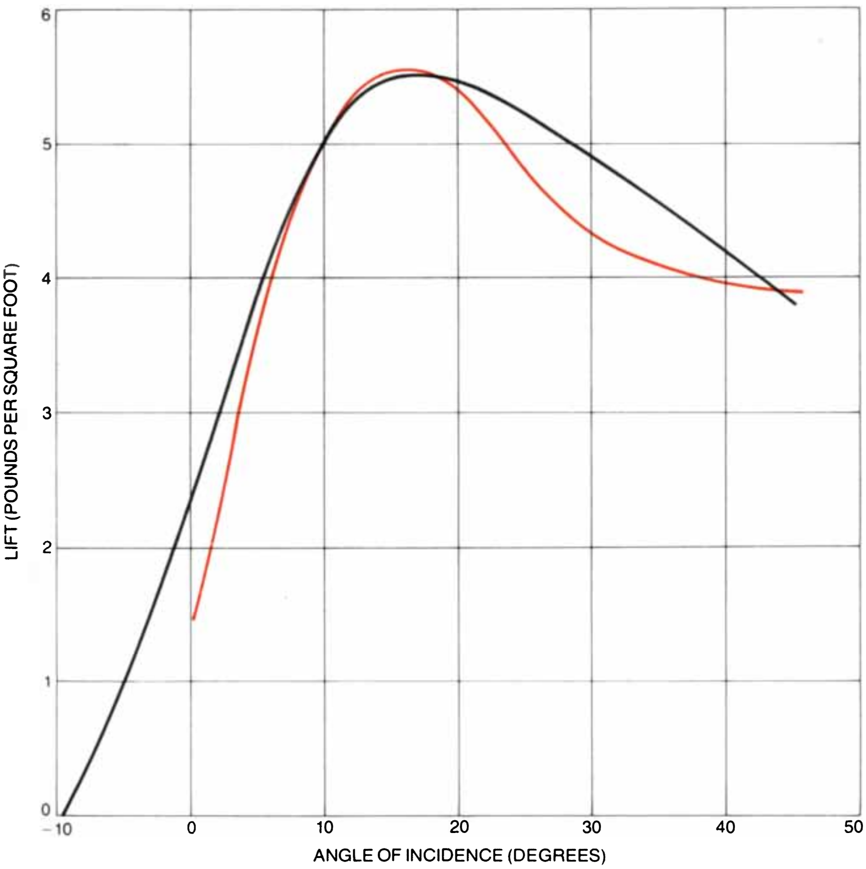
While the engine was being run to test instruments on November 28 a propeller shaft cracked. It had to be replaced, and so Orville returned to Ohio to make two new shafts of stronger material. On the train leaving Dayton for the trip back to Kitty Hawk he read newspaper accounts reporting that Langley's flying machine had been destroyed on launching the day before, December 8. The competition was out of the race.

With all repairs finished the Flyer was prepared for tests on December 14. Wilbur won a coin toss and became the pilot for the first flight. The launching rail was laid on a slight incline so that gravity would aid the launching into a weak breeze. In a slight crosswind the aircraft took off but soon stalled, striking the ground 60 feet from the end of the rail. The forward control surfaces and their supporting structure were damaged but were easily repaired.

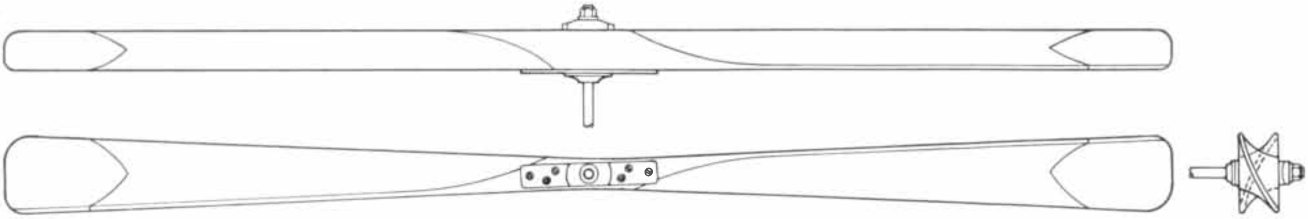
Orville was pilot for the second test, which took place on December 17. The wind was steady at 27 miles per hour, which was a dangerously large fraction of the Flyer's cruising speed of from 30



FORCES ON A GLIDER WING include the weight (*W*), the lift and the drag. In the procedure employed in about 1900 to design a wing surface the glider was assumed to be flying steadily (here toward the left) at a velocity (*V*) along a path making the angle θ below the horizon. A small portion of the weight (*W sin* θ) acts to pull the glider forward, just balancing the drag. The lift was defined as being perpendicular to the flight path, supporting the other component (*W cos* θ) of the weight. The lift and the drag are functions of the angle of incidence (α).



DATA ON LIFT obtained by Lilienthal in 1889 (*black*) and by the Wrights in 1901 (*color*) are compared. The Wrights had to contend with the fact that their first intensive flight tests, which they made with their 1900 glider, showed that their wings were generating less lift than Lilienthal's data had led them to expect. Their first solution for correcting the situation was to increase the camber of their wings; they then carried out wind-tunnel tests. They eventually discovered that their results were quite close to Lilienthal's when the angle of incidence of the wing was between five and 10 degrees and that the source of the error lay elsewhere.



WRIGHTS' PROPELLER DESIGN, which they worked out themselves after finding that the published data on marine propellers offered little useful information, was based on the principle that a small segment of the blade could be treated as a wing section. The marine data suggested that the most efficiency for a given amount of power would be obtained by passing the largest possible amount of air through the propeller. This concept accounted for the rather large diameter (8.5 feet) of the two propellers on the Wrights' Flyer of 1903.

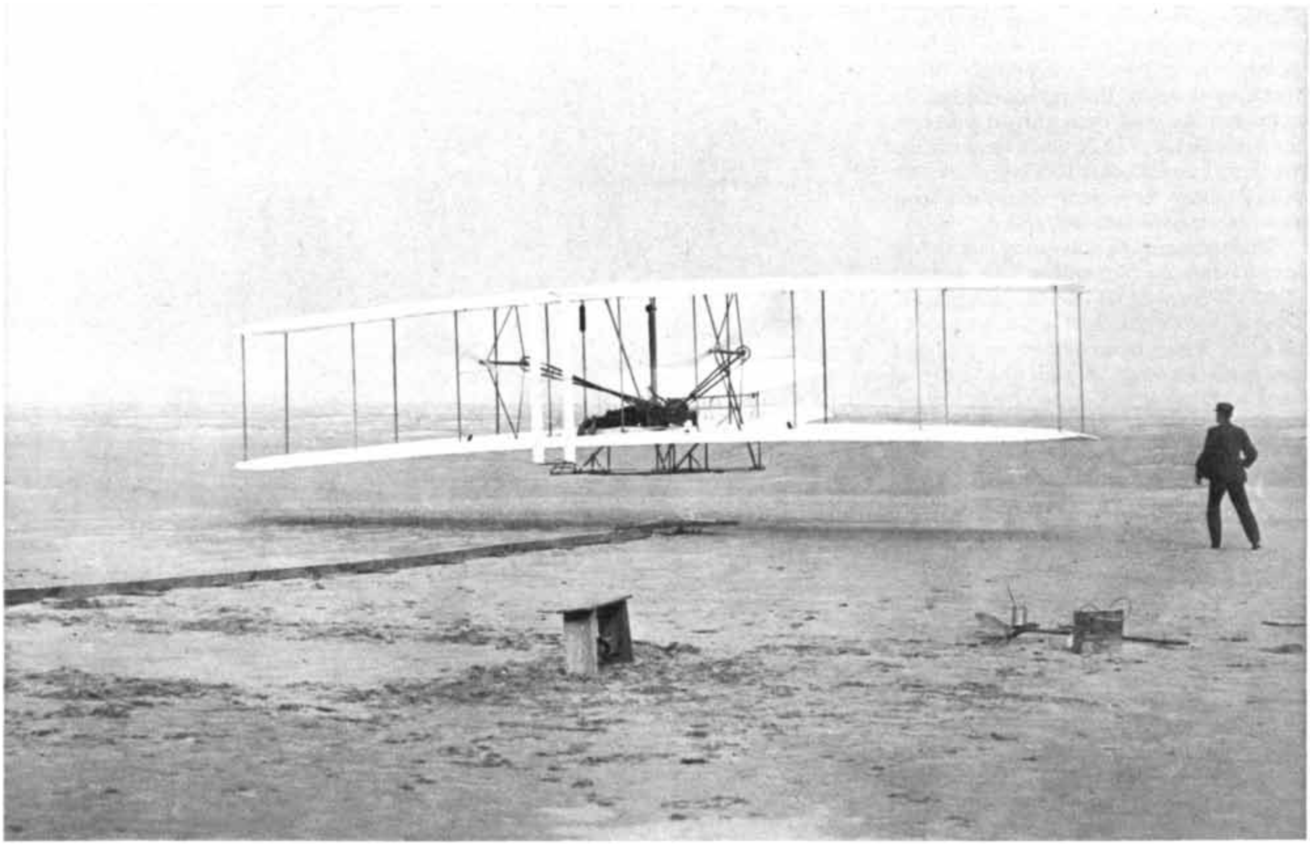
to 35 m.p.h. Wilbur guided the starboard wing during the launching, with the rail now positioned on a level stretch of sand, and Orville took off. His flight covered 120 feet at an average ground speed of 7 m.p.h. This successful flight is the subject of one of aviation's most moving and best-known photographs, which shows the plane flying a few feet off the ground with Wilbur running alongside.

Each brother made two flights that day before the Flyer was blown across the sand and damaged beyond easy repair. After logging a total of 98 seconds in the air the first powered, man-carrying airplane never flew again. It is now on permanent exhibit at the Smithsonian's Air and Space Museum.

For two more years the Wright brothers worked alone to perfect their new airplane. Their progress was not generally known. In the aeronautics community only Chanute was not skeptical of their claim to have the first truly successful flying machine. In 1906 the Wrights were granted a U.S. patent for their system of lateral control, involving both deflection of the wing surfaces and movement of the vertical tail.

Only after they had completed satisfactory licensing agreements in the U.S. and Europe were the brothers willing to show their airplane (an improved version of the 1903 Flyer) in public. Wilbur first flew publicly near Le Mans in France in August, 1908; a few days later Orville gave a demonstration at Fort Myer, Va. The flying skills of the brothers startled the world, and the Wrights became internationally celebrated.

Their airplanes of 1908 had the same basic configuration as the 1903 Flyer, but the design could not survive. Within less than two years aircraft with aft tails showed superior performance, and the Wrights were forced to abandon the canard. They contributed little thereafter to the development of aircraft, but their achievements in the period from 1899 to 1908 remain unparalleled.



FIRST FLIGHT of the 1903 Flyer was photographed at Kill Devil Hills near Kitty Hawk by John T. Daniels of the Kill Devil Life Saving Station, who was using the Wrights' camera. The date was December 17, 1903. Orville is the pilot and Wilbur is running alongside. The flight, which was one of four flights made by the Wrights that day, covered 120 feet at an average speed of seven miles per hour over the ground. Wilbur made the last and longest flight of the day, staying aloft for 59 seconds. After a total flight time of 98 seconds that day the Flyer was never taken into the air again. It is now on display at the Air and Space Museum of the Smithsonian Institution.

Compartments in Animal Development

Flies, and maybe other animals too, seem to be composed of a number of compartments: homologous units within which key genes execute decisions committing several clones of cells to a line of development

by Antonio García-Bellido, Peter A. Lawrence and Gines Morata

The development of an animal is such a commonplace miracle that its reliability tends to be taken for granted. Yet man makes nothing so complex as an ant or a fruit fly, and if he did, it would surely be subject to errors of construction and assembly; it is a long way from a blueprint to a finished product. The blueprint on which accurate animal development depends is encoded in the DNA of the chromosomes. The code is essentially an instruction tape made up of a linear sequence of nucleotides, the subunits of DNA. How is this information in one dimension translated into structure in three dimensions? The search for principles that govern the process of morphogenesis is one of the most active areas in developmental biology today.

In many viruses (for example the tobacco-mosaic virus) the process whereby genetic information is elaborated into the structure of a virus particle is now understood. The nucleotide sequence of the gene for the protein outer coat of the virus is translated into a sequence of amino acids that itself specifies the folding, and thus the three-dimensional shape, of the coat-protein molecule. That molecule in turn is so formed that a number of identical molecules can stack together in a specific way: they have matching surfaces and "cozy corners" that specify the position and orientation of other identical molecules as they are added to the growing virus. The ultimate form of the virus particle can therefore be traced back rather directly to its origin in the nucleotide sequence.

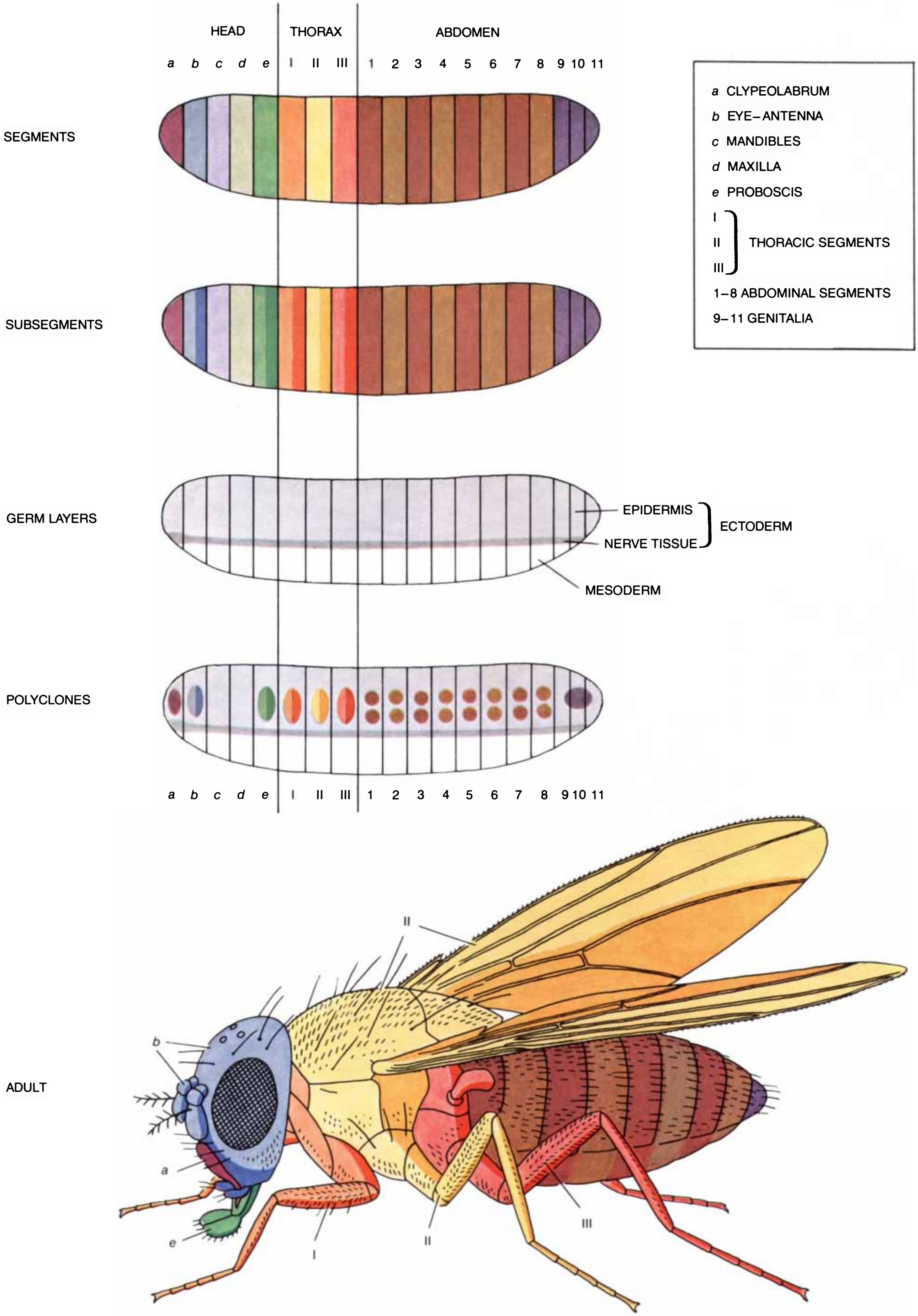
The morphogenesis of an animal is obviously a more complex process. The egg is subdivided into cells that, although they are initially similar, become increasingly differentiated as they divide and proliferate; structures are formed not so much by the assembly of finished parts as by the progressive elaboration of cells and tissues at particular sites. And yet some common principles may operate in the development of viruses and in the development of animals. In rather different ways both viruses and animals appear to be modular in construction. Viruses are built by the assembly of identical subunits, and experiments in our laboratories suggest that animals may be made up of a number of different but fundamentally homologous units, or "compartments," that are in effect variations on a basic theme. Our hypothesis is that different combinations of special genes are active within each compartment and determine its individual structure.

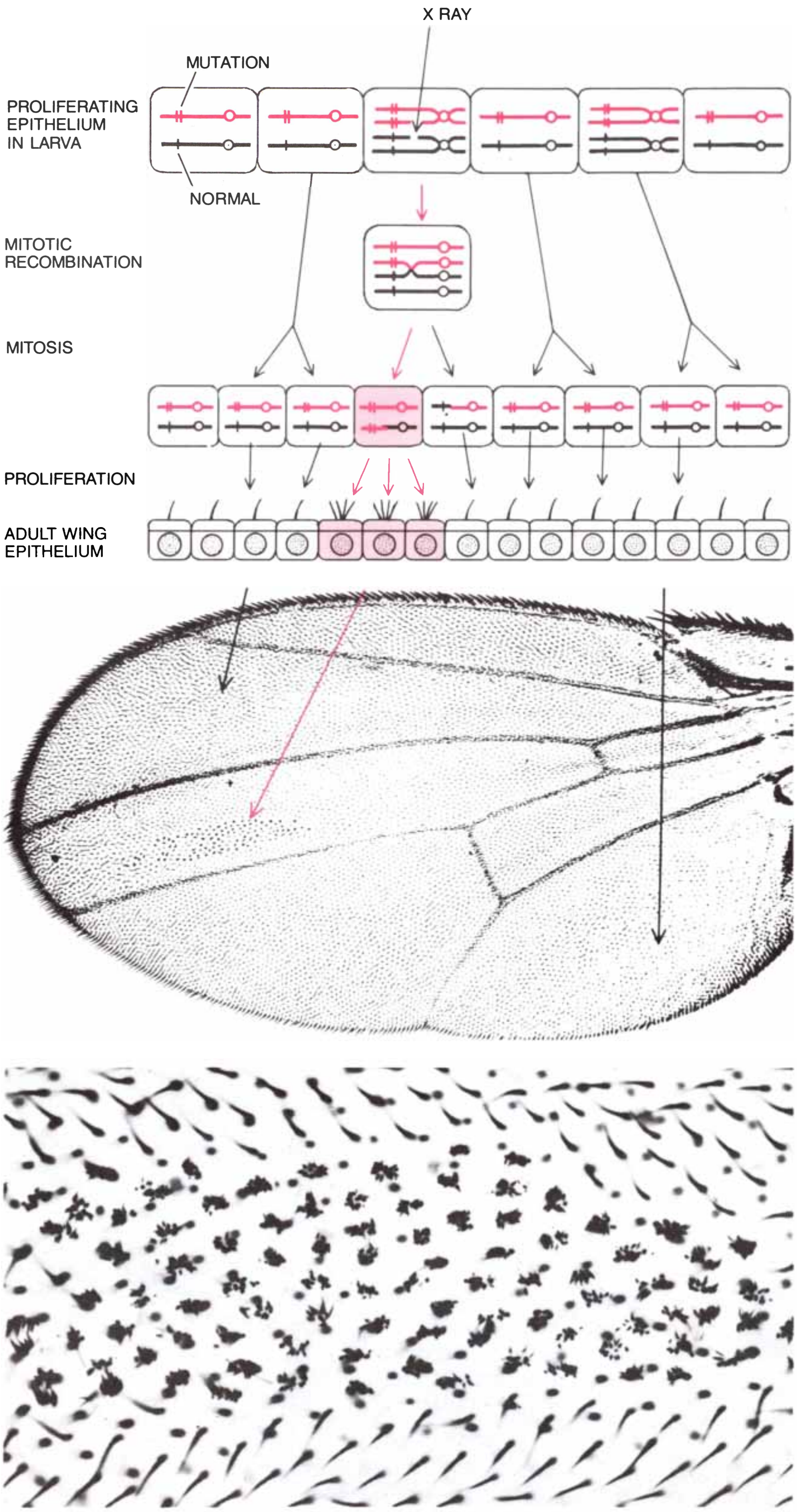
All animals are made from two-dimensional sheets of cells, which can become folded and thickened. Fundamentally, therefore, the attempt to understand morphogenesis can be reduced to a problem in two dimensions: How does information in the chromosomes control the organization and structure of cell layers? Insects are particularly favorable organisms to study because they express the problem in its simplest form. Their cuticle, or outer shell, is secreted by the epidermis, an intricately convoluted layer that is only one cell thick. Another advantage of insects is that each cell secretes a particular piece of the cuticle above it.

In the fruit fly *Drosophila melanogaster* a large number of mutations are known that change the structure or color of the cuticle in specific ways. For example, on the wing of normal flies the cells each secrete a single hair, whereas in flies with the mutation called *multiple wing hairs* each cell secretes a group of hairs. At the Institute of Genetics of the Spanish Research Council in Madrid and the Medical Research Council Laboratory of Molecular Biology in Cambridge we have developed techniques, based on the innovative work of Curt Stern of the University of California at Berkeley, for indelibly marking individual cells in the developing embryo or larva of flies by means of these mutations. We breed strains of flies in which each cell is heterozygous for the mutation: it has one chromosome bearing the mutation and one bearing the normal gene. Because the mutation is recessive and is expressed only when a cell is homozygous (that is, when both chromosomes of the pair carry the mutation) these flies have a normal appearance.

By exposing a fly embryo or larva to X rays we cause breaks in some chromosomes, and occasionally the breaks are in just the right places so that pieces of chromosome "cross over" in the process of cell division and a cell becomes homozygous for the mutation. As the cell divides it generates a clone of daughter cells, all of which have two chromosomes bearing the marker mutation and therefore express the mutant characteristic. The clone gives rise to a recognizable patch of mutant cuticle on the adult fly. By irradiating hundreds of embryos

COMPARTMENTS in the development of the fruit fly *Drosophila melanogaster* are indicated by means of schematic diagrams of the embryo and the adult fly. The embryo, a peripheral epithelium of some 5,000 cells descended from a single nucleus in the egg, becomes divided into segments, each destined to give rise to a segment of the adult. Some of the segments are now known to divide into anterior (*light shades*) and posterior (*dark shades*) subsegments, each populated by a distinct "polyclone" of cells. The epithelium divides into germ layers: ectoderm, comprising epidermis and nerve tissue, and mesoderm. Some epidermal cells are selected to form imaginal disks (*colors*) of adult polyclones, from which adult parts will develop; the rest of the epidermal cells form larval parts. (Embryonic segments *c* and *d* may not form adult parts.) The adult segments represent the basic "compartments" in insect development, according to the authors' hypothesis. The anterior and posterior polyclones construct a second-order set of anterior and posterior compartments and may in turn be subdivided, by a series of decisions made by "key" genes, into polyclones populating still more limited compartments.


HEAD
THORAX
ABDOMEN
a b c d e I II III 1 2 3 4 5 6 7 8 9 10 11
SEGMENTS
SUBSEGMENTS
GERM LAYERS
EPIDERMIS
NERVE TISSUE
ECTODERM
MESODERM
POLYCLONES
a b c d e I II III 1 2 3 4 5 6 7 8 9 10 11
ADULT
a CLYPEOLABRUM
b EYE–ANTENNA
c MANDIBLES
d MAXILLA
e PROBOSCIS
I II III THORACIC SEGMENTS
1–8 ABDOMINAL SEGMENTS
9–11 GENITALIA
b
a
e
I
II
II
III



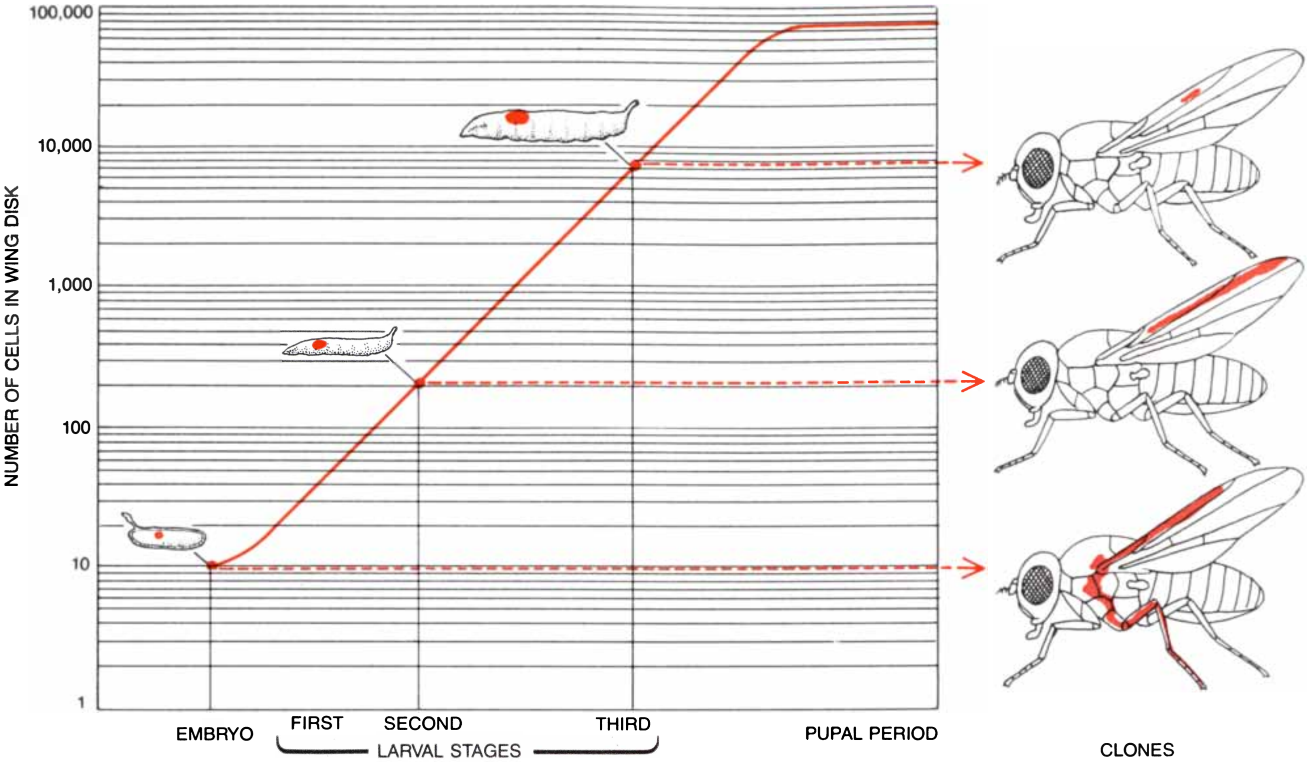
CELL LINEAGE is studied in flies that are heterozygous for a recessive mutation affecting the cuticle, in this case a mutation that makes clumps of small hairs instead of a single hair. One chromosome (*color*) of a homologous pair carries the mutation; the other does not, and the recessive mutation is not normally expressed. The fly embryo or larva is exposed to X rays. Irradiation may break chromosomes in such a way that in the course of mitosis, or cell division, pieces of the chromosomes "cross over" and a cell becomes homozygous for the mutation, that is, both homologous chromosomes bear the mutation. That cell and its descendants (*light color*) produce abnormal hairs. A patch of abnormal hairs on the wing reveals such a mutant clone and delineates its extent, as is shown in photomicrograph (*middle*) and enlargement (*bottom*).

or larvae we can mark single cells at random, and by examining the adult flies we can build up in piecemeal fashion a picture of the lineage of cells in the course of development.

In thinking about how animals develop one might visualize two contrasting modes of cell lineage: fixed and indeterminate. If cell lineage were fixed, each cell in the embryo would always develop to form a specific part of the adult, and every embryo would develop identically. If cell lineage were indeterminate, any cell could develop to form any part of the adult, and cell pedigrees would vary freely from embryo to embryo. Analysis of cell lineage in the wing of *Drosophila* shows that neither of these extreme models is correct. When we plot on a drawing of a wing the extent of a large number of clones produced by the method we have described, we find that the clones overlap. This means that even though the final structure of each wing is the same, the cells that produce it do so in different ways in different wings; the wing does not develop by a fixed lineage. The marked clones are distributed over the wing, showing that growth takes place everywhere (not as it does in plants, where cells divide only within a special zone). And the number of cells in each clone varies considerably, even in the same wing and with no relation to position, showing that the number of cell divisions is not fixed.

Cell lineage is not indeterminate either, however. There is a limit to what different cells can do. In some experiments, designed to test developmental potential, we induce marked cells in a background of unmarked cells that divide particularly slowly; when marked cells are induced in the early embryo, they have enough time to outgrow the slowly dividing unmarked cells, and they could fill the entire wing several times over. Yet they never do; they never fill more than half of the wing (and half of the leg on the same side). The extraordinary fact is that all these marked clones respect exactly the same boundary line, a line that subdivides the second thoracic segment of the adult, which includes the wing and the leg, into two regions. Examination of the maps of smaller clones made by conventional techniques reveals that they respect the same boundary: although the small clones may appear in all regions of the wing, no individual clone ever crosses the boundary. The large clones simply delineate the line more graphically.

It must be, therefore, that when the clones are formed in the early embryo, what will be the second thoracic segment is already subdivided into two small groups of cells, one group destined to make the anterior region of the segment and the other destined to make the posterior region. We call the regions compartments. Each compartment is



CLUES TO GROWTH of the wing and the leg are derived from studies of clones induced at successive stages of development. If the small group of cells (*dark color*) in the embryo that will develop to form a wing and a leg is irradiated, the induced clones are few (because the target is small), but they can be large, extending from the wing to the leg (*light color*). In the larva the cells that will form each adult part proliferate within discrete imaginal disks. As the number of cells in the wing disk increases, irradiation induces clones with higher frequency, but each clone is smaller and populates a more limited region of the adult because a cell irradiated later has fewer descendants.

made by a set of complete clones, which we call a polyclone, that develops from a few founder cells. The polyclone has a precisely defined destiny, but the individual cells in the polyclone have ill-defined roles. As we have indicated, a rapidly dividing cell can form nearly all of the compartment while the other cells in the polyclone form correspondingly less of it.

Evidence is accumulating that other parts of the fly are also made up of compartments. Such compartments have been identified in all three thoracic segments and in the head, the abdomen and the genital organs. Some of the compartments are bounded by discontinuities in the cuticle, such as the borders of the various segments that constitute an adult fly, but this is not always the case. For example, there are no natural landmarks that delineate the anterior-posterior boundary in the wing; the boundary is invisible without the genetic marking technique. Because it is easier to study the external surface of the fly, we are still ignorant about the cell lineage of the internal organs. We do not know, for example, whether there are compartments in the gut, the musculature or the central nervous system.

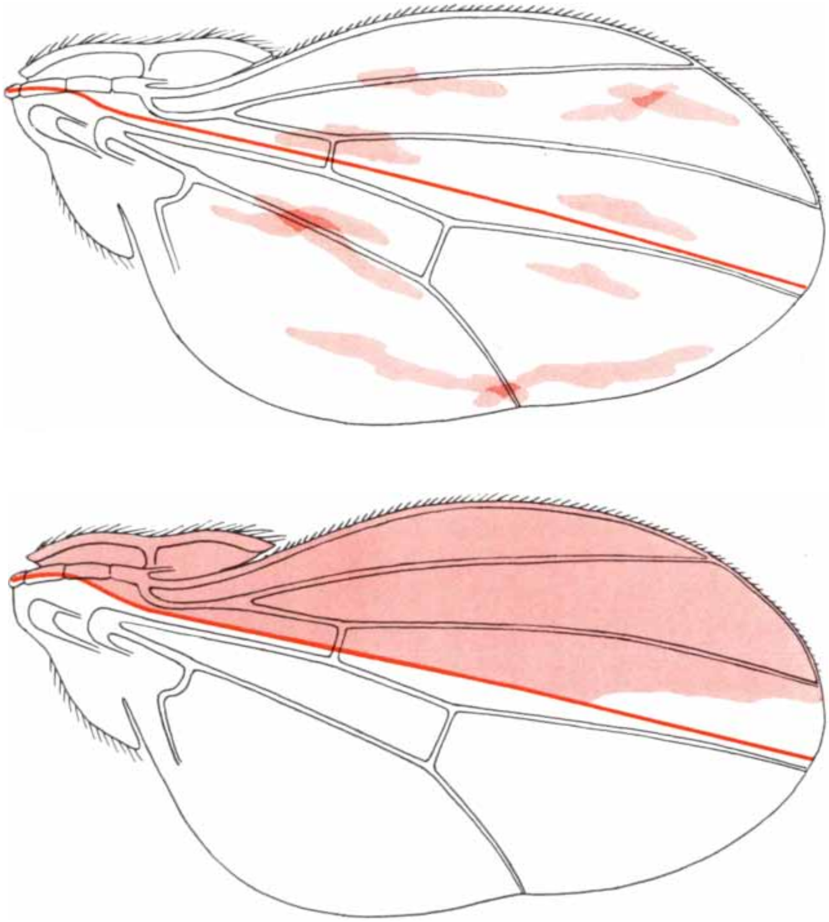
Cell-marking experiments have established two important facts about the construction of compartments. One is that polyclones are established progressively. We have mentioned that the second thoracic segment arises from anterior and posterior polyclones formed at a very early stage in the embryo. Soon afterward there is a "decision" separating the polyclones that will form the adult wing from the ones that will form the leg. Still later, within the developing wing, both the anterior and the posterior polyclone become subdivided into daughter polyclones that will make the dorsal and ventral parts of the wing. The full story of this progression is still not known, but we have worked out one possible developmental tree, with each fork representing a decision point at which a polyclone is subdivided [*see top illustration on page 110*].

The other established fact is that the same decision can be made in different segments. The anterior-posterior subdivision takes place in all three thoracic segments and in the head, and there is some evidence that the dorsal-ventral decision takes place in at least two thoracic segments. Each end point in the tree can therefore be specified by a binary code that identifies the decisions made along the way. Within the second thoracic segment, for example, the code for "anterior leg, dorsal" and the code for "anterior wing, dorsal" differ by only one element, that is, by one decision: the decision made when the wing cells and the leg cells separated.

How are these decisions made and what is their nature? Very little is known about how polyclones become subdivided. What is clear is that the decision is geographical, in that a coherent block of cells in a particular region of, say, the developing wing becomes allocated to the dorsal wing polyclone. The founder cells of the polyclone are not themselves a clone; that is, there is not a single ancestral cell that gives rise to them—and only to them. Analysis of cell lineage also tells us that the decision is final (at least in the normal course of development) because all the progeny of each determined cell contribute only to the appropriate compartment.

We now turn to the role of genes in the lineage of compartments. It has been known for a long time that certain "homoeotic" mutations in *Drosophila* transform an entire region of the fly. Each of these mutations gives rise to a specific transformation and has a specific site in the chromosome, and so a different gene must mutate in each case.

One homoeotic mutation, *Antennapedia,* changes the antenna into a leg; *ophthalmoptera* transforms part of the eye into wing tissue. A group of closely linked homoeotic mutations called the bithorax series has been studied intensively by Edward B. Lewis of the California Institute of Technology. Mutations in the *bithorax* gene transform the anterior part of the third thoracic segment into an anterior second thorac-



CELL LINEAGE in *Drosophila* is not fixed. Plotting typical wing clones induced in different flies shows that the clones overlap; descendants of each marked cell make different parts of the adult in each case (*top*). The lineage is not completely indeterminate either, however. Even when marked cells are made to grow relatively fast, they never fill more than about half of a wing (*bottom*), either the anterior half (as here) or the posterior half. Examination of small clones (*top*) reveals that they respect the same boundary. Development is compartmentalized.

ic segment. The transformation affects both the dorsal and the ventral part of the segment, but only as far as a limiting line that coincides precisely with the boundary between the anterior and the posterior compartment. Mutations in the *postbithorax* gene change the posterior third thoracic segment into a posterior second thoracic segment; again the limit of the transformation coincides with the boundary between the anterior and the posterior compartment. Mutations in the *engrailed* gene affect the posterior regions of the wing and the leg, making them similar to the corresponding anterior parts; once again the affected areas coincide with compartments.

The nature of the *bithorax, postbithorax* and *engrailed* mutations suggests a generalized conclusion: Compartments are the realm of action of at least some homoeotic genes. To put it another way, polyclones are the groups of cells on which homoeotic genes operate. There is now evidence that the normal *engrailed* gene may be important wherever the property "posterior" is required, because the mutation (which impairs the function of the *engrailed* gene) affects the posterior parts not only of the wing but also of all three legs and of the eye-antenna segment without affecting the anterior parts.

These observations have led us to the idea that the tree of binary decisions actually describes the progressive acquisition of a genetic address for each successively smaller compartment. A key regulatory gene operates in an "on-off" mode at each decision point, and the same gene operates to effect homologous decisions in different parts. For example, positive action by the *engrailed* gene would implement the decision to "be posterior." This gene would have to be active in all posterior compartments and would be switched off in all anterior compartments; its absence or malfunction (as in the *engrailed* mutation) would explain the transformation of a posterior region into an anterior one.

This hypothesis predicts that each homoeotic mutation should change only one property, and that is what we observe. For example, when *engrailed* transforms posterior to anterior, the dorsal-ventral and thorax-wing decisions remain unaffected. Moreover, one would expect that homoeotic mutations should act in combination. When flies are mutant for both *postbithorax* and *engrailed,* the effects of the mutations should be additive, and that too is observed.

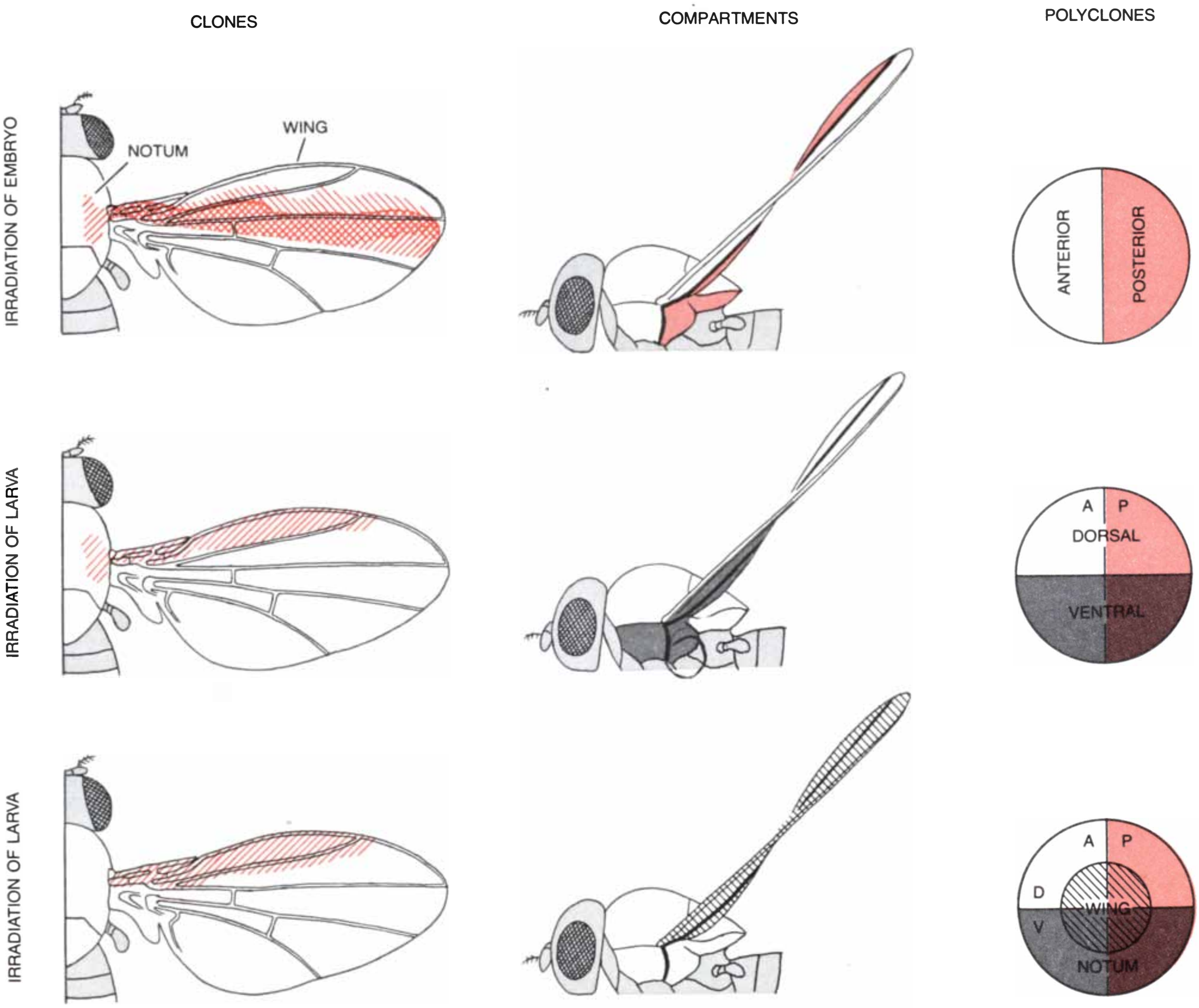
We believe that the combination of active homoeotic genes provides a genetic address that is kept by each individual cell as it develops, and that as the cell divides the same address is passed on to its offspring. One piece of evidence for this is that the genetic address can be changed by removing one of the key genes. Take *engrailed,* for example. By the crossing-over technique described above we make a clone of cells in the wing homozygous for the *engrailed* mutation. In the anterior part of the wing such an operation has no effect on development—because, we believe, the gene damaged by the *engrailed* mutation would in any case be switched off in that polyclone. In the posterior region, however, the gene is required (to maintain the decision to "be posterior"), and its inactivation in one cell changes the genetic address of that cell and its descendants. The result is that the clone of cells develops anterior structures. This experiment shows that what a cell does depends primarily on its own genetic address.

The behavior of an *engrailed* mutant cell, however, depends not only on its address but also on its position within the compartment as a whole. The anterior edge of the wing bears specialized stubby bristles that are not found on the posterior edge. When a mutant clone includes a portion of the posterior edge of the wing, these stubby bristles are found along the edge portion made by the clone, but they are not made by other cells in the clone (nor, of course, by neighboring edge cells outside the clone). This shows that the ultimate pattern depends on the position of cells in the organ as well as on their genetic address. Both the cells that are mutant for *engrailed* and the normal posterior cells surrounding them "know" their position in the wing compartment, and they respond by interpreting that positional information differently. It appears that the positional information is the same in the anterior compartment of the wing as it is in the posterior compartment, but the interpretation differs (as it would when, if the same grid were placed on maps of two different cities, a single map reference might designate very different sites in the two cities even though the sites were in the same square on both maps).

The clones of cells that are mutant for *engrailed* do not respect the boundary between anterior and posterior compartments (although they do respect other compartment boundaries). Mutant cells that originate in the posterior polyclone can invade anterior territory, something normal posterior cells never do. We take this to mean that a "label" is added to posterior cells distinguishing them from anterior ones, and that mixing of cells with the label and cells without it is minimal. We suggest that the *engrailed* gene is ultimately responsible for labeling posterior cells; cells defective in that gene lack their posterior label and can therefore mix with, and infiltrate among, anterior cells. Every polyclone may have a unique combination of labels and individual mixing properties. (Experiments by the late Ernst Hadorn and his group at the University of Zurich established that cells from different parts of the fly normally sort themselves out after having been artificially mixed.) We suggest that the purpose of the labels may be twofold: to keep different polyclones from intermixing and to assist in the matching process whereby different compartments are fitted together correctly.

How might the homoeotic key genes function? So far there are only a few hints; not even the nature of the gene products is known. One might have expected each key gene to regulate its own unique set of genetic elements. If that were the case, most of the genes in *Drosophila* would function in only one compartment or subset of compartments. The fact is that most mutations do not respect compartment boundaries. The mutation *white,* known primarily for its effect on the *Drosophila* eye, also has effects on pigments in the excretory tubules, the testes and elsewhere; mutations affecting the hairs of the cuticle change the hairs on almost the entire surface of the fly. It is probable, then, that the key homoeotic genes operate on the same set of subsidiary structural genes.

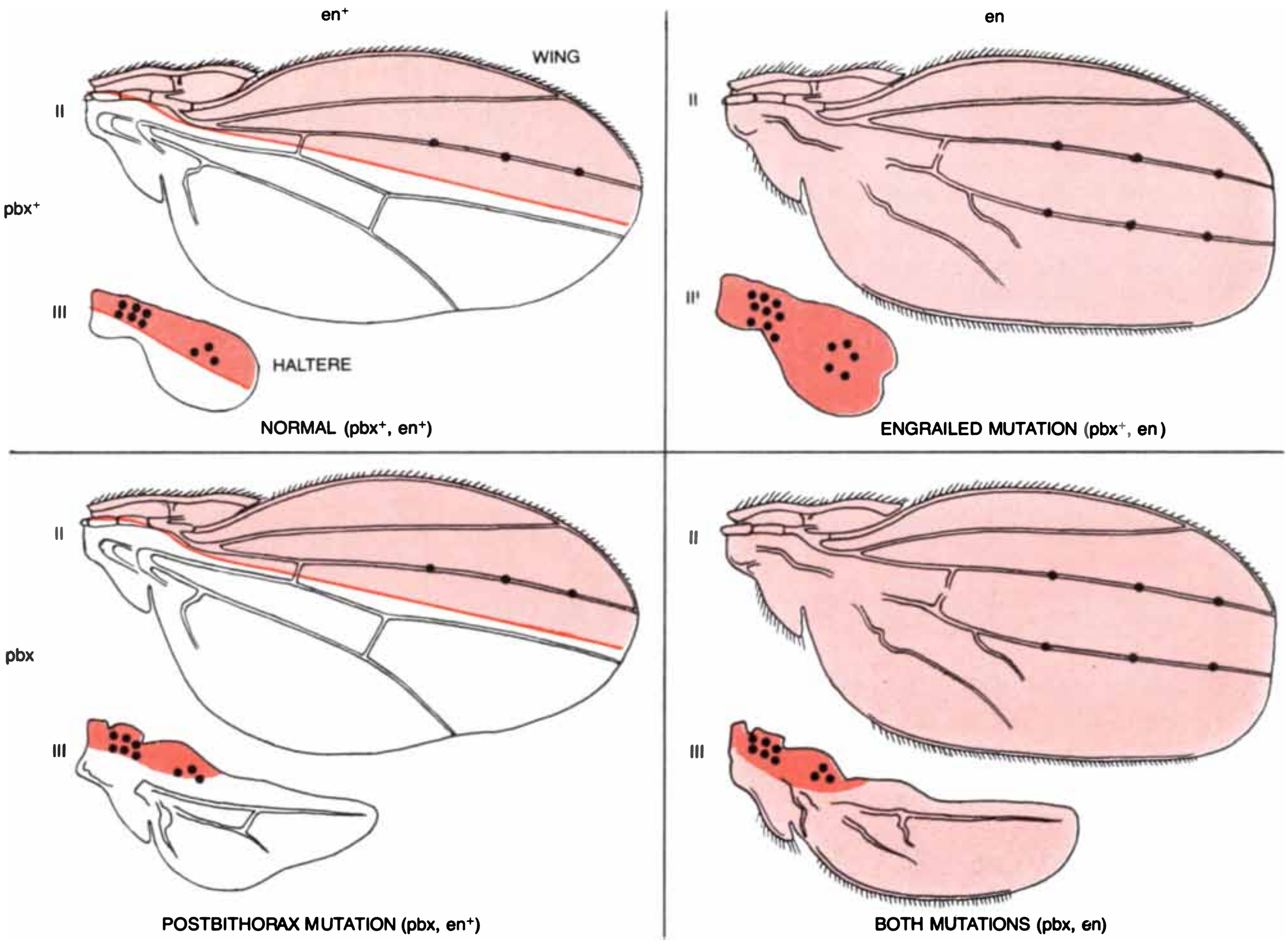
In summary, we have developed a hypothesis that describes the genetic strategy followed in the building of an insect. We suggest that each piece of the insect—a compartment made by a particular group of cells—is specified by a genetic address, in effect a binary zip



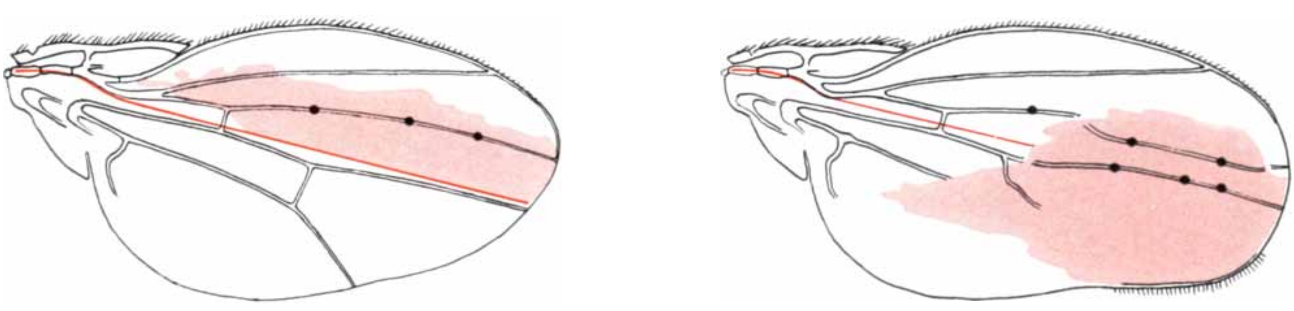
POLYCLONES IN WING DISK build compartments in the wing and the notum, the "fuselage" of the thorax. Irradiation of the embryo (*top*) induces a large clone (*color hatching, left*) that can extend throughout either the anterior (*white*) or the posterior (*color*) compartment (*center*), but not into both compartments because the cells are already divided into anterior and posterior polyclones (*right*). Clones induced in the larval period (*middle*) are further confined to either the dorsal (*white*) or the ventral (*dark gray*) region of the anterior or posterior compartments (*center*); the wing disk has become divided into four polyclones (*right*). At about the same time the clones become still further confined (*bottom*) to either notum (*white*) or wing (*black hatching*). The preexisting polyclones have become subdivided again; each of eight polyclones (*right*) has a different combination of three properties: anterior or posterior, dorsal or ventral, wing or notum.

code representing the decisions of key regulatory genes. The address derives from a cell-lineage tree in which at each fork (where groups of cells rather than single cells are affected) a switch turns a gene on in one branch and leaves it off in the other branch. The state of the switch is maintained in all daughter cells, so that the genetic address carried by a particular cell line becomes longer at each fork. The final binary code in an adult cell contains a history of the decisions made by the cell's ancestors. This strategy gives rise to differently situated compartments that have different genetic addresses.

The model makes sense because a small number of genetic control elements suffice to specify a large number of different compartments. One can see how in principle evolution could have



HOMOEOTIC MUTATIONS are additive, as is shown by the effects (here somewhat schematized) of the mutations *engrailed* (*en*) and *postbithorax* (*pbx*) on the second (*II*) and third (*III*) thoracic segments. Normally (*top left*) the second thoracic segment forms a wing with anterior (*color*) and posterior (*white*) compartments and the third thoracic segment forms a similarly divided balancing organ, the haltere. The mutation *engrailed* transforms posterior segments into anterior ones (*top right*); *postbithorax* transforms the posterior compartment of the third segment into that of the second segment, making the haltere a wing (*bottom left*). When both are present, *postbithorax* makes a haltere-wing; *engrailed* changes the posterior wing compartment of the haltere-wing into an anterior wing compartment (*bottom right*).



EFFECTS of *engrailed* can be studied in mutant clones made in normal wings. The cells of these *engrailed* clones are marked by a second mutation that distinguishes mutant cells even where *engrailed* has no effect. An *engrailed* clone (*color*) in the anterior compartment develops normally and respects the anteroposterior boundary (*left*). An *engrailed* posterior clone, however, displays anterior patterns and structures, which are confined to the clone (*right*). The mutant pattern is appropriate to each cell's position: there are typical anterior bristles at the posterior edge, and only there. Moreover, the clone does not respect the compartment boundary line. The mutant cells may lack a "label," which is present on normal cells of the posterior polyclone, that prevents the intermingling of anterior and posterior cells.

its effects on reiterated units such as the identical segments of a primitive insect. Differences among segments might be small initially, but as soon as a genetic element became responsible for an individual segment, that segment could evolve independently of other segments. Whereas the number of segments and the overall genetic logic would remain relatively stable during the course of evolution, the morphology of each individual segment could change more rapidly. Homoeotic mutations in some other insects support this concept. There are transformations in mosquitoes and beetles, for example, the directions of which have their exact counterparts in *Drosophila,* even though the particular structures involved are quite different.

We should emphasize that even if these hypotheses are correct, they leave many questions unanswered. We do not understand how the underlying archetypal pattern of segments is established. Positional information appears to be critical for the genetic decisions made within polyclones and for determining the pattern, shape and size of individual compartments. The importance of positional information in the development of various animals has been established [see "Pattern Formation in Biological Development," by Lewis Wolpert; SCIENTIFIC AMERICAN, October, 1978], but the mechanisms that enable cells to behave in a way that is appropriate to their position are still not understood.

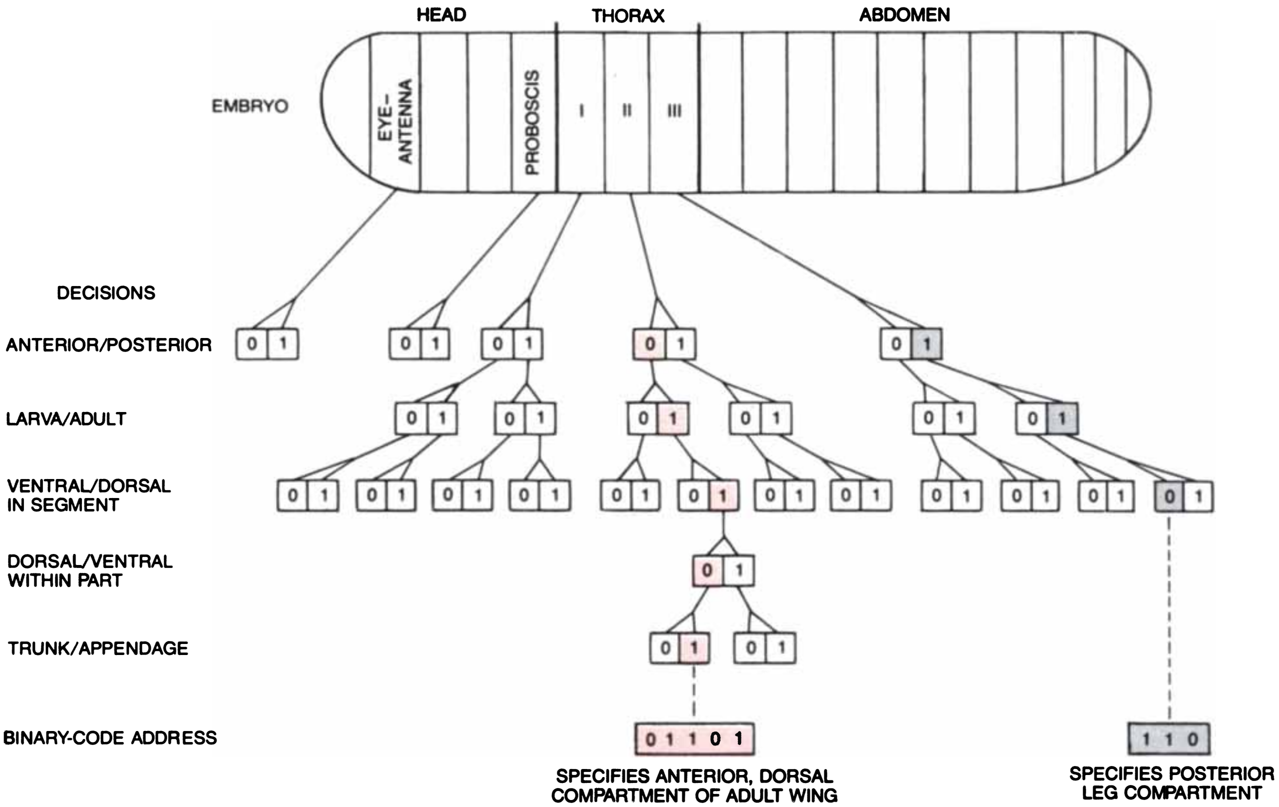
Is there evidence for compartments in the development of animals other than insects? Certainly there are reiterated elements in a wide range of animals, from worms to vertebrates. In the case of one vertebrate, the mouse, enough evidence has been gathered on cell lineage so that one may profitably pose the question: Do mice have compartments? To consider that question one should have in mind the diagnostic criteria that characterize development by means of compartments: (1) Small groups of cells (on the order of tens of cells) generate large, defined regions of the adult; (2) the decision as to which cells constitute a block of founder cells is geographical, depending on cell position rather than on cell ancestry; (3) there is a binary tree of developmental decisions; (4) different cell-recognition properties characterize different polyclones; (5) homoeotic mutations can switch the development of a polyclone and thereby substitute one organ for another.

There is evidence, particularly from the investigations of Richard L. Gardner and his colleagues at the University of Oxford, that the first four diagnostic criteria are met in the course of early development in the mouse. The Oxford group's methods of labeling cells are different from ours and involve more experimental intervention. Essentially their approach is to mix two different embryos or to transplant cells from one embryo into another. Then they follow the fate of the different cells in the mosaic embryo, sometimes with the help of fluorescent antibodies that bind only to cells from one of the two donors.

In mice (as in other mammals) the fertilized egg cleaves a few times to give rise to the morula, a ball of about 16 to 32 cells. As the morula develops into a blastocyst its cells come to form two distinct polyclones that give rise to two distinct regions: the cells on the inside constitute the inner cell mass and the rest of the cells form a peripheral cell layer called the trophoblast. This observation satisfies the first of the criteria listed above. Much current research is focused on just when cells become irrevocably committed to forming one or the other of these two disparate structures. For us what is significant is that in normal development two discrete cell types develop from one type of cell, and that the decision appears to be geographical, thereby satisfying the second criterion. Subsequently the inner cell mass itself divides to form two polyclones, satisfying the third criterion. Again the decision is geographical, because the cells close to one margin of the inner cell mass become the primitive endoderm (which will give rise to specific parts of the extraembryonic membranes), whereas the rest of the cells form the primitive ectoderm (which will give rise to the fetus and some other parts).

It is possible to transplant single cells of the inner cell mass from one blastocyst into blastocysts of a different genotype (from a rat blastocyst to a mouse blastocyst, for example) and to follow their daughter cells. Every transplanted cell enters one of two mutually exclusive regions, either the parts made by the primitive endoderm or the parts made by the primitive ectoderm. This shows that there are two discrete cell types in the inner cell mass from which the transplanted cell was taken, and only two. The ability of each donor cell type to enter the correct part of the host embryo and to colonize only the tissues that part normally forms suggests that the two cell types have different mixing properties, satisfying the fourth criterion.

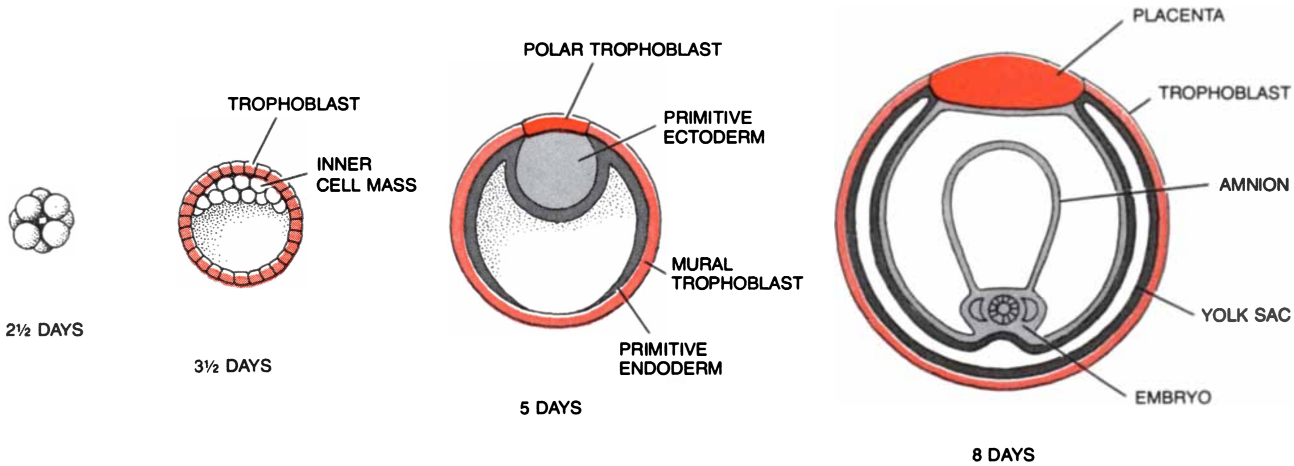
In spite of this evidence for the rigid determination of discrete cell types in the early development of the mouse, studies with mouse mosaics having different coat colors and other markers show that the embryonic cells do undergo considerable mixing before the mouse is born. The final result is thus quite different from what happens in the growth of the *Drosophila* epidermis, in which cells constituting a clone usually remain together. This distinction between insect development and mammalian development may not, however, be true in all tissues at all times. In mice the



SUCCESSIVE STEPS of polyclone subdivision appear to constitute a binary tree of "on" (*1*) and "off" (*0*) decisions by a number of key regulatory genes. The diagram is speculative and shows the authors' current view of the tree. Homologous decisions are made in different segments. The anterior-posterior decision appears to be made in at least the five segments where it is indicated. Further decisions have been identified as shown. The mutation *engrailed* presumably affects the initial decision for posterior as opposed to anterior; other key genes are not known. The succession of decisions builds up a binary-code "genetic address," which specifies one particular compartment.

clones appear to be more coherent in the early embryo than they are in the newborn animal, and in insects there is much more mixing in the epidermis of the abdomen and in the excretory tubules than there is in the epidermis of the head and the thorax. We suspect that mixing within a polyclone may be unimportant, since all the cells of the polyclone are genetically equivalent.

In the mouse, as in *Drosophila*, the process whereby existing polyclones are subdivided into daughter polyclones is still being investigated. So far there is no evidence for homoeotic mutations in mice. What we do know is that mice and flies have enough developmental features in common to allow us to pose the question: Do insects, mice and man all develop according to a similar genetic strategy, expressed in compartments?



MAMMALIAN DEVELOPMENT may involve compartments. The fertilized egg divides to form a small ball of cells, the morula (*left*). At the blastocyst stage ($3\frac{1}{2}$ days) the proliferating cells have formed two polyclones: the inner cell mass and the trophoblast. After five days four polyclones can be distinguished: polar trophoblast, mural trophoblast, primitive endoderm and primitive ectoderm. These cell layers form particular structures and probably subdivide into other polyclones before embryo can be distinguished at about eight days.

Now JAL offers full-fare passengers special privileges before, during and after the flight.

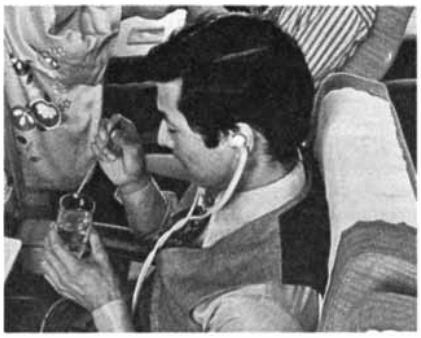
JAL feels that the full-fare passenger deserves better than just a couple of free in-flight gestures. That's why JAL's Executive Service starts working for you the minute you book. Ask your Travel Agent, Corporate Travel Dept. or Japan Air Lines for further details. And find out what it is like to be pampered Japanese style, all the way.

We never forget how important you are.
JAPAN AIR LINES

Bilingual business cards, English on one side, Chinese or Japanese on the other, at a nominal fee.

Special Executive check-in at selected airports.

True Japanese hospitality in the Executive Service section.



Free earphones—for in-flight entertainment.

Free alcoholic beverages—cheers!

1st Class offers the Sky Sleeper Service, a private bed at an additional charge, on all 747 Transpacific flights.

JAL's Executive Hotel Service offers a room at a special rate and allows check-out as late as 6 p.m.

At hotels all over the Orient, JAL's Hospitality Desks give any assistance you may need.

The Executive Service Lounge in the Imperial Hotel, Tokyo, "your office away from the office."

Executive Service

The Bag Model of Quark Confinement

Quarks appear to be real, and yet they have not been observed in isolation. One hypothesis for why they have not been is that they are confined in bags analogous to the bubbles in a liquid

by Kenneth A. Johnson

Subatomic particles are classified in two broad categories: the leptons (the electron, the muon, the tau particle and the neutrinos associated with each of them), which do not interact strongly and appear to have no constituent parts, and the hadrons (including the proton, the neutron and the pion), which do interact strongly and show signs of an inner structure. The force that holds neutrons and protons together in the atomic nucleus is one manifestation of the strong interaction between hadrons. That interaction is one of the four forces of physics, the others being the electromagnetic force, the weak force and gravitation.

More than 200 kinds of hadron have been identified over the past few decades. They too have been divided into two categories, depending on how they decay. The baryons are particles that decay ultimately into the proton, and the mesons are particles that decay entirely into leptons and photons (quanta of electromagnetic energy) or into proton-antiproton pairs.

The hadrons all appear to be combinations of the constituents named quarks. The quark model amounts to an impressive simplification of nature because at present just five "flavors," or kinds, of quark (and in most cases just three flavors) can account for the properties of the multitude of hadrons. The strong interaction binds the quarks together to form baryons and mesons. On the quark model baryons are made up of three quarks, and mesons are made up of a single quark-antiquark pair. What is intriguing about quarks, whose flavors are fancifully labeled "up," "down," "strange," "charmed" and "bottom," is that none has ever been observed in isolation, in spite of many attempts to find one. To explain this phenomenon I and a group of co-workers (Allen Chodos, Robert L. Jaffe, Charles B. Thorn and Victor F. Weisskopf) at the Massachusetts Institute of Technology have developed a theoretical model in which the quarks making up a hadron are imprisoned in what may be called a bag or bubble. The model serves to clarify many aspects of the strong interaction between quarks.

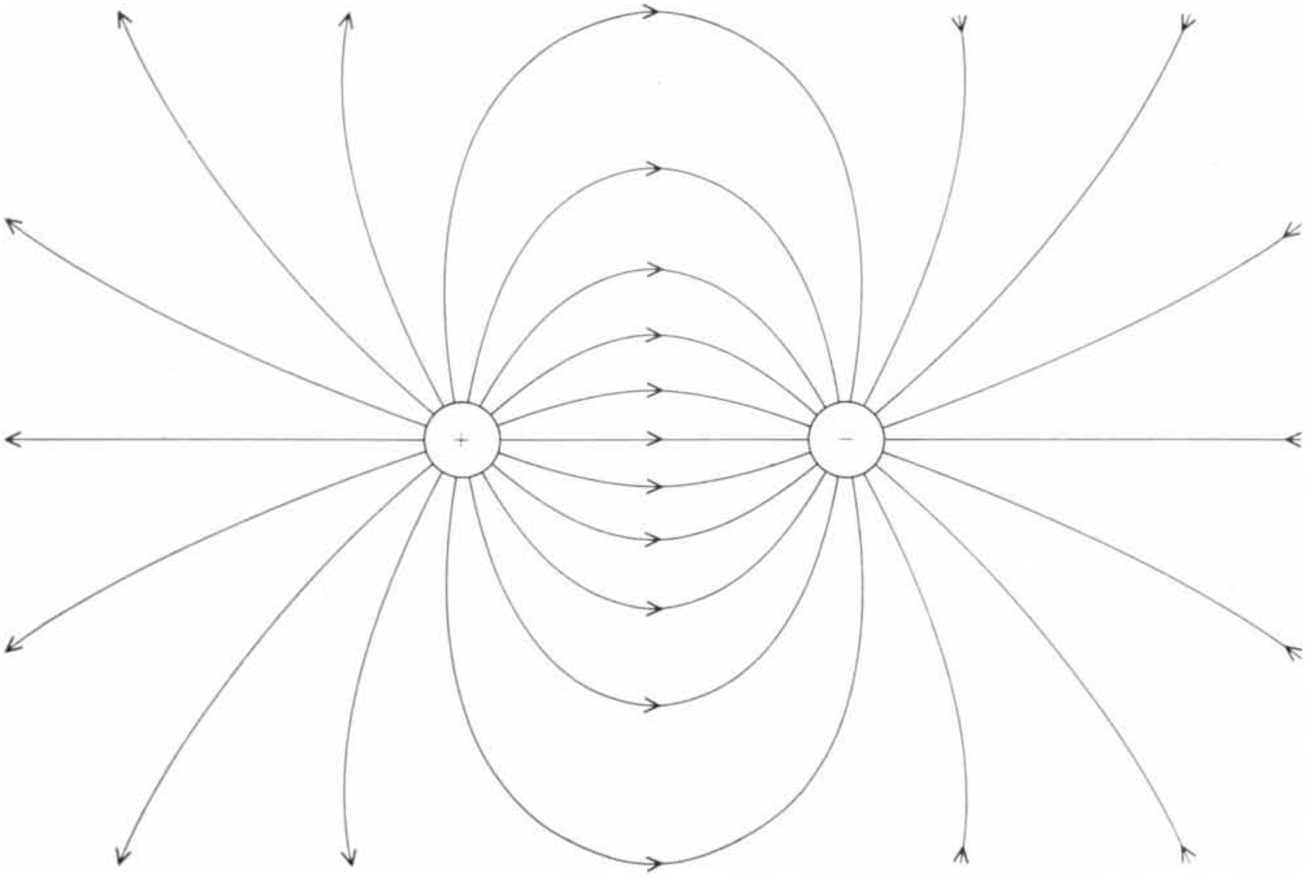
When Murray Gell-Mann and George Zweig of the California Institute of Technology introduced the quark model in 1963, quarks were not universally regarded as real entities that might be found in the laboratory. The possibility was considered that they might just be theoretical constructs for bringing order to the unwieldy collection of hadrons. The quarks were thought to be peculiar objects because they carry only a fraction (+2/3 or −1/3) of the electric charge of the proton, whereas all other known physical entities carry an integral multiple of that charge. Moreover, the quarks seemed to violate the exclusion principle, the quantum-mechanical equivalent of the intuitive notion that no two things can be in the same place at the same time.

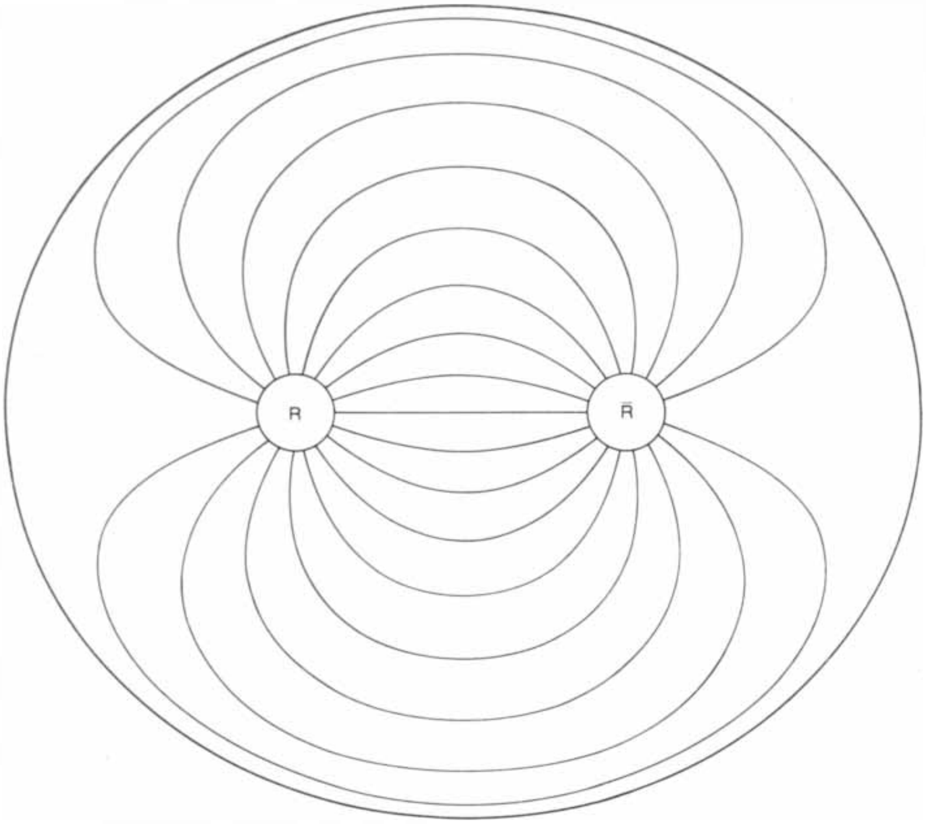
In the late 1960's the status of quarks as the fundamental constituents of all hadrons was dramatically confirmed by a series of experiments done by groups led by Jerome I. Friedman and Henry W. Kendall of M.I.T. and by Richard E. Taylor of the Stanford Linear Accelerator Center (SLAC). High-energy electrons were inelastically scattered off protons, and the scattered distribution indicated the existence of particles within the protons. These particles behaved the way quarks were posited to behave and so they were identified as quarks. The scattering experiment demonstrated not only the physical reality of quarks but also their pointlike nature. Therefore all the fundamental constituents of matter, quarks as well as leptons, seem to act like dimensionless points.

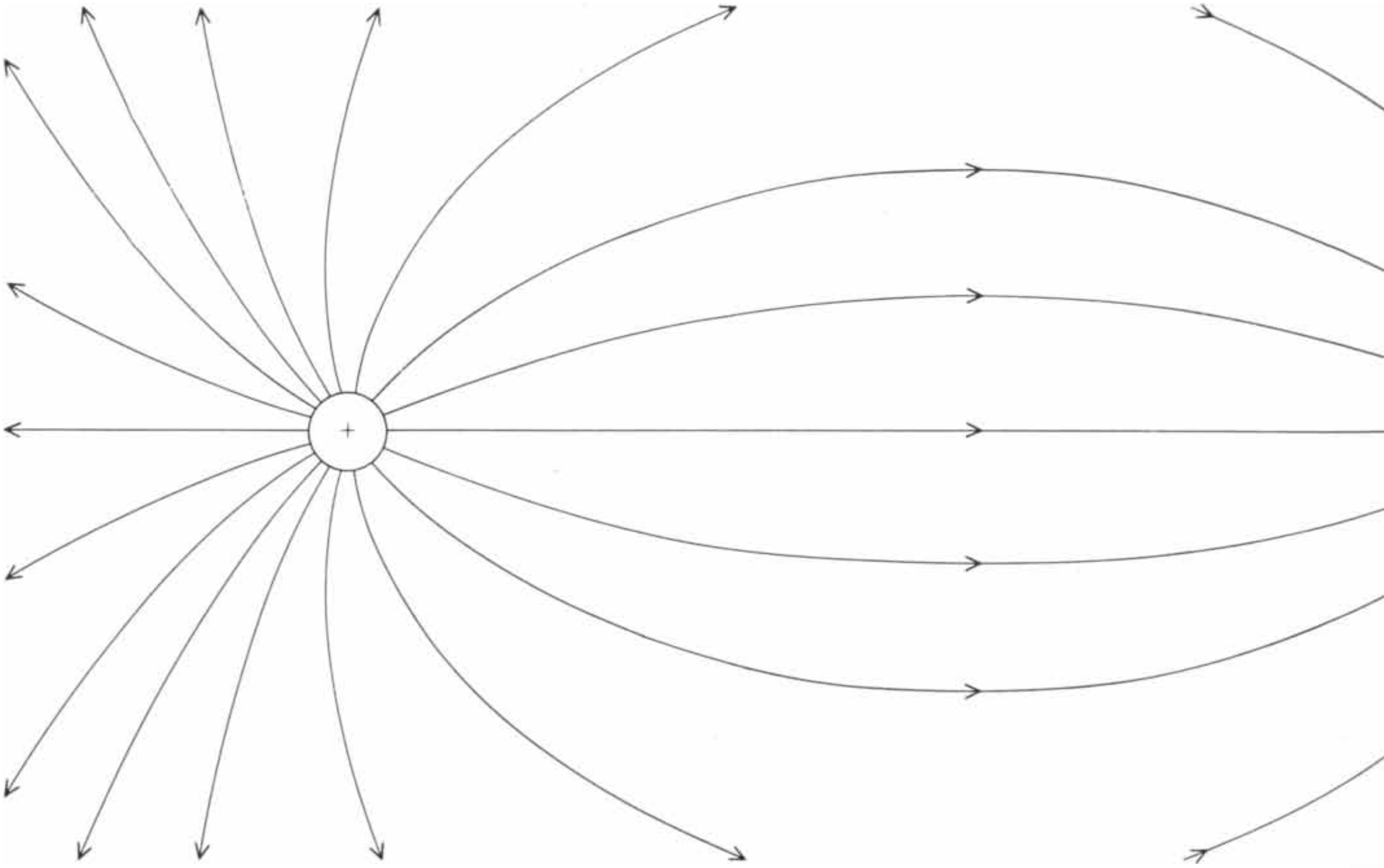
The fact that three quarks make up a proton accounts for their fractional charge (the proton being composed of two "up" quarks, each with a charge of +2/3, and one "down" quark, with a charge of −1/3). Moreover, the introduction of the new quantum number, or physical property, labeled "color" makes it possible to reconcile the behavior of the quarks with the exclusion principle. Yet since no quark has ever been seen in the free state, quarks are still peculiar objects. It is relatively easy to remove an atom from a molecule, an electron from an atom or a proton from an atomic nucleus. Only a relatively small amount of energy is needed to free any one of these constituent particles. Huge amounts of energy have not succeeded in removing even one quark from one hadron.

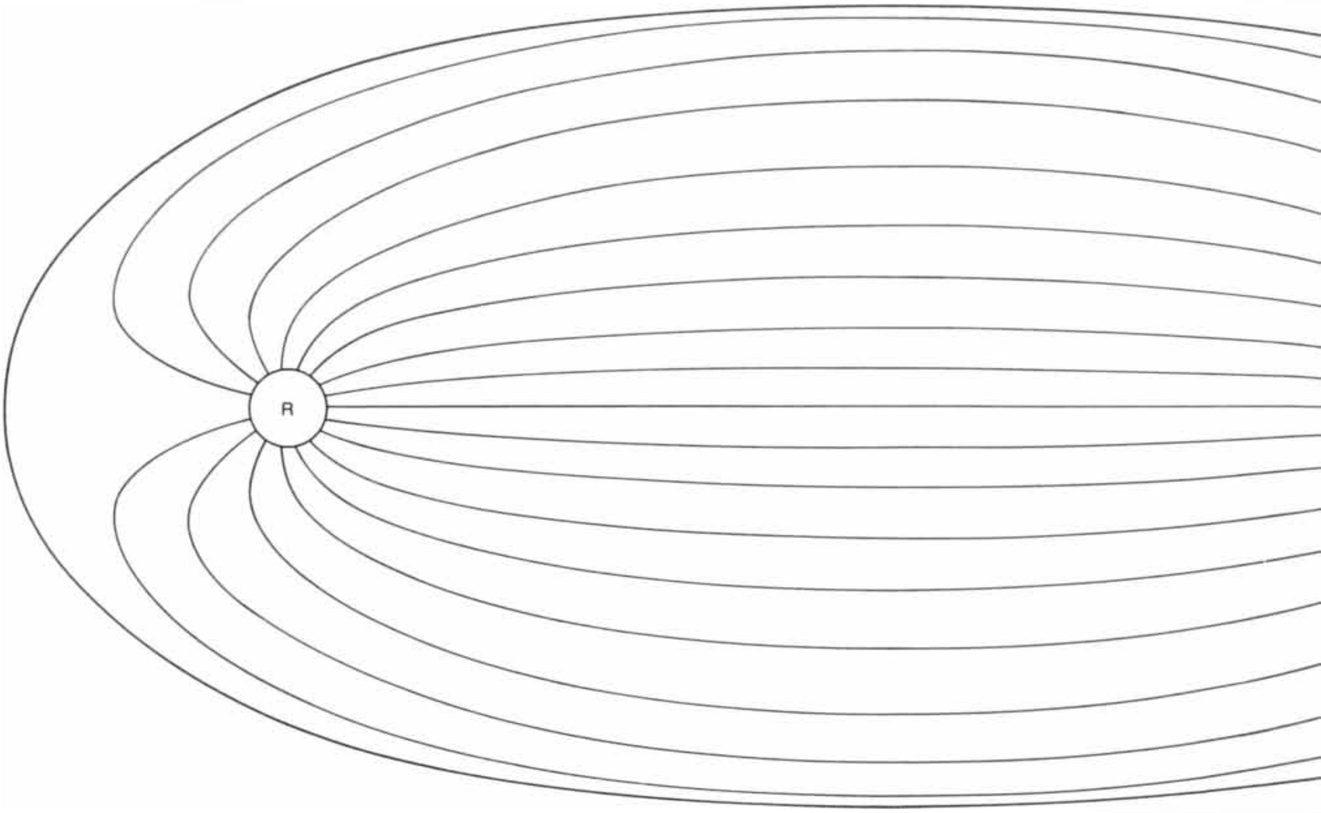
At first some physicists believed a

LINES OF ELECTRIC FORCE (*top*) generated by opposite electric charges are curves in space indicating the direction of the electric force a positive charge would be subjected to if it were placed on the curves. The number of lines in a given spatial region indicates the magnitude of the force; the greater the number of lines, the stronger the force. Electric-field lines fill all space, but at a distance from the charges the field-line density is low and therefore the force is too. Lines of "color" force (*bottom*) link the quark-antiquark pair of a typical hadron: any subatomic particle that is subject to the interaction known as the strong force and that shows signs of an inner structure. Color is the designation of a property of quarks that has nothing to do with visual color. Each kind of quark comes in three colors: "red," "blue" and "yellow." The antiquarks have anticolors: "antired," "antiblue" and "antiyellow." All hadrons are "white," that is, they are colorless averages of the three colors and three anticolors. Two-quark hadrons consist of a quark of one color, say red (*R*), and the antiquark of the corresponding anticolor, say antired ($\bar{R}$). In the bag model of hadrons quarks are confined to a bag or bubble within a hadron, and so the color-field lines are also confined. The fact that the lines are parallel intensifies the force acting between the charges because the lines all pull in the same direction.


+
−


R
R̄



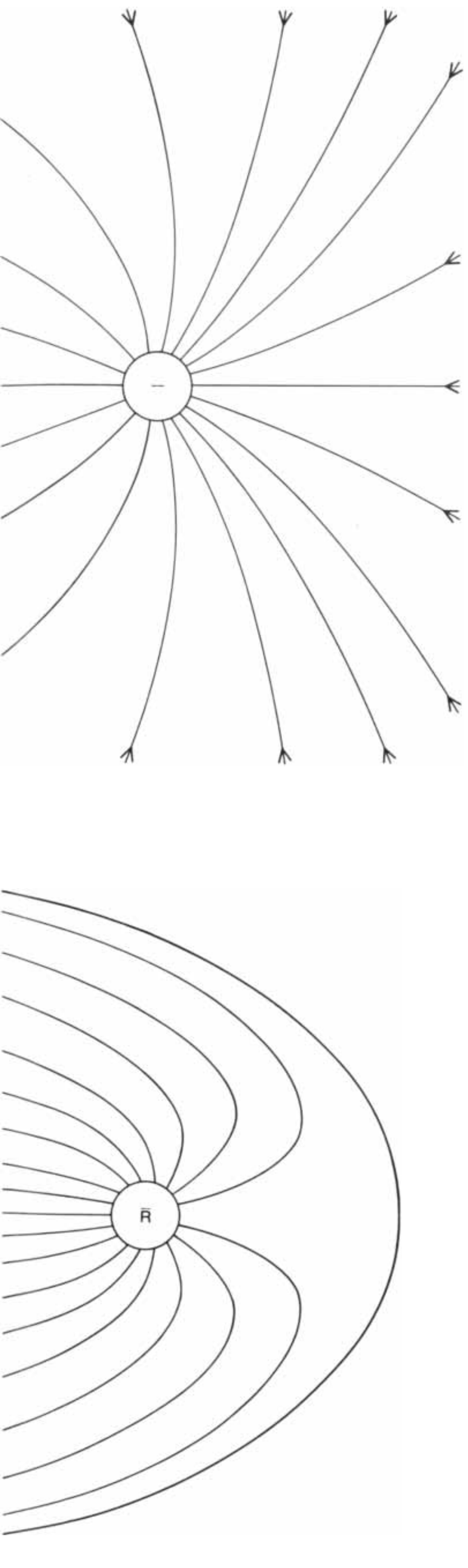

R

quark would be liberated as soon as particle accelerators were upgraded to higher energies. That has not happened, and quarks are now thought to be permanently confined to the interior of hadrons. In the past decade theorists have proposed several models of quark confinement; the bag model is only one of them. The bag model is gaining adherents because it may be closely related to what most theorists consider to be the best candidate for the theory of strong interactions, namely quantum chromodynamics, or color dynamics.

I want to introduce the bag model by invoking an analogy suggested by one of my graduate students, David Shalloway. Imagine a world in which the atmosphere consists not of air but of water. Energy could be expended to boil the water, creating bubbles of water vapor. Inside the bubbles the water-vapor molecules would behave, to a first approximation, like the weakly interacting particles of an ideal gas. In this hypothetical world water vapor could exist only inside the bubbles. A molecule of water vapor that returned to a liquid from a bubble would no longer be water vapor; it would cease to exhibit the properties of a particle of ideal gas.

The concept of hadronic matter in the bag model is similar. Hadrons are bubbles in space that contain quarks. Inside a bubble the quarks move about freely and independently, just as the water-vapor molecules move about inside a bubble. If an attempt is made to remove a quark from the bubble, the confining strong interaction between the quarks takes over. Like the water-vapor molecules, quarks can exist only in the bubbles enclosing them. In other words, hadrons act like ideal quark-vapor bubbles embedded in an atmosphere of ordinary space where quarks cannot exist.

Just as it takes a certain amount of latent heat to boil water, so it takes energy to "boil" a vacuum, that is, to convert energy into matter. Indeed, particle accelerators can be regarded as vacuum boilers. A large fraction of the kinetic energy of the accelerated particles is deposited on a small target. Most of the energy goes into making quarks and antiquarks that combine to form the hadrons found in the debris of the reaction. In the bag model some of the energy also goes into the vacuum, pushing aside ordinary space in order to inflate the bubbles enclosing the quarks. The amount of energy needed to inflate a bubble is called the bag constant (B) and has a value of about 55 million electron volts per cubic fermi (one fermi being 10^{-15} meter).



SEPARATION OF CHARGES affects the force between them. When two opposite electric charges (*top*) are separated, the field lines bend more. The electric force gets weaker because only the component of the field lines that is lying along the direction between the charges has a net effect on the force. The force drops off as the square of the separation between the charges. When two quarks (color charges) are separated (*bottom*), the lines do not bend but continue to pull in the same direction. As a consequence of the fact that the color-field lines are parallel the color force between a quark and an antiquark is a constant.

If a particle is confined to a region of space that extends Δx in any direction, then according to the uncertainty principle of quantum physics the particle's momentum has a spread of about $\hbar/\Delta x$, where $\hbar$ is Planck's constant divided by 2π. This means that when quarks are confined to a bag of finite size, they move with a momentum on the order of $\hbar/\Delta x$ and hence exert a pressure on the inside surface of the bag. At some equilibrium value the pressure of the quarks on the inside surface balances the confining pressure B that keeps the matter inside the bag. It is this equilibrium value that determines the size of the hadrons. The pressure of ordinary space balancing the "uncertainty-principle pressure" of the quark vapor is somewhat analogous to the pressure of water balancing the pressure of the water vapor in a bubble.

It may look as though the problem of quark confinement has been trivially solved by simply postulating that the quarks are imprisoned in bags or bubbles. Yet a bag containing only a single quark would in effect be a quark, albeit a "fat" one. Such a particle would have the same charge and other properties of a single quark, and so the bag model as I have presented it so far does not account for the failure to observe an isolated quark. To account for quark confinement I must introduce the dynamical property of color. Each flavor of quark comes in three colors: "red," "blue" and "yellow." The antiquarks have anticolors: "antired," "antiblue" and "antiyellow." Quark color has nothing whatsoever to do with visual color. The word color is used because the way different colored quarks combine in quantum mechanics is reminiscent of the way visual colors combine. The best way to think about color (and also about flavor) is simply to imagine it as an extra variable associated with quarks. Just as a quark has a spin angular momentum of $1/2\hbar$ that can be oriented either up or down along any spatial axis, so quarks have flavors and colors that can also have different discrete values.

All the hadrons are "white," that is, they are colorless averages of the three colors. Roughly speaking, ordinary baryons consist of a red, a blue and a yellow quark, since these colors average to white, and ordinary mesons consist of a quark of one color and an antiquark of the corresponding anticolor. One way of establishing the truth of quark confinement is to show that the property of color can never be observed, because to observe a color would be to see a free quark. It is now believed the strong interaction is based on quark color in much the same way that the electromagnetic interaction is based on electric charge. The other interactions (the electromagnetic, the weak and the gravitational) are all "color-blind."

The dynamical effects of quark color can best be elucidated by analogy to a much older concept: electric charge. Michael Faraday introduced the picture of a charge as a source of lines of electric force. A line of force is a curve in space that indicates the direction of the electric force a positive charge would be subjected to if it were placed on the curve. The number of lines in a given spatial region indicates the magnitude of the force; the greater the number of lines, the stronger the force. In Faraday's picture the lines of force begin on positive charges and end on negative charges because a positive charge is repelled by its own kind of charge and attracted by its opposite. The number of lines beginning or ending on a charge is a measure of the strength of the charge. Hence large charges generate strong electric forces.

Quark color can be thought of as a kind of charge. In electrodynamics there is only one type of charge (which can be either positive or negative), but in chromodynamics there are several types, corresponding roughly to the allowed color combinations. The color force acts not only to move the quarks but also to change their color, and so the color field itself must carry color. (An electric field acts only to move electric charges, not to change their sign or magnitude; this phenomenon was given mathematical generality in a theory of electrodynamics developed by C. N. Yang of the State University of New York at Stony Brook and Robert L. Mills of Ohio State University.) Quarks interacting in this way are consistent with the results of experiments where high-energy electrons were inelastically scattered by protons, the same experiments that revealed that protons are made up of quarks. (This consistency was pointed out by David J. Gross and Frank Wilczek of Princeton University and H. David Politzer of Cal Tech.) Fortunately it is not necessary to understand the mechanism of color transmutation in order to appreciate the relation between quark color and quark confinement.

Let me pursue the analogy between electric charge and color charge [*see illustrations on pages 113 through 115*]. It is helpful to think of a line of force that begins on one charge and ends on another as a kind of rubber band connecting the charges, because tension along the line of force is in part responsible for the attractive force between the charges. Adjacent lines exert a pressure on each other, and so the lines do not combine to form a single straight line linking the charges. The net force between the charges is the sum of the forces associated with each line. As the electric charges are separated the lines bend more. The force gets weaker because only the component of the field lines that lies along the direction between the charges has a net effect. The force drops off as the square of the separation between the charges. This inverse-squared relation is known as Coulomb's law.

One major difference between an ordinary electric field and the color field as described by the bag model is that the color lines of force connecting the quarks do not exist everywhere but are confined to the bag. At the surface of the bag where there are no quarks (and hence no color charges) the lines cannot begin or end, and since the lines cannot leave the bag, they must be tangent to its surface. At the surface the outward pressure of the color-field lines is not balanced by adjacent lines, since there are none, but by the confining bag pressure B. As a result the field lines are kept parallel to one another.

The fact that the lines are parallel intensifies the force acting between the charges because the lines all pull in the same direction. As the quarks (color charges) are separated the lines do not

		LEPTONS		QUARKS	
FIRST FAMILY	PARTICLE	e (ELECTRON)	ν_e (ELECTRON NEUTRINO)	u (UP)	d (DOWN)
	CHARGE	-1	0	2/3	$-1/3$
	MASS	$.5 \times 10^{-3}$	0? ($<.6 \times 10^{-7}$)	$\sim .5 \times 10^{-2}$	$\sim 1 \times 10^{-2}$
	POSITED	——	1929	1963	1963
	DISCOVERED	1898	1954	——	——
SECOND FAMILY	PARTICLE	μ (MU)	ν_μ (MUON NEUTRINO)	c (CHARMED)	s (STRANGE)
	CHARGE	-1	0	2/3	$-1/3$
	MASS	.11	0? ($<.57 \times 10^{-3}$)	~ 1.7	$\sim .3$
	POSITED	——	1957	1970	1963
	DISCOVERED	1936	1962	1974	1947
THIRD FAMILY	PARTICLE	τ (TAU)	ν_τ (TAU NEUTRINO)	t (TOP OR TRUTH)	b (BOTTOM OR BEAUTY)
	CHARGE	-1	0	2/3	$-1/3$
	MASS	1.78	0? ($<.25$)	?	~ 5
	POSITED	——	1975	1977	——
	DISCOVERED	1975	?	19??	1977
FOURTH FAMILY?		?	?	?	?
FIFTH FAMILY ?		?	?	?	?
. . . ?		?	?	?	?

FAMILIES OF ELEMENTARY PARTICLES each consist of two kinds of lepton (a particle that is not subject to the strong force and that shows no signs of an inner structure) and two "flavors," or kinds, of quark, each in three different colors. Like the leptons, the quarks seem to act like dimensionless points. The charge is measured in units of the charge of the proton, and the mass is given in terms of its energy equivalent in billions of electron volts (GeV). Although no quark has been seen in isolation, a discovery date is listed for each flavor that marks when a two-quark hadron consisting entirely of that flavor was found. The family-classification scheme is based on the fact that the quarks in a given family interact with each other in the same way that the quarks in any other family do. The leptons have been assigned to the families in order of increasing mass. Three families are known, and there is no reason blocking the discovery of still others.

bend but continue to pull in the same direction. The number of lines also remains the same. It is therefore a direct consequence of the fact the lines are parallel that the color force between a quark and an antiquark is a constant. The magnitude of the force is independent of the distance between the quarks, except when they are extremely close together.

The constant force between oppositely colored quarks is what keeps them together permanently. The magnitude of the force, which is the net tension in the color-field lines connecting the charges, is about 15 tons. (By way of contrast, the electric force of the proton on the electron in a hydrogen atom is about 10^{-11} ton.) The force of confinement is truly strong. Quarks are confined because the constancy of the force between a quark and an antiquark means that an arbitrarily large amount of energy is required to separate them. For example, to separate the quark and the antiquark in a meson by one inch would require the energy needed to make 10^{13} proton-antiproton pairs.

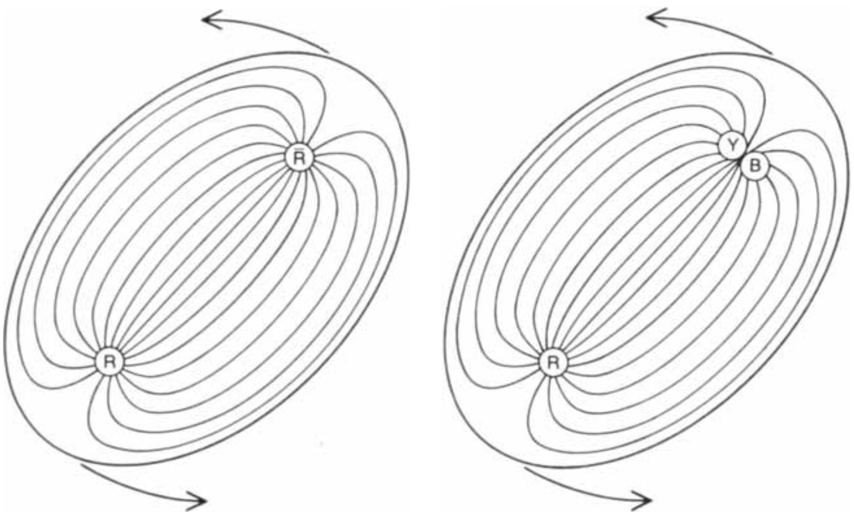
The only composites of quarks that have a finite mass are those in which the color-field lines all terminate inside a finite volume. Since the lines must begin and end on color charges, such composites must be colorless combinations of quarks. In other words, all hadrons are white, in agreement with observation. The problem of quark confinement has been solved by reducing it to color confinement. It is a consequence of the bag model that colors are confined. The way colors combine accounts for the fact that the baryon consists of only three quarks and the meson consists of only a single quark-antiquark pair. The simplest white composite consists of a quark and an antiquark of opposite colors. The next-simplest consists of a red, a blue and a yellow quark. It is impossible to make a colorless combination of quarks unless the number of quarks minus antiquarks is either zero or a multiple of three.

Chromodynamics in conjunction with the bag model can explain many aspects of hadron bahavior that once seemed paradoxical. For example, if a quark and an antiquark were spun, they would recede from each other like two skaters spinning in a circle holding the ends of an elastic rope, except that with quarks the rope consists of the color lines connecting the opposite color charges. When the system is spun, its angular momentum increases, the color lines stretch and energy is pumped into the color field the lines represent. As a result the mass of the system increases. The relation the model predicts between spin angular momentum and mass has been verified experimentally for all the hadrons [*see bottom illustration on this page*]. The experimental discovery of this relation was a main motivation for another model of quark confinement, the string model, in which two quarks cannot be separated because they are linked by a "string." The bag model provides a physical explanation of the nature of the string.

The derivation of the relation between mass and angular momentum on the basis of the bag model indicates why the same relation applies to mesons and baryons. The relation depends only on the color charge of the quarks at the ends of the particle. Since a baryon is

	NAME	QUARK FLAVORS	SPINS	CHARGE	MASS (GEV)
MESONS	π^+	$u\bar{d}$	↑ ↓	+1	.14
	K^+	$u\bar{s}$	↑ ↓	+1	.50
	ρ^+	$u\bar{d}$	↑ ↑	+1	.76
	ϕ	$s\bar{s}$	↑ ↑	0	1.04
	J/ψ	$c\bar{c}$	↑ ↑	0	3.10
	D^{0*}	$c\bar{u}$	↑ ↑	0	2.00
	S^*	$s\bar{s}$ OR $(u\bar{u} + d\bar{d})$	↑ ↓ ↑ ↓	0	1.00
	γ	$b\bar{b}$	↑ ↑	0	9.46
BARYONS	p^+	uud	↑ ↑ ↓	+1	.938
	n	ddu	↑ ↑ ↓	0	.940
	Δ^{++}	uuu	↑ ↑ ↑	+2	1.2
	Σ^+	uus	↑ ↑ ↓	+1	1.4
	Ω^-	sss	↑ ↑ ↑	−1	1.65

QUARK CONTENTS of some prominent hadrons are listed here. The known quarks come in five flavors named "up" (*u*), "down" (*d*), "strange" (*s*), "charmed" (*c*) and "bottom" (*b*). The antiquarks are designated $\bar{u}$, $\bar{d}$, $\bar{s}$, $\bar{c}$ and $\bar{b}$. The hadrons have been divided into two classes depending on how they decay. The baryons are hadrons that decay ultimately into a proton; they are all made up of three quarks. The mesons are hadrons that decay entirely into photons and leptons or into proton-antiproton pairs; most mesons are made up of a single quark-antiquark pair. The arrows in the spin column indicate the direction in which the quarks are spinning.



MESONS AND BARYONS have the same relation between angular momentum and mass, and the bag model is able to explain why that is the case. If the quark and the antiquark in a meson were spun (*left*), they would recede from each other like two skaters spinning in a circle holding the ends of an elastic rope, except that with quarks the rope consists of the color-field lines connecting the opposite color charges. When the system is spun, the angular momentum increases, the color lines stretch and energy is pumped into the color field the lines represent. As a result the mass of the system increases. The relation depends only on the difference in color between the quarks at each end of the particle. Since a baryon (*right*) is white, the two quarks (yellow and blue) at one end have a net color (antired) that is the opposite of the color (red) of the quark at the other end. Thus the quarks have the same color as the meson's antiquark.

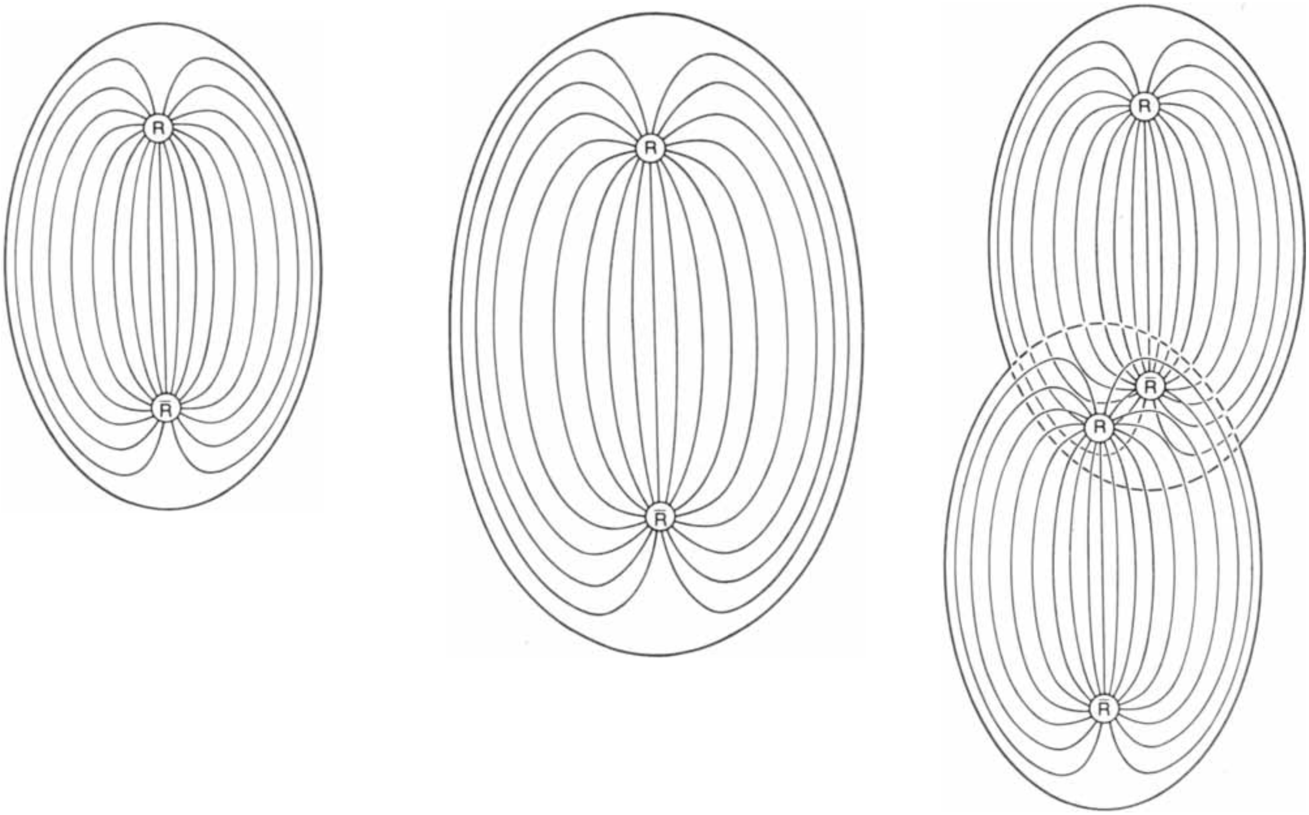
white, the quarks at one of its ends have a combined color that is the opposite of the color of the quark at the other end. In other words, they must have the same net color as the antiquark in the meson.

The concept of color can also account for the production of hadrons. An electron and a positron (a positively charged electron) can be created as a pair, and so can a quark and an antiquark. Such a process requires a "polarizing field" acting to separate the color charges of a quark-antiquark pair in a meson. A new pair will appear spontaneously in the color field of the quark and antiquark of the meson. The color lines link the quark of the meson with the newly created antiquark and the antiquark of the meson with the newly created quark. Both quark-antiquark pairs are colorless, so that color lines do not join them. Because a constant force does not act between the pairs they can separate. In other words, the strong interaction can transform one meson into two others, for example a ρ meson into two pions.

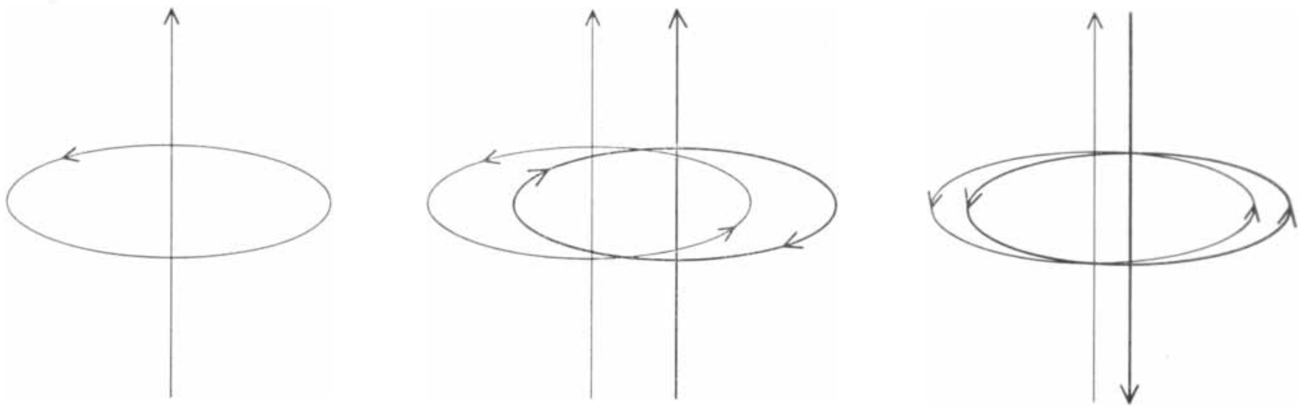
The concept of color-field lines has also been successful in describing the self-annihilation of an electron and a positron at very high energies. If the particles carry a sufficiently large energy, they can generate an electromagnetic field strong enough to polarize the vacuum and create an oppositely charged quark-antiquark pair. If the pair consists of quarks of low mass, they fly apart just about as fast as the electron and the positron come together but not necessarily along the same axis. The quark and the antiquark are of course connected by color-field lines. As the quark and the antiquark rapidly move apart pairs of colored quarks that form mesons are created by the strong color field. The pressure of the color-field lines confines the diverging quark-antiquark pair to a tubular region of space about 10^{-15} meter thick. On the basis of the uncertainty principle the component of the pair's momentum that lies in the direction of the width of the tube is found to be the equivalent of only a few hundred million electron volts. Since the quark and the antiquark are rapidly diverging, the component of momentum along the axis defined by their motion is much higher. As a result the emerging hadrons also move primarily in this direction. Such hadron "jets" have been seen recently at SLAC and at other laboratories.

Electricity and magnetism are interrelated phenomena, as when electric charges in motion generate magnetic fields. Since color charge was introduced in analogy to electric charge, perhaps there is a kind of color magnetism. Indeed there is, and one of its consequences was one of the earliest successes of chromodynamics. This consequence was discovered independently by Howard Georgi, Alvaro De Rújula and Sheldon L. Glashow of Harvard University and by a group of us from M.I.T. It is a well-known empirical fact that currents flowing in the same direction through two parallel wires generate a magnetic force that pulls the wires together. By the same token currents flowing in opposite directions generate a magnetic force that pushes the wires apart.

Since quarks have spin angular momentum, they act as moving color charges and hence generate a color magnetic field. A spinning quark can be thought of as a color-current loop in which a fixed color charge circulates in a plane at right angles to the direction of the quark spin. Since the quarks in a meson are of opposite color, if the spins are parallel, then the color currents flow in opposite directions and so the quarks magnetically repel each other. On the other hand, if the spins are antiparallel, then the color currents flow in the same direction and so the color-magnetic force pulls the quarks toward each other. That makes particles whose quark spins are aligned more massive than those whose spins are opposed. This consequence of chromodynamics has been observed for all the hadrons. For



TRANSFORMATION OF A MESON (*left*) into two mesons requires a "polarizing field" acting to separate the color charges of the quark pair in the meson (*middle*). A new quark-antiquark pair appears spontaneously in the polarizing field between the quark and the antiquark of the meson (*right*). The color lines link the quark of the meson with the newly created antiquark and the antiquark of the meson with the newly created quark. Both quark-antiquark pairs are colorless, and so color lines do not join them. Because a constant force does not act between the pairs (as it does between the quark and the antiquark in each pair), they can separate and form two mesons.



COLOR-MAGNETIC FIELD is generated by a moving quark (color charge) just as an ordinary magnetic field is generated by moving electric charges. A quark has spin angular momentum; it is therefore always spinning and can be treated as a color-current loop in which a fixed color charge circulates (counterclockwise for spin up) in a plane at right angles to the direction of the quark spin (*left*). Since the quarks in a meson are of opposite color, if the spins are parallel, then the color currents flow in opposite directions and so the quarks magnetically repel each other (*middle*). If the spins are antiparallel, then the color currents flow in the same direction and so the color-magnetic force pulls the quarks toward each other (*right*). That makes hadrons whose spins are aligned more massive than those whose spins are opposed.

example, the ρ meson whose quark spins are parallel is heavier than the π meson whose spins are antiparallel.

The fact that the atomic nucleus is made from a collection of neutrons and protons rather than directly from quarks is most likely a consequence of the magnitude of the color-magnetic forces. The energy of a conglomeration of up and down quarks is lower when they first cluster into colorless lumps (neutrons and protons), which in turn bind together (albeit relatively more weakly) to form an atomic nucleus. The mass density of the quarks in a hadron is only about twice as great as the mass density of the nucleons (protons or neutrons) in the atomic nucleus. This means that if nuclear matter could somehow be squeezed so that the mass density was increased by at least a factor of two, it could be energetically favorable for all the nucleons to lose their individual identity by merging to form a quark fluid. The transition from nuclear matter to quark matter is probably not a phase transition, such as the change from a gas to a liquid as the temperature is lowered, but a more gradual transition, such as the change from a monatomic gas to a diatomic gas as the temperature is lowered.

It would be extremely difficult to squeeze nuclear matter in the laboratory. It is believed, however, that nuclear matter is compressed by huge gravitational forces in a neutron star, the cold remnant of the explosion of a supernova. If the density of the squeezed neutron matter exceeds the density of the quarks in the neutron, the neutrons will merge to form a quark fluid. Hence the dense central region of a neutron star might actually consist of quark matter.

This possibility has important consequences for the stability of the neutron star. The matter of the star is squeezed by gravitational mass, mostly the mass of neutrons. Resistance to the squeeze is provided by the repulsive interaction of the neutrons and by the kinetic energy of the cold neutrons. It is thought the kinetic energy of the trapped quarks accounts for most of the neutron mass, and so the pressure would increase immensely if the neutrons merged and liberated the quarks. This effect would appreciably increase the allowed mass of a stable neutron star. Work is under way to discover the properties of quark matter so that the allowed mass can be calculated.

Since the color-field model describes so many features of the properties and interactions of hadrons, particle physicists have hoped that quark confinement would be a necessary consequence of chromodynamics. Is it possible the bag model of hadrons could follow from chromodynamics? The answer is yes. To understand how this might be so consider a possible analogy to superconductivity: the absence of electric resistance in a metal at a temperature near absolute zero. An applied magnetic force cannot penetrate a superconducting body. The magnetic field induces electron currents in a thin layer at the surface of the body; these currents in turn give rise to a magnetic field that is opposed to the applied field and is just large enough to cancel the applied field within the body. Under some conditions, however, the applied magnetic field can penetrate the body by creating in it a region of normal conductivity, in which lines of magnetic flux can be trapped. This phenomenon is called the Meissner effect.

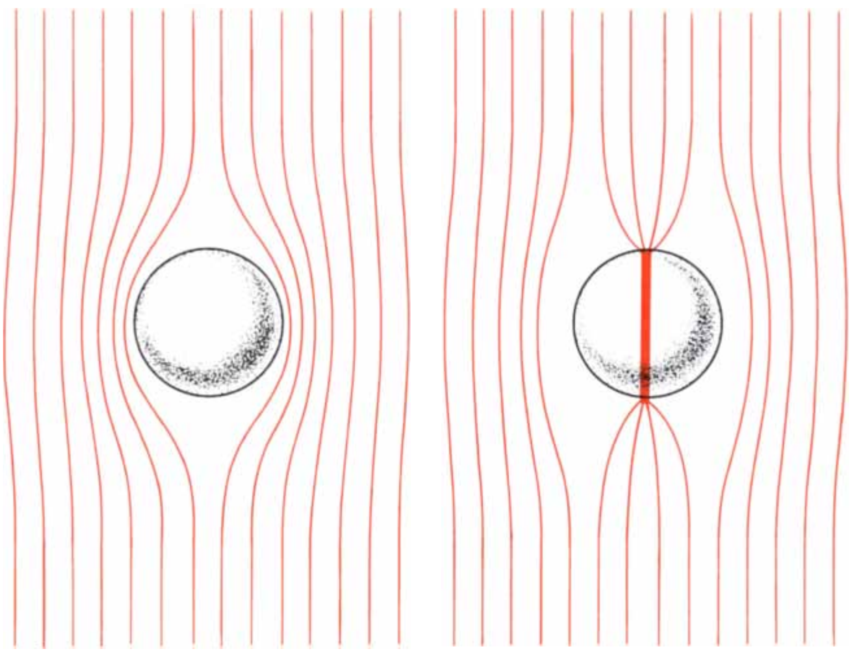
The Meissner effect may have an analogue in chromodynamics. Perhaps the vacuum in quantum chromodynamics expels color-field lines just as a superconductor expels magnetic-field lines. In that case color fields might exist only where they expended energy to create a region of "normal vacuum" (analogous to the region of normal conductivity of the Meissner effect) in the surrounding vacuum. The color-expelling property of the surrounding vacuum would serve the same function as the bag pressure in the bag model, namely keeping the color-field lines confined to the inside of the hadrons. Charles G. Gross, Curtis G. Callan and R. Dashen of Princeton have made detailed calculations that support such a dual-vacuum model. Another model of this kind has been investigated by Holger B. Nielsen and M. Minomiya of the Niels Bohr Institute in Denmark. These models are at the forefront of continuing work on quark confinement.

Quantum chromodynamics and the bag model have been fairly successful in providing a mechanism for permanently sequestering quarks inside hadrons and hence for explaining many aspects of the strong force between quarks. Whether or not the mechanism succeeds ultimately depends on how well it works when it is exploited to make more refined and detailed calculations of the properties of hadrons. The phenomenon of quark confinement is perhaps the most paradoxical aspect of the quark model. A definitive explanation of the phenomenon will eliminate the last major detraction from the many successes of the quark hypothesis.

It is time to take stock of the picture of subatomic particles that has emerged from the quark model. Quarks are considered to be the elementary constituents of matter since they seem to interact like dimensionless points, but are they truly elementary? The history of science provides many examples of physical entities that were once thought to be elementary but that later turned out to be composite. Atoms, atomic nuclei, protons and neutrons were all thought to be indivisible at various times in this century until further experimentation unearthed their constituent parts. Is it possible that quarks will suffer the same fate? And what about the

	PARALLEL SPINS			ANTIPARALLEL SPINS		
ORBITAL ANGULAR MOMENTUM	PARTICLE	QUARK CONTENT	MASS (GEV)	PARTICLE	QUARK CONTENT	MASS (GEV)
0	Δ^+	uud	1.23	p	uud	.94
0	Σ^{+*}	uus	1.38	Σ^+	uus	1.19
0	Ξ^{-*}	dss	1.54	Ξ^-	dss	1.32
0	ρ^+	$u\bar{d}$	.77	π^+	$u\bar{d}$	.14
1	A_2^+	$u\bar{d}$	1.31	B^+	$u\bar{d}$	1.23
2	g^+	$u\bar{d}$	1 69	A_3	$u\bar{d}$	1.64
0	K^{-*}	$u\bar{s}$	.89	K^-	$u\bar{s}$	.49
1	K^{-**}	$u\bar{s}$	1.42	"Q"	$u\bar{s}$	1.3
2	K^{-***}	$u\bar{s}$	1.78	L^-	$u\bar{s}$	1.77

MASS EFFECTS OF COLOR MAGNETISM are shown here for various hadrons. The table clearly indicates that other things being equal a particle whose quark spins are parallel is more massive than one whose spins are antiparallel. The orbital angular momentum is given in units of Planck's constant divided by 2π. When the quarks rotate, they move away from each other as a result of the centrifugal force to which they are subjected in the rotating reference frame; the greater the rotation rate, the larger the separation. The color-magnetic interaction gets weaker with increasing separation. This means that as orbital angular momentum increases (and hence as rotation rate increases) the masses in the parallel and antiparallel states converge.



PROPERTIES OF A SUPERCONDUCTOR (a metal at temperatures near absolute zero that has no resistance to electric current) might serve as an analogue to how quark confinement could follow from chromodynamics, or color dynamics. An applied magnetic field cannot penetrate a superconducting body (*left*). The magnetic field induces electron currents in a thin layer at the surface of the body; the currents in turn give rise to a magnetic field that is opposite to the applied field and just large enough to cancel the applied field within the body. The only way the applied magnetic field can penetrate the body is for it to create a region of normal conductivity, in which lines of magnetic flux could be trapped (*right*). The vacuum in quantum chromodynamics might expel color-field lines just as a superconductor expels magnetic-field lines. In that case color fields might exist only where they expended energy to create in the surrounding vacuum a region of "normal vacuum" analogous to the region of normal conductivity. The color-expelling property of the surrounding vacuum would serve the same function as the bag in the bag model, namely keeping the color-field lines, and hence the quarks, confined.

leptons, which also are regarded as being elementary? Might they too have constituent parts?

The possibility that quarks or leptons have a substructure is all but ruled out by quantum mechanics and the theory of relativity. Quarks and leptons are thought to be no larger than 10^{-17} meter. For the electron this upper limit in size comes directly from experiments. On the basis of the uncertainty principle it is possible to calculate the momentum of a hypothetical constituent entity confined to a region of space 10^{-17} meter in diameter. The theory of relativity indicates that the kinetic energy of the constituent would be greater than its momentum multiplied by the speed of light. The kinetic energy of the constituent confined to a region 10^{-17} meter in diameter would increase the mass of the quark or lepton of which it was a part; the least energetic constituent would contribute a mass of at least 20 times the mass of the proton. Therefore it is extremely improbable that the electron, say, has constituents. For that to be the case there would have to be an accidental and nearly perfect cancellation of the binding energy of the constituent particles by their kinetic energy. In physics such accidents rarely happen, and so it is most unlikely that electrons (as well as other leptons and quarks) are composite in this sense.

If quarks and leptons are indeed elementary, nature cannot continue to complicate the picture of matter by creating new kinds of particles that are even more fundamental. Instead it seems to be complicating the picture by creating new flavors of quarks and new species of leptons. The known quarks and leptons can be classified by grouping them into families. Each family consists of four particles whose total electric charge is zero, namely two leptons (a particle and its associated neutrino) and two quark flavors (in the three different colors: red, blue and yellow). Three families have been identified so far. The first consists of the electron, the electron neutrino, the "up" quark and the "down" quark. The second consists of the muon, the muon neutrino, the "charmed" quark and the "strange" quark. The third consists of the tau particle, the tau neutrino, the "bottom" quark and an undetected but theoretically expected "top" quark.

The members of the first family play the largest role in naturally occurring phenomena. The other elementary particles, born chiefly in particle accelerators, have an ephemeral existence and are responsible only for extremely subtle effects in ordinary matter. They decay into members of the first family by the weak interaction.

This classification scheme is based on the fact that the four particles in a given family interact with one another in the same way that the four particles in any other family do. For example, within a family one quark flavor can be transformed into the other by the weak interaction. (That happens in ordinary radioactivity, where an up quark is changed into a down quark or vice versa.) By the same token one lepton can be transformed into the other. Quarks can also interact across family lines, although this interaction is considerably weaker. There are no known cases of leptons crossing family lines, of leptons changing into quarks or of quarks changing into leptons.

The members of the first family had all been identified by the late 1960's, when the up and down quarks were detected within the proton and the neutron. The particles in the second family were discovered between 1936 and 1974. The discovery in 1974 of the charmed quark, the last member of the second family, lent tremendous support to the quark model. Four years earlier the necessary existence of the charmed quark had been deduced from theoretical considerations by Glashow, John Iliopoulos and Luciano Maiani. The quark was subsequently discovered by Samuel C. C. Ting and his co-workers at the Brookhaven National Laboratory and by Burton D. Richter and his co-workers at SLAC. The third family came on the scene in 1975 with the discovery of the tau particle. Although it is not known with certainty that the tau's neutrino is distinct from the electron's neutrino, most particle physicists believe it is. Unless the top quark is extremely massive, it should be discovered within the next five years.

There is no theoretical reason blocking the discovery of still more families of particles. Some theorists believe insights into this possibility could come eventually from unified theories of weak and electromagnetic interactions. In the late 1960's one such theory was proposed independently by Steven Weinberg of Harvard and by Abdus Salam of the International Centre for Theoretical Physics in Trieste. Their proposal successfully accounts for many properties of the weak interaction. In its current form the proposal is fully compatible with the family-classification scheme, but it offers no explanation for the existence of the families. Although more far-reaching proposals that incorporate Weinberg's and Salam's theory provide such an explanation, they lack empirical support. Moreover, no compelling explanation has been offered for why the particles in a given family have the masses they do. Whereas a definitive model of quark confinement and a confirmation of the elementary nature of quarks and leptons would close a long chapter in particle physics, namely the search for the ultimate constituents of ordinary matter, the possible proliferation of families is opening another chapter.

The Animals of the Burgess Shale

The fossils of a rock formation in western Canada are a rich sample of an animal community in the mid-Cambrian. Some of the animals are ancestors of those living today; others are unique and bizarre

by Simon Conway Morris and H. B. Whittington

By far the most numerous fossils representing the first abundant life on the earth are the hard parts of various marine animals without backbones: shells and similar fragments of external skeleton. This makes for a lopsided fossil record. For example, of the 30 or so phyla of animals living today more than half are made up of species with few hard parts or none. As a result the descent of these phyla remains largely undocumented by fossil evidence.

Fortunately the situation is not completely lopsided. A few geological deposits have been discovered that as a result of exceptional circumstances contain exquisitely preserved fossils of animals that are partly or entirely soft-bodied. Here we shall describe one such deposit: the Burgess Shale of western Canada. The great age and the rich variety of the marine invertebrates in the Burgess Shale make it perhaps the best-known of all such deposits. In addition to describing the Burgess Shale fauna we shall attempt to reconstruct the kind of underwater environment these organisms inhabited early in Paleozoic times: some 530 million years ago.

In the fall of 1909 the Secretary of the Smithsonian Institution, Charles Doolittle Walcott, was searching for fossil-bearing rock formations in British Columbia. Following a footpath that ran across the western slope between Wapta Mountain and Mount Field in the southern part of the province, Walcott literally stumbled over a block of shale that had fallen onto the path from the slope above. Examining the easily split rock, he was astonished to find the fossil impressions of a number of soft-bodied organisms preserved in its layers. In a letter to a colleague in Toronto dated November 27, 1909, he dryly reported that he had spent "a few days collecting ... in the vicinity of Field and found some very interesting things."

Walcott returned to the spot the following year to search upslope for the shale stratum that had been the source of his fallen rock. His search was successful: he found two fossil-bearing shale exposures separated by a vertical distance of some 70 feet. He did shallow quarrying in both; the lower exposure proved to be the richer of the two. He shipped back to the District of Columbia thousands of fossil specimens that he removed from what he called his Phyllopod Bed. (The term, little used by paleontologists today, refers to certain fossil arthropods, or joint-legged animals, that are probably ancestral to living crustaceans.)

As Walcott's own work and decades of study by others have shown, the fossils of the Burgess Shale include a great abundance of marine invertebrates: more than 120 species. Some of them belong to the Phylum Porifera: the sponges. This phylum of primitive animals is the only one in the subkingdom of parazoans, a category higher than the subkingdom of the one-celled protozoans but lower than the subkingdom of the many-celled metazoans. Perhaps 10 other species represent metazoan phyla that were unknown before they were found in the Burgess Shale; they are not present elsewhere in the fossil record. The scores of other species that lack hard parts can be assigned to one or another phylum of metazoans with living relatives as follows:

Coelenterates: the phylum that includes such living marine animals as jellyfishes, sea pens and corals. The Burgess Shale coelenterate species number perhaps four.

Echinoderms: the phylum that includes, among others, starfishes, sea cucumbers and crinoids, or sea lilies. At least four species of Burgess Shale echinoderms are recognized.

Mollusks: the phylum that includes, among others, oysters and clams, squids and octopuses and the primitive chitons (of the Class Amphineura). Three Burgess Shale molluscan species are recognized.

Arthropods: the phylum that includes, among a great many others, lobsters, shrimps, crabs and barnacles (all of the Class Crustacea) and the less familiar terrestrial animal *Peripatus* (a member of the Class Onychophora). The Burgess Shale arthropods include several representatives of the long extinct trilobite class, a peripatus-like animal that was aquatic rather than terrestrial, and about 30 other species of arthropods.

Priapulids: a minor phylum of unsegmented marine worms. The living genus, *Priapulus,* gives the phylum its name. Seven species of these now obscure bottom burrowers flourished in the Burgess Shale muds.

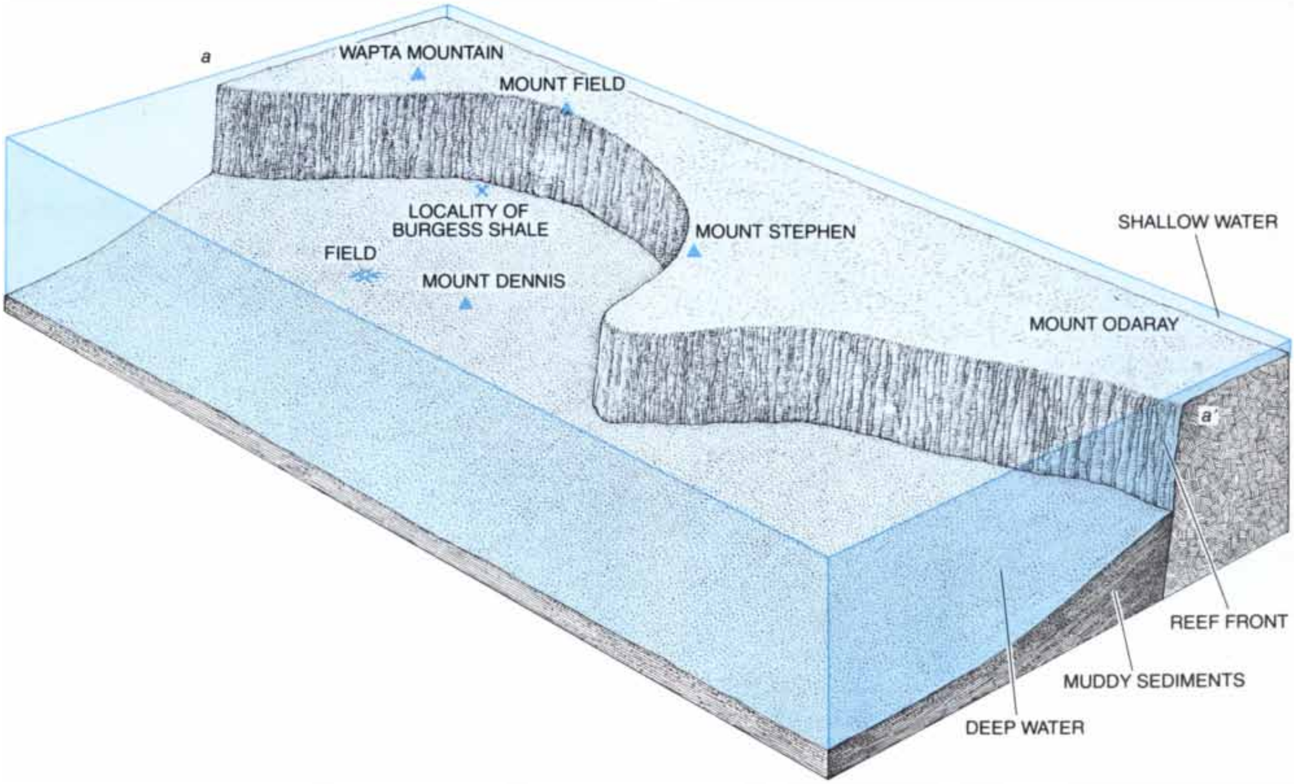
Annelids: the phylum that includes earthworms, leeches and a less familiar but very large class of marine worms, the polychaetes. The annelid phylum is represented in the Burgess Shale by six species.

Finally, we find among the Burgess Shale fauna one of the earliest-known invertebrate representatives of our own conspicuous corner of the animal kingdom: the chordate phylum. Among its living representatives (other than vertebrates) are the sea squirts and the peculiar marine animal *Amphioxus.* The chordates are represented in the Burgess Shale by the genus *Pikaia* and the single species *P. gracilens.*

Such a remarkably preserved soft-bodied fauna, representing eight known and 10 or more previously unknown phyla that flourished in the Middle Cambrian, is by itself of great interest to students of the fossil record. In addition to this intrinsic interest, however, the Burgess Shale invertebrates, with their specialized adaptations, have an even wider importance in clarifying the early evolution of the animal kingdom. The only earlier-known soft-bodied animals are representative of late Precambrian time, 700 to 600 million years ago and therefore at least 70 million years earlier than Middle Cambrian times. These are the Ediacara animals, first discovered some 30 years ago in the Ediacara Hills of southern Australia and recognized since then in a number of other places throughout the world. The Ediacara

FOUR ANIMALS that lived in the ocean in Middle Cambrian times, some 530 million years ago, are seen in these fossils. At the top left is a trilobite, *Olenoides,* one of the many animals whose anatomy has been preserved in remarkable detail in the silts that solidified to form the Burgess Shale of British Columbia. The specimen is 5.5 centimeters long. Unlike most arthropods, or joint-legged animals, *Olenoides* had unspecialized limbs. At the top right is another Burgess Shale arthropod, *Waptia.* When this bottom-feeding animal was extended, it was four centimeters long. At the bottom left is *Opabinia,* one of about 10 animal species found in the shale that belong to previously unknown phyla. It had five eyes and steered its seven-centimeter body with a vertical tail fin as it swam close to the sea floor in search of food. At the bottom right is one of the many unsegmented marine worms that inhabited the sea floor. It is *Selkirkia,* one of the priapulid phylum. With its projecting proboscis it measured five centimeters. A successful group in the Cambrian, priapulid worms are now rare.

LOWER QUARRY, named the Phyllopod Bed by Charles Doolittle Walcott of the Smithsonian Institution, who first sampled the Burgess Shale, shows patches of winter snow. The view looks south. This and a higher shale exposure were requarried for fossils in 1966–67.

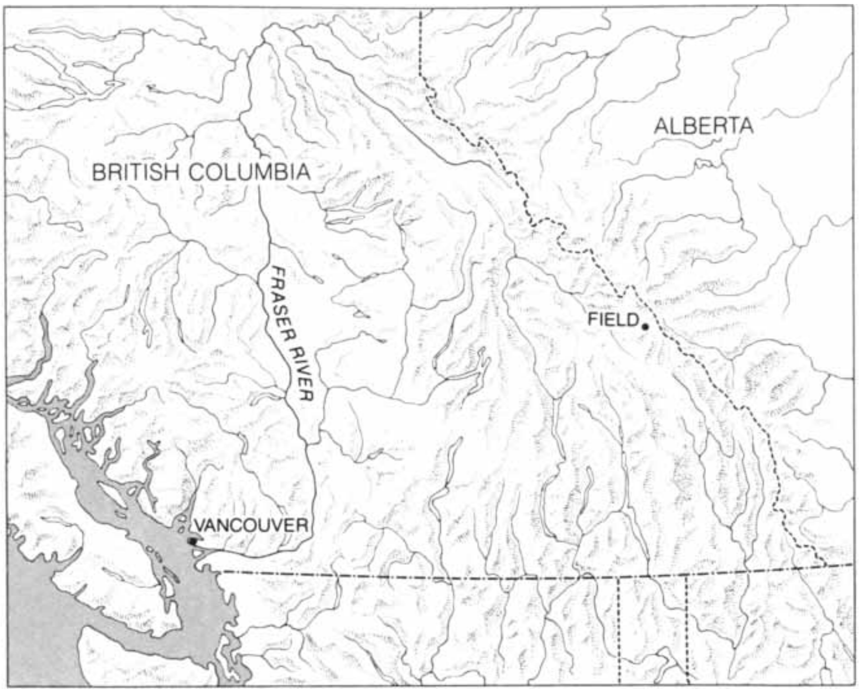


BURGESS SHALE OUTCROP, marked by the colored *X* in this block diagram of the Middle Cambrian seascape, is a small portion of an extended sea-bottom deposit of silts accumulated at the foot of a deeply embayed algal reef that rose vertically some 530 feet above the shallowest silts. The reef did not rise above sea level but was covered by shallow water. The vertical scale in the block diagram is exaggerated by a factor of five, and the distance (*a–a'*) from north to south along the reef is some nine miles. Later uplift and dissection gave rise to the peaks of the Rockies along the border of Alberta and British Columbia, indicated by colored triangles. The reconstruction of the sea floor and the reef is based on the work of I. A. McIlreath of Petro-Canada, one recent investigator of the unique formation.

fauna stands in marked contrast to the Burgess Shale fauna both in the kinds of animals represented (chiefly coelenterates) and in these earlier animals' limited range of specialization.

The event that separates the impoverished Ediacara fauna from the Burgess Shale fauna is an explosive evolutionary diversification of multicelled animals that took place near the beginning of Cambrian time. The fossils of the Burgess Shale thus give us a unique glimpse into the results of this sudden metazoan adaptation relatively soon after it occurred.

In spite of the work done by Walcott and others significant gaps remain in what is known about the Burgess Shale paleoenvironment and how its fauna was preserved. A fuller appreciation of these gaps stimulated a reinvestigation of the site by the Geological Survey of Canada, beginning more than a decade ago. The authorities of the Yoho National Park in British Columbia and the Parks Canada administration in Ottawa granted special permission to collect material from the shale outcroppings. Walcott's quarries were reopened in 1966 and 1967 under the direction of J. D. Aitken of the Geological Survey of Canada. Both the new material collected during these two seasons and a part of the great Burgess Shale collection amassed by Walcott some 60 years earlier then came to us at the University of Cambridge for analysis.



BURGESS SHALE FORMATION is situated some 350 miles northeast of Vancouver near the town of Field, B.C. The fossil-rich formation was found accidentally by Walcott in 1909.

What kind of environment did the Burgess Shale fauna inhabit? Studies done by I. A. McIlreath of Petro-Canada and W. H. Fritz of the Geological Survey of Canada show that the animals lived on or in a muddy bottom where sediments had accumulated at the base of a gigantic reef. This structure, made up of material secreted by algae, rose vertically for hundreds of feet from a deep-water basin that was gradually being filled with sediments. Scattered outcrops of the reef front can still be traced for miles across British Columbia. The bottom waters of the basin were apparently limited in circulation, rich in hydrogen sulfide and poor in oxygen. The various invertebrates flourished where the muddy sediments were banked high enough against the reef to be clear of the stagnant bottom waters, about 530 feet below sea level.

The reef-front sediments were not stable. Studies of the shale by D. J. W. Piper of Dalhousie University show that periodic slumping resulted in the flow of mud into the deeper anaerobic waters of the basin. These flows wiped away all the surface tracks and subsurface burrows made by the Burgess Shale fauna. Because the animals trapped in the torrents of mud died during or shortly after their burial they could not leave new traces. This means that the way of life of each species must be deduced from a study of their organs of locomotion and from comparisons with living invertebrates of the same kind.

At the same time the catastrophic burials, in anaerobic deposits of fine silt where scavengers could not survive, greatly favored the preservation of the animals' soft parts. As the mud gradually compacted and became hard rock the buried carcasses were flattened and the soft parts were transformed into thin films of calcium aluminosilicate. In general the films are rather dark, but certain parts of most specimens are preserved as highly reflective areas.

Paradoxically, although the animals' soft parts are wonderfully preserved, signs of rotting after burial can often be detected. Many specimens are associated with a black-stained area, a result of the body contents of the carcass seeping out into the surrounding mud. In extreme cases the fossil of a worm consists only of a hollow bag of cuticle because practically all the animal's internal organs have been destroyed by decay. In some worms a subtler indication of decay is the pulling away of body-wall muscles from the cuticle.

Burial in a mud flow has other important effects. For one thing, many of the animals came to be buried at all angles; the shale bedding has therefore preserved them in a variety of orientations that reveal much more of the animals' anatomy than simple horizontal burial does. For another, the fluid sediments that penetrated between the appendages of animals such as arthropods and polychaetes during the turbulent flow of silt were eventually reduced to thin layers of shale. Judicious work with a microchisel enables one to remove these fine layers, thereby revealing further details of a specimen's anatomy that would otherwise remain hidden.

The composition of the Burgess Shale fauna upsets the conventional notion of what makes up a typical assemblage of Cambrian animals. The fossils found at most Cambrian localities are the exoskeletons of such arthropods as trilobites, the shells of various members of the brachiopod phylum and of such echinoderms as the extinct plate-shelled Eucrinoid class. Animals such as these account for barely 20 percent of the invertebrate genera in the Burgess Shale. Is it justified, then, to regard the Burgess Shale assemblage as the Cambrian norm, at least with respect to the fauna of deeper waters, and to view the other Cambrian faunas as being skewed by the selective fossilization of animals with hard parts?

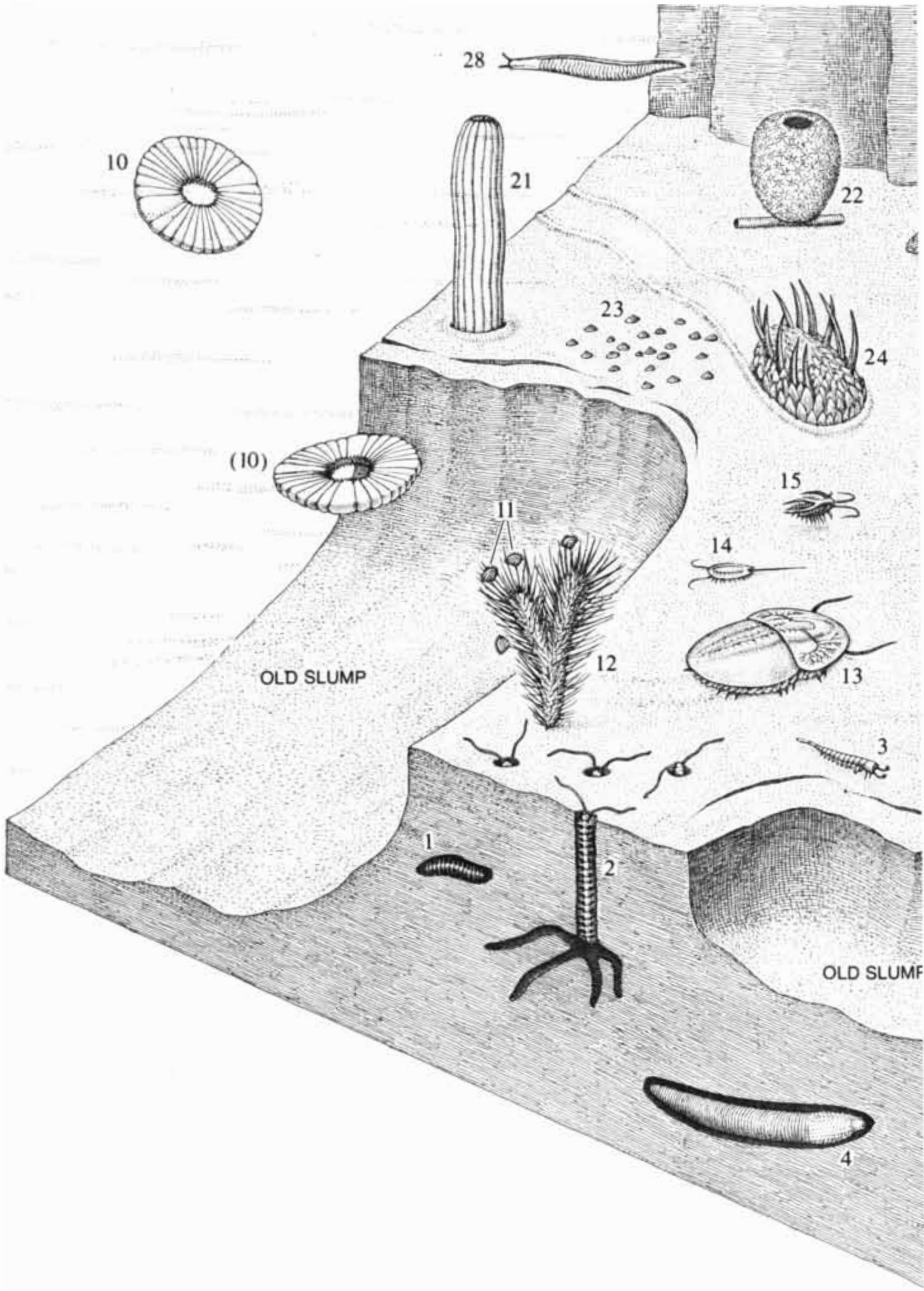
Since the Burgess Shale represents a single environment that has been frozen for a split second of geologic time, no firm answer can be given to the question. Several factors nonetheless suggest that the Burgess Shale fauna was not untypical of Middle Cambrian times. The scattered occurrence of species similar to those from the Burgess Shale in other Cambrian rocks hints at the existence in this period of a widespread soft-bodied fauna. Furthermore, in some Cambrian fossil assemblages certain rather peculiar species were able to flourish because access by sea to the area of deposition was limited. The Burgess Shale, on the other hand, lay at the edge of the open sea and would have been exposed to colonization by marine larvae floating in from other areas. This circumstance adds weight to the hypothesis that the Burgess Shale fauna approaches the Cambrian norm.

In this connection it should be noted that representatives of certain modern invertebrates that have almost certainly had a long geological history are absent from the Burgess Shale. No species of the platyhelminth phylum, the flatworms whose living members include flukes, tapeworms and planarians, are present. There are also no species of another worm phylum, the Nemertea, which includes the modern proboscis worm, and none of still another, the sipunculid phylum. It may be that such worms are not represented because the reef-front environment was not suitable for them.

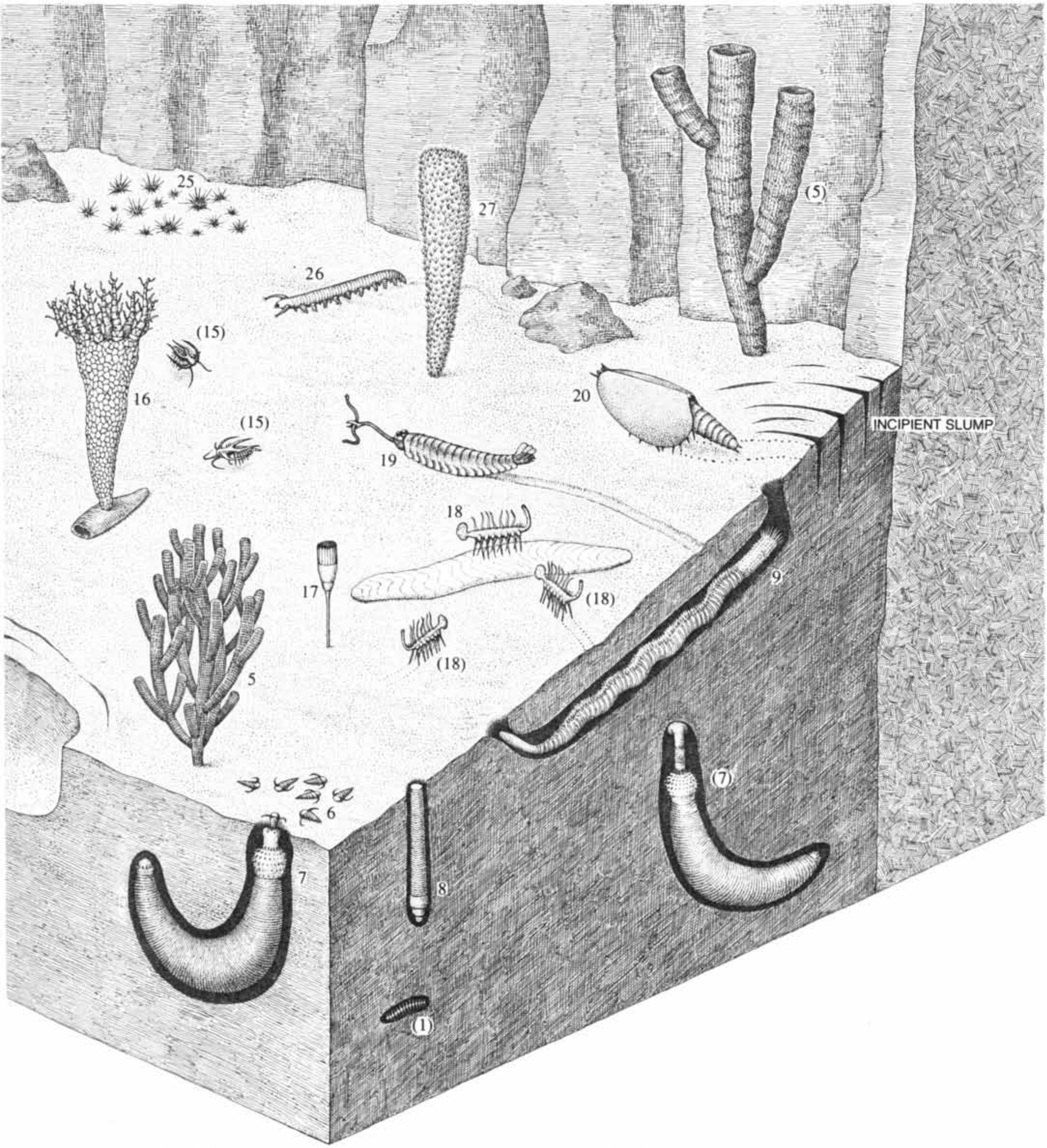
Most of the species found in the Burgess Shale can be placed in the ecological framework of a bottom-dwelling marine community that thrived on the muddy sea floor between intervals of slumping. The mud supported an active group of burrowing invertebrates, with priapulid worms predominant. Attached to the sea floor and growing to various heights were a variety of sponges representing at least 15 genera; they fed on food particles suspended in the water. Actively patrolling the sea-floor surface or plowing through the mud in search of food were many species of arthropods. Certain brachiopods occupied a peculiar niche: they attached themselves to the elongated spicules of one of the sponges, *Pirania*. For the brachiopods the advantages are obvious: they lived somewhat above the turbid waters of the sea floor and could capture food particles such as the sponges fed on at these higher levels.

In addition to this community of fixed and mobile surface dwellers and burrowers a number of free-swimming species inhabited the waters along the reef front. Of these animals there are only tantalizing glimpses, in the form of rare specimens buried by chance in the slumped sediments. The different members of this pelagic fauna probably lived at different depths; some among them may have been species swept into the reef-front area from the open sea.

At most Cambrian fossil localities the mineralized exoskeletons of trilobites, the most familiar of all Paleozoic arthropods, are in the majority. In the Burgess Shale, however, trilobites—with one exception—are comparatively unimportant. The exception is *Olenoides,* which is of great significance because in several specimens the appendages have



UNDERWATER SCENE where silts slope down from the face of the great reef and the Burgess Shale fauna lived is shown in an idealized reconstruction. No attempt has been made to show the animals in numbers proportional to their fossil abundance. The fauna are identified by number, starting at the bottom left; only about a fifth of the species fossilized in the shale are shown. Most of the immobile animals of the sea floor are sponges: *Pirania* (*12*), seen with brachiopods attached to its spicules; *Eiffelia* (*22*); the gregarious *Choia* (*25*); a gracile species of *Vauxia* (*5*), with a more robust species at the top right, and *Chancelloria* (*27*). Three other immobile animals are *Mackenzia* (*21*), a coelenterate; *Echmatocrinus* (*16*), a primitive crinoid, seen attached to an empty worm tube, and *Dinomischus* (*17*), one of the Burgess Shale species that represent hitherto unknown invertebrate phyla. The burrow-dwelling animals are *Peronochaeta* (*1*), a polychaete worm that fed on food particles in the silt; *Burgessochaeta* (*2*), a second polychaete that captured food with its long tentacles; *Ancalagon* (*4*), a priapulid worm pos-

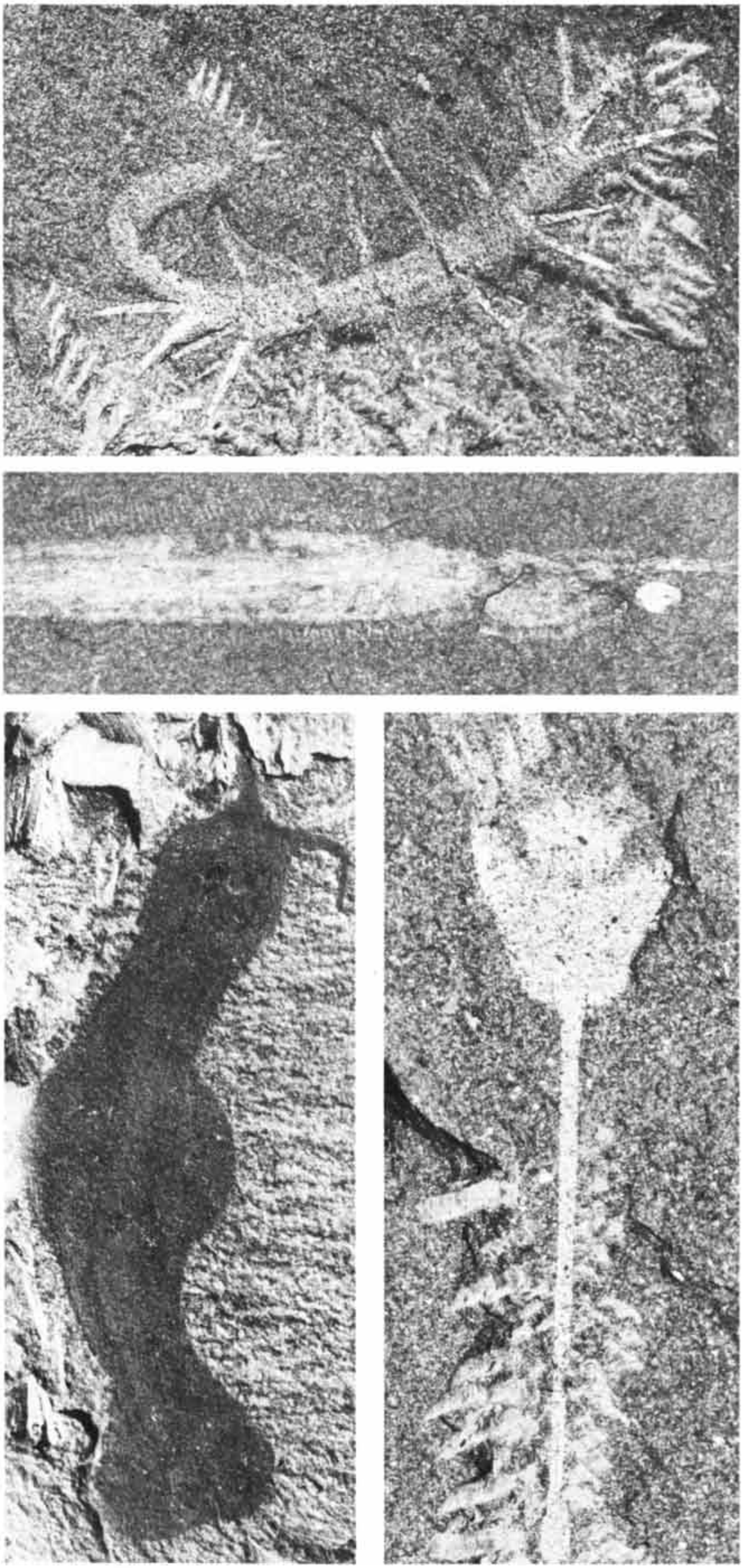


sibly ancestral to some modern parasites; *Ottoia* (*7*), another priapulid, seen at the center feeding on the mollusk *Hyolithes* (*6*) and at the right burrowing; *Selkirkia* (*8*), a third priapulid, seen here in a burrow front end down, and *Louisella* (*9*), a fourth priapulid that inhabited a double-ended burrow and undulated its body to drive oxygenated water over its gills. *Peytoia* (*10*) is a free-swimming coelenterate shaped like a pineapple ring. The sea-floor-dwelling mollusks in addition to *Hyolithes* are *Scenella* (*23*), its soft parts hidden under "Chinese hat" shells, and *Wiwaxia* (*24*), with its covering scales and defensive spines, seen here plowing a trail through the silt. Among the many arthropod genera of the sea floor are *Yohoia* (*3*), with its distinctive grasping appendages; *Naraoia* (*13*), an atypical trilobite that retained some larval characteristics; *Burgessia* (*14*), with its long tail spine; *Marrella* (*15*), which may have swum just above the sea floor; *Canadaspis* (*20*), an early crustacean, and *Aysheaia* (*26*), a stubby-legged animal suggestive of the living land dweller *Peripatus*. Other representatives of new phyla seen in addition to *Dinomischus* are *Hallucigenia* (*18*), one preparing to feed on a dead worm and two others approaching it, and *Opabinia* (*19*), seen here grasping a small worm with its single bifurcated appendage. Finally, seen swimming alone at the top left, is *Pikaia* (*28*), the sole representative of the chordate phylum in this Middle Cambrian fauna. *Pikaia* probably used its zigzag array of muscles to propel itself above the sea floor. The phylum of chordates includes the subphylum of vertebrates, which evolved later.

been preserved in detail. *Olenoides* had a pair of slender antennae in front and a pair of cerci, or antennalike structures, in back. The limbs along the length of the animal, up to 16 of them, were all similar in construction. The coxa, a large unit closest to the body, carried a battery of ferocious-looking spines. Attached to the coxa were two appendages; one was a filamentous gill and the other was a walking leg. *Olenoides* could seize and shred soft food, such as small worms, and pass the fragments along to its mouth. The forward antennae and the rear cerci no doubt supplied the animal with information about both food and potential predators. The fact that the primitive limbs of this trilobite are all similar is in marked contrast to the arrangement in many fossil and living arthropods whose limbs are variously modified and specialized.

About 40 percent of the Burgess Shale fauna consists of arthropods. Both in the number of species and the number of individual specimens the soft-bodied representatives of the phylum outrank the hard-shelled trilobites. Many of these "nontrilobites" have had their appendages preserved in remarkable detail; some of them must have been effective predators and scavengers. The most abundant is *Marrella*, an arthropod with a wedge-shaped head that bore two pairs of long hornlike spines curving to the rear. *Marrella* sensed the sea-floor environment with a pair of antennae and swept food particles toward its mouth with adjacent feathery appendages. Its score or more of side limbs were jointed, and a filamentous gill branched from each limb.

The next most abundant of the nontrilobites is *Canadaspis*. All but the hind part of its body was concealed by a double shell. Careful removal of this covering reveals the underlying appendages, which are remarkably like those of certain living crustaceans. Still another arthropod was one of the larger predators of the sea floor. This is *Sidneyia*, an animal whose distinctive limbs recall those of the living horseshoe crab, *Limulus*. It has been possible to identify some of the gut contents of *Sidneyia* as being fragments of brachiopod shells, evidence



UNIQUE ANIMALS of the Burgess fauna include the four seen in the photographs at the left. At the top is *Hallucigenia*, shown in the illustration on the preceding two pages. Second from the top is *Nectocaris*, a streamlined animal with conspicuous fin rays. At the bottom left is *Amiskwia*, a gelatinous worm with prominent fins. At the bottom right is *Dinomischus*, the stalked animal also shown reconstructed on the preceding two pages. Each of these species and six or more others represent hitherto unknown phyla of invertebrates.

that this arthropod was able to crush hard-bodied prey.

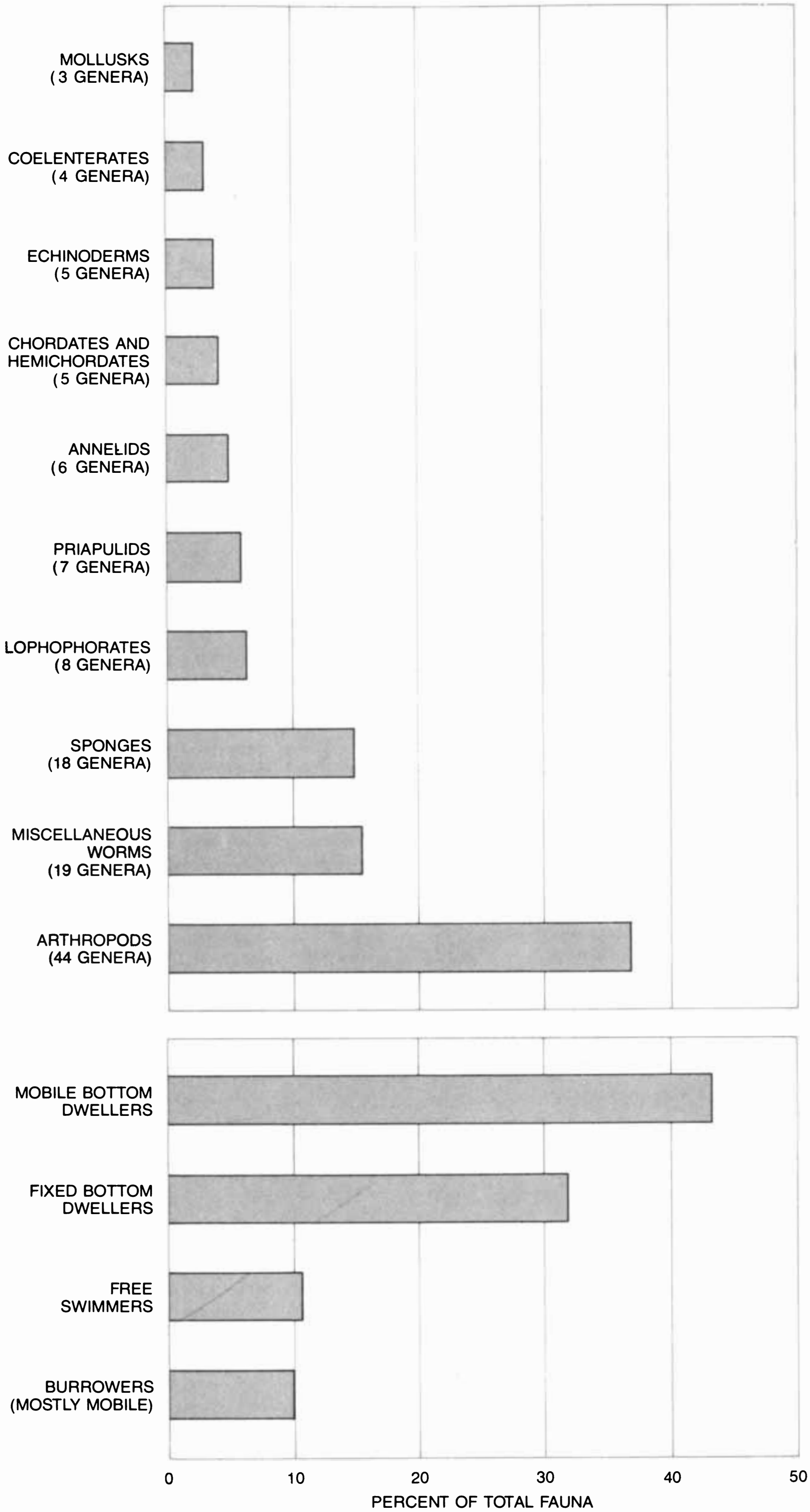
Sidneyia was not the largest of the Burgess Shale arthropods. The shale contains the impressions of isolated large limbs; if they are in proportion, they may have belonged to an animal as much as a meter long. The name *Anomalocaris* has been assigned to these fossils, which possibly represent one of the largest of all Cambrian invertebrates.

The most interesting of all the Burgess arthropods is *Aysheaia,* an animal with a pudgy body and stubby limbs. When Walcott first published a photograph of this fossil half a century ago, a number of zoologists wrote to him to point out how much this Middle Cambrian invertebrate resembled *Peripatus,* an animal with eight genera of relatives (comprising two families) in the small class of onychophores within the arthropod phylum. *Peripatus* was a land animal and *Aysheaia* was a marine form; nevertheless, *Aysheaia* surely represents the kind of ancestor that could have given rise to such living arthropods as myriapods and insects.

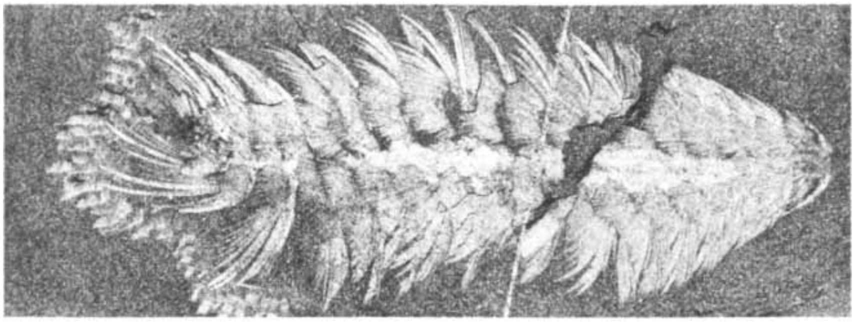
Although some of the nontrilobite arthropods in the Burgess shale, such as *Aysheaia,* are reminiscent of later forms, most of them cannot be placed in any recognized group. They have no obvious relatives either among the other Burgess Shale species or among the arthropods of later times. Because they exhibit a surprisingly wide array of anatomical features, indicating a high degree of specialization, they are evidence of a hitherto unsuspected adaptive radiation of arthropods in Cambrian times. It appears that the numerous stocks that arose during this period of rapid evolution were mostly unsuccessful. It is interesting to note that the animals that were to become dominant in later geological history generally have only a minor position in the various Cambrian faunas; a hypothetical observer would have been hard-pressed to predict just which groups had the flexibility necessary for long-term biological success.

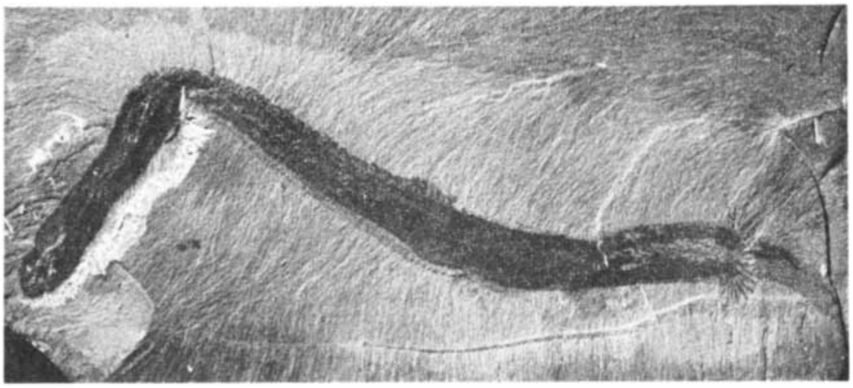
Of the Burgess Shale animals other than arthropods the representatives of six phyla are particularly noteworthy. Among the echinoderms the class of holothurians, the group that includes the sea cucumbers, was once thought to be widely represented. Now only one animal, *Eldonia,* is so classified. Unlike the great majority of the species in its class, *Eldonia* had a jellyfishlike body and a pair of oral tentacles. These animals probably swam through the water in shoals, using their tentacles to capture food. Another Burgess Shale echinoderm, the sea lily *Echmatocrinus,* is the earliest crinoid in the fossil record; as might be expected, it shows a number of primitive features.

The species of the coelenterate phy-



BURGESS SHALE GENERA currently number 119. The percent of the total assigned to various phyla is indicated in the upper part of this bar chart. Nearly 40 percent of the total are arthropods; only 14 of the 44 arthropod genera are trilobites. Worms other than priapulids and annelids (19 genera) and sponges (18 genera) make up another 30 percent of the total; mollusks are the most poorly represented. In terms of habitat, as the lower set of bars indicates, more than 40 percent of the Burgess Shale animals wandered the sea floor and more than 30 percent were rooted in the silt. Most of the burrowing animals also moved freely, although some remained fixed. Burrowers were slightly outnumbered by animals that swam above the sea floor.





TWO PHYLA OF WORMS in the Burgess Shale fauna are the familiar annelids and the less common priapulids. A typical Burgess Shale annelid is *Canadia,* a polychaete worm (*top*); its setae, bundles of fine bristles that were organs of locomotion, are preserved in detail. A typical priapulid worm (*bottom*) is *Louisella,* also reconstructed in illustration on pages 126 and 127.

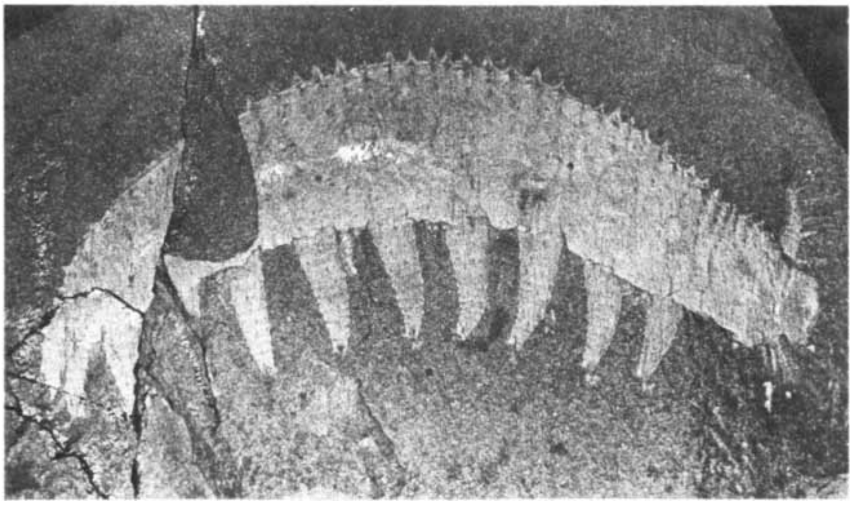
lum are among the most primitive of the metazoans. For example, in the late Precambrian fossil assemblage from the Ediacara Hills the coelenterates predominate. In contrast, the several Burgess Shale coelenterates, some resembling jellyfishes and others resembling sea pens, seem to have played a rather limited role in the community. On the other hand, the Burgess Shale sponges, the most primitive of all the animals present, were prominent members of the community. They were abundant and varied in form; some species grew on the sea floor in thickets.

The various Burgess Shale "worms" were mainly assigned by Walcott to the annelid phylum in general and to the class of polychaetes in particular. It is now realized that many of them belong to other phyla. Nevertheless, it is among the polychaete worms that some of the most spectacular examples of soft-body preservation are to be found: the setae, or bundles of fine bristles, that were these animals' organs of locomotion have been particularly well preserved as bright, reflective films in the shale. One of the polychaetes, *Canadia,* apparently did not live in a burrow but spent much of its time swimming close to the sea floor. Another, *Burgessochaeta,* was probably a more typical burrower, taking refuge in the muddy bottom and searching for food around the burrow entrance with its long tentacles.

Today the priapulid phylum is of interest only to a handful of specialists. These worms, however, were an important group in Cambrian times, and two priapulids present in the Burgess Shale are particularly noteworthy. One of them, *Ottoia,* is the most abundant of the group. It has been preserved in such detail that muscles are clearly visible and the gut content of some specimens can be analyzed. *Ottoia* fed on two kinds of shellfishes: brachiopods and hyolithids. The hyolithids, possibly members of the mollusk phylum, had a conical shell that was capped by a protective operculum, or lid, when the animal was fully withdrawn. The teeth of *Ottoia* were not strong enough to break open the shell, and so the hyolithids were swallowed whole and their soft parts were digested as the shells passed through the priapulid's gut unscathed. These shellfishes were not *Ottoia's* only food. A unique specimen contains within its gut the remnants of another worm of the same species, showing that (as with some living priapulid worms) *Ottoia* could be cannibalistic.

Parasitologists take considerable interest in another Burgess Shale priapulid. It is *Ancalagon,* which may be ancestral to the living group of spiny-headed worms, the Acanthocephala, that seem to have been parasites for millions of years. These parasitic worms have no gut and absorb nourishment through their body wall while they are lodged in the intestine of their host. If evolution is hypothetically reversed and the worms are reendowed with the organs necessary for a free-living existence, the reconstructed animal is remarkably like *Ancalagon.*

Two other supposed worms, once considered to be polychaetes, are *Wiwaxia* and *Pikaia.* The body of *Wiwaxia* was covered with large scales. Long spines that curved upward and outward along the animal's back evidently were protection against predators. That the spines actually were protective is indicated by the fact that in some specimens of *Wiwaxia* they have been snapped off. An inhabitant of the sea floor, *Wiwaxia* nourished itself by scraping off fragments of food with a rasping organ. The rasp resembles the radula, or horny



ANCESTRAL ARTHROPOD with a striking resemblance to the living onychophore *Peripatus* is this remarkably preserved invertebrate, *Aysheaia.* Cambrian arthropods such as *Aysheaia* could have been ancestral to such living members of that phylum as the myriapods and insects.

toothed tongue, of certain living mollusks. Is *Wiwaxia* a primitive mollusk? If it is, the details of its remarkably preserved anatomy will throw new light on the early evolution of this highly successful phylum of invertebrates.

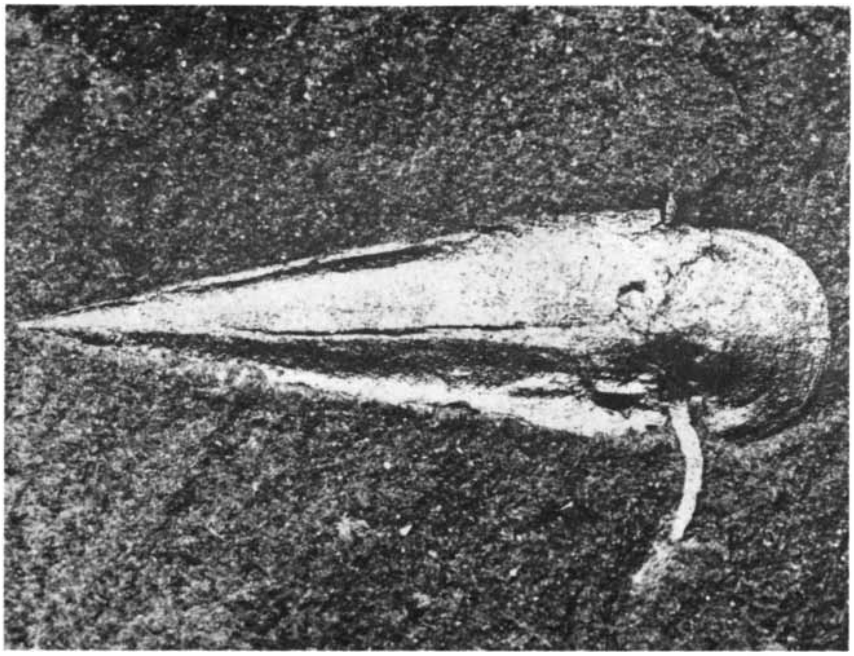
What about *Pikaia*, formerly considered a polychaete worm? Some 30 well-preserved specimens show a prominent rod along the animal's back that appears to be a notochord, the cartilagelike stiffening organ that gives the chordate phylum its name. In addition to this key anatomical feature the blocks of muscle in *Pikaia* form a zigzag pattern that is comparable to the musculature of the primitive living chordate *Amphioxus* and of fishes. Although *Pikaia* differs from *Amphioxus* in several important respects, the conclusion that it is not a worm but a chordate appears inescapable. The superb preservation of this Middle Cambrian organism makes it a landmark in the history of the phylum to which all vertebrates, including man, belong. There are possible instances of even earlier chordates from Lower Cambrian formations in California and Vermont but none is as rich in detail.

Perhaps the most intriguing problem presented by the Burgess Shale fauna is the 10 or more invertebrate genera that so far have defied all efforts to link them with known phyla. They appear to be the only known representatives of phyla whose existence had not even been suspected. Their origins must lie in Precambrian obscurity, where the initial metazoan diversification began. The peculiarity of these novel animals is exemplified by the aptly named *Hallucigenia*.

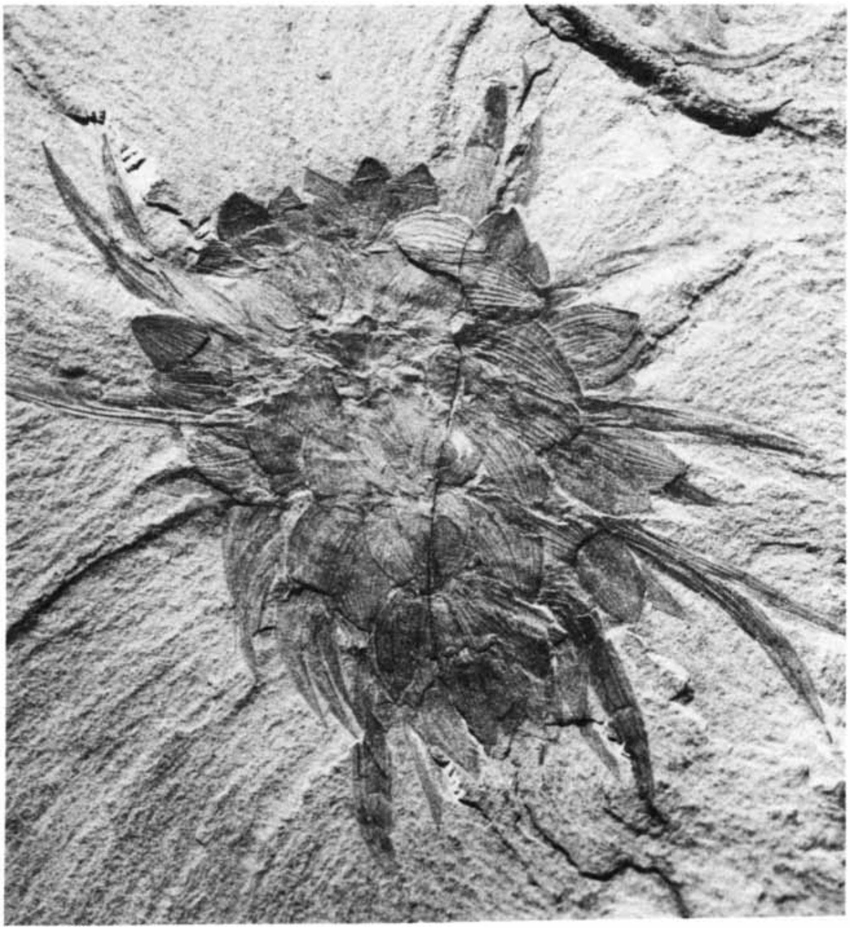
This animal propelled itself across the sea floor by means of seven pairs of sharply pointed stiltlike spines. Seven tentacles arose from the upper surface of the animal's body; at the end of each tentacle was a pair of strengthened tips. Did the tentacles gather food? If they did, did each tentacle act as an individual mouth with a direct connection to the animal's alimentary canal? There are more questions than answers, but a valuable clue to the animal's behavior is preserved in a specimen from a Harvard University collection. There one can see more than 15 individual *Hallucigenia* associated with a large worm. There seems little doubt that, having detected the carcass of the worm, these odd animals had congregated to scavenge it.

Compared with *Hallucigenia* a second unique animal, *Opabinia*, seems almost orthodox. Its five eyes were arranged across its head, so that it was probably able to avoid predators with ease as it swam close to the sea floor, steering itself with a vertical tail fin. *Opabinia* fed by capturing prey with a grasping organ that projected forward.

Alternative approaches to problems



MOLLUSK REPRESENTATIVE, ***Hyolithes,*** **had a cone-shaped shell that was capped by a protective lid. One of the burrowing worms,** ***Ottoia,*** **preyed on these mollusks but was not able to break the shell open. The worm digested** ***Hyolithes'*** **soft parts and excreted its shell.**

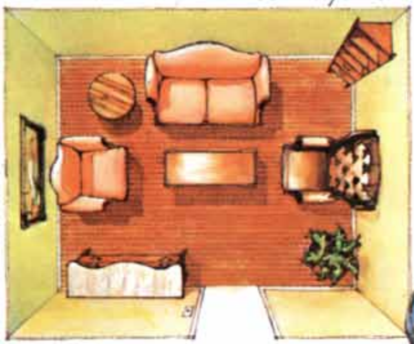


PROBABLE MOLLUSK, ***Wiwaxia,*** **with its cover of large scales and array of long protective spines, was first placed among the polychaete worms of the Burgess Shale. Its rasplike feeding organ, similar to a mollusk's radula, suggests that it belongs to the molluscan phylum instead.**

Who took The Crown Jewel of England? Solve the mystery– you could win $25,000!

The sleuth who finds the truth may win a $25,000 first prize, $5,000 second prize, or one of five $1,000 runner-up prizes.

The scene is the drawing room of a 17th century manor house in the heart of the Kent countryside.



The Drawing Room

The Butler

There are five people in the manor house.

Although no one knows it, The Crown Jewel of England is about to be taken.

The Contessa is whispering to the Squire.

The man in the heavy boots sits holding his favorite drink -a Beefeater Gin and Tonic.

The person seated opposite the Brigadier is enjoying a Beefeater Gimlet.

The Butler enters with a Beefeater Gibson for the person seated to the right of Lady Trumbull.

The Brigadier mumbles to himself to buy a bottle of Beefeater Gin - The Crown Jewel of England - on the way home.

Suddenly, the lights go out.

Lady Trumbull

The Contessa

The Gibson Girl swoons into the waiting arms of the Martini Man.

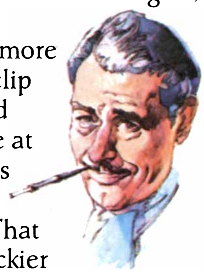
Lady Trumbull faints.

No one could hear any footsteps.

It should be easy to deduce who turned out the lights, dear reader.

But now, more importantly, clip out the coupon and tell us (for a chance at $25,000) who takes The Crown Jewel of England. That is a far trickier problem, and no one, not even you, is above suspicion. Good luck and good hunting!

The Brigadier



The Squire

Who took The Crown Jewel of England?

OFFICIAL RULES: (No purchase required.) **1.** Clip out this entire coupon. On the coupon hand print your name, address and tell us who you think took The Crown Jewel of England. Mail in a hand addressed envelope no larger than 4⅛" x 9½" (#10 envelope) to Beefeater Gin "The Crown Jewel of England" Sweepstakes, P.O. Box 9241, Blair, Nebraska 68009. **2.** IMPORTANT: In order to be eligible for a prize, you must hand print on the lower left hand corner of your mailing envelope who you think took The Crown Jewel of England. **3.** Entries must be received by August 31, 1979. Enter as often as you wish, but each entry must be mailed separately. **4.** Winners will be determined in a random drawing from among all correct entries received. If fewer than seven (7) correct entries are received, random drawings will be held from among all entries received to award prizes not awarded via correct entry. Random drawings are under the supervision of the D.L. Blair Corporation, an independent judging organization whose decisions are final. **5.** This sweepstakes is open to residents of the United States of legal drinking age in their state of residence at time of entry. Employees of Kobrand Corp., its distributors, retailers, advertising and promotion agencies and their families are not eligible. This sweepstakes is void in Ohio, Utah and Texas, and wherever prohibited by law. Limit one prize per family. No substitution of prizes is permitted. All Federal, State and local laws and regulations apply. For a list of prize winners, send a SEPARATE self-addressed, stamped envelope to: Beefeater Gin "The Crown Jewel of England" Winners' List, P.O. Box 6279, Blair, Nebraska 68009. **6.** All entries must be on Official Entry Coupons. For each additional coupon you wish send a SEPARATE self-addressed, stamped envelope to: Beefeater Gin "The Crown Jewel of England" Coupon Request, P.O. Box 7104, Blair, Nebraska 68009. Your request must be received by July 31, 1979.

Clip out this coupon and tell us who you think took The Crown Jewel of England. Write your answer below and also on the lower left hand corner of your mailing envelope. Please hand print. (NO PURCHASE REQUIRED)

Mail to: BEEFEATER GIN "THE CROWN JEWEL OF ENGLAND"
P.O. BOX 9241, BLAIR, NEBRASKA 68009

Who took The Crown Jewel of England? ______________

Name ______________

Address ______________

City ______ State ______ Zip ______

BEEFEATER® GIN. The Crown Jewel of England.

SA

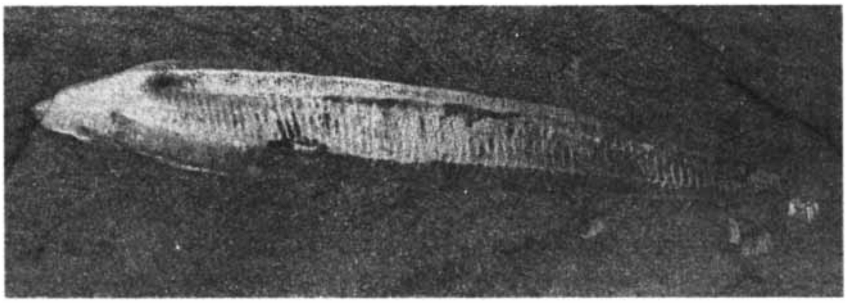
of functional design are evident among these unusual invertebrates. For example, for a worm with a fluid-filled body cavity one problem is that muscular contraction in one part of the body will distort the shape of the rest of the body. In annelid worms the problem has been solved by dividing the body cavity into a series of watertight compartments. *Banffia*, a unique Burgess Shale worm, developed an alternative solution. The stiffened front half of its body was separated from the more saclike back half by a prominent constriction at the midpoint. The constriction appears to have damped the hydrostatic fluctuations set up by the locomotor muscles of the animal's front half, thereby minimizing the distortion of its unstiffened back half.

Some representatives of new phyla have been preserved by the dozen. Others, particularly the free-swimming inhabitants of the higher water levels that would seldom be trapped by slumping mud, are quite rare. One such animal is the worm *Amiskwia;* judging by its prominent fins, it was probably quite an active swimmer. Another animal, *Nectocaris*, a fast-swimming predator, had enormous eyes and evidently propelled its streamlined body by rapid lateral flicks of its body. Prominent dorsal and ventral fins, stiffened by numerous fin rays, helped to keep the animal stable as it was swimming.

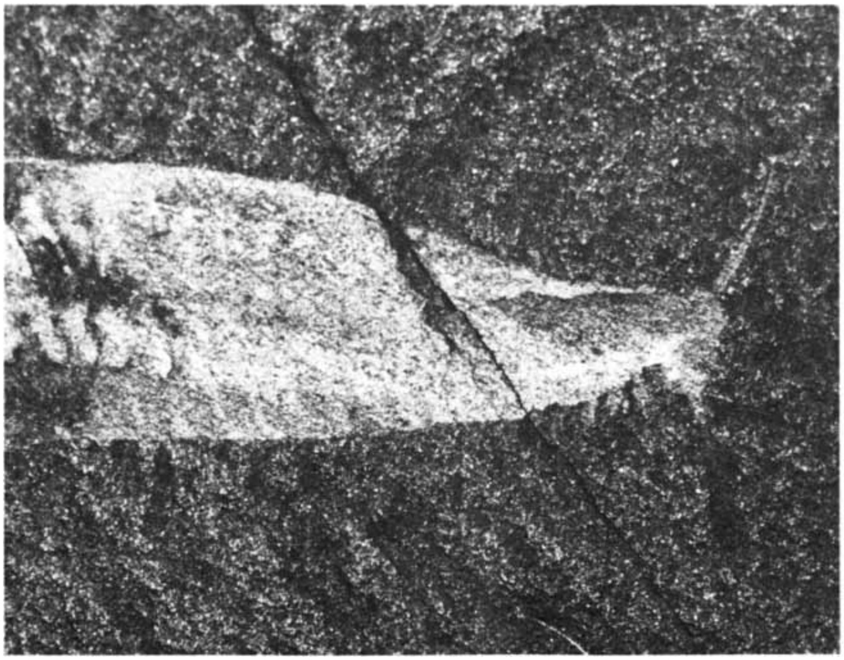
Conodonts, or "cone teeth," are enigmatic fossils that resemble tiny teeth; they are found in formations ranging in age from the latest Precambrian to the Triassic, a span of almost 400 million years. Although they look like teeth, they cannot have acted as such because they show no signs of wear. What soft-bodied animal had conodonts and for what purpose has long been an unanswered question. Another rare pelagic invertebrate preserved in the Burgess Shale, *Odontogriphus*, may be that animal. The tentacular feeding apparatus of the animal, another unique representative of a hitherto unknown phylum, incorporates a set of minute conical objects that appear to be conodonts. Since conodonts cannot have acted as teeth, the hypothesis has been advanced that they were some kind of support for the feeding tentacles. Was the feeding apparatus of *Odontogriphus* and of animals like it the source of the conodonts so copiously distributed throughout the Paleozoic and the earliest Mesozoic fossil record? Possibly so.

As more is learned about the Burgess Shale fauna the picture of Cambrian life will gain a new perspective, particularly with respect to the explosive evolution of the metazoans. For example, the wide range of arthropods, with their distinctive and different groupings of anatomical characteristics, is already such that a single phylum seems too small to hold them all. The adaptive radiation of the Cambrian invertebrates can be seen as the initial response to the availability of a very wide variety of marine ecological niches. Hence many Cambrian animals seem to be pioneering experiments by various metazoan groups, destined to be supplanted in due course by organisms that are better adapted. The trend after the Cambrian radiation appears to be the success and the enrichment in the numbers of species of a relatively few groups at the expense of the extinction of many other groups.

An additional possibility is suggested by the Burgess Shale fauna itself. Some groups of major stature in Cambrian times, such as the priapulid worms, may have fared badly against later competitors and only escaped extinction by migrating into marginal niches that were either unattractive or unavailable to other metazoans. One such manifestation of movement into a marginal niche is the scaling down of body size. This miniaturization may well be how some priapulids managed to survive. An alternative escape route is to become parasitic; the priapulids that appear to have given rise to the parasitic spiny-headed worms could be an example of the alternative. In any event the Burgess Shale fauna affords both a marvelous glimpse of evolution in action during this brief interval of Middle Cambrian times and a stern reminder of how impoverished and distorted the fossil record is. The study of these soft-bodied animals illuminates many hitherto unsuspected aspects of the history of life.



STIFFENING ROD, or notochord, runs partway along the back of the early chordate *Pikaia*. The animal's head, seen in more detail in the illustration below, is at the right. The pattern of its musculature resembles that of fishes and of the living primitive chordate *Amphioxus*. A reconstruction of this free-swimming chordate appears in the illustration on pages 126 and 127.



FRONT END of *Pikaia* is seen enlarged in this photograph, making visible the animal's pair of sensory tentacles and behind them a short row of small appendages. Earlier Cambrian formations preserve the remains of possible chordates, but none compare with *Pikaia* in detail.

Save up to 15% on these 4 outstanding titles

Animal Physiology

Roger O. Eckert, University of California, Los Angeles
with chapters by
David Randal, University of British Columbia

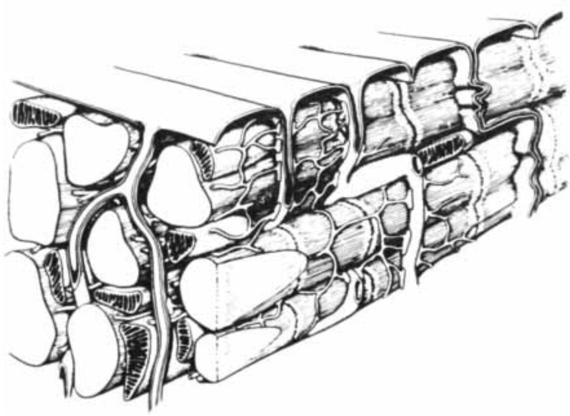
Animal Physiology clearly illustrates the major physiological concepts that will provide students with an in-depth understanding of living systems. The text avoids encyclopedic detail, emphasizing instead the physical and chemical principles and mechanisms that living systems use to meet their functional needs. *Animal Physiology* is presented in a broadly based biological context, hence the examples were selected from a wide spectrum of animal life.

Each chapter concludes with a summary, a set of exercises, suggested readings, and references. The text is beautifully illustrated in two colors, contains an extensive glossary, and incorporates related material in boxes throughout the text. It is designed for use in general, animal, and comparative physiology courses.

"Outstanding in-depth coverage, illustrations, and up-to-date information. I liked the book . . ."

—S. B. Chaplin, University of Missouri

558 pages, 588 illustrations.
Regularly $23.00. **Special Price: $19.55**



Fractals

Form, Chance, and Dimension

Benoit B. Mandelbrot, IBM Fellow,
Thomas J. Watson Research Center

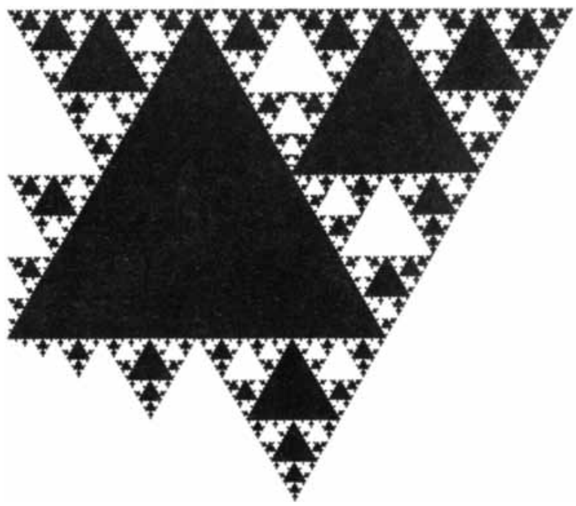
In nature fractals are represented by highly irregular shapes, like the coastline of Britain or the pattern of Brownian motion. Mandelbrot invented the term fractals and is a leader in applying the mathematics used to describe Brownian motion to other problems in science. Irregular structures which can be described by this approach include hierarchial stellar clustering, river networks, turbulence, and soap films. Each has the property that when suitably magnified, the resulting pattern is similar to the original. Mandelbrot discusses various ways of calculating the measure and dimension of these objects. He draws on classical works in real analysis and probability, as well as on more speculative insights of the philologist George Zipf.

Because of its wide range of applications, this book is exciting not only to mathematicians, but also to scientists in any field. The book's striking illustrations and unique ideas make it delightful reading for anyone who enjoys viewing the world with a fresh perspective.

"Mandelbrot's small volume is a good bet to become a classic of its deep yet elementary genre."

—Phillip Morrison, SCIENTIFIC AMERICAN

365 pages, 68 illustrations.
Regularly $18.50. **Special Price: $15.72.**



The Visual Perception of Motion in Depth

It now appears that information concerning motion in depth is processed in two distinct channels, which converge on a single motion-in-depth stage of the human visual perception system

by David Regan, Kenneth Beverley and Max Cynader

A person negotiating a busy city street performs an impressive feat of visual judgment. The total visual information available to the human brain is contained in the pair of two-dimensional images formed when the lenses of the eyes focus incoming light on the retinas. On the basis of these rather blurred two-dimensional images the pedestrian must judge the position and direction of oncoming cars and people as well as of his own body with respect to the continuously changing three-dimensional world around him. How does the human visual system process the information in those images, making possible precise eye-to-limb coordination of the kind that is needed in such a task as maneuvering through traffic?

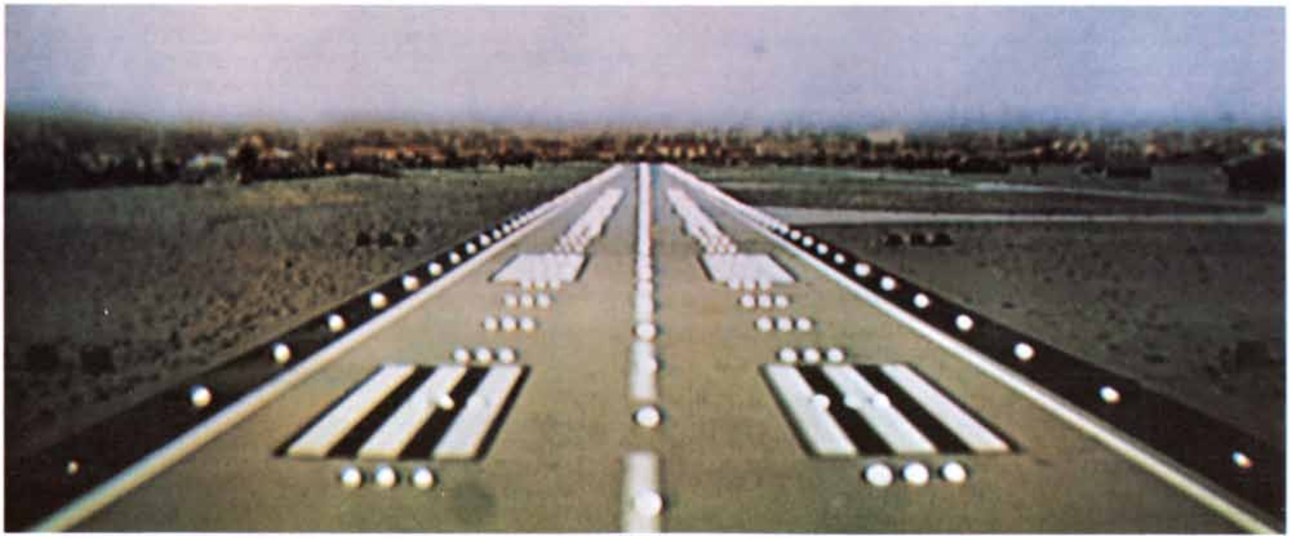
In spite of the impressive advances that have been made in the study of visual perception over the past 15 years there is still only a fragmentary understanding of how eye-to-limb or eye-to-hand coordination in three-dimensional space is achieved when either the observer or the objects around him are in motion. Perhaps most perplexing is the question of how motion in depth, which obviously cannot be incorporated in a two-dimensional retinal image, is perceived. There are, however, cues to motion in depth present in such an image. For example, an object that is moving away from the observer appears to get progressively smaller. J. J. Gibson of Cornell University has pointed out that it is these "stimulus correlates" of motion that serve as the effective stimuli to the neural organizations in the brain that mediate motion perception. How, then, are the stimulus correlates of motion in depth perceived and interpreted by the human visual system?

One way to approach the question is to consider not the everyday demands placed on the visual system but rather some of the most challenging tasks presented to it: tests of eye-to-hand coordination in which extremely precise judgments of motion must be made at top speed and on the basis of few cues. For example, in the game of cricket the batsman must hit a hard ball that is bounced directly at him with a speed of up to 100 miles per hour. Harriet G. Williams of the University of Wisconsin has shown that in baseball a skilled batter can make more reliable estimates of the ball's trajectory if he can see the pitcher's legs and feet. The cricket batsman undoubtedly employs similar cues to make a swift decision on how to place his feet in preparation for hitting the ball. In cricket, however, there is an added difficulty. The ball is made with a raised, stitched seam that can cause it to change direction when it bounces. Therefore any final adjustments in position the batsman makes must be based on visual judgments made after the ball has left the ground. Assuming that the ball travels at 90 miles per hour and bounces 30 feet from the batsman, he has only about .23 second in which to make those judgments and carry out whatever adjustments are called for. A similarly demanding visual task is the everyday business of an airline pilot, because the final stages of a routine landing are carried out under visual guidance. In order to make an exact judgment of his motion with respect to the landing field the pilot depends on the flow pattern he sees as the plane approaches the ground, that is, the manner in which objects in his visual field appear to flow radially away from the landing point.

During the past 10 years we have carried out a series of psychophysical experiments to determine how an individual's visual system processes cues to motion in depth as he engages in these challenging tasks of eye-to-hand coordination. Using modern electronic techniques for generating visual stimuli, we have studied these cues to motion in depth in combination with (and sometimes in competition against) a variety of related stimuli. As a result we have been able to construct a theoretical model of the operations of the visual system that underlie the perception of motion in depth. Before describing these experiments and the conclusions we have drawn from them, however, it is necessary to briefly discuss a current, widely accepted view of how the visual system responds to a stimulus.

The visual pathway of the brain breaks a retinal image down into a number of abstract features that are processed separately by parallel channels. Although this concept is not new, it did not elicit much interest until a possible physiological basis for sensory channels was revealed by modern techniques of neurophysiological testing (in particular the use of microelectrodes to record the activities of single neurons, or nerve cells, in the visual cortex of anesthetized animals, and more recently the "2-deoxyglucose" method developed by Louis Sokoloff and his colleagues at the National Institute of Mental Health to identify the neurons activated by a particular stimulus). The neurophysiological experiments disclosed the existence of classes of neurons that respond best to specific stimulus features.

It was the discovery of these putative information filters that gave rise to the present understanding of how "sideways" motion, or movement across the retina, is perceived. As early as 1911 the psychologist A. Wohlgemuth hypothesized that the visual system incorporated mechanisms sensitive to motion in a particular direction. Then in 1962 David H. Hubel and Torsten N. Wiesel of the Harvard Medical School established that in the visual cortex of cats there are classes of directionally specific neurons: neurons that fire at a high frequency when a properly oriented stimulus moves across the visual field in one direction but fire at a much reduced rate when the stimulus moves in the opposite direction. Classes of such directionally specific neurons have been identified in several species of animals, and it is now



LANDING AN AIRPLANE is an exacting test of eye-to-hand coordination, requiring that a pilot make swift, precise judgments of his own motion in depth with respect to a stationary landing field. As is shown in this sequence of photographs of the screen of a flight simulator, one of the few cues to direction of motion available to the pilot is the flow pattern he sees through his cockpit window: the contours of the visual field appear to flow radially away from a focus that corresponds to the touchdown point, so that as objects move toward the edges of the field they seem to move faster and to get larger and their texture seems to coarsen. Motion in depth cannot be reproduced in the two-dimensional retinal images from which the human brain derives all visual information. Hence it is cues such as flow patterns that serve as the effective stimuli for the neural mechanisms that mediate the perception of such motion (*see bottom illustration on page 143*).

The Pentax MX.

From the people who've made more major advances in 35mm SLR's than any other company.

At Pentax, major breakthroughs are a matter of routine. In fact, the actual look of today's 35mm SLR is the direct result of Pentax photo-engineering.

Which makes the Pentax MX part of a long tradition of innovation.

We were first to mass-produce 35mm SLRs combining the pentaprism and instant-return mirror. We were first to use TTL (through-the-lens) light metering. We introduced the first Super-Multi-Coating for 35mm lenses. And we were first to offer automatic electronic exposure in a 35mm SLR.

So it's no coincidence that Pentax is first in total sales of 35mm SLRs worldwide.

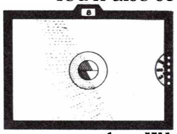
And it's no surprise, that in the Pentax MX, we've built a totally reliable system camera with features and capabilities far beyond its price range.

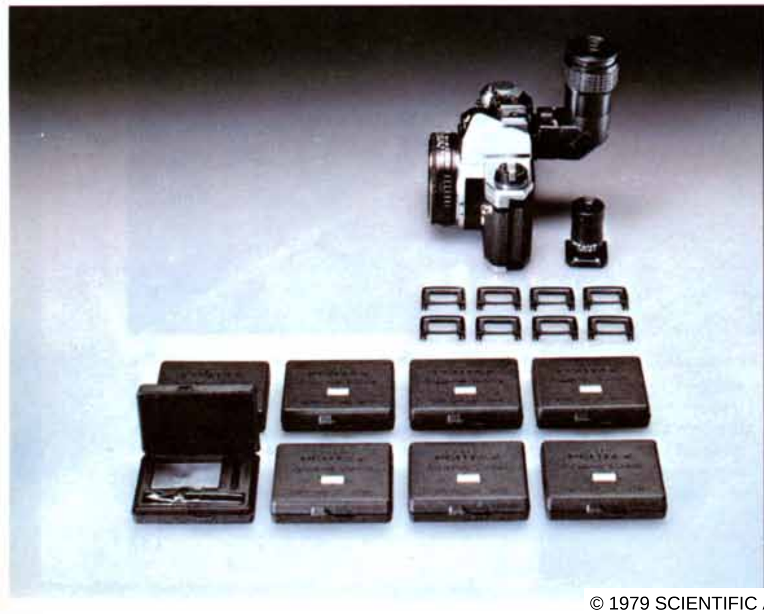
The MX was built for pros by pros to be the smallest and lightest total-system 35mm SLR in the world. It's so light, you can add either the 2-fps Winder MX or the 5-fps Motor Drive MX, and the whole system will still weigh less than other SLRs *without* a winder or a drive unit.

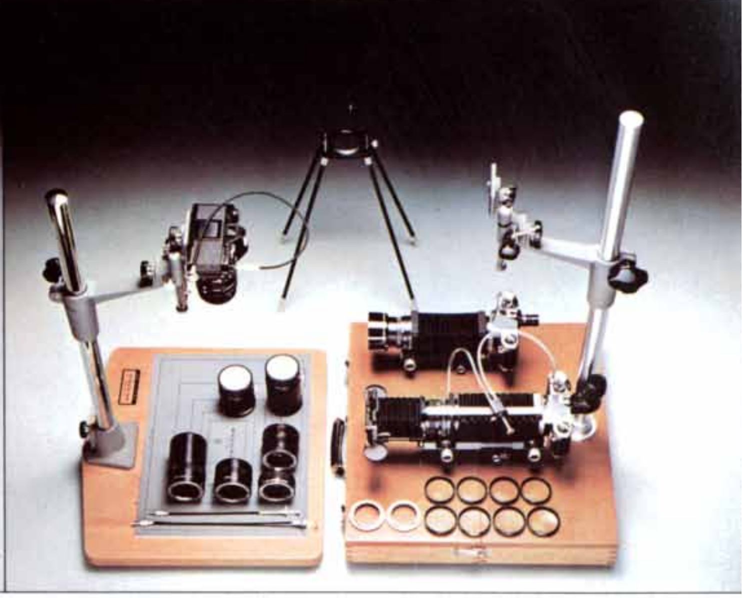
THE PANORAMIC VIEWFINDER

The image in the viewfinder is so big and bright it snaps right into focus regardless of lighting conditions. Inside the finder you'll see aperture setting, shutter speed, and the two adjacent shutter speeds.

You'll also see our incredible five-dot LED exposure readout. When the green dot in the center is lit, you have perfect exposure. When one of the two yellow dots is lit, you're a half stop under- or over-exposed. And when one of the two red lights is on, you're a full stop (or more) over or under. With all that information right in the viewfinder,



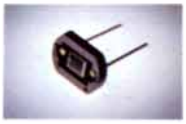




you can work easily without having to move the camera away from your eye. And the solid-state LEDs are rugged. You can't jar them out of alignment, as you might with many matched needle systems.

GPDs—A BETTER WAY TO METER LIGHT

The light sensors in the MX are a major technological breakthrough in themselves, developed by Pentax in cooperation with Nippon Electric. They're called Gallium Arsenide Phosphide Photo Diodes, or GPDs. They react 1,000 times faster than CdS cells in low light situations, and they respond instantly to any radical light change, no matter how rapid. And yet, unlike SPD cells, GPDs are virtually immune to infrared light. In the future, you'll probably see GPDs in most other SLRs. At Pentax, we use them right now.



COMPACT LENSES

A lot of SLRs have been made smaller lately, but unfortunately, their lenses haven't. Pentax decided to design 18 ultra-thin, ultra-light M-series lenses for use on the MX. Of course, our regular Pentax lenses also fit the MX. So there are over 40 interchangeable Pentax lenses available, from 1,000mm telephoto to 17mm fisheye. All have the Pentax bayonet mount, which is becoming the standard of the industry. And all have Super-Multi-Coating, a process we invented and perfected.

MORE INNOVATIONS

Loading the MX is a real treat. Our "Magic-Needle" takeup spool is probably the most efficient loading system you've ever used.

Also, we make nine interchangeable focusing screens which can be changed in seconds.

MOTOR DRIVE MX

When attached to your Pentax MX, the Motor Drive MX lets you fire off consecutive exposures at up to 5fps. And it's available with three different power sources.

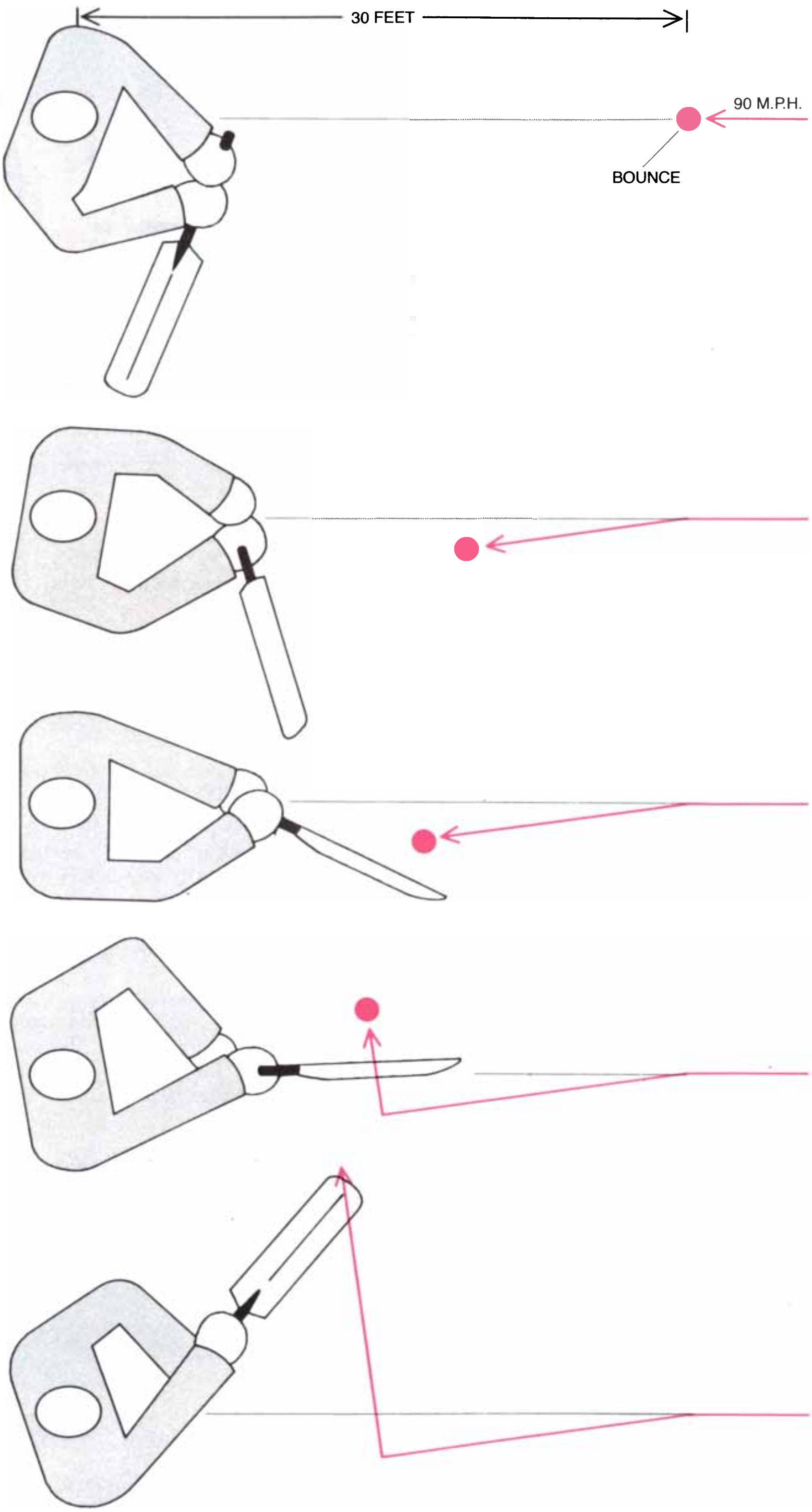
DATA BACKS AND 250 EXPOSURE BACK

The back of the MX accepts the Pentax Bulk Film Magazine and data back. If you have a lot to shoot, the Bulk Film Magazine holds 250 frames. With the Dial Data MX, your MX becomes a data camera. On each exposure, your choice of alpha/numeric data is imprinted automatically. And our Data MX camera comes complete with a built-in clock, so you can imprint the time, date and more on each frame.

All in all, the Pentax MX is a formidable camera, the heart of a superb, sophisticated, professional system. It's the result of years of meticulous research and exhaustive testing. We wouldn't have it any other way. And we don't think you would either.

For more information, write: Pentax MX, P.O. Box 2585, Littleton, CO 80161

PENTAX®



HITTING A RAPIDLY MOVING BALL in the game of cricket is another difficult test of eye-to-hand coordination that provides few cues to motion in depth. In cricket the batsman must determine how to hit a hard ball that is bounced toward him by the bowler. The bowler's body position and movement provide some cues to the ball's direction of motion, but when the ball bounces, it may change direction sharply. Assuming that the ball bounces 30 feet from the batsman and travels at 90 miles per hour, the batsman has only .23 second during which to make final judgment of its motion. During that time his principal cues to ball's direction of motion in depth are changing size and relative velocities of left and right retinal images of ball.

widely presumed they are also present in the human visual system, giving support to the concept of a visual channel that is preferentially responsive to sideways motion.

The reliability of the visual judgments of motion in depth made by cricket or baseball players led us to suppose that information about motion in depth is handled quite separately from information pertaining to sideways motion, that is, handled in a distinct channel of its own. The properties of this hypothetical channel, we realized, would have to explain how some ballplayers and pilots can perform reliably even after losing the vision of one eye. These individuals evidently obtain motion-in-depth information through one eye alone, although individuals with normal visual equipment undoubtedly obtain a great deal of motion-in-depth information from the interaction of the two eyes, that is, from comparison of the images seen by the left and right eyes.

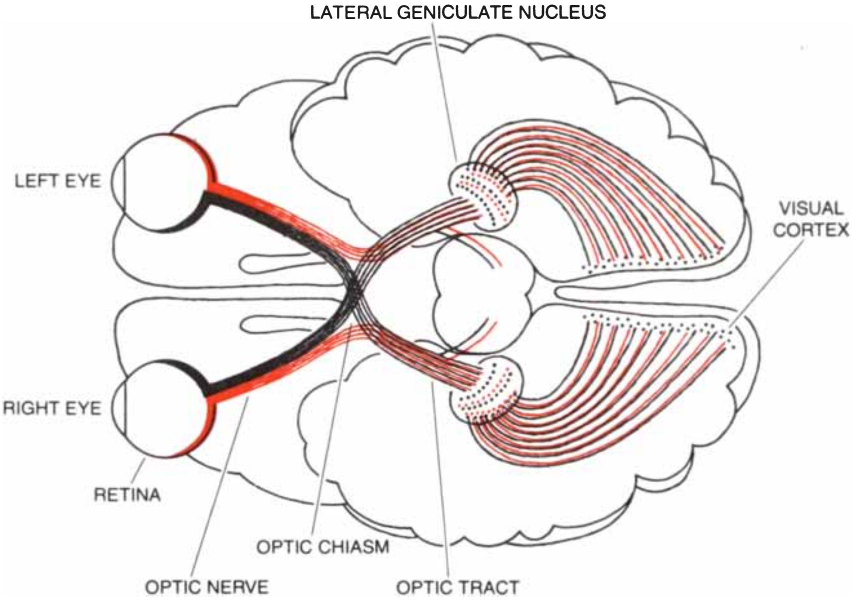
There appear to be two different, although not unrelated, mechanisms for the processing of monocular and binocular cues to motion in depth. Consider first the monocularly available stimulus correlates of motion. For example, changing the size of a retinal image can create a compelling illusion of motion in depth. (Indeed, this phenomenon is often exploited in space-adventure films to create the illusion that stars, planets and spacecraft are traveling rapidly in depth.) Moreover, changing image size is one of the important cues to motion in depth available to the cricket batsman. Therefore we began this part of our investigation by asking whether there was a channel in the human visual system that responded preferentially to changing-size stimuli.

In our first changing-size experiments we made use of the classic psychophysical procedure known as selective adaptation. In this procedure the visual sensitivity of subjects to a variety of stimuli is measured before and after the subjects are exposed to a strong level of a particular "adapting" stimulus. If the adapting stimulus fatigues a particular channel in the nervous system, after prolonged exposure the subjects will have more difficulty detecting a stimulus that is processed by this channel. Thus by testing a number of stimuli against each adapting stimulus it is possible in effect to dissect the visual system: to sort out the neural organizations involved in different perceptual tasks.

In our changing-size experiments subjects were made to adapt to a stimulus square generated on the screen of a cathode-ray tube. Each side of the square could be made to oscillate so that two distinct modes of stimulation could be presented to the subject: when the opposite sides of the square were made to move in opposite directions, the size of the square oscillated; when the opposite sides were made to move in the same direction, the position of the square oscillated. A subject's visual sensitivity to either stimulus was measured as the minimum amplitude of oscillation he could detect. (In addition the stimulus square could be generated either as a bright figure on a dark background or as a dark figure on a bright background, but that choice turned out to have little effect on the perception of changing size: adaptation to bright changing-size squares affected the subject's sensitivity to dark changing-size squares and vice versa.)

We began by measuring subjects' visual sensitivity to both changing image size and changing image position before they were adapted to any stimulus. Then we directed them to gaze at a square showing a large-amplitude oscillation in size. When the subjects' visual sensitivity was measured again after 20 minutes of exposure to the changing-size stimulus, we found that their ability to detect changing size was significantly reduced, although their ability to detect changing position was hardly affected. This result pointed to the existence of a distinct changing-size channel. It certainly did not reflect the subjects' adaptation to the lighting changes on the screen: when we had subjects adapt to a flickering light, we found that the adaptation had little effect on their ability to detect either changing size or changing position.

There was still the possibility that the reduction in subjects' sensitivity to changing-size stimuli was simply due to the fatigue of directionally selective motion filters. To test that explanation we adapted subjects for 20 minutes to a changing-position stimulus: a square whose position oscillated with the same amplitude and frequency as the changing-size adapting stimulus. This time we found only small reductions in the subjects' sensitivity to either changing size or changing position. It is important to understand that in this control experiment each side of the adapting square moved exactly as it had in the preceding, changing-size experiment. The changing-position and changing-size stimuli differed only in the relation between the movement of opposite sides. Thus we had discovered a channel that responded exclusively to the oppositely directed movement of opposite sides of a retinal image. Moreover, the selective response persisted even when the sides of the test square were rather far apart, separated by a visual angle of up to 1.5 degrees. (Here, as in many experiments in visual perception, the distance between two points is measured as the angle between the lines of sight from the observer to each of the points.)

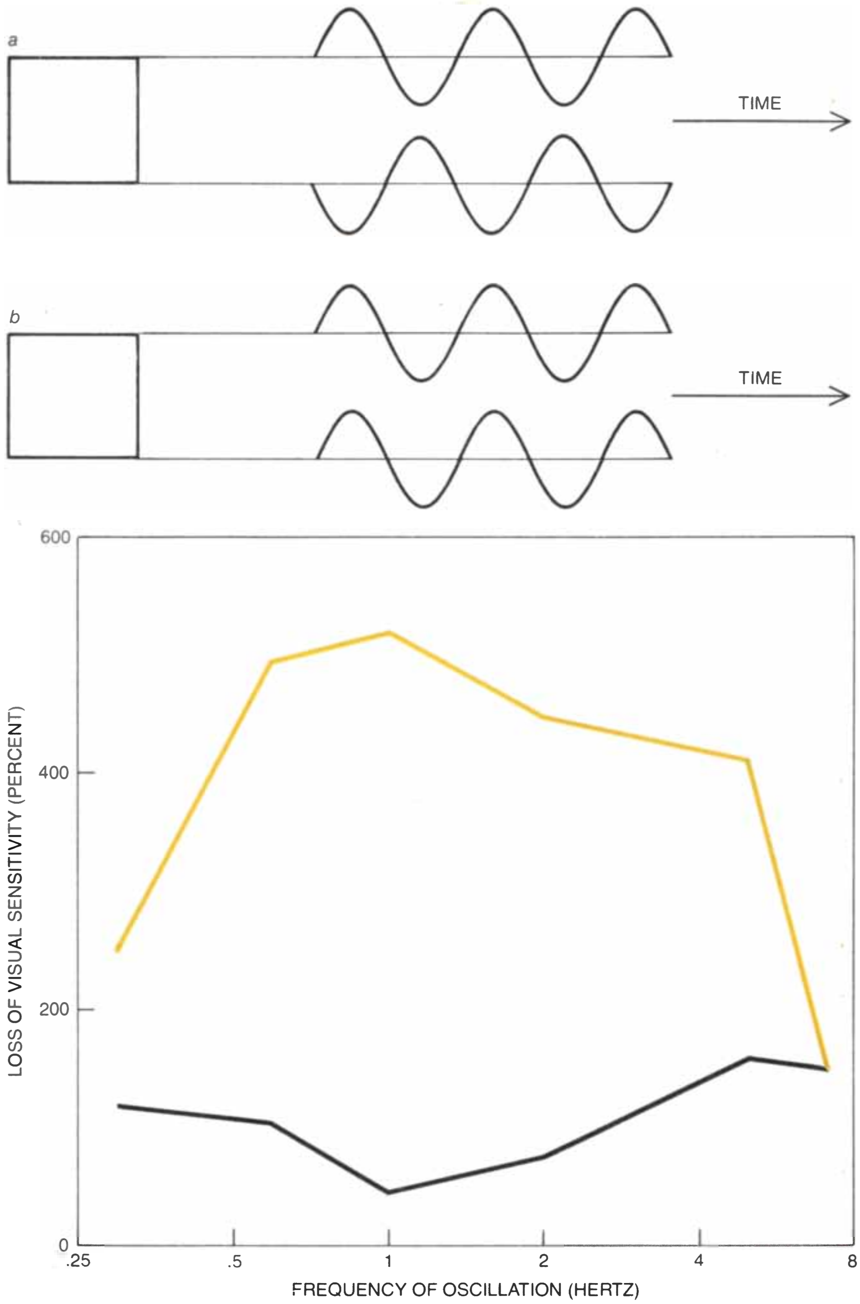


HUMAN VISUAL PATHWAY from the retina to the cerebral cortex consists essentially of six different layers of neurons, or nerve cells: three of them in the retina, one in the cluster of cells called the lateral geniculate body and two in the visual cortex. The fibers from the retinal neurons in each eye are bundled together to form the optic nerve, and the two optic nerves meet at the optic chiasm, where roughly half of the fibers from each eye cross over to the opposite side of the brain. Some of the neurons of the visual cortex are activated by stimuli viewed with a single eye, some by stimuli viewed with either eye and some by a definite relation between stimuli viewed binocularly, or with both eyes. In addition many cortical neurons have been found to respond preferentially to certain stimuli, such as motion in a particular direction across the retina. The existence of such neurons supports the hypothesis that different features in a retinal image are processed separately, by parallel visual channels that are sensitive to one type of stimulus and comparatively insensitive to other types. For example, information pertaining to sideways motion, or motion across the retina, seems to be processed in a distinct channel. It now appears that underlying perception of motion in depth are channels sensitive to monocularly available cue of changing image size and channels sensitive to binocularly viewed cue of relative velocities of left and right retinal images. Human brain is here viewed from below.

At this point we were able to construct a simple psychophysical model of the way the brain processes changing-size information. As the size of a retinal image, say, increases, the movements of opposite sides stimulate pairs of neural systems sensitive to motion in opposite directions. Whenever such a pair of directionally selective motion filters are activated simultaneously, they generate a signal that activates a neural mechanism sensitive to increasing size. The signal from this changing-size filter can generate two different responses. It can either activate a higher-level stage for the perception of changing size or it can feed into a motion-in-depth stage. How, then, does the visual system determine whether a signal from the changing-size filter reflects the changing size of an object or its motion in depth? We obtained some insight into this question when we attempted to generate a negative aftereffect of changing image size.

Many different visual stimuli elicit a negative aftereffect, and the most widely accepted explanation of that phenomenon is based on the idea that neural mechanisms can become fatigued after prolonged stimulation. A well-known example of a negative aftereffect of motion is experienced by someone who stares at a waterfall for a few minutes. When he then switches his gaze to a stationary part of the landscape, it will appear to be moving slowly upward. To elicit a negative aftereffect of changing size we directed subjects to gaze at a test rectangle in which the vertical sides moved toward each other at a constant speed while the horizontal sides remained stationary; after one second this contracting figure disappeared from the screen, reappearing .25 second later at its original size to begin a new cycle. When subjects had been exposed to this adapting stimulus for several minutes, it was replaced with a test rectangle whose sides were stationary. A negative aftereffect of changing size was indeed evident: the test rectangle seemed to be expanding continuously along the horizontal axis.

Once again we were able to rule out the possibility that our results were due to the fatigue of directionally selective motion filters. By contracting the test



TEST SQUARE WITH OSCILLATING SIDES was projected on the screen of a cathode-ray tube to determine whether the human visual system has a distinct channel for processing changing-size information. As is shown at the top, two distinct modes of visual stimulation were possible, that is, opposite sides of the square could be set to move either in opposite directions (*a*) so that the size of the square oscillated or in the same direction (*b*) so that the position of the square oscillated. Subjects' visual sensitivity to either type of stimulus was measured by determining the minimum amplitude of oscillation that could be detected. To establish the existence of a changing-size channel the psychophysical technique known as selective adaptation was employed. First, subjects' sensitivity to changing size and changing position was measured, and then the subjects were exposed for 20 minutes to a strong level of the changing-size stimulus: a square whose sides oscillated in opposite directions with a large peak-to-peak amplitude. In some cases adaptation to a particular stimulus will fatigue whatever neural mechanism is sensitive to the stimulus; then by exposing adapted subjects to a variety of stimuli it is possible to determine the specific sensitivities of that mechanism. In this instance when subjects' sensitivity to changing size and changing position (over a range of frequencies of oscillation) were retested after the adaptation period, a selective loss of sensitivity to changing image size was evident. As is shown in the graph at the bottom, subjects experienced up to five-fold losses of sensitivity to changing size (*color curve*), whereas sensitivity to changing position was almost unaffected (*black curve*). This selective adaptation to oppositely directed motion of opposite sides is too large to be credited to fatigue of directionally selective (sideways) motion filters. These findings and others have resulted in construction of new model of neural activities involved in processing changing-size information (*see top illustration on opposite page*).

rectangle along the horizontal axis it was possible to cancel the changing-size aftereffect, and the rate of contraction required provided us with a measure of the magnitude of the aftereffect and a way to track its decay. As long as the rectangle was less than 1.5 degrees of visual angle in width, the magnitude of the changing-size aftereffect was much larger than that recorded for the classic sideways-motion, or waterfall, aftereffect. Thus the negative aftereffect we had elicited seemed to be largely due to the fatigue of the visual system's changing-size filters.

The compelling illusion that the test rectangle was changing size was rarely accompanied by any illusion that it was moving in depth. We were surprised to find, however, that once the changing-size aftereffect had died away completely another longer-lasting aftereffect was clearly visible: the test rectangle seemed to be moving in depth without changing size. (The magnitude of each aftereffect dropped off exponentially with time, although the decay of the motion-in-depth aftereffect was much slower.)

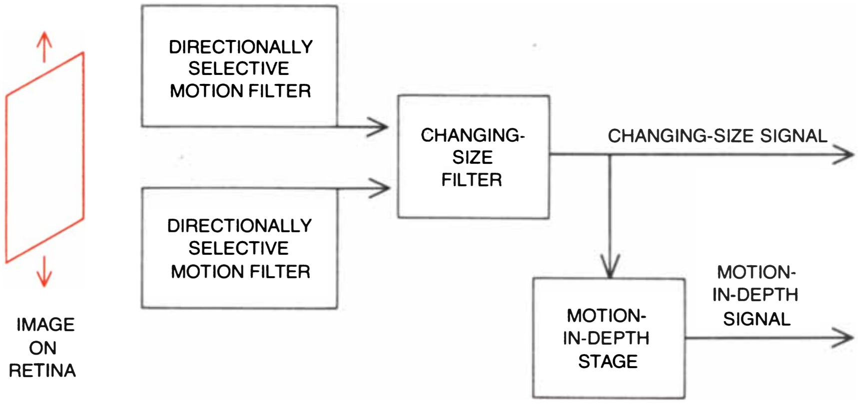
We observed an interesting departure from this pattern when the adapting stimulus was adjusted so that it contracted not in just one direction but in both directions. Then the motion-in-depth aftereffect was usually clearly visible as soon as the adaptation period ended; for most subjects no changing-size aftereffect was visible at all. In other words, the changing-size aftereffect was generated only by the adapting stimulus whose shape was also changing. It is this finding that may give a clue to the way the changing-size channel decides how to interpret the ambiguous message of, say, an expanding retinal image. Although in the real world this may mean either that the object is growing or that it is moving closer, the motion-in-depth interpretation is both more probable and more likely to call for a swift motor response. Our aftereffect findings lead us to speculate that as long as an object's shape remains constant the visual system (possibly by means of some neural biasing system) responds as though changing size were caused by motion in depth. This might be described as a "best guess" solution. For an animal (including the human one) there would be little survival advantage conferred by a visual system that submitted for leisurely intellectual judgment the question of whether a predator was approaching rapidly or swelling rapidly!

We wondered whether the changing-size channel might also be responsible for the human visual sensitivity to flow patterns, the major stimulus correlate of motion available to a pilot landing an airplane. To investigate this possibility we generated on the screen of the cathode-ray tube a flow pattern consisting of short line segments that alternately moved radially away from a focal point and back toward it. The line segments were lengthened and shortened in proportion to their distance from the focus, and the direction of flow was reversed at five-second intervals. During a 10-minute adaptation period subjects were instructed to gaze directly at a "fixation" mark placed at some distance from the focus of the flow pattern.

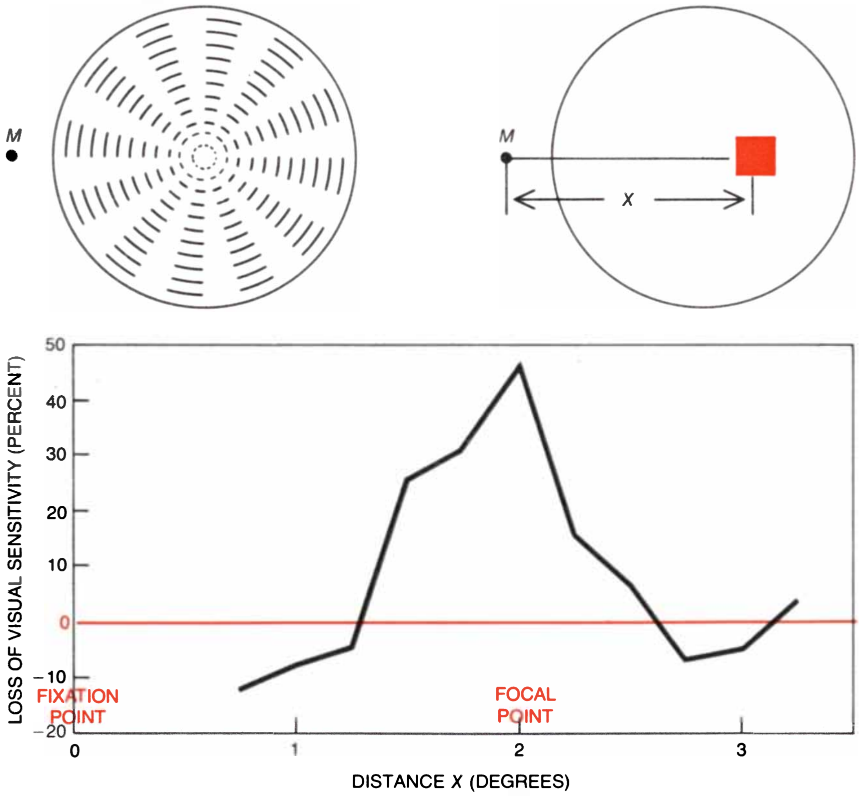
Before and after the adaptation period, and when there was no flow pattern on the screen, we measured subjects' sensitivity to changing-size squares situated at different distances from the point that corresponded to the focus of the flow pattern. (It is important to understand that a flow pattern does not shift when an observer looks away from its focus; for this reason a pilot can judge his direction of flight even when he is looking to one side of the visual field.) We found that once subjects had been adapted to the flow pattern it was much more difficult for them to perceive changes in the size of the test square. In other words, the changing-size channel was sensitive to flow-pattern stimuli. Moreover, the reduction in visual sensitivity to changing size was greatest when the test square was placed precisely at the point that had previously been occupied by the focus of the flow pattern. This result may explain Gibson's finding that pilots are able to judge the direction of flight of a descending airplane by looking at a film made from the plane as it was landing. Changing-size filters may help to locate the focus of the flow pattern and so provide a basis for judging the direction of flight of the plane.

Both changing image size and flow patterns are compelling cues to motion in depth, and both are available to one eye alone. The strength of these monocular cues may explain how some pilots (such as the celebrated Wiley Post) and ballplayers can continue to perform effectively after they have lost the vision of one eye. Turning to the subject of binocular cues, however, it is not surprising to learn that they too can play an important role in the perception of motion in depth. It has long been established that binocular vision plays an important part in stereopsis, or the perception of position in depth, particularly as it relates to nearby objects. Our experiments indicate, however, that the perception of stereoscopic motion—the perception of motion in depth that results from interactions of signals from the two eyes—and the perception of stereoscopic position are accomplished by quite different neural mechanisms. To explain how they differ we should first describe the mechanism by which position in depth is perceived.

Even in the time of Leonardo da Vinci it was understood that a human being's



CHANGING-SIZE CHANNEL is activated by an increase or decrease in the size of an image (*color*) on the observer's retina. As is shown in this model of the information flow in the channel, the movement of each pair of opposite sides in the image of a square stimulates a pair of opposing directionally selective motion filters. When such a pair of filters are stimulated simultaneously, they activate a changing-size filter: a neural mechanism that is preferentially sensitive to changing size. The signal put out by that mechanism can lead either to the perception of changing size or, by way of a motion-in-depth stage of visual perception system, to perception of motion in depth. It is not yet clear how visual system chooses between the alternate routes.



RESPONSE OF CHANGING-SIZE CHANNEL TO FLOW PATTERNS was determined by adapting subjects to the flow pattern shown at the top left. At five-second intervals the direction of flow of the pattern was reversed, so that the line segments first lengthened and moved radially away from the focus, or center, of the pattern and then shortened and moved back toward it. During the 20 minutes the pattern was on the screen of a cathode-ray tube subjects were instructed to direct their gaze to a fixed point *M* on the screen that did not necessarily coincide with the focus. Before and after the adaptation period subjects were exposed to a changing-size square (*color*), placed at a variable distance *X* from *M*, as is shown at the top right. (The distance *X* is measured in degrees of visual angle: the angle between a subject's lines of sight to a pair of points.) These measurements revealed that visual sensitivity to a changing-size square placed sufficiently close to the focal point was indeed depressed, even, as is shown in the graph at the bottom, after an oblique viewing of the flow pattern. Hence the changing-size channel does indeed seem to be activated by a flow pattern. The effect of adaptation was strongest when the square was closest to the point on the screen previously occupied by the focal point of the flow pattern. This finding suggests that the changing-size filters may also serve to locate the focus of a flow pattern, which would indicate a pilot's direction of flight.

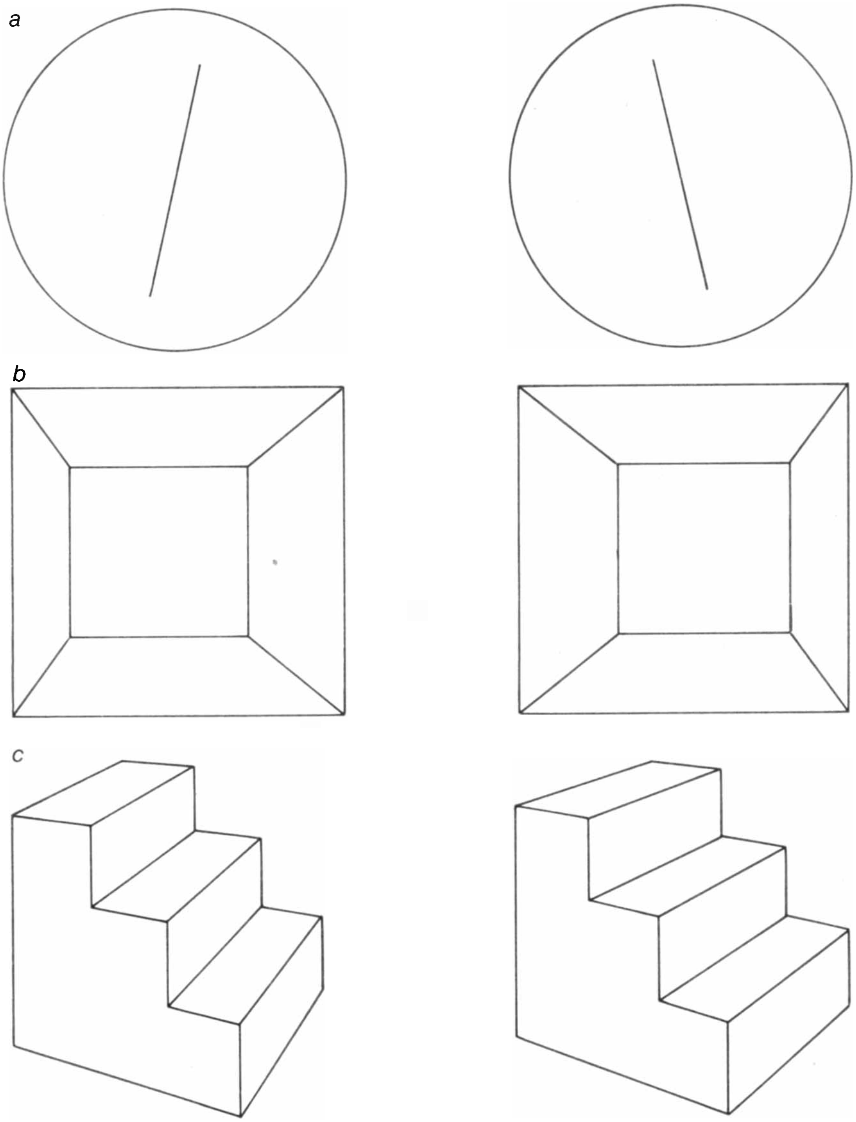
left and right eyes, their pupils being separated by some two and a half inches, have slightly different views of the visual world and that as a result no two-dimensional drawing or painting of a solid object can ever look entirely realistic. Nevertheless, it was not until the 19th century that a clear understanding emerged of how the brain utilizes the geometrical difference between the left and right retinal images of a solid object. In 1838 the British physicist Sir Charles Wheatstone demonstrated to the Royal Society of London the role of this geometrical difference, called binocular disparity, in the precise location of objects in visual space. Wheatstone made pairs of drawings of objects as they would be seen by each eye separately and presented them to the appropriate eyes by means of the instrument he named the stereoscope: a handsomely made device of brass and polished wood that was not essentially different from the stereo viewers of today.

Looking through the stereoscope, most people were able to "fuse" the pairs of drawings they saw into a single image that appeared to extend in depth. In this way Wheatstone demonstrated that a geometrical difference between the left and right retinal images alone is sufficient to convey the perception of depth. Within a year of Wheatstone's demonstration to the Royal Society photography became a practical reality, and by using pairs of daguerreotypes taken from slightly different positions he was able to show that his discovery held not only for line drawings but also for faithful two-dimensional representations of solid objects. Over the past two decades Bela Julesz of Bell Laboratories has extended Wheatstone's work to show how the visual system makes a point-by-point comparison of the retinal images of an object in order to respond to the object as a whole.

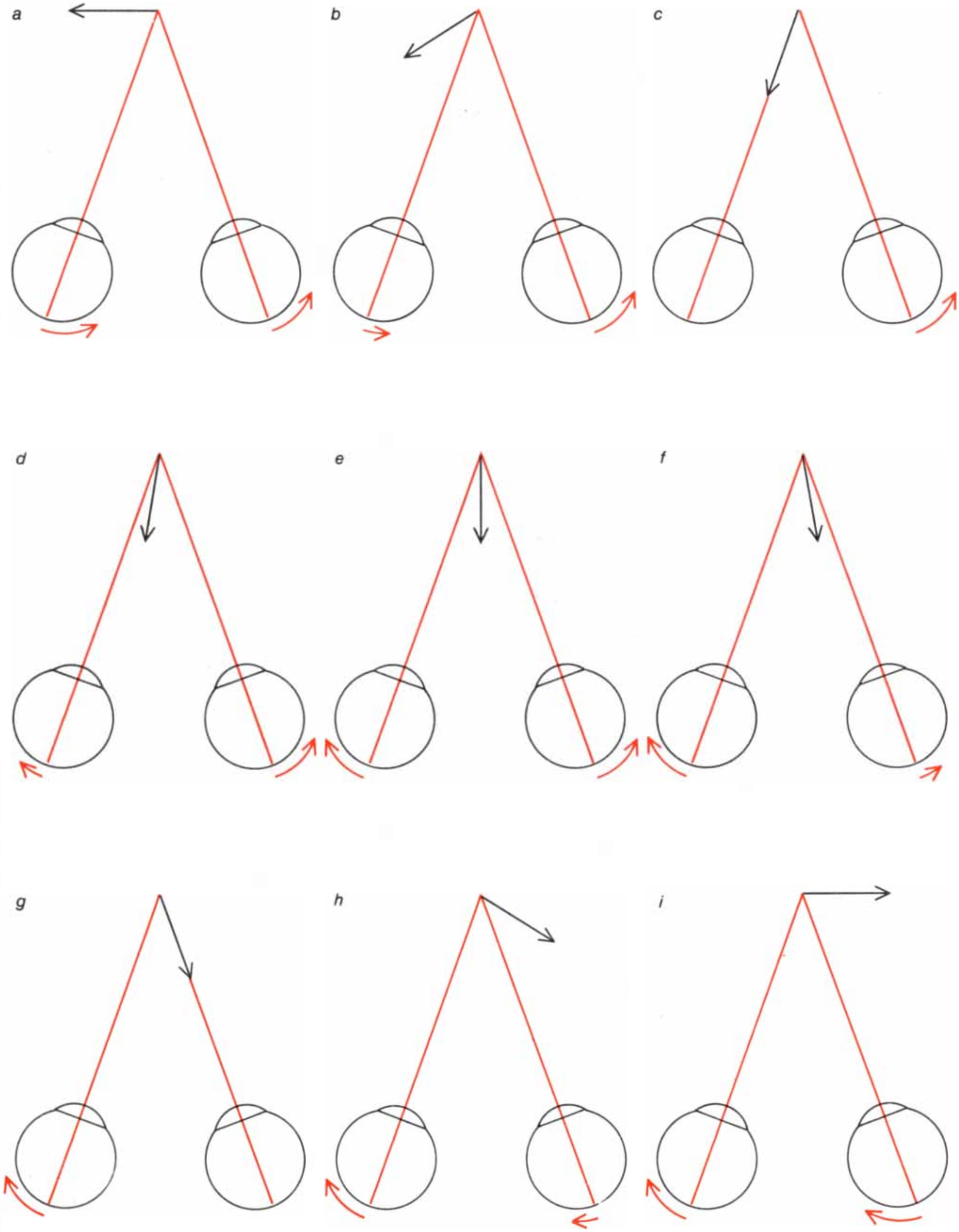
Thus the channel that mediates the perception of depth and solidity responds strongly to the stimulus of binocular disparity. A neural basis for this stereoscopic-position channel was discovered in 1967 by Horace B. Barlow, Colin Blakemore and John D. Pettigrew of the University of California at Berkeley, P. O. Bishop of the Australian National University and Tosaku Nikara of the University of Sydney. Working with microelectrodes to record the activity of single nerve cells in the visual cortex of the cat, they discovered classes of binocularly driven neurons that responded most strongly to pairs of optically fused stimulus bars with a specific binocular disparity. Such neurons, they found, could be highly selective in the sense that they fired only when the disparity of a pair of stimulus bars fell in a narrow range.

Assuming that such classes of disparity-sensitive neurons exist in the human brain, do they mediate the perception of motion in depth as well as the perception of position in depth? Consider the problem confronting the cricket batsman who in order to predict the future location of an approaching ball must judge its instantaneous velocity as well as its instantaneous position. As the ball moves toward him it activates a series of disparity-sensitive neurons, creating in effect a sequence of snapshots of the ball. Neural circuits, taking into account the eyes' exact angle of convergence, could in principle compute from these snapshots the ball's speed and direction of motion. Because the responses of the static-position channels are rather slow, however, such an arrangement would be expected to function in a rather sluggish and imprecise way, whereas the performance of ballplayers demonstrates that in reality such judgments can be made swiftly and accurately.

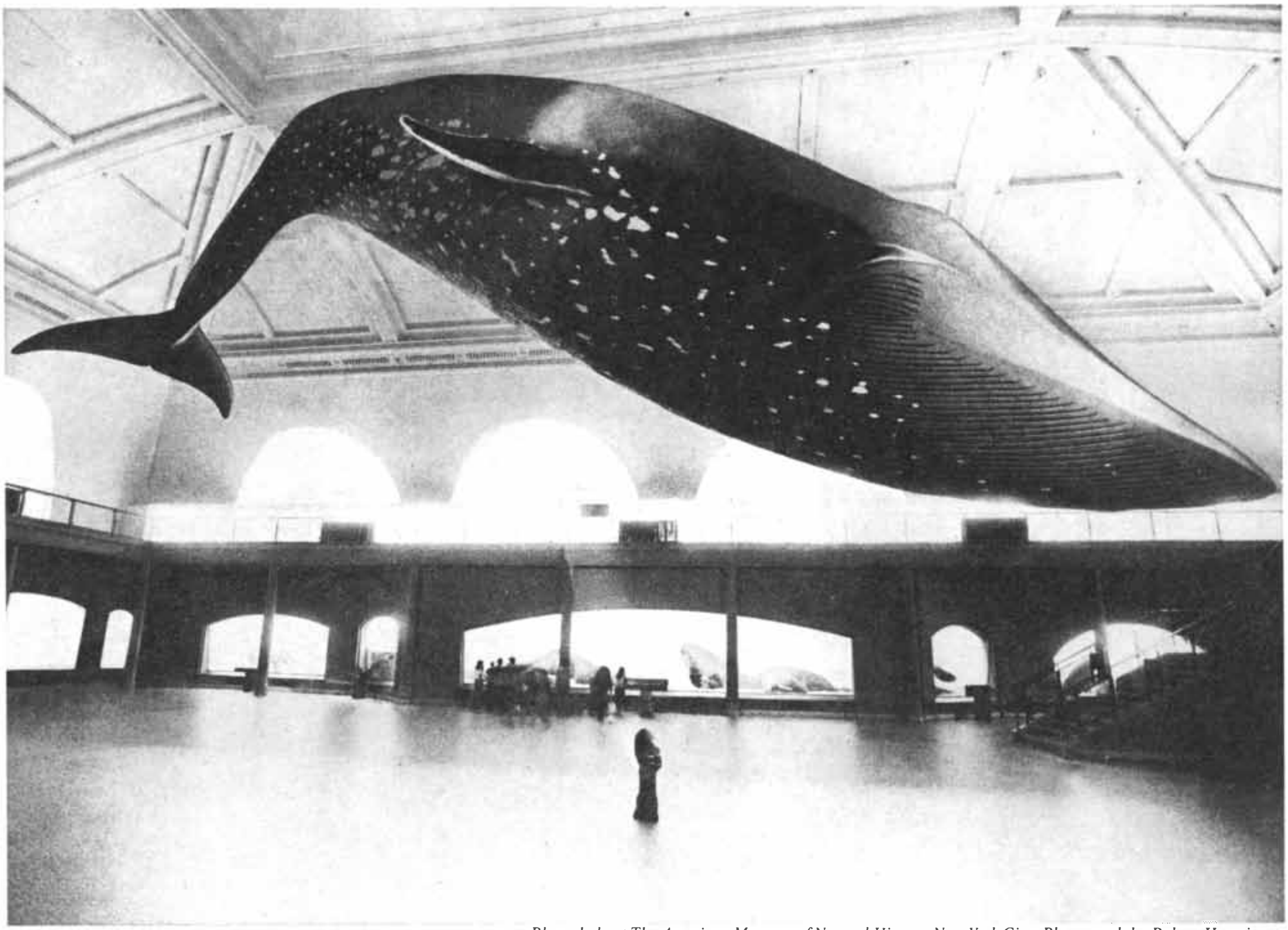
In 1973 two of us (Regan and Beverley) pointed out that the trajectory of an object moving in depth might be determined with more speed and precision directly from the relative velocities of the left and right retinal images of an object. For example, consider again the cricket player. If the left and right retinal images of an oncoming ball are



STEREOPSIS, or the perception of solidity and static position in depth, can be generated entirely by binocular disparity, that is, by the slight geometrical differences in the views of an individual's left and right eyes. For example, the pairs of drawings shown here depict a line (*a*), an outline of a cube (*b*) and an outline of a staircase (*c*) as these objects would be seen by the left and right eyes separately. When such pairs of images are presented to the appropriate eyes by means of a stereo viewer, most observers are able to "fuse" them so that they see respectively a single line tilted in depth, a single three-dimensional cube and a single three-dimensional staircase. Thus binocular disparity is an adequate stimulus for the binocularly driven stereoscopic position-in-depth channel. The authors' experiments suggest that binocularly viewed cues to motion in depth are processed by a distinct stereoscopic motion-in-depth channel.



RELATIVE VELOCITIES of the left and right retinal images of an object provide a cue to the object's direction of motion in depth. For example, if the left and right images of the object are moving in the same direction (*a, b, h, i*), the object will pass wide of the observer's head, whereas if only one image is moving (*c, g*), the object will hit the observer directly in one eye, and if the images are moving in opposite directions at the same speed (*e*), it will hit him directly between the eyes. In addition, if the left image is moving slower than the right one, then object will pass closer to left eye than to right, and if right image is moving slower, then object will pass closer to right eye than to left.



Blue whale at The American Museum of Natural History, New York City. Photograph by Robert Huntzinger

"If man doesn't preserve the last of the great whales, museums may soon be their only habitat." *Dick Cavett*

The blue whale is the largest animal that has ever lived on this earth. It can grow up to 100 feet long and weigh 160 tons.

Whales are still over-exploited

Since 1900 we humans have killed well over 300,000 blue whales so now there are only a few thousand left.

Four other species of great whales are being destroyed. Despite international demands to stop the slaughter, whalers — mainly Japanese and Russian—have already reduced some species to 20 percent of their former population.

Whalers can turn to other trades but the great whales can never be recreated.

We may have much to learn from them

These peaceful creatures have highly developed brains and a close family and social organization. They can communicate over great distances, and dive to great depths.

We cannot yet explain their full role in the ecology of the seas, but if they are wiped out and we find their role was a vital one, it will be too late.

World Wildlife Fund is campaigning to save the life and resources of the seas — for our own sakes and those of our children.

You can help

Send for our free information kit or send your tax-deductible contribution to: World Wildlife Fund, Department SA-3, 1319 18th Street, Northwest, Washington, D.C. 20036.

THE SEAS MUST LIVE

World Wildlife Fund

Imagination and Research on the Frontiers of Science

Books of Special Interest to Readers of *Scientific American*

GRAVITATIONAL CURVATURE

An Introduction to Einstein's Theory

Theodore Frankel, University of California, San Diego

This fresh introduction to gravitational curvature captures the geometric essence of the subject in a modern mathematical setting. Replacing sometimes tedious analytical computations with geometric arguments, the book presents new derivations for Einstein's equations of gravitation along with other results. The derivations presented are at once clearly motivated and mathematically clean and simple. *Gravitational Curvature* includes new and attractive derivations for the deflection of light, Schwarzschild's exterior and interior solutions, and the Oppenheimer-Volkoff equations important in the study of stellar structure.

The book stresses the global aspects of cosmology and is suitable for independent study. The reader should have a background in Riemannian geometry and basic physics, or some familiarity with relativity theory. Background chapters, with derivations, are included on special relativity, continuum mechanics, and electromagnetism.

1979, 172 pages, 29 illustrations, cloth $18.50, paper $8.95

CANCER

Science and Society

John Cairns, Mill Hill Laboratory, London

"The best book that I have read explaining cancer to the intelligent layman." —Bruce N. Ames, University of California, Berkeley

A common cause of death in the industrialized Western world, cancer is one of the most feared of all diseases. Dispelling the misinformation surrounding cancer, John Cairns reviews its nature, its causes, its place among other diseases, the current state of research, and the prospects for its final conquest. He convincingly argues that cancer is largely due to controllable variables in the human environment and should therefore be preventable.

Written clearly but never simplistically, *Cancer: Science and Society* is for the general public as well as for physicians, medical students, and students of biology. References to the technical literature on the subject are given throughout the book for those who wish to read in more depth.

1978, 199 pages, 52 illustrations, cloth $13.00, paper $7.50

FRACTALS

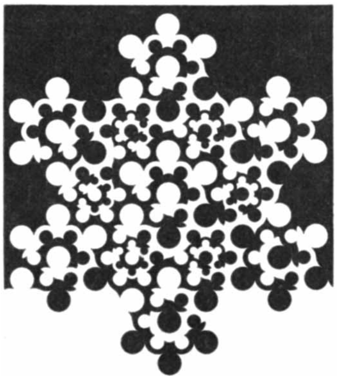
Form, Chance, and Dimension

Benoit B. Mandelbrot, IBM Fellow, International Business Machines Corporation, Thomas J. Watson Research Center

"The new and fascinating world of patterns presented here is all the justification this book needs. Read it, or better yet, look at the pictures. It is inconceivable that it would fail to give some food for thought." —*American Scientist*

Fractals: Form, Chance, and Dimension is an intriguing investigation that claims the geometry of fractals is necessary to describe geometric shapes in nature. Mandelbrot examines fractal applications for describing the surface of the earth, the clusterings of stars, and the shapes of clouds, snowflakes, and other forms appearing in nature.

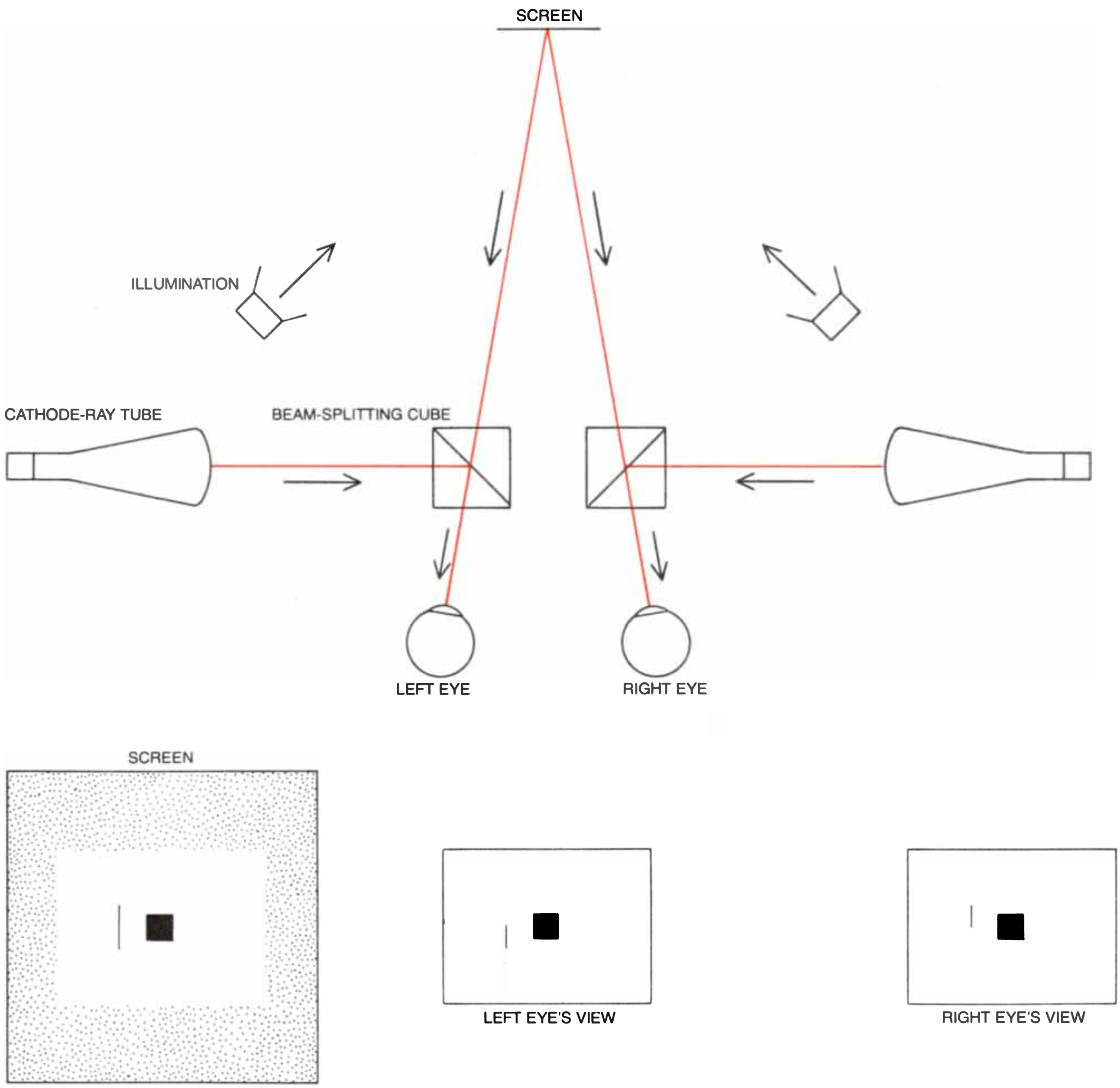
1977, 365 pages, 68 illustrations, cloth $18.50



W. H. Freeman and Company
660 Market Street, San Francisco, CA 94104

moving in the same direction, the ball will pass wide of his head, whereas if the images are moving in opposite directions the ball will—if it is unimpeded—hit him directly between the eyes. Moreover, if the left retinal image is moving slower than the right one, the object will pass closer to the right eye than to the left one, and so on. In this way the relative speeds and the relative directions of the retinal images might provide a reliable indication of the ball's motion in depth. In other words, the relative velocities of the left and right images of the ball, rather than the rate of change of the binocular disparity, might serve as the stimulus for a distinct stereoscopic-motion channel.

To investigate this possibility we conducted selective adaptation experiments with motion-in-depth stimuli generated with a modern electronic version of Wheatstone's stereoscope. Independently oscillating patterns were generated with a pair of cathode-ray tubes and presented separately to a subject's left and right eyes. The speed and direction of motion of the patterns (in some cases fields of horizontally oscillating dots and in others horizontally oscillating bars) could be adjusted, and the optically fused image appeared to be on a



ELECTRONIC STEREO VIEWER was utilized to investigate the stereoscopic perception of motion in depth. In these psychophysical experiments independently moving stimuli such as bars, rectangles or fields of dots were generated on a pair of cathode-ray tubes and presented separately to an observer's left and right eyes. The fused stimulus (here a rectangle) appeared against a white square on the screen opposite the observer. When the stimuli for the two eyes were made to oscillate in opposite directions, the fused stimulus seemed to be moving in depth along a line passing between the eyes, and when the stimuli were made to oscillate in the same direction, the stimulus seemed to be moving in depth along a line passing wide of the head. Fine dots covering outer area of the screen were helpful in keeping a subject's eyes converged on screen; when that convergence was accurate, thin lines projected separately to eyes appeared exactly in line.

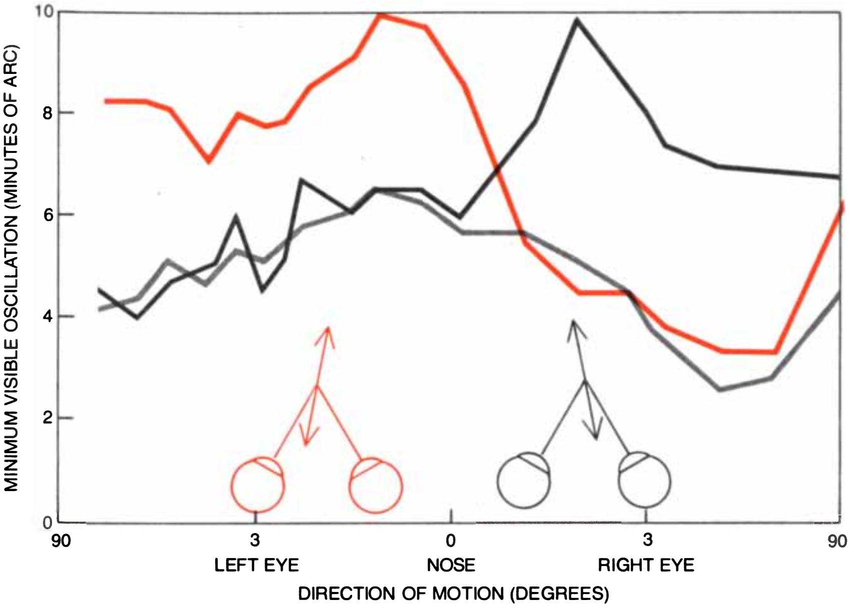
screen facing the subject. When the two patterns were set to move in, say, opposite directions, the resulting image resembled a "3-D" motion picture: subjects saw an expanse of dots or a bar that seemed to move back and forth in depth.

The impression of motion in depth was quite strong at the outset, but after several minutes it weakened until at the end of 10 minutes all motion in depth seemed to have died away. In other words, we had created a condition of temporary blindness to motion in depth. This result supported the existence of a stereoscopic-motion channel, since sideways motion could be viewed about equally well before and after the adaptation period. The motion-in-depth channel seemed to be more or less independent of the established channels that are sensitive to binocularly viewed sideways motion.

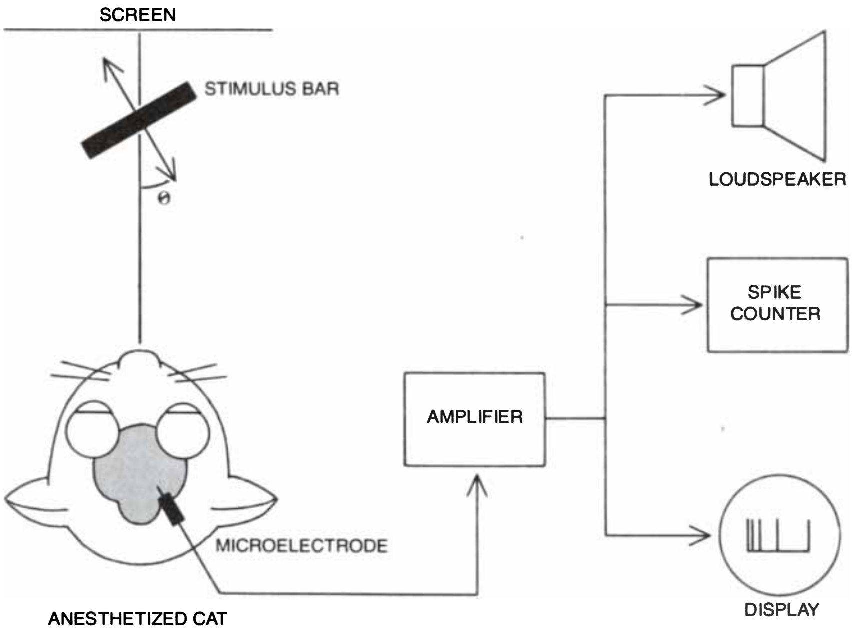
We were also intrigued to discover that the sizable effect of the adaptation to stereoscopic motion was specific to the direction of motion of the adapting stimulus. For example, in subjects who were exposed to oscillations along a line passing to the left of the nose, visual sensitivity was reduced for all the trajectories passing to the left of the nose but not for those passing to the right of the nose. Hence there seemed to be at least two different directionally specific stereoscopic-motion channels.

This observation suggested that those channels were selectively sensitive to the ratio of the velocities of the left and right retinal images and were relatively indifferent to absolute speed. To determine how many such "ratio-tuned" channels exist in the visual system we then tried adapting subjects to 13 different directions of motion in depth. Our results could be explained in terms of only four channels, each comprising two (opposite) directions of motion. These channels turned out to differ greatly in their degree of directional selectivity. The channels sensitive to motion passing between the eyes are highly selective, responding only to trajectories within specific 1.5-degree ranges. Thus those channels account for only about six degrees of visual space. The channels for motion missing the head respond to movements in a much wider range of directions, accounting for the remaining 354 degrees of visual space.

The view that information pertaining to position in depth and information pertaining to motion in depth are processed in different channels is supported by evidence of a quite different type gathered by one of us (Regan) and by Whitman Richards of the Massachusetts Institute of Technology. Richards has developed a technique, called stereoperimetry, for measuring the distribution of the sensitivity to depth over a subject's visual field. With this tech-



STEREOSCOPIC-MOTION CHANNELS that respond preferentially to motion in a specific direction were identified when subjects were adapted to a field of dots moving in depth. First, subjects' normal sensitivity to motion in depth along a number of different directions was measured (*gray curve*). Then subjects were adapted for 20 minutes to a field of dots oscillating along a specific trajectory passing to the left of the nose. (This binocularly viewed stimulus was generated by the electronic stereo viewer shown in the illustration on the opposite page.) Subjects' sensitivity to motion in depth after the adaptation period (*color curve*) was reduced for all directions to the left of the nose and for none to the right; the converse was also true (*black curve*). This finding indicates that the visual system contains at least two binocularly driven neural mechanisms sensitive to a particular direction of motion in depth. Subsequent experiments have revealed a total of four such stereoscopic-motion channels. In this experiment direction of motion was measured as angle between trajectory and line from center of viewing screen to subject's nose; amplitude of oscillation is given in minutes of arc of visual angle. Spacing along horizontal axis of graph has been distorted to emphasize certain directions of motion.



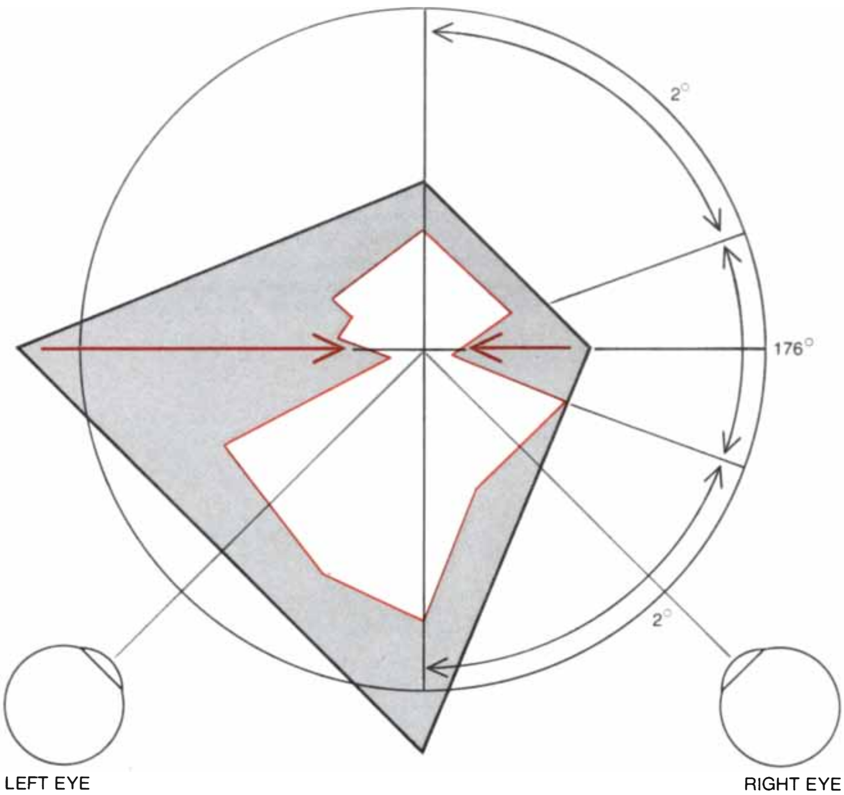
EXPERIMENTAL SETUP for recording the activity of single neurons in the visual cortex of the cat is shown. The tip of a microelectrode was placed close to a neuron in area 18 of an anesthetized cat's cortex, and a pair of moving bars were presented separately to the animal's eyes. A human observer verified that from the cat's vantage the bars fused into a stimulus bar that seemed to be oscillating in depth. By varying the relative motions of the individual bars it was possible to change the direction of motion θ of the oscillating stimulus bar. Number of spikes, or times a neuron fired, in response to movement along a particular trajectory was recorded with counting device and continuously monitored by means of oscilloscope and a loudspeaker.

nique it has been possible to identify subjects whose visual fields include areas that are "blind" to motion in depth but display normal sensitivity to position in depth. (More recently Richards has discovered individuals who show the converse defect.)

To search for a physiological basis for these channels, in particular classes of neurons responding selectively to stereoscopic motion in depth, two of us (Regan and Cynader) used microelectrodes to record the firing patterns of single neurons in area 18 of the visual cortex of anesthetized cats. A pair of stimulus bars that oscillated independently from side to side were presented separately to the animal's left and right eyes. (A human observer checked to see that the bars appeared as a single image moving back and forth in depth.) By adjusting the relative velocities of the two bars it was possible to vary the direction of motion in depth (in a single horizontal plane), so that a neuron's response to motion in directions throughout a 360-degree range could be tested.

About a third of the neurons tested in this way proved to be sensitive to the direction of motion in depth. Many of the same neurons turned out to be relatively insensitive to binocular disparity. In other words, these neurons emphasized motion information at the expense of position information. It is important to understand that with this experimental setup the direction of motion in depth could not be judged with one eye alone. (In fact, motion in depth could not even be seen with one eye alone.) The direction of motion of the stimulus image was available to the cat's brain only through the comparison of the signals from the left and right eyes. Therefore our results reflected the interaction of signals from the left and right eyes.

To obtain a measure of this binocular interaction we also recorded the response of the same neurons over the same full range of stimulus directions, first with the cat's left eye covered and then with its right eye covered. For one type of neuron tested the sum of the individual firings for the left and right eyes was always higher than the binocularly stimulated firing, demonstrating that the signals from the two eyes acted to inhibit each other [*see illustration on this page*]. Most important, the strength of this inhibitory interaction varied greatly depending on the direction of motion of the stimulus. In general the interaction was strongest when the image appeared to be moving sideways, and it lessened only for a remarkably narrow range of directions of motion in depth: no more than four degrees around the trajectory passing between the eyes. In other words, the neurons fired strongest in response to trajectories aimed directly at the head. Neurons were also discovered that achieved a similar directional selectivity by means of a different mechanism. For this type of neuron the binocular interaction, instead of suppressing firing for sideways motion, facilitated firing in the preferred direction, so that the response to binocular stimulation was much greater than the sum of the monocularly stimulated responses.



SELECTIVE RESPONSE OF NEURONS to movement along a particular direction in depth is demonstrated by this polar graph of the firing pattern of a single neuron in the visual cortex of the cat. In the graph the radial distance from the center of the circle represents the total number of times the neuron fired in response to each movement of the stimulus bar. The color curve shows the neuron's firing in response to the simultaneous stimulation of the left and right eyes over a 360-degree range of directions. The neuron fired most strongly when the bar moved toward the head and hardly at all when the bar moved sideways, supporting the psychophysical findings pointing to the existence of directionally specific stereoscopic-motion channels. The black curve shows the sum of the values obtained when the left and right eyes were stimulated independently. As the difference between the two curves (*gray shading*) indicates, for this type of neuron the eyes interact to inhibit each other in a way that clearly depends on the direction of motion in depth. In this case the inhibition was strongest when the target was moving sideways, as is shown by the colored arrows. Hence the binocularly driven firing of the neuron was quite weak except in narrow eight-degree range centered on trajectory passing between eyes. As is shown by curved black arrows, angular measures in graph have been distorted accordingly.

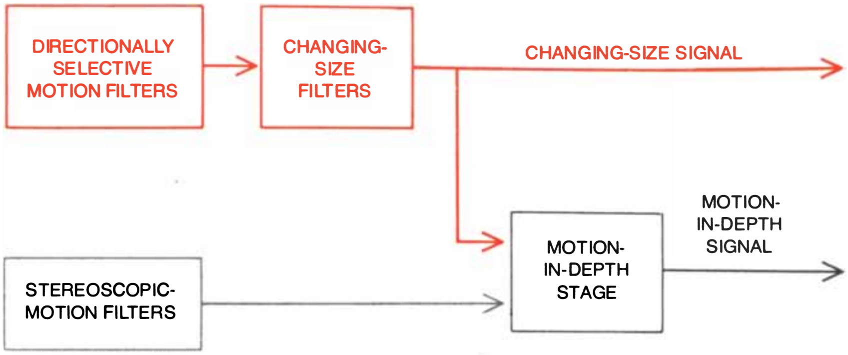
Of the many neurons tested in the cat's visual cortex several were found corresponding to each of the different stereoscopic-motion channels revealed by our psychophysical experiments on human subjects. In particular some neurons responded best to a narrow range of trajectories passing through or close to the head; of these neurons some were specific to motion directed toward the head and others were specific to motion directed away from it. Other neurons fired in response to motion in a much broader range of directions (often more than 90 degrees), responding to trajectories that passed wide of the head. These findings about area 18 of the visual cortex of the cat have recently been confirmed for area 17 of the visual cortex of the monkey by Gian F. Poggio and William H. Talbot of Johns Hopkins University.

(We were also intrigued to find that compared with the responses of neurons sensitive to static disparity, the responses of neurons sensitive to stereoscopic motion were fairly independent of the physical distance between the stimulus and the animal's head. It is tempting to speculate that a neuron responsive to stereoscopic motion in depth signals something along the lines of "I don't know exactly where it is, but it is going to hit you." The activity of such neurons may explain how a squash player is able to dodge a ball that is hit toward his head even though he only sees it from the corner of his eye and so cannot judge its exact position in space.)

Since a compelling impression of motion in depth can be generated either by changing-size stimuli alone or by stereoscopic-motion stimuli alone, we wondered whether there was any qualitative difference in the perception of motion in depth evoked by such different visual inputs. We began our investigation of this question by establishing, as we had in earlier experiments, a motion-in-depth aftereffect of changing size. Subjects were adapted to an expanding stimulus rectangle so that when they were presented with a test rectangle whose sides were stationary, it appeared to move away in depth. We then tried to cancel this motion-in-depth aftereffect, by means not of a changing-size stimulus but of a stereoscopic-motion stimulus: identical test rectangles moving apart horizontally were presented independently to the left and right eyes, so that their fused image seemed to be moving in depth toward the subject's head. Subjects with normal visual sensitivity to stereoscopic motion found that the motion-in-depth aftereffect of changing size could be canceled quite easily in this manner. (Some experimentation was required, however, to find the exact relative velocities of the retinal images for which the aftereffect was canceled: if this input was too weak, the test rectangle appeared to move away from the subject in depth, and if it was too strong, the test rectangle appeared to move toward him.)

This finding indicates that changing-size stimulation and stereoscopic-motion stimulation generate visual signals that converge on the same motion-depth stage of the visual perception system. Further evidence for this theoretical model was provided when we generated a stimulus square for which the relative velocities of the left and right retinal images indicated motion directly toward the head and the changing size indicated motion directly away from the head. In this situation the cue of relative velocity is equivalent to that of changing binocular disparity. By adjusting the rate of change of this disparity it was possible to cancel the impression of motion in depth. (Interestingly, at this cancellation point the stationary square appeared to be continuously decreasing in size.)

In the real world changing size and changing disparity are almost never opposed as they were in these experiments. With the contrived experimental situation, however, it is possible to make a direct comparison of the two visual inputs, and this comparison proves to be revealing about real moving objects. There is a constant ratio between the magnitudes of changing size and changing disparity observed for a moving object, a ratio that is determined geometrically by the width of the object and the separation of the observer's eyes but that surprisingly is almost independent of viewing distance. When we normalized our data with respect to this ratio, we found that an object's velocity in depth and the length of time it is observed have a considerable effect on the relative effectiveness of changing size and changing disparity as stimuli for motion-in-depth perception. For example, for a slowly moving object glimpsed briefly, say for .2 second, the changing-size input would tend to be more effective, whereas for a rapidly moving object viewed for a longer time, say a full second, the binocular, stereoscopic-motion channel would tend to be more effective. This is a property of the human visual system that is not predictable from geometry. (Another important factor is the way individuals differ in their natural sensitivity to these two types of stimuli. For some people changing disparity can be up to nine times more effective than changing size, whereas for others it is an almost totally ineffective stimulus.)



MOTION-IN-DEPTH STAGE of the visual perception system can be driven either by changing image size or by the relative velocities of the two retinal images, as is shown in this model of the neural activities that lead to the perception of motion in depth. The model explains why certain illusions of motion in depth that result from adaptation to a changing-size stimulus can be canceled by adjusting either rate of change of size or relative velocities of the retinal images.

These findings may shed some light on the question of how a pilot's flying performance would be affected if he suddenly lost binocular vision. C. E. Lewis and his colleagues at the National Aeronautics and Space Administration's Flight Research Center in Edwards, Calif., conducted heroic experiments in this area, suddenly depriving jet pilots of the vision of one eye during landings. Surprisingly, they found that landing performance suffered very little in these circumstances, and in some cases it even improved. It is of course possible that the pilots made more of an effort after losing binocular vision, but this explanation seems less likely in view of the fact that comparatively inexperienced private pilots, whose skills were presumably tested to the limit under normal conditions, showed that they too could land equally well with one eye. Our less adventurous laboratory experiments suggest that there is no simple answer to the question of how a sudden loss of binocular vision affects the perception of motion in depth, because the ratio of visual sensitivities to the binocular stimulus of changing disparity and the monocular stimulus of changing size depends on the size of the object viewed, on its velocity, on the inspection time and very much on the individual.

The visual world seen by a pilot or a ballplayer is much richer than the configurations of visual stimuli we utilized in the experiments described here, and so it is legitimate to question the relevance of the laboratory findings to those real-world situations. Our approach, which is similar to that adopted by many (but by no means all) visual scientists over the past 100 years, is based on the belief that carefully designed experiments can break down the operation of the visual system into elements that are more or less independent, and that a knowledge of these elements will advance the understanding of how vision operates in the complex everyday world.

Even in the challenging tasks undertaken by pilots and ballplayers the motor response to a complex visual scene is often simple. The signals generated by a number of different inputs converge to give rise to a single motor output such as the moving of an airplane throttle or the swinging of a baseball bat. If a better understanding is to be reached of how eye-to-hand coordination is achieved in everyday tasks, it will be necessary to study not only the responses to individual stimulus correlates of motion but also the interactions of those responses when several such cues are present at the same time.

THE AMATEUR SCIENTIST

How to build a simple seismograph to record earthquake waves at home

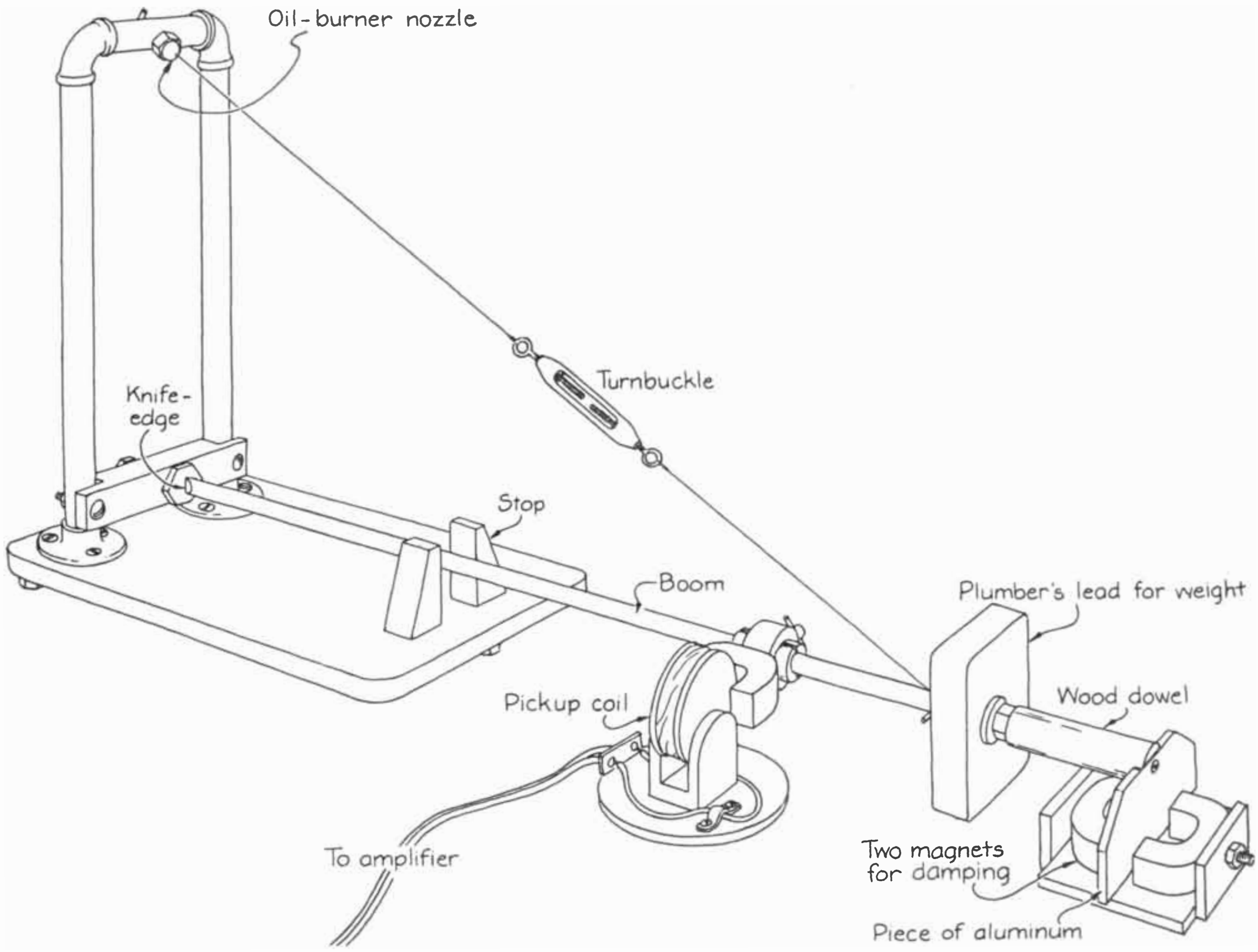
by Jearl Walker

This month I shall describe a seismograph built by James D. Lehman of James Madison University in Harrisonburg, Va. He has sent me his design for making the apparatus. The machine can be built quite easily and is sufficiently sensitive to record North American earthquakes of magnitude 4.8 or more on the Richter scale and earthquakes elsewhere of magnitude 6 or more. Lehman has obtained records of an underground test of a nuclear device in Nevada and a severe earthquake in Turkey.

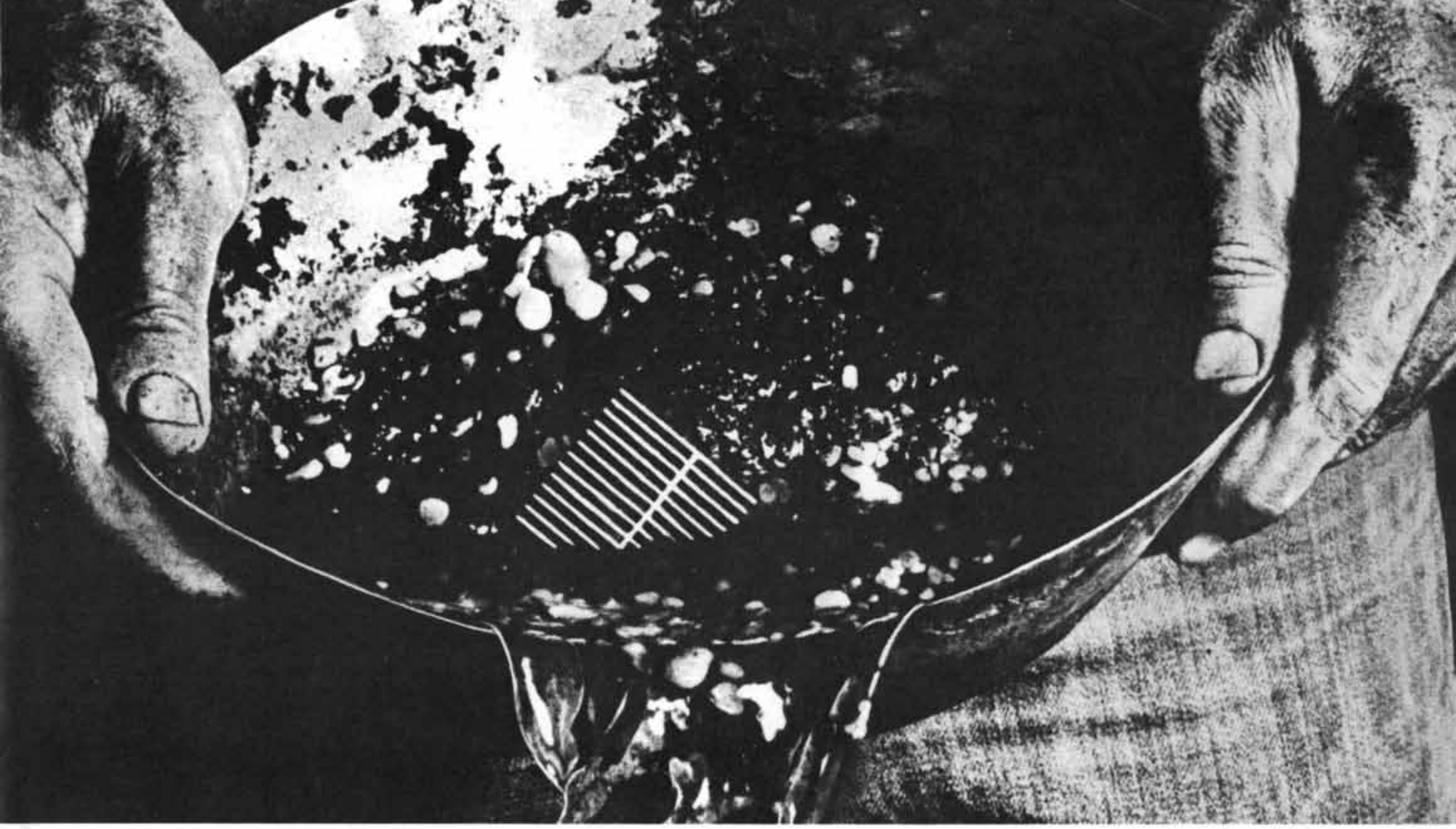
The main objective in the design of any seismograph is to translate a slight movement of the earth into a larger motion that can be put on some kind of permanent record, which can then be analyzed. The waves from a distant earthquake do not all arrive at the same time; some take longer paths through the earth and some travel more slowly. Hence the record should be plotted as a function of time so that the arrival of the various components of the waves can be noted and timed. The difference in arrival times of the types of waves can be utilized to determine the distance of the earthquake from the instrument.

From a variety of possible seismograph designs Lehman has chosen a cantilever type coupled to a strip-chart recorder. Basically his seismograph consists of a weighted boom, a horseshoe magnet mounted on the boom and a coil of wire positioned between the poles of the magnet. When seismic waves pass through the room, the magnet and the coil move with respect to each other, thereby changing the magnetic flux through the coil. The voltage generated in the coil by the variation in flux is then amplified and recorded on the strip-chart recorder. A wiggle of the floor therefore ends up as a larger wiggle on the paper. Since the paper is moving through the recorder at a specified rate, the clock time and the duration of the wiggle can be determined from it.

The arm of the boom rests on a knife-edge that is held against a frame. The other end of the boom is supported by a wire attached to the top of the frame. The two support points for the boom



The seismograph built by James D. Lehman



Using a computer is easy... for onComputing readers.

Read onComputing to find out . . .

- **What a microcomputer can do**
- **How to get started**
- **What's new in personal computers**
- **Where to buy your computer**
- **How to use your computer**

The editors of **onComputing** realize that much of the material written about computers is not at all suited to the person who just wants to use a computer as a tool for business, education, home entertainment, laboratory work, or other applications. **onComputing** is dedicated to helping the computer user understand the capabilities of a microcomputer—in non-technical language.

onComputing is entertaining and informative. It contains practical articles on how to get started, including what you'll need for your application and what it will cost. It features book reviews, product reviews, information on what's new in personal computers, where to buy a personal computer, and—how to use it.

onComputing is a totally new publication. It is issued quarterly and contains articles from some of the best known names as well as from competent amateurs. It is edited and produced under the guidance of an experienced staff of computer experts. The articles in **onComputing** have never appeared in any other publication. They are all fresh, informative, and valuable reading for anyone interested in using a computer—for fun or profit!

onComputing™
GUIDE TO PERSONAL COMPUTING
Start your subscription today.

EVERY THREE MONTHS **onComputing** will bring the latest developments in the field of personal computing: use, applications, books, selection—all in an easy-to-read style.

onComputing Subscription Department, P.O. Box 307, Martinsville, NJ 08836

REGULAR subscription rate.

☐ U.S. 1 yr. (4 issues) @ $8.50 ☐ Canada & Mexico 1 yr. (4 issues) @ $10.00

FOREIGN (to expedite service, please remit in U.S. funds drawn on a U.S. bank.)

☐ Europe (and all other countries, except above) 1 yr. @ $12.00 — surface delivery

☐ Bill VISA ☐ Bill Master Charge ☐ Bill me (North America only)

Card Number ______________________ Expiration ________

Signature ______________________ Name (Please print) ______________________

Street/Apartment Number ______________________

City ______________ State/Province/Country ______________ Postal Code ______________

7179

(the knife-edge and the point where the wire is attached to the frame) are not on a vertical line. Therefore when the boom is set in oscillation by a seismic wave, it will swing horizontally in a pendulumlike motion around its rest position. If such a motion is allowed, about all that the seismographic record will show is the natural swinging of the boom excited by passing seismic waves. The periods and arrival times of the individual waves will be lost.

To avoid such loss Lehman has designed the seismograph to damp the boom's swing quickly. This damping is accomplished by causing a piece of aluminum mounted on the far end of the boom to swing through a magnetic field. The motion generates currents in the aluminum that in turn generate magnetic fields of their own. The magnetic fields induced by the motion oppose further motion of the boom, which soon stops swinging.

If the natural swing is stopped, what the seismograph records is the relative velocity between the boom and the floor as the boom and the pickup coil move with respect to each other because of their different inertial masses. The current created in the pickup coil is proportional to the velocity of this motion.

If there is no damping, the boom will be most sensitive to seismic waves that have the same period as the natural pendulum period of the boom. With damping it will be most sensitive to waves with periods about half the length of the boom's natural pendulum period. An instrument built to respond to a seismic-wave period of about a second is called a short-period seismograph, and one that responds to periods of from 10 to 20 seconds is a long-period seismograph. Lehman adjusts his boom for a natural period of from 12 to 18 seconds, and so he has a long-period instrument.

The main frame of the seismograph consists of a base and two upright sections of pipe. The base is best made of metal for rigid support, but seasoned wood or thick pressed board could also serve, provided you guard against warping from the humidity in the room. Lehman has a metal base of about 40 by 25 centimeters supported by three 3/8-inch machine nuts he has glued to it, two at the rear and one in the middle of the front. The three-point support of these nuts should provide a level base free of any rocking. Shims of thin metal can be put under any of the legs to tilt the base as desired when the seismograph is prepared for operation.

The boom consists of a steel rod 5/16 inch in diameter and 75 centimeters long. One end is ground to a knife-edge. The other end is threaded for about eight centimeters. At the threaded end a nut is screwed on as far as the threads go and then a piece of plumber's lead weighing five to six pounds is slipped over the end, followed by another nut. The two nuts are screwed against the lead, holding it in place on the rod.

The uprights of the seismograph are attached to the base by means of two half-inch pipe flanges, which are bolted to the base so that their edges are about half an inch from its edge. Large wood screws or quarter-inch machine bolts will serve to fasten the flanges to the base, but make certain that the fasteners do not interfere with the legs supporting the base. Countersinking the heads of the fasteners may be necessary.

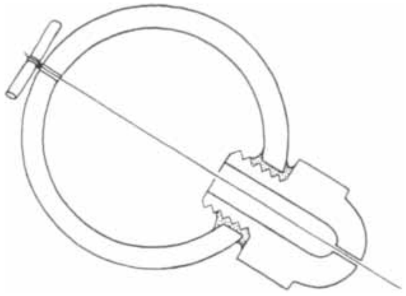
Two lengths of half-inch pipe, each 18 inches long and threaded at both ends, are screwed into the flanges. Crosspieces will run horizontally between these two uprights at two places: where the knife-edge of the boom rocks against the frame and where the wire supporting the other end of the boom is attached. The lower crosspiece is a metal bar bolted onto the upright pipes. The upper crosspiece is a half-inch pipe that is 3½ inches long and threaded at both ends. It is attached to the upright pipes by two half-inch pipe elbows.

Midway across the higher crosspiece the wire supporting the boom must be attached. Lehman uses a nozzle from an oil burner as a pivot point for the wire. The standard type of nozzle has a diameter of 9/16 inch. Take out the inside filter because all that is needed is the tiny hole through which the wire is to run. Assemble the base, the flanges, the uprights and the upper crosspiece and determine where the wire from the boom should be attached. You will want it midway across the crosspiece and on the pipe so that when the rig is assembled, the wire will form an angle of from 30 to 40 degrees with the horizontal. Mark the proper location on the cross pipe and then drill a 9/16-inch hole in it. Glue the nozzle into the hole, run the wire through the nozzle and through the pipe and then anchor the wire on the other side of the pipe.

Lehman suggests that you use either No. 26 Nichrome wire or piano wire. One length of the wire will come down from the nozzle and another piece will come up from the far end of the boom, where it is attached through a hole drilled in the boom. The two ends are attached and tightened through a turnbuckle about halfway up.

The lower crosspiece is a flat length of metal bolted to the upright pipes about two inches above the base. To it is attached a piece of hard metal such as the head of a bolt ground flat. The knife-edge of the boom must rest on the hard metal plate with the knife-edge vertical. When the seismograph is finally assembled, the nozzle on the upper crosspiece should be about a centimeter forward of the vertical from the knife-edge plate. An adjustable metal plate would be helpful here.

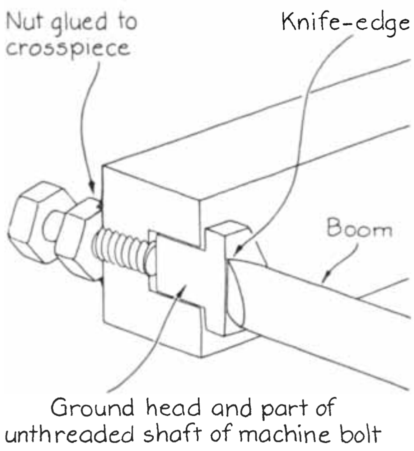
Attach to the boom a horseshoe alnico magnet. One can be obtained from the Edmund Scientific Company (6982 Edscorp Building, Barrington, N.J. 08007). Lehman has a magnet with a 25-pound pull. The magnet should have a gap of about 7/8 inch. In the gap will rest the pickup coil from which signals about the relative motion of the magnet and the coil are fed to the electronic equipment. Lehman uses a rod clamp and pieces of wood to attach the magnet to the boom near the piece of lead.



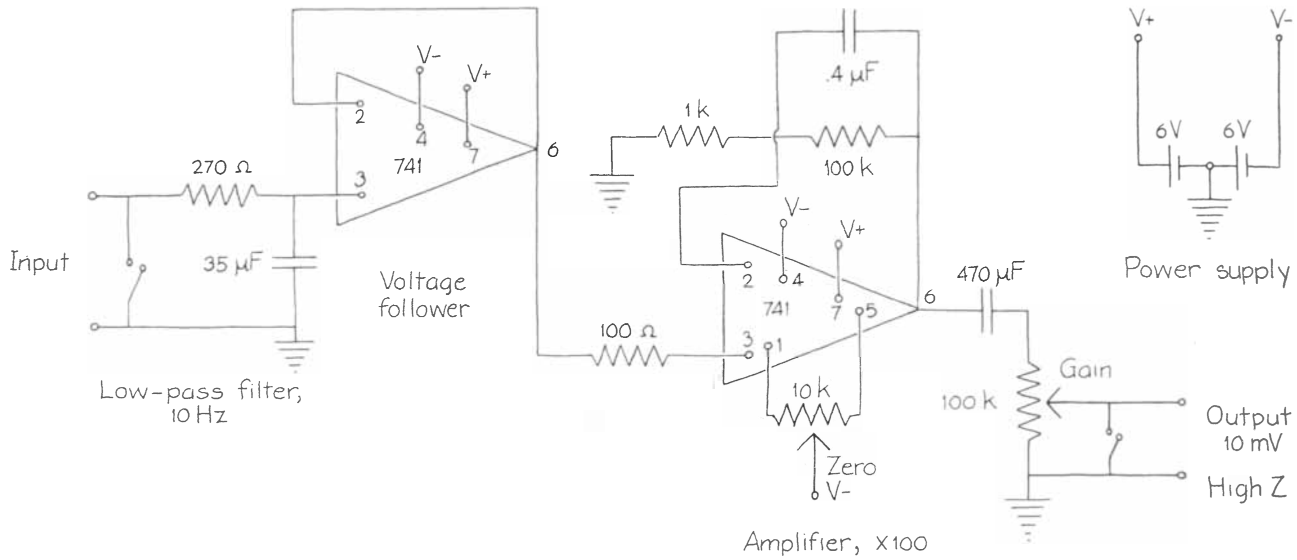
Nozzle rig for the support wire

The pickup coil is made by winding 10,000 or more turns of scramble-wound No. 34 magnet wire on a coil form. If you have an Edmund magnet (No. 41949), the coil form should be 3/4 inch wide and three inches in diameter, so that the pickup coil can fit from a quarter inch to a half inch into the gap of the magnet with sufficient clearance on the sides. Lehman made his coil form by gluing two circular pieces of 1/16-inch plastic to a wood cylinder an inch in diameter. The form was mounted on a drill press and then the wire was carefully wound onto the form while the press was run at low speed. (Be careful about running the press; it is not difficult to get your fingers caught in the wire.) Lehman brought the ends of the coil out to the sides of the form and then ran lamp cord from the ends to the amplifier. The cord should not be much longer than 10 feet; otherwise it will have too much electrical resistance.

The coil is mounted on a stable base



Positioning of the knife-edge



Circuitry of the amplifier

that has the same type of three-leg support as the main base. Be sure to avoid installing any material that can become magnetized. For the base you could use a piece of aluminum with brass hardware to fasten the coil in place.

At the far end of the boom you will need another magnetic field in order to damp the swing of the boom. Lehman mounts two of the Edmund magnets on a wood base, orienting them so that they attract each other. He screws a wood dowel on the threaded end of the boom. On the outer end of the dowel he hangs a flat piece of aluminum (three inches on a side) that is either glued to the wood or held in place by a brass wood screw. The idea is to have the aluminum hang down between the two facing magnets so that when the boom swings, the aluminum swings through their magnetic field. The eddy currents created in the aluminum set up magnetic fields of their own that oppose the motion and thereby damp out the swing of the boom.

The entire seismograph is covered to protect it from air currents that arise either from convection in the room or from the air-conditioning system. The top of the covering should be glass or clear plastic so that you can inspect the apparatus.

The voltage on the pickup coil is read through an amplifier. You could buy a commercial direct-current amplifier or build the one shown in the illustration above. Lehman says that his circuit is normally operated to multiply the signal by a factor of 10. The circuit is mounted on fiberboard and placed inside a small box. The 60-hertz noise from the wiring of the room is filtered out in the input stage of the amplifier, as is any mechanical noise in the frequency range of 10 hertz or more. The seismic signals fall in the range from one to 1/15 hertz and therefore pass through the filter. The first 741 gate acts as a voltage follower for the voltage on the pickup coil. The second 741 gate acts as a hundredfold amplifier. Lehman says that two six-volt batteries of the lantern type can power the amplifier for several months.

The output of the amplifier has been designed to match the 10-millivolt level of Lehman's strip-chart recorder (a Heath IR 18M model). He says that the time base of the paper flow through that model is so good there was no need for an external clock system to put time marks on the paper. Felt-tip pens were found satisfactory for the inking. The ink cartridges could be reused by employing a medicine dropper to refill them with ink. One day's run typically consumed about two-thirds of a cartridge.

If you want to buy another type of graphic recorder, Lehman emphasizes the importance of finding one with independent controls for the paper flow and the coordination of the pen and the paper. The penholder should be universally jointed, and the chart should be easily removable for rewinding and other adjustments.

Lehman obtained standard 10-inch chart paper for his machine, running it through at the rate of a foot (30 centimeters) per hour so that the time spacings on the paper were a convenient one inch per five minutes (.5 centimeter per minute). By running the paper through on both sides Lehman cuts the cost of paper in half. He estimates that a full year of daily recording needs only $10 worth of paper.

To set up the seismograph detach the boom and its wire support and then level the base by placing metal washers or shims as needed under the legs. A bubble level is helpful in this operation. The nozzle opening on the upper crosspiece should be about a centimeter off the vertical from the place where the knife-edge will eventually rest on the lower crosspiece. To achieve this arrangement you may have to change the lower crosspiece or put shims under the two rear legs on the base.

Add the boom and attach its wire support to the turnbuckle so that the boom is horizontal and the knife-edge is vertical on its plate. Give the boom a slight push to check its period of swing, which should be in the range from four to eight seconds. Also check its rest position; it should be horizontal and perpendicular to the metal plate on which the knife-edge rests. At the other end of the boom you may want to add stops about a quarter of an inch from the rest position of the pendulum on each side to prevent the pendulum from swinging wildly as you make adjustments to the seismograph. With the stops in place the rest position should, of course, be midway between them.

Leave the seismograph idle for several days so that the framework can adjust to the stress and the wire can stretch. Then adjust the boom so that its natural swing period is from 12 to 18 seconds. To increase the swing period to this value put a bushing either under the metal plate on which the knife-edge rests or under the front leg of the base. With such a step you reduce the distance by which the nozzle is off the vertical from the knife-edge point. If you were to tilt the assembly back far enough for the nozzle to be exactly on the vertical from the knife-edge, the boom would no longer swing like a normal pendulum because there would be no restoring force tending to return it to its rest position. With the assembly in proper alignment the nozzle point is off the vertical from the knife-edge, and the boom constantly tends to return to its rest position because of its own weight. The more the

SCIENCE/SCOPE

<u>International electronic message service will be tested</u> by the U.S. Postal Service through a contract with the Communications Satellite Corporation (COMSAT). The system will provide overnight service -- and in some instances same-day service -- in linking the U.S. with London, Paris, Frankfurt, Amsterdam, Brussels, and Buenos Aires. A Hughes-built Intelsat IV-A satellite will be used in a one-month feasibility test to transmit facsimile mail solicited from banks and large corporations. Messages originating in New York or Washington, D.C., will be transmitted via land lines to an earth station at Etam, West Virginia. There they will be beamed to the satellite, back down to the overseas earth station, then by land to an INTELPOST (International Electronic Post) service center.

<u>Old pastel masterpieces may retain their beauty</u> with help from modern electronics, specifically from an electrostatic plate that prevents the chalky pigment from flaking. The plate consists of positive and negative circuits laminated in plastic. When placed behind a drawing and given a 10,000-volt potential, it generates a field of static electricity that holds the pastel particles to the surface of the paper. The unit and painting are enclosed in a dust-free frame to prevent contamination. Hughes built the plate at the request of the Conservation Center of the Los Angeles County Museum of Art. The Armand Hammer Foundation provided funds.

<u>All three types of U.S.-built wide-body jetliners</u> will carry Hughes passenger entertainment and service equipment now that Pan American World Airways has ordered the system for its new Lockheed L-1011s. The system, transmitting multiplexed signals at rates up to 5 million bits per second, provides stereo music and movie sound tracks, plus reading light and flight attendant call service. An earlier system is standard equipment on all McDonnell Douglas DC-10s. Also, 14 operators of Boeing 747s, citing high reliability and low operating costs, have chosen the system as either original or replacement equipment.

<u>Hughes is seeking electronics engineers</u> for research, design, development, and production of surface-based and undersea systems. These include modern phased-array radars; advanced digital communications; microwave systems; computer systems; liquid crystal, CRT, TV-type displays; advanced signal and data processing equipment; A/D and D/A converters; and sonar arrays and processing equipment. If you qualify, send your resume to Hughes Aircraft Company, Ground Systems Group, Professional Employment, Dept. SE, 1907 W. Malvern, Fullerton, CA 92634.

<u>Major developments toward an optical filter that can be tuned electronically</u> to specific wavelengths of light have been reported by Hughes scientists. The device is tuned by a microprocessor that varies the electric field distribution onto an electro-optic crystal. One filter with a lithium-tantalate crystal has been operated across the visible light spectrum from deep blue to deep red. Another has been tuned into infrared wavelenths. The device promises to find important uses in pollution monitoring, multispectral imaging, and monitoring color consistency in commercial products, including paints and dyed fabrics.

Creating a new world with electronics

HUGHES

HUGHES AIRCRAFT COMPANY
CULVER CITY, CALIFORNIA 90230

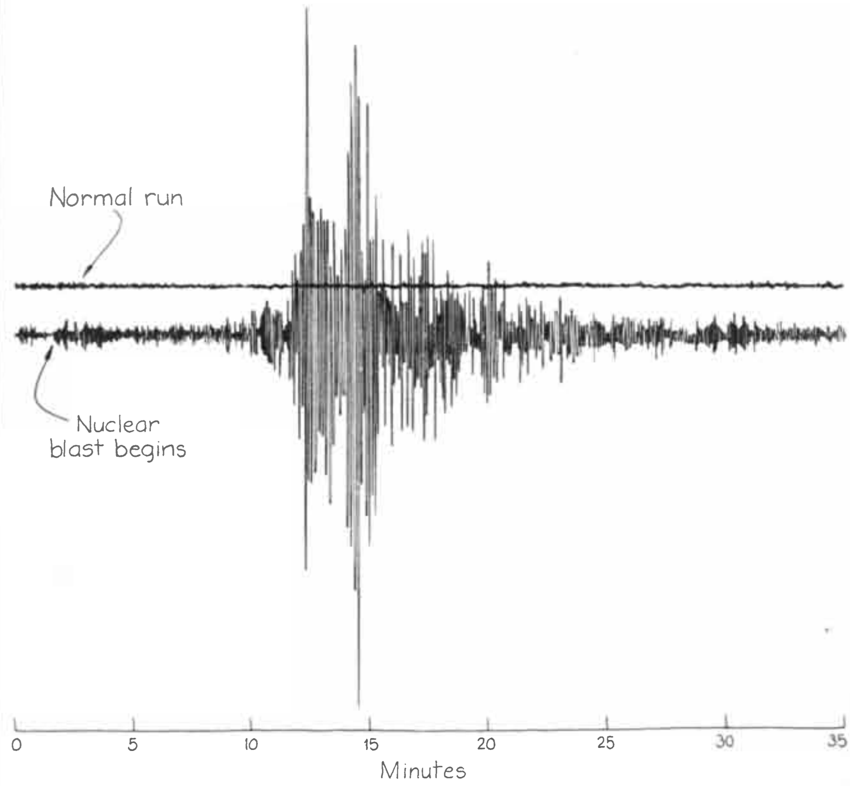
nozzle point is off the vertical from the knife-edge, the greater the restoring force is and the smaller the natural swing period of the boom will be. You should adjust the alignment of the nozzle and the knife-edge so that the natural period is from 12 to 18 seconds. The boom will thus be sensitive to seismic waves of slightly shorter period.

The next task is to adjust the damping. First check it by placing the damping magnets in their proper position around the vertical aluminum plate and then swinging the end of the boom over a distance of about a centimeter. When you release the end, it will swing past the rest position. With proper damping the overshoot should be about two millimeters. You can adjust the amount of damping by adjusting the gap between the magnets. Closing the gap increases the magnetic field and the damping. You could also adjust the damping by varying the natural swing period of the boom. For example, a longer period would mean less restoring force, so that (with the same magnetic field) the damping would increase.

These steps constitute a rough tuning of the damping. Now position the pickup coil with respect to the magnet attached to the boom so that they are both at the proper height from the floor and so that the coil is about a quarter of an inch inside the gap of the magnet when the boom is in its rest position. Then cover the entire assembly and adjust the gain on the amplifier until you can see background noise recorded on the strip-chart recorder. Lehman says that if you then approach the seismograph, the tilting of the floor by your weight should give you a readout on the strip-chart recorder with an amplitude of about three inches.

To check the damping walk up to the seismograph, wait a few seconds and then walk away. The paper on the recorder should show a pair of peaks for your advance and a pair for your retreat. The first peak in each pair is due to the tilting of the floor as you advance. The second peak, which is smaller, represents the overshoot of the boom past its rest position. If the boom is properly damped, the ratio of the two peaks should be between 6 : 1 and 10 : 1. If it is not, you will have to adjust the damping. Lehman points out that this may take some patience until you finally get the damping just right, but once all the adjustments are made the instrument will not need further attention for weeks or even months.

The best place for your seismograph would be in a room undisturbed by either people or thermal changes. Do not put the instrument where sunlight will fall on it or the floor, because the resulting variations in the stress of the floor will show up in your readings. Even some unexposed floors may show thermal variations because of sunlight striking the building. Lehman has had good



Lehman's record of a nuclear test

success with one of his instruments on a gravel floor in an unfinished part of his campus. Since the seismograph is most sensitive to waves passing perpendicularly to the boom and since most earthquakes occur on an east-west azimuth, you should set up your seismograph with the boom running north and south.

When you first begin to record, you will notice a constant background noise in the readings. Much of it is from microseisms, which are small oscillations of the earth that are still largely unexplained. Investigators have long tried to link microseisms to natural phenomena such as the beating of surf on a cliff-lined shore or the presence of a cyclone over the ocean. Although at times the connection between microseisms and such a cause has seemed clear, at other times the microseisms have seemed totally unrelated to the proposed cause.

In some of the data Lehman has sent me one can identify several other kinds of noise in the records. One day a brief set of spikes revealed a blast at a local quarry or construction site. A winter cold front appeared on another record as a set of continuous but random spikes. On at least two occasions Lehman has been able to display the effects of hurricanes in the Atlantic.

You will probably be less interested in these kinds of waves in your recordings than in seismic waves. Such waves are of four main types (and several other less important types and subclassifications I shall not mention). Two types propagate through the body of the earth but in different ways and at different speeds. One of them is the transverse wave designated *S*. The material participating in the wave motion oscillates perpendicularly to the direction of travel of the wave. The term transverse is meant to convey that sense of oscillation. Most of the waves with which you are familiar are transverse waves, for example waves on a guitar string or on the surface of a body of water.

The other type of wave that travels through the earth is the longitudinal wave designated *P*. Here the material participating in the wave motion oscillates back and forth in the direction in which the wave is traveling. The term longitudinal is meant to convey the fact that the oscillations are parallel to the wave direction.

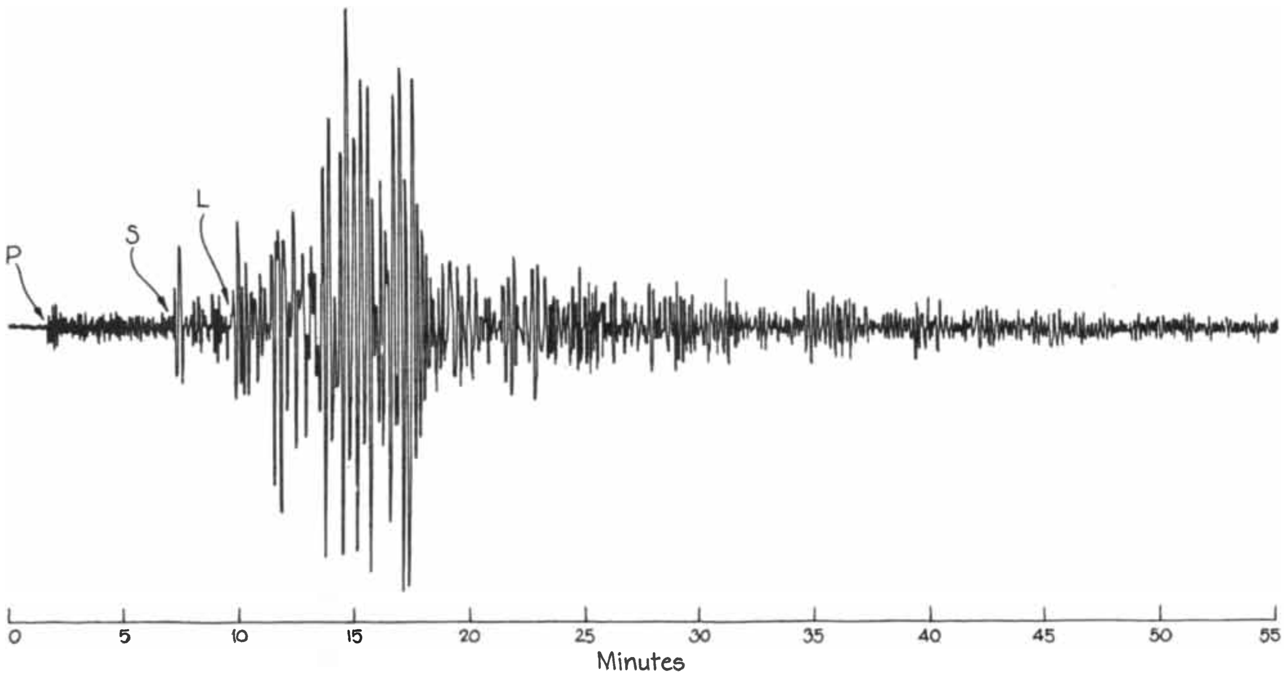
The remaining two general types of seismic waves propagate along the surface of the earth. With one type, the Love wave (*L*), the material participating in the wave motion oscillates transversely to the direction of travel of the wave and horizontally with respect to the earth's surface. With the other type, called Rayleigh waves (*R*), the material moves in elliptical paths perpendicular to the surface and aligned along the direction of travel of the wave.

If all these waves arrived at your seismograph at the same time, you would not be able to distinguish them, but they arrive at different times because they either take different routes to your instrument or travel at different speeds. The first to arrive are the *P* waves because they both take a direct route through the earth and travel faster than the *S* waves. Next come the *S* waves, followed by the two types of surface waves, which are delayed by traveling over a longer path along the surface of the earth.

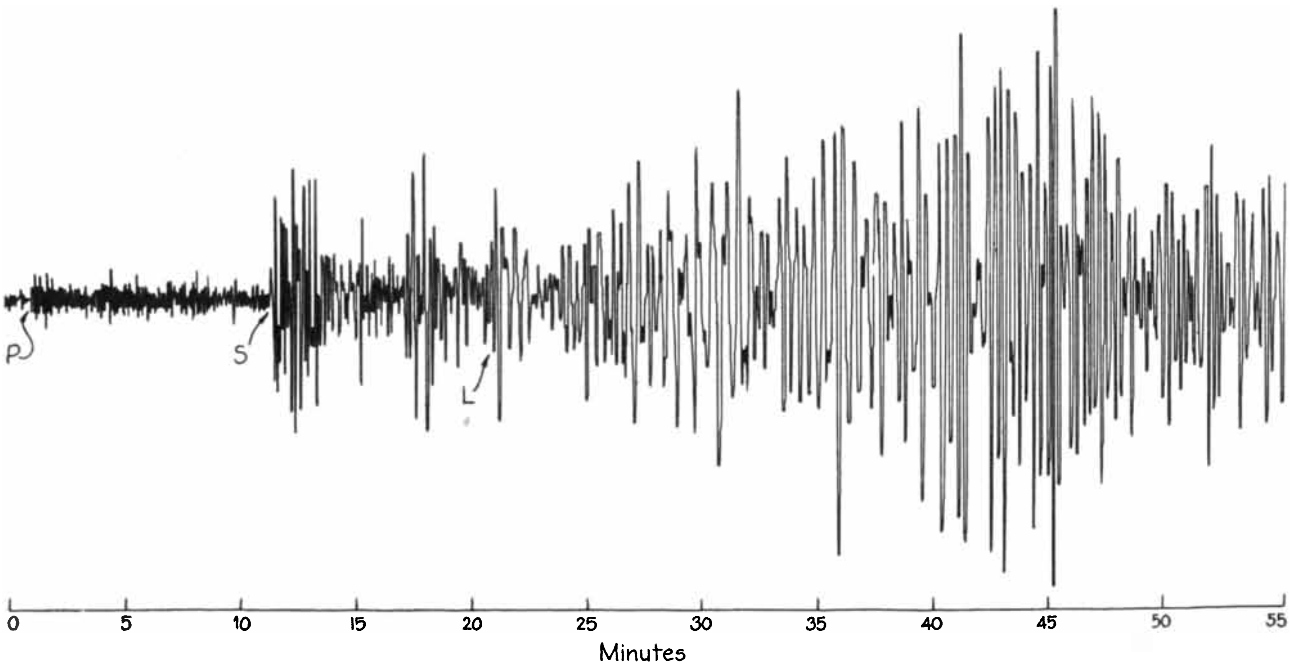
The amount of time between the arrival of the *P* and the *S* waves can be used to determine the distance between you and the center of the earthquake. Tables relating the difference in arrival times and the corresponding distance are available in seismograph references such as the book by Markus Bath cited in the bibliography for this issue [*page 162*]. The distance is often quoted in terms of the angular distance between the focus of the earthquake and the seismograph as measured from the center of the earth. For example, if the difference in arrival times is about seven minutes 11 seconds, the focus is about 50 degrees away from you.

With a single instrument you can only guess at the direction. Rough determination of the direction can be made with two instruments oriented perpendicularly to each other (one seismograph on a north-south axis, the other east-west). Precise determinations of the distance, direction and depth of the focus are obtained only when several widely separated seismograph stations record the event. The difference in arrival times of the waves at the stations provides the clue to where the focus is.

Several periodicals can help you to



Recording of an earthquake in the Galápagos on March 29, 1976



The record resulting from an earthquake in Turkey on November 24, 1976 determine the locations of major earthquakes. *Preliminary Determination of Epicenters* and *Earthquake Information Bulletin* are both available from the Superintendent of Documents, Government Printing Office, Washington, D.C. 20402. *Earth Science Event Reports* can be obtained from the Center for Short-lived Phenomena, 185 Alewife Brook Parkway, Cambridge, Mass. 02138. Another valuable reference is *Principles Underlying the Interpretation of Seismographs,* Special Publication No. 254 (revised in 1966), which can be obtained from the U.S. Coast and Geodetic Survey, Rockville, Md. 20852. Other references are in the bibliography for this issue. You should also have a globe in order to measure the distance to earthquake epicenters. Coordinated Universal Time (UTC) can be got by radio on stations WWV and WWVH.

Background noise and earthquakes are not the only things you may record. Underground nuclear explosions also send seismic waves through the earth and along its surface. Lehman has recorded a number of such explosions that were set off in Nevada, seven in the first three months of 1976 alone.

Lehman sent me several examples of earthquake data, two of which I find particularly striking. One is a record of the earthquake that occurred in the Galápagos on March 29, 1976, at 0539 : 33 UTC. The quake had a shallow focus, was of magnitude 6.7 and apparently resulted from slippage along the boundary between the Cocos and the Nazca tectonic plates. Six minutes 57 seconds after the quake began the *P* waves reached Lehman's recorder. They had periods of from one second to three seconds and were recorded for more than five minutes. It took the *S* waves 12 minutes 28 seconds to arrive, since they had longer periods and greater amplitudes than the *P* waves. The surface waves took about 15 minutes to arrive and had periods of from 12 to 15 seconds. Oscillations continued for about an hour. The difference in the arrival times of the *P* and *S* waves (five minutes 31 seconds) indicated that the epicenter was about 2,440 miles (3,930 kilometers) from Lehman's apparatus.

In another example Lehman recorded the waves from a devastating earthquake in Turkey on November 24, 1976, at 1230 UTC. It was the worst quake in the Anatolian fault zone since 1939. Lehman recorded the first *P* wave at 1235 : 10 UTC and the first *S* wave at 1245 : 30. Activity was still visible an hour and a half after the first *P* wave arrived.

Lehman is prepared to answer questions on the construction or operation of the seismograph. He is particularly interested in exchanging seismic records. Lehman's address is Physics Department, James Madison University, Harrisonburg, Va. 22807.

BIBLIOGRAPHY

Readers interested in further explanation of the subjects covered by the articles in this issue may find the following lists of publications helpful.

MATHEMATICAL GAMES

GÖDEL'S PROOF. Ernest Nagel and James R. Newman. New York University Press, 1958.

THE BACH READER: A LIFE OF JOHANN SEBASTIAN BACH IN LETTERS AND DOCUMENTS. Edited by Hans T. David and Arthur Mendel. W. W. Norton & Co., Inc., 1966.

SEMANTIC INFORMATION PROCESSING. Edited by Marvin L. Minsky. The MIT Press, 1968.

THE PROMPT AND DELAYED EFFECTS OF NUCLEAR WAR

LONG-TERM WORLDWIDE EFFECTS OF MULTIPLE NUCLEAR-WEAPONS DETONATIONS. National Academy of Sciences, Washington, D.C., 1975.

THE EFFECTS OF NUCLEAR WEAPONS. Edited by Samuel Glasstone and Philip Dolan. U.S. Departments of Defense and Energy, Government Printing Office, 1977.

SOVIET CIVIL DEFENSE. Report issued by the Director of Central Intelligence, Bureau of Public Affairs, Office of Public Communications; July, 1978.

THE EPSTEIN-BARR VIRUS

THE ONCOGENICITY OF EPSTEIN-BARR VIRUS. G. Miller in *The Journal of Infectious Diseases,* Vol. 130, No. 2, pages 187–205; August, 1974.

THE EPSTEIN-BARR VIRUS AND NEOPLASIA. George Klein in *The New England Journal of Medicine,* Vol. 293, No. 26, pages 1353–1357; December 25, 1975.

RECENT PROGRESS IN EPSTEIN-BARR VIRUS RESEARCH. M. A. Epstein and B. G. Achong in *Annual Review of Microbiology,* Vol. 31, pages 421–445; 1977.

THE IMMUNOLOGICAL APPROACH TO STUDY OF POSSIBLY VIRUS-INDUCED HUMAN MALIGNANCIES USING THE EPSTEIN-BARR VIRUS AS EXAMPLE. W. Henle and G. Henle in *Progress in Experimental Tumor Research, Vol. 21: Viruses and Human Cancer,* edited by Yohei Ito. S. Karger, 1978.

THE CENTRAL PARSEC OF THE GALAXY

THE GALACTIC CENTER. J. H. Oort in *Annual Review of Astronomy and Astrophysics,* Vol. 15, pages 295–362; 1977.

INFRARED OBSERVATIONS OF THE GALACTIC CENTER, I: NATURE OF THE COMPACT SOURCES. E. E. Becklin, K. Matthews, G. Neugebauer and S. P. Willner in *The Astrophysical Journal,* Vol. 219, No. 1, Part 1, pages 121–128; January 1, 1978.

THE INFRARED EMISSION OF THE GALACTIC CENTER. G. H. Rieke, C. M. Telesco and D. A. Harper in *The Astrophysical Journal,* Vol. 220, No. 2, Part 1, pages 556–567; March 1, 1978.

THE ORIGINS OF THE FIRST POWERED, MAN-CARRYING AIRPLANE

THE WRIGHT BROTHERS. Fred C. Kelly. Harcourt Brace and Company, 1943.

THE PAPERS OF WILBUR AND ORVILLE WRIGHT. Edited by Marvin W. McFarland. McGraw-Hill Book Company, Inc., 1953.

SIR GEORGE CAYLEY, INVENTOR OF THE AEROPLANE. John L. Pritchard. Horizon Press, 1962.

THE INVENTION OF THE AEROPLANE 1799–1909. Charles H. Gibbs-Smith. Faber and Faber, 1966.

WILBUR AND ORVILLE WRIGHT: A CHRONOLOGY COMMEMORATING THE HUNDREDTH ANNIVERSARY OF THE BIRTH OF ORVILLE WRIGHT, AUGUST 19, 1871. Arthur G. Renstrom. United States Library of Congress, Science and Technology Division, 1975.

COMPARTMENTS IN ANIMAL DEVELOPMENT

GENETIC CONTROL OF WING DISC DEVELOPMENT IN *DROSOPHILA*. A. García-Bellido in *Symposium on Cell Patterning: Ciba Foundation Symposium 29.* Associated Scientific Publishers, 1975.

COMPARTMENTS AND POLYCLONES IN INSECT DEVELOPMENT. F. H. C. Crick and P. A. Lawrence in *Science,* Vol. 189, No. 4200, pages 340–347; August 1, 1975.

COMPARTMENTS IN THE WING OF *DROSOPHILA:* A STUDY OF THE *ENGRAILED* GENE. P. A. Lawrence and G. Morata in *Developmental Biology,* Vol. 50, No. 2, pages 321–337; June, 1976.

HOMOEOTIC GENES, COMPARTMENTS AND CELL DETERMINATION IN *DROSOPHILA*. G. Morata and Peter A. Lawrence in *Nature,* Vol. 265, No. 5591, pages 211–216; January 20, 1977.

HOMOEOTIC AND ATAVIC MUTATIONS IN INSECTS. A. García-Bellido in *American Zoologist,* Vol. 17, No. 3, pages 613–629; Summer, 1977.

THE BAG MODEL OF QUARK CONFINEMENT

MASSES AND OTHER PARAMETERS OF THE LIGHT HADRONS. T. DeGrand, R. L. Jaffe, K. Johnson and J. Kiskis in *Physical Review D,* Vol. 12, No. 7, pages 2060–2076; October 1, 1975.

STRINGLIKE SOLUTIONS OF THE BAG MODEL. K. Johnson and C. B. Thorn in *Physical Review D,* Vol. 13, No. 7, pages 1934–1939; April 1, 1976.

THE CONFINEMENT OF QUARKS. Yoichiro Nambu in *Scientific American,* Vol. 235, No. 5, pages 48–60; November, 1976.

QUARK CONFINEMENT. R. L. Jaffe in *Nature,* Vol. 268, No. 5617, pages 201–208; July 21, 1977.

THE ANIMALS OF THE BURGESS SHALE

THE ENIGMATIC ANIMAL *OPABINIA REGALIS,* MIDDLE CAMBRIAN, BURGESS SHALE, BRITISH COLUMBIA. H. B. Whittington in *Philosophical Transactions of the Royal Society of London,* Series B, Vol. 271, pages 1–43; June 26, 1975.

FOSSIL PRIAPULID WORMS. Simon Conway Morris in *Special Papers in Paleontology, No. 20.* The Paleontological Association of London; December, 1977.

THE BURGESS SHALE (MIDDLE CAMBRIAN) FAUNA. Simon Conway Morris in *Annual Review of Ecology and Systematics,* in press.

THE VISUAL PERCEPTION OF MOTION IN DEPTH

CONTRIBUTIONS TO THE PHYSIOLOGY OF VISION: ON SOME REMARKABLE, AND HITHERTO UNOBSERVED, PHENOMENA OF BINOCULAR VISION. Charles Wheatstone in *Philosophical Transactions of the Royal Society of London,* Series A, Vol. 128, pages 371–394, 1838; Vol. 142, pages 1–18, 1852.

THE PERCEPTION OF THE VISUAL WORLD. James J. Gibson. Houghton Mifflin Company, 1950.

ILLUSORY MOTION IN DEPTH: AFTEREFFECT OF ADAPTATION TO CHANGING SIZE. D. Regan and K. I. Beverley in *Vision Research,* Vol. 18, No. 2, pages 209–212; 1978.

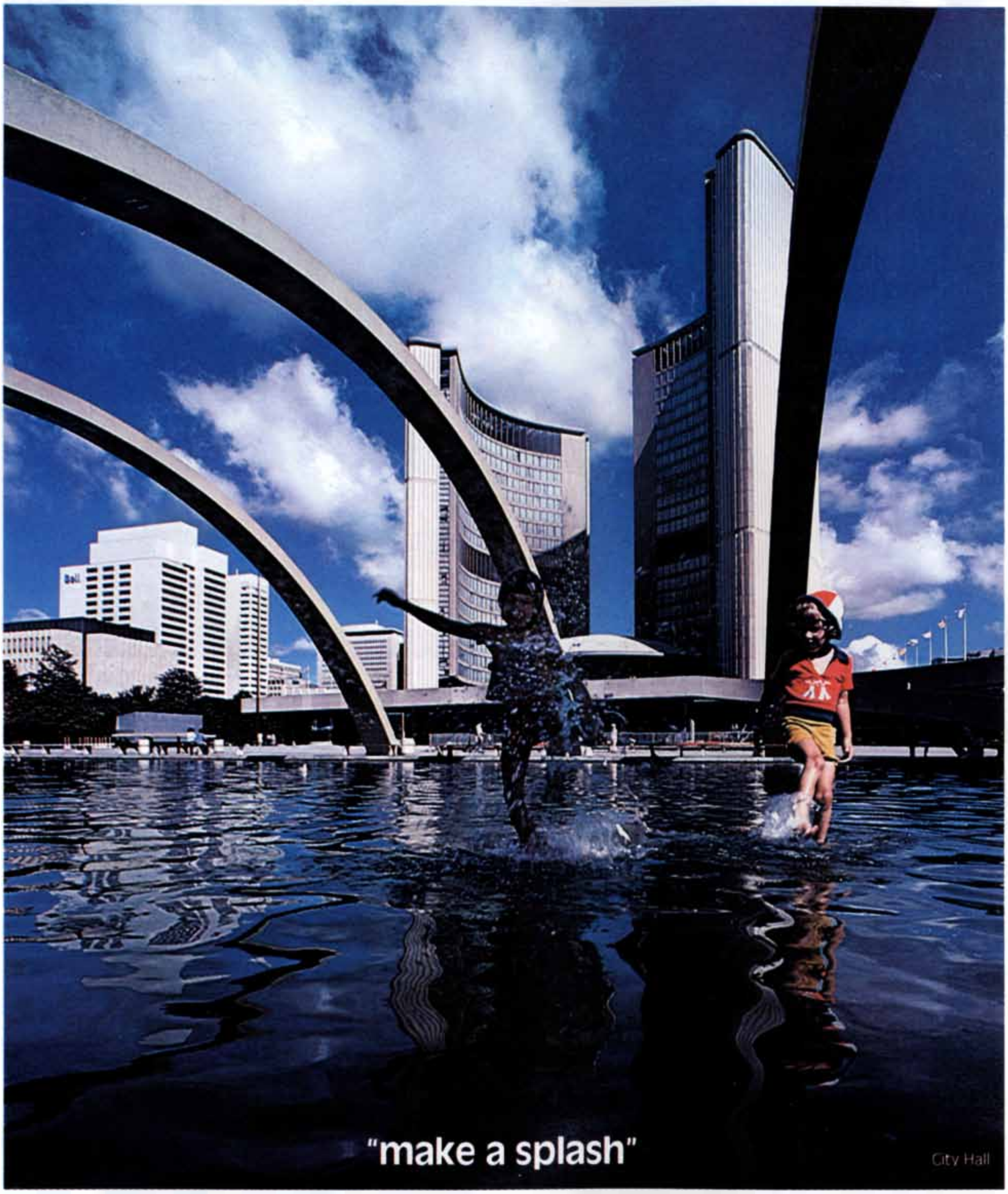
NEURONES IN CAT PARASTRIATE CORTEX SENSITIVE TO THE DIRECTION OF MOTION IN THREE-DIMENSIONAL SPACE. Max Cynader and D. Regan in *Journal of Physiology,* Vol. 274, pages 548–569; 1978.

THE AMATEUR SCIENTIST

ELEMENTARY SEISMOLOGY. Charles F. Richter. W. H. Freeman and Co., 1958.

EARTHQUAKE INTERPRETATIONS. Ruth S. Simon. Colorado School of Mines, 1969.

AN INTRODUCTION TO SEISMOLOGY. Markus Bath. John Wiley & Sons, Inc., 1973.


"make a splash"
City Hall

Lockheed knows how.

Shielding the space shuttle from fiery temperatures.

When NASA's space shuttle streaks through a 2,300°F reentry into Earth's atmosphere, the occupants won't feel the heat.

That's because Lockheed-developed insulation will form a shield of some 34,000 small 'tiles' covering about 70% of the spacecraft. The tiles shed heat so efficiently that a piece still glowing white-hot can be held safely with bare fingers.

Each tile is minutely matched to exact points on the shuttle's surface. And fitting them precisely is like assembling the world's biggest three-dimensional jigsaw puzzle.

Most important: the tiles will insulate the shuttle against not just one, but about 100 returns through the abrasive atmosphere.

Looking at the origin of the universe.

In 1983, a 43-foot-long telescope system will be slipped into orbit 300 miles above Earth. And man may then look back 14 billion years to where our universe perhaps began.

The NASA/Lockheed Space Telescope will detect objects 50 times fainter and 7 times deeper into space than those ever before seen. It will lock onto stars, galaxies, and other space phenomena with absolute accuracy for as long as 30 to 40 hours. And it will perform ultraviolet and infrared measurements impossible from Earth.

Potential rewards are huge. Scientists may discover new worlds and new energy sources. They may even find ways to protect Earth's delicate environment in a new, uncharted age.

Scouting the sun's uproar.

Tucked into NASA's OSO-8 Orbiting Solar Laboratory, a Lockheed-developed Mapping X-Ray Heliometer has successfully finished three years of amassing vital data on the sun's ever-varying X-ray radiation.

work for man.

The NASA Space Shuttle. Lockheed serves it with a wide range of technical support activities in addition to the heat insulation described below.

The heliometer was part of the most complete scientific instrument payload ever sent up to study solar and cosmic X-ray sources. And though planned for only one year's work, the OSO-8 package operated in its 350-mile-high orbit for better than three years.

The heliometer experiment again clearly demonstrated Lockheed's unique capability in independent, basic research. And the data will probably unveil dramatic solar facts.

They should set guidelines for better predicting of solar flares and yield new knowledge of exactly how the sun's X-ray emissions affect Earth's weather and climates.

Recorders that talk from other planets.

We've now put 'eyes' on Mars, viewed Venus and Jupiter close up, and will get intimate looks at Saturn in late 1980.

Special Lockheed recorders, each no bigger or heavier than a gallon bottle of water, help scientists see the outer planets at close range. Locked within a space vehicle, each recorder can store and transmit more than a half-billion bits of information. Those signals are then turned by Earth-station computers into sharp photos and precise scientific data.

Lockheed recorders have flown on all Apollo moon missions, five Mariner space probes, the flights to Mars, Venus, and Jupiter, and are now headed toward Saturn and beyond.

And if the search that far away reveals no life, there's still the rest of the universe.

The unmatched record of achievement.

Nothing underscores Lockheed's leading place in space technology better than the record of the Agena spacecraft.

The versatile Agena, veteran of more than 300 flights, can count more successful launches than those made by any other spacecraft, and it has been used on nearly half of all United States space missions.

In making space work for man, Agena is one more proof that Lockheed knows how.

Lockheed


The GLENLIVET
AGED 12 YEARS
Discover more.
Let your mature taste lead you from the finest premium Scotch on up to the most expensive 12-year-old Scotch in the world. The Glenlivet. The ultimate in Scotch.
Most premium Scotch is blended and depends on several whiskies for taste and smoothness.
Yet the unblended character of The Glenlivet is noble enough to stand alone. In this distinctive Scotch whisky you'll experience an exceptional smoothness and full-bodied richness, unmatched in all other premium Scotch.
Try the taste beyond premium Scotch.
The Glenlivet
Unchanged since 1824.
The GLENLIVET
12 YEARS OLD
Unblended all malt Scotch Whisky
Distilled and Bottled in Scotland
GEORGE & J. G. SMITH LIMITED
THE TASTE BEYOND PREMIUM SCOTCH