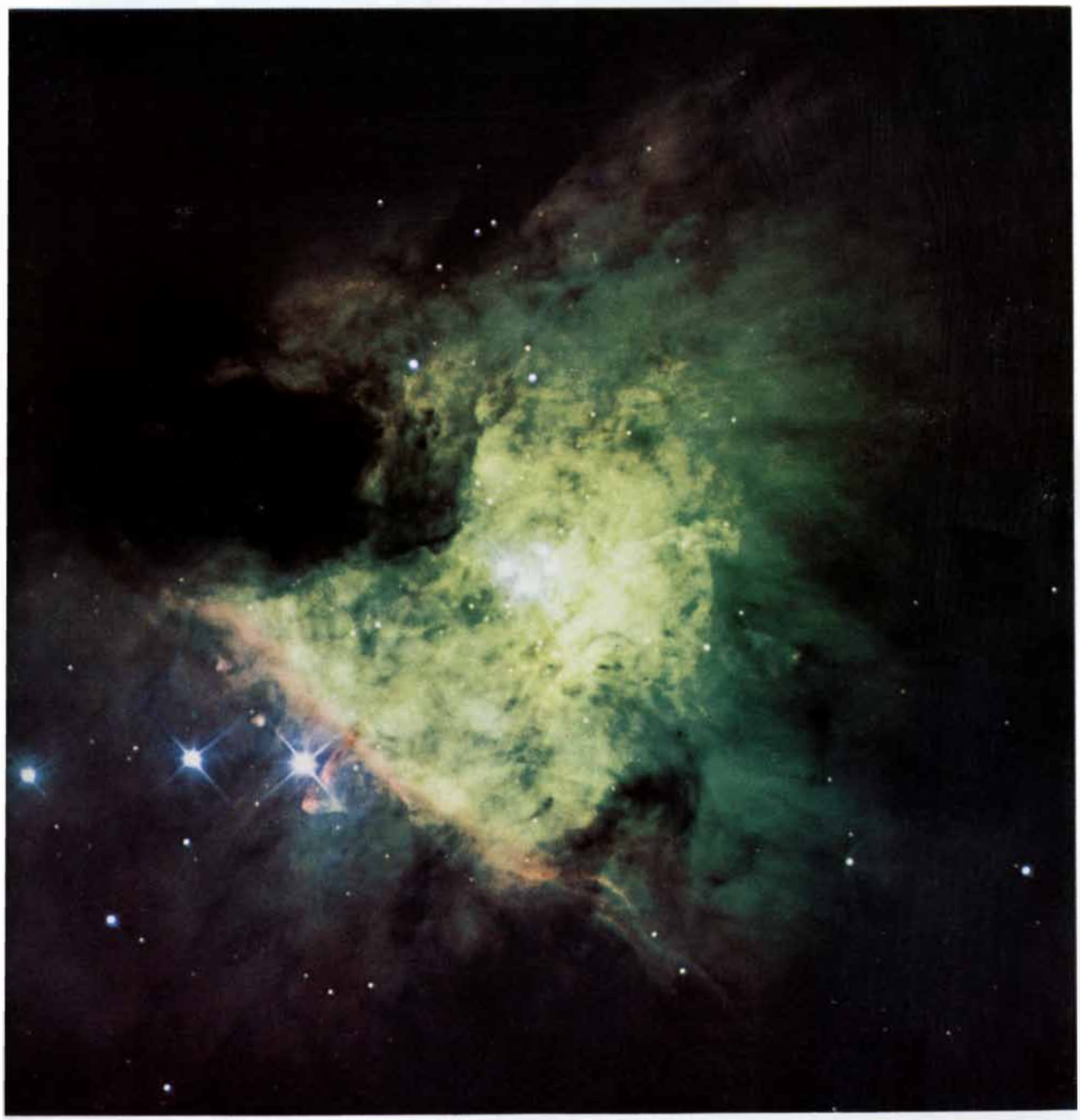
# SCIENTIFIC AMERICAN



EMISSION NEBULAS

*$1.00*

*October 1974*

# Born Free.

Cars are like people.

Some seem filled with inhibitions, concessions, compromises.

Then there's the other kind. The kind that exudes charisma. Confidence. Free spirit.

Buick's new Skyhawk, as its name suggests, is of the latter category.

It sits there, rakish and purposeful, square-headlamped and aerodynamic, tautly suspended on its radial-shod wheels. Like any confident design, Skyhawk does not beg you to love it. It does not over-promise or boast. It simply *is.* And its presence inspires affection.

Yet for all its machismo, for all its tendency to look like some exotic European prototype, Skyhawk is as untemperamental as an old friend.

Its engine is a Buick-built, 3.8 liter V-6 that is designed to be both responsive and frugal. High-Energy Ignition (no points or ignition condenser to replace) is standard, as is a 4-speed transmission, plus front and rear stabilizer bars in the suspension.

And to further demonstrate its utilitarian talents, Skyhawk is a hatchback. Which opens to reveal seating for four, a generous luggage compartment, an expanse of natty houndstooth upholstery and a big tachometer. All standard, of course.

There's more, but space limits. Suffice to say if you've stayed with us this long, you're probably the kind of person Skyhawk was designed to attract.

In which case the two of you really ought to meet each other.

At your Buick dealer.

BUICK

***Dedicated to the Free Spirit in just about everyone.***

GM


# "Gold be damned. I'm putting my money in the mattress."

**More good things have been decided over Grand Marnier than any other drink in the world.**

For free recipe booklet, write Carillon Importers, Ltd., 745 Fifth Ave., NYC 10022. Product of France, made with fine cognac brandy. 80 proof.

Established 1845 **SCIENTIFIC AMERICAN** October 1974 Volume 231 Number 4

ARTICLES

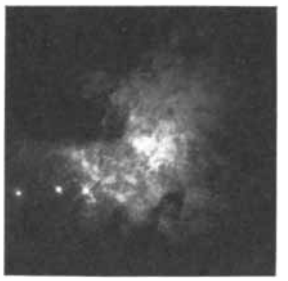


DEPARTMENTS





PUBLISHED MONTHLY BY SCIENTIFIC AMERICAN, INC., 415 MADISON AVENUE, NEW YORK, N.Y. 10017.  SECOND-CLASS POSTAGE PAID AT NEW YORK, N.Y., AND AT ADDITIONAL MAILING OFFICES. AUTHORIZED AS SECOND-CLASS MAIL BY THE POST OFFICE DEPARTMENT, OTTAWA, CANADA, AND FOR PAYMENT OF POSTAGE IN CASH. SUBSCRIPTION RATE: $12 PER YEAR, U.S., POSSESSIONS AND CANADA; $15 PER YEAR, ALL OTHER COUNTRIES. EUROPEAN SUBSCRIPTION COPIES MAILED C/O PUBLICATIONS DISTRIBUTION SERVICE, ST. KATELIJNEVEST 14, B2000 ANTWERP, BELGIË.



## THE COVER

The photograph on the cover, made with the 120-inch reflecting telescope at the Lick Observatory, shows the central region of the Great Nebula in Orion. The photograph was made by the dye-transfer process, in which three black-and-white negatives made through color filters are combined to produce a color print. Such a print reproduces colors more accurately than standard color emulsions do, and therefore better distinguishes the wavelengths emitted by the various kinds of atoms that make up the nebula (see "The Structure of Emission Nebulas," by Joseph S. Miller, page 34). The four bright stars in the center of the nebula are known as the Trapezium. The luminosity of the nebula is mainly due to the brightest of the four, the energetic photons of which ionizes the atoms of the tenuous gas around it. The green glow is radiated by doubly ionized oxygen atoms, that is, atoms from which two electrons have been removed. The three bright stars to the southeast (*lower left*) do not appear to contribute to the ionization.

## THE ILLUSTRATIONS

Cover photograph by Joseph S. Miller, Lick Observatory

| Page | Source |
|---|---|
| 22–23 | EROS Data Center, U.S. Geological Survey, Department of the Interior |
| 24–33 | Graphic Presentation Services, Inc. |
| 35–37 | Lick Observatory |
| 38–42 | Allen Beechel |
| 44 | Ellen R. Dirksen, University of California at San Francisco (*top*); Birgit Satir, University of California at Berkeley (*bottom*) |
| 45 | George V. Kelvin |
| 46 | Peter Satir, University of California at Berkeley (*top*); George V. Kelvin (*bottom*) |
| 47 | Birgit Satir, University of California at Berkeley (*top left and center*); Peter Satir, University of California at Berkeley (*top right*); George V. Kelvin (*bottom*) |
| 48 | George V. Kelvin (*top*); Peter Satir, University of California at Berkeley (*bottom*) |
| 49 | Peter Satir, University of California at Berkeley (*top*); George V. Kelvin (*bottom*) |
| 50 | Peter Satir, University of California at Berkeley (*top*); Fred D. Warner, Syracuse University (*bottom*) |
| 51 | Keith E. Summers, National Heart and Lung Institute (*top*); George V. Kelvin (*second from top*); Charles J. Brokaw, California Institute of Technology (*bottom*) |
| 52 | Peter Satir (*left*), George V. Kelvin (*right*) |
| 64–80 | Lorelle M. Raboni |
| 83 | Sol Mednick |
| 84–85 | Ralph Morse |
| 86–87 | Ilil Arbel |
| 88 | Ralph Morse |
| 90 | Ilil Arbel |
| 92–97 | Tom Prentiss |
| 101–106 | Alan D. Iselin |
| 110 | Julian and Eunice Whittlesey |
| 112 | Michael H. Jameson, University of Pennsylvania |
| 113–117 | Eric Mose |
| 118 | Michael H. Jameson, University of Pennsylvania |
| 120–125 | Ilil Arbel |
| 127–130 | Jerome Kuhl |

# LETTERS

Sirs:

In Judge Bazelon's article ["Psychiatrists and the Adversary Process," by David L. Bazelon; SCIENTIFIC AMERICAN, June] he fails to point out that the adequacy of the adversary process to achieve justice is controversial within the legal profession, and that the verdict often goes to the most skillful lawyer rather than the party with the just cause. Nor does he say that the acceptable methods of this process include "distortion (if not misrepresentation) and the manipulation of bias [which] defeat[s] the possibility of finding truth or reality." ("Lawyers, Truth, and the Zero-Sum Game." James Marshall in *Notre Dame Lawyer*, Volume 47, pages 919–926; April, 1972.) Nor that "the objective of each lawyer is to win his client's case at any cost.... The principles which underlie what is being fought about become relatively unimportant. What is of supreme importance is to win." (*Trial by Combat in American Courts.* David Dressler. Yale University Press, 1964.) However, Judge Bazelon demonstrates this in his article. The mere existence of his own Appellate Court affirms the legal position that "the definition of [due process] and the appraisal of its adequacy are matters for [judicial] determination." Yet he misrepresents the comparable position of the American Psychiatric Association as "an adamant statement of what I must call professional mystique."

Another distortion is his disingenuous characterization of Rouse as "a patient confined without his consent [who] sued for release or for adequate treatment of his disability." This half-truth omits that when arrested, 18-year-old Rouse was carrying a fully loaded .45-caliber pistol, several hundred rounds of ammunition, two electric power drills, hacksaws and other "home hobbyist" equipment. Rouse's mother insisted that he had been mentally ill for some time and needed treatment, not jail. The reader is not told that Rouse was acquitted on the grounds of insanity and that the law mandated his hospitalization. Nor does Judge Bazelon tell us that Rouse had been showing unequivocal improvement in group therapy when he refused further treatment. Nor are we told that the hospital staff and the lower-court judge tried to persuade Rouse to continue the effective group therapy, nor that the case was heard after a year of persistent refusal of treatment.

Judge Bazelon tries to convince the reader that psychiatrist witnesses obstruct due process, as if they could require lawyers and judges to abrogate their legal skills or knowledge. Sworn witnesses must answer all legitimate questions put to them with "the truth, the whole truth, and nothing but the truth." If they do not, they face contempt or perjury proceedings. However, judges do sit quietly while lawyers neglect to ask crucial questions or to object to improper ones. Both sides use psychiatrist witnesses and the jury determines which side presents the more reasonable and more credible testimony.

Judge Bazelon focuses on the failure of his *Durham* rule as a case in point. Here too we get half-truths. In a law-journal article, whose audience is more sophisticated about the realities of the law, Judge Bazelon admitted that the poor "quality of the legal representation available for indigent defendants" was implicated in the failure of the *Durham* rule, although he does not even hint at it here. ("Implementing the Right to Treatment." David L. Bazelon in *The University of Chicago Law Review*, Volume 36, pages 742–754; Summer, 1969.)

**Scientific American,** October, 1974. Vol. 231, No. 4. Published monthly by Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017; Gerard Piel, president; Dennis Flanagan, vice-president; Donald H. Miller, Jr., vice-president and secretary; George S. Conn, treasurer; Arlene Wright, assistant treasurer.

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

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

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

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

☐ Please enter change of address.

☐ Please enter subscription.
U.S. and Canada:
☐ 1 year, $12 ☐ 2 years, $22 ☐ 3 years, $30
All other countries:
☐ 1 year, $15 ☐ 2 years, $27 ☐ 3 years, $37

NAME

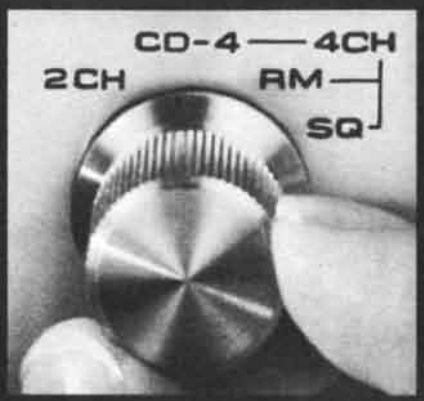
NEW ADDRESS

OLD ADDRESS

# Quadraphonic at its very best.

The fabulous Pioneer QX-949 4-channel receiver has everything going for it. To start, it has built-in circuitry for every source of 4-channel sound reproduction — discrete CD-4, plus SQ and RM matrixes. No matter what the quad program source, the QX-949 plays its flawlessly, without adapters, add-on decoders or demodulators.

But, it's a powerful 2-channel stereo receiver, too, thanks to Pioneer's exclusive Power Boosting circuit. In 4-channel, it delivers 40 watts RMS x 4, 8 ohms, 20-20,000 Hz, four channels driven. In 2-channel, it escalates to 60 watts RMS x 2, across the 20-20,000 Hz spectrum, with both




channels driven. THD and IM distortion is only 0.3% in both cases.

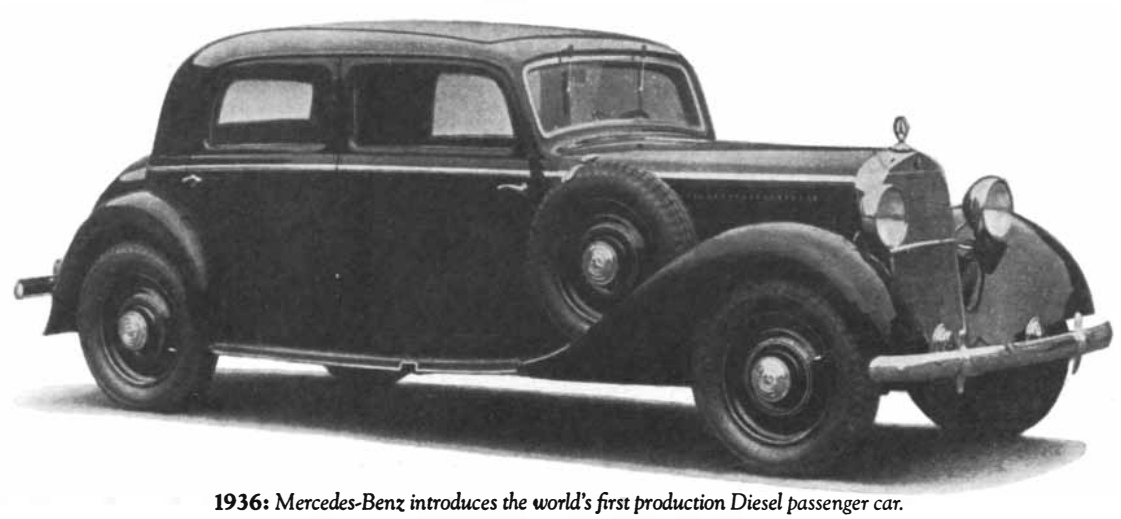
The QX-949 also features a scope-type 4-channel level indicator so you can actually see what you hear. The advanced circuitry also includes a Dolby adaptor input/output and a 4-channel broadcasting multiplex output terminal. In addition, the QX-949 has terminals for a wide range of program sources to suit every listening interest.

Pioneer has always meant the best in high fidelity. Continuing this tradition of excellence — the QX-949 is quadraphonic at its very best. $749.95, including walnut cabinet. Other Pioneer quad receivers from $499.95.

U.S. Pioneer Electronics Corp., 75 Oxford Drive, Moonachie, New Jersey 07074
West: 13300 S. Estrella, Los Angeles 90248 / Midwest: 1500 Greenleaf, Elk Grove Village, Ill. 60007 / Canada: S. H. Parker Co.

PIONEER*
when you want something better

# Mercedes-Benz cannot introduce the Diesel passenger car as a new idea. We did that 38 years ago.



**1936:** *Mercedes-Benz introduces the world's first production Diesel passenger car.*

## We've sold well over a million Diesel automobiles-more than 60,000 in America alone. What we've learned is that before you offer a Diesel to the world, you'd better know exactly what you're doing.

In 1936, the world's first series-produced Diesel passenger car rolled out of our Stuttgart, Germany, plant. The experience we have accumulated since then has paid dividends.

The Mercedes-Benz Diesel of today is a motor vehicle unique in all the world. It delivers extraordinary mileage. Its durability is legendary. And it already meets the tough 1975 Federal clean air standards for automobile engines.

Diesel power introduced new economy and reliability to the automobile. It also fostered new problems. Some manufacturers judged the enigmas greater than the rewards. Mercedes-Benz was not one of them.

We alone held to our convictions concerning the Diesel, refining it year by year over the past four decades.

Today, our Diesel engines are unique. Two sophisticated developments are among the reasons why:

**1. Multistage fuel injection.** With this system, the injection nozzles and pump vary the amount of fuel entering the combustion chambers to meet load conditions.

**2. Precombustion design.** Two interconnected chambers are used for each cylinder. Air and fuel are mixed and burned in two stages, providing a longer power pulse for smoother running.

Result: uniquely smooth, quiet,

clean-burning Diesel power.

Engineering advances like these take years to perfect. Small wonder that the specialized development of Diesel automobiles has discouraged other manufacturers. Why has Mercedes-Benz alone consistently championed the idea of the Diesel passenger car for so long? Consider the following:

### Extraordinary fuel economy

In a strictly supervised test conducted by *Motor Trend* magazine, a 240D—even with automatic transmission—recorded an honest 24.4 miles per gallon. This test was a combination of actual urban and highway driving—no "economy run" or "proving ground" tests.

No full-sized sedan, foreign or domestic, delivers this kind of mileage. And even the domestic mid-sized sedan which came out best in the *Motor Trend* tests was still nearly 30% behind the Mercedes-Benz 240 Diesel Sedan.

### How to avoid being a guinea pig

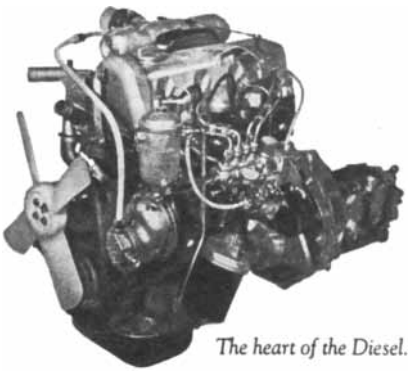
*Buy a Mercedes-Benz Diesel. We alone have sold and serviced Diesels in America for over a decade. For us, Diesels are hardly a novelty. Highly skilled mechanics familiar with our Diesels work in our 390 dealerships all over the U. S. So, when it's time for the little routine maintenance your Mercedes-Benz Diesel requires, you're an old friend, not a guinea pig.*

Another plus: because Diesel fuel is often less expensive than gasoline, Diesel owners get more miles from cheaper gallons.

### Economy car

Operating economies depend on far more than fuel appetites. Periodic maintenance and reliability are equally critical.

How does the Mercedes-Benz Diesel measure up? Any car needs some routine servicing, but with a Mercedes-Benz Diesel you can forget about tune-ups. Diesels need no spark plugs, no points, no distributor, and no carburetor.



*The heart of the Diesel.*

No conventional ignition or carburetion systems—no tune-ups, ever. And parts that aren't there can't break down.

### Amazing resale value

There is one area where a Mercedes-Benz Diesel is just like any other Mercedes-Benz. Resale value. The automobile industry's most authoritative sources agree that any Mercedes-Benz has a better resale value than *any* full- or mid-sized sedan made in America.

Because the 240D is so new, there has not been time to gather the proven figures which previous Mercedes-Benz Diesels can boast. But consider the history of the 220D, the 240 Diesel's immediate predecessor. As an average owner, if you had bought a 220D two years ago, it would now be worth over 90% of the price you paid. If your Diesel were now three years old, it would still have retained more than 80% of its original value.

### An absolute nonconformist

Obviously, the Mercedes-Benz Diesel is an iconoclast in its approach to modern driving problems. Like all Mercedes-Benz automobiles, it carries its dissidence to every element of its construction.

The Mercedes-Benz 240D is a full 5-passenger sedan. It has interior space and trunk room on a par with the finest domestic luxury cars. It has disc brakes on all four wheels. It has a separate suspension system for each wheel. It embodies the decades of safety innovation and advanced engineering that are so much of the Mercedes-Benz philosophy.

If you seek exceptional mileage, maintenance economy, dependability and resale value—without the "economy car" penalties that usually go along with them—it's time to consider a Mercedes-Benz Diesel.

# Mercedes-Benz

**Engineered like no other car in the world.**


**1974:** *The most advanced Diesel automobile sold in America.*

Nor does he tell us that in 18 years *Durham* was not freely adopted in any other Federal court jurisdiction, nor does he offer his colleagues' reasons for this rejection.

It is not surprising that some consider Judge Bazelon's construction of the law naïve when he says, regarding *M'Naghten*, "according to this holding of a British court in 1843. . . ." After the M'Naghten trial was completed the House of Lords asked the 15 judges of the Queen's Bench to clarify England's law regarding criminal insanity by answering five questions. The judges, with one dissent, did this in conference, with none of the legal trappings or judicial aids of court process. They were analogous to a Congressional committee, advisory or investigative, but were not empowered to make legally binding decisions. Their answers, the M'Naghten rules, formulated in an atmosphere of political and civil turbulence, ignored and obliterated some liberal elements in the common law of England of that time.

Judge Bazelon speaks righteously of the responsibility to accord due process to all. However, he quotes some anonymous "leading forensic psychiatrist" damning a hospital staff conference that decided a controversial diagnosis by a five-four decision as a "star-chamber proceeding." This conference released the individual for public trial, where psychiatrist witnesses for the defense could challenge the staff diagnosis openly. A major defining characteristic of the star-chamber proceeding is the use of anonymous witnesses, such as Judge Bazelon's, who cannot be questioned about testimony, personal involvement, credibility or even qualifications as a "leading forensic psychiatrist." Does Judge Bazelon allow this kind of evidence in his court?

Judge Bazelon may have convinced some readers of *Scientific American* that psychiatrists obstruct legal process and are responsible for the failure of *Durham*. His article suggests to me that his scholarship and his credibility are deficient and not what should be expected from a presiding judge sitting so close to the Supreme Court. I do not know why he used half-truths and anonymous testimony. Perhaps Judge Bazelon does.

The problems of legal insanity should concern all citizens. The solutions require thoughtful consideration of all contributing factors, not shortsighted scapegoating. Lawyers, psychiatrists and judges are in need of better education and understanding in this area. I suggest that the method of the university seminar is better suited to this purpose than are articles with tactics redolent of the Roman gladiator circus.

JACQUES M. QUEN, M.D.

Department of Psychiatry
The New York Hospital–
Cornell Medical Center
New York

Sirs:

It would be pointless to use this space to respond to Dr. Quen's *ad hominem* arguments or to engage in a controversy over the details of individual cases. For example, he thinks it important that the case of *Rouse* v. *Cameron* involved a person who was heavily armed when arrested and who was unwilling to be treated. Those facts relate only to the particular decision to be made in *Rouse: What* constitutes adequate psychiatric treatment under a statute that gives a right to such treatment? But the question I raised in the article was: *How* should courts go about deciding issues involving psychiatric expertise? Whether the treatment offered to Rouse was in fact adequate to his peculiar needs is entirely irrelevant to my question.

Dr. Quen does make one observation, however, that deserves further attention. He joins many others, including myself, who have suggested that the adversary process has serious inadequacies as a method for resolving controversies. (Dr. Quen and other readers may wish to refer to articles I have published in the legal literature taking the adversary process to task for its shortcomings both in theory and in practice. For example, "The Defective Assistance of Counsel, *University of Cincinnati Law Review*, Volume 42, page 1, 1973; "New Gods for Old: 'Efficient' Courts in a Democratic Society," *New York University Law Review*, Volume 46, page 653, 1971.) There is a continuing inquiry within the legal profession about these defects, and various improvements have been tried. To date no model has appeared that does not entail even greater defects. But the inquiry has revealed weaknesses and helped to remedy some of them, or at least to mitigate their effects.

Important and interesting as a discussion of the role of lawyers and judges in the adversary process may be, this point is only tangential to an article titled "*Psychiatrists* and the Adversary Process" (emphasis added). Nowhere does Dr. Quen attempt to answer the questions I raised about the limitations of psychiatric expertise. Instead he con-

# Texas Instruments SR-50 scientific calculator.

## A technological milestone.

## In sheer math power. And in price $149.95

From the moment you open this box you'll begin to work with mathematics as never before.

Now, right at your fingertips, you have an unmatched combination.

In performance and value.

Accuracy and efficiency...

# Compare the SR-50. Feature for feature. Function for function. Dollar for dollar.

Algebraic keyboard. Accuracy to 13 significant digits. Display rounded to 10 digits. Single function keys for finding powers, roots, factorials, logarithmic, trigonometric, and hyperbolic functions. For sum and store. For converting answers to scientific notation. Unmatched capability and performance for just $149.95.

**Lots of choices, few standouts.**
Today, scientific calculators are available in a wide range of prices and capabilities. New models, new brands, new manufacturers have recently entered the market. But, when you want the best in a full-function slide-rule calculator, the choices narrow down fast.

Still, only two names and three models stand out. The SR-50 outperforms the HP-35 and costs less – about $75 less at the time of this writing. The HP-45, on the other hand, offers some features not found on the SR-50, but cost $175 more.

It boils down to basic value analysis. And, in that light, the SR-50 is unmatched.

| FUNCTIONS | SR-50 | HP-35 | HP-45 |
|---|---|---|---|
| LOG, LN | yes | yes | yes |
| TRIG (ARC, SIN, COS, TAN) | yes | yes | yes |
| HYPERBOLIC (ARC, SIN, COS, TAN) | yes | no | no |
| DEG ↔ RAD CONVERSION | yes | no | no |
| DEG/RAD MODE SELECTION | yes | no | yes |
| DEC ANGLES ↔ DEG/MIN/SEC | no | no | yes |
| POLAR COORD ↔ RECT COORD | no | no | yes |
| $Y^x$,$e^x$ | yes | yes | yes |
| $X^2$ | yes | no | yes |
| $\sqrt{X}$ | yes | yes | yes |
| $\sqrt[x]{y}$ | yes | no | no |
| 1/X | yes | yes | yes |
| X! | yes | no | yes |
| EXCHANGE X WITH Y | yes | yes | yes |
| METRIC CONVERSION CONSTANTS | no | no | yes |
| %, AND Δ% | no | no | yes |
| MEAN AND STANDARD DEVIATION | no | no | yes |
| STORE, RECALL | yes | yes | yes |
| Σ TO MEMORY | yes | no | yes |
| **FEATURES** | | | |
| ROUNDING TO TEN DIGITS | yes | no | no |
| ALGEBRAIC NOTATION (SUM OF PRODUCTS) | yes | no | no |
| DEG/RAD SWITCH | yes | no | yes |
| MEMORY (OTHER THAN STACK) | 1 | 1 | 9 |
| FIXED DECIMAL OPTION | no | no | yes |
| KEYS | 40 | 35 | 35 |
| SECOND FUNCTION KEY | no | no | yes |


**An extension of your professional capability.** The SR-50 delivers the math power today's technical environment demands. It can help you get more complex mathematical work done faster. With greater accuracy.

It can save you valuable time and magnify your talents. An SR-50 lets you use your technical imagination with more freedom. You have time to explore more options – use "brute force" approaches such as successive approximations, educated guesswork, trial and error.

And, at $149.95, an SR-50 is within practically anyone's budget. Whether you are at the top of your profession or a beginning student. In fact, at this price, it becomes a practical consideration for productivity-minded managers and administrators to equip virtually every professional in their organization with an SR-50 scientific calculator.

And, you can get them now from TI, individually or in volume quantities, within short lead times.

trives to confuse the issue by attacking not what I said about psychiatrists but rather what I did *not* say about the inadequacies of lawyers and judges.

Dr. Quen's only positive suggestion, relegated to the final sentence and given no explication, is the use of "the university seminar" to advance education and understanding. If he means to say that we need ongoing inquiry to explore new approaches, I quite agree. To be sure, I consider this article, as well as many of my previous writings and opinions, to be a part of that process. Articles and court opinions are never meant to take the place of discussions (in universities or elsewhere), but they are clearly relevant to any discussion of these issues.

But if articles are no substitute for seminars, neither are seminars a substitute for court decisions. Dr. Quen seems to be implying that the university seminar (or, more broadly, purely scientific research and debate) is a better way than the adversary process to reach conclusions about psychiatric questions in the courtrooms. Certainly if behavioral scientists reach a consensus about a particular problem, the decisions required of courts and juries are simplified. If such decisions were purely technical matters, perhaps they could be left to the "experts" to resolve.

But courts must decide controversies as they arise in individual cases, long before the experts may reach consensus. And these are not merely technical decisions, neatly isolated from questions of social policy and community values.

The law does not ask simply: "Was the criminal defendant insane when he killed his victim?" It asks instead: "Was the defendant insane in such a way that he should not be held criminally responsible?" The first question might be called purely technical except that there is no consensus about insanity in the psychiatric community. But the second question clearly demands that social and moral judgments be taken into account in resolving the scientific controversy. At that point the obstacles to resolution by a neutral expert become insurmountable and the question becomes one for judges and juries as representatives of the community.

That brings us back to square one: In the absence of expert consensus, how are courts to make the decisions that society demands?

DAVID L. BAZELON

Chief Judge
United States Court of Appeals
Washington

# The Effects of Ice on Scotch

*How fast a drink of Scotch whisky over rocks loses its flavor depends on the proof of the Scotch and the richness of its blend. These two factors are optimized for "on the rocks" Scotch drinkers in 90-Proof Famous Grouse, a venerable old brand from Scotland only recently introduced to America.*

by Allen Mac Kenzie

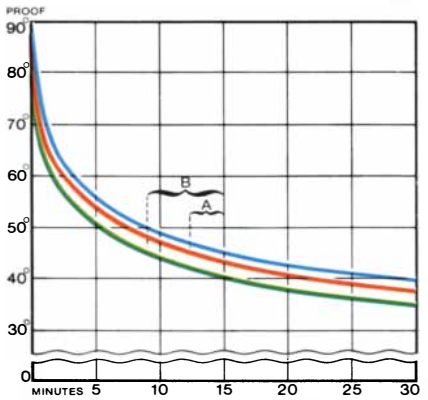
In countries where Scotch has been consumed for centuries, ice and whisky rarely mingle. But on this side of the Atlantic, the picture is quite different. While a small percentage of American Scotch drinkers take it neat, better than 35% drink it "on the rocks." The rest of us add varying amounts of water, club soda, etcetera. And ice. Always plenty of ice —the great American drink requisite.

It would seem then that the American Scotch devotee, particularly our on-the-rocks fancier, has a right to raise a serious question: *Is the Scotch I drink ideally suited to enjoyment over ice?*

### Pursuing a Perfect Proof

Let's turn our attention first to the proof at which Scotch whisky is bottled. Consider the hypothesis that there is indeed a better proof for on-the-rocks Scotch drinking than that of the brand you currently favor.

Practically every Scotch sold in this country is bottled at 80, 86, or 86.8 Proof. So at the instant you pour Scotch over ice, it contains between 40% and 43.4% alcohol by volume. (Proof is double the percentage of alcohol. ) The chilling effect of the ice is accompanied by dilution. And when your drink has been properly cooled —in 30 seconds to a minute—you achieve what one Scotch connoisseur refers to as "the ideal sip." From then on, the Scotch drinker's enjoyment typically runs downhill, as the drink loses its freshness.




**DILUTION BY ICE OF SCOTCH AT THREE DIFFERENT PROOFS (72°F).**
**80 Proof 86.8 Proof 90 Proof**

While there is no way to preserve that fresh Scotch flavor indefinitely, we submit that you can sustain the freshness substantially longer with 90-Proof *Famous Grouse*. If you have never heard of this brand, we are not surprised. It is a well established name in Scotland, but only recently introduced to America. So far as we know, *Famous Grouse* is the only Scotch now available in this country at 90 Proof.

### A Revealing Experiment

To demonstrate the merits of a slightly higher proof, we performed a simple experiment: 50 millilitres of Scotch (about 1.7 ounces) was chilled with 100 cc of ice. The ensuing dilutions at 80, 86.8 and 90 Proof are charted in the graph at left.

You'll notice that after 15 minutes on the rocks, the proof of *Famous Grouse* is diluted to a level which occurs after 12½ minutes when the Scotch is 86.8 Proof, and after 9 minutes when it is 80 Proof. In essence, the *Famous Grouse* brand has remained about 2½ minutes fresher than 86.8-Proof Scotch (Interval A on graph), 6 minutes fresher than 80-Proof Scotch (Interval B). If you "nurse" a drink beyond 15 minutes, the advantages of 90 Proof Scotch are even more pronounced.

Proof, of course, is not the only influence on the flavor of a blended Scotch. The proportion of malt to grain whiskies, origins of the malts, aging methods—these are also important factors determining the relative richness of Scotch flavor.

The makers of *Famous Grouse*—Matthew Gloag & Son of Perth, Scotland —have been producing Scotch in the same family for six generations. And they have performed their most noble feat in the rich blend they created for *Famous Grouse* Scotch. Its flavor—so remarkable at the outset—holds firmly to its character during prolonged contact with ice.

Knowledge of Scotch, however, cannot be indefinitely pursued in the abstract. Your learning process must ultimately include a leisurely sip of *Famous Grouse* on the rocks. For Scotch drinking is one of those pleasures enjoyed most, not in the pursuit, but in the conquest. Scotland's greatest bard, Robert Burns, said it best:

*"Gie me a spark o' Nature's fire,*
*That's a' the learning I desire."*


90 PROOF. BLENDED SCOTCH WHISKY. BOTTLED IN SCOTLAND
IMPORTED BY AUSTIN, NICHOLS & CO., LAWRENCEBURG, KY.

# 50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

OCTOBER, 1924: "For many years it has been realized that a star must grow hotter as it contracts, but this contraction was thought to be limited by the density at which the atoms of the gas were packed so closely together that little or no free space remained between them. A recent paper by Professor Eddington suggests that the contraction may go much further than had been thought. Inside a star the temperature must be measured in millions of degrees, Eddington points out. Under these conditions all the outer electrons of atoms would be knocked loose and the remaining residue would be very much smaller. From investigations of X rays the diameters of such atomic residues can be estimated, and Eddington concludes that no influence of their close packing would be perceptible until the density became at least 10,000 times that of water. One great advantage of the new theory is that it explains such a star as the companion of Sirius, which is very much fainter than the sun but nearly as massive and a good deal hotter on the surface. If it is as bright per square mile as this implies, it would be only three times the diameter of the earth and 50,000 times as dense as water. This last deduction has always been supposed to be absurd. If Eddington shows that it is not, he will have done us all a service."

"The superpower scheme is planned to do the same thing with electric power that the engineer did more than 100 years ago with water—gather it in a few great centralized power stations and 'pipe' it to the user. The subcommittee of the Northeastern Superpower Committee, under Secretary Hoover, has reported in favor of building large centralized electric plants and feeding the output into interconnected transmission lines for distribution to every kind of power user. The vast electric power reservoir thus formed will serve the section of the country extending east of the Alleghenies from Maine to Maryland. To obtain the maximum economy the power will be developed in large plants with a capacity of from 300,000 to 750,000 horsepower."

"The long, grueling flight of the Army aviators in their attempt to circumnavigate the globe by airplane is nearly completed. Of the three airplanes that departed Seattle on the first stage of the flight for Alaska, the Pacific isles and Japan, two have survived the journey as far as Iceland, where the fliers are now waiting to attempt the hop to Greenland. When they come safely through this hazard, they will surely complete their circumnavigation."

"Biochemistry is undoubtedly one of the most fascinating, but at the same time one of the most complicated, branches of chemistry. Through the study of fermentation, however, science has reached a clearer understanding of biochemical processes. For many centuries it was known that in the presence of certain organic substances, which were called ferments, chemical changes in organic matter took place. The ferments were divided into ferments of cellular structure and ferments without cellular structure, or enzymes. Some of the greatest chemists of the 19th century, however, were firm in their belief that there is no difference between the kinds of substances. The proof of this theory was provided by Eduard Buchner when he discovered that the reactions by which sugar is split up into alcohol and carbon dioxide take place in the absence of living ferment cells. Recent researches have established that the ferments are merely catalytic agents. Furthermore, they do not confine their activity to the breaking down of complicated chemical combinations into simpler ones: they are also capable of building up more complex compounds from simpler ones. In one respect the ferments do differ from the majority of other catalytic agents. Each ferment is adapted to act as catalyst in one particular set of reactions only, in some cases discriminating between sugars that differ from one another only by minute differences in molecular structure."

"It has been found possible to concentrate radio energy more or less in one direction by the use of wave lengths much shorter than those used in ordinary commercial broadcasting. Marconi and Franklin have developed a system in which the wave length is six to 10 meters. The aerial is very small, consisting of a short wire with the transmitter inserted at the center. This is placed vertically and is at the focus of a parabola, from which the energy is reflected and concentrated into the form of a beam."

"Cellophane is a perfectly transparent substance that can be made in extremely thin sheets. All of it hitherto has come from France, but now announcement has been made that an American company has devised a process for making cellophane here. The process consists in taking pure wood pulp and treating it first with caustic soda and then with carbon disulphate. The product thus obtained is forced through a slit-like aperture into a coagulation bath, where it solidifies into a solid, transparent mass."

SCIENTIFIC AMERICAN

OCTOBER, 1874: "New York has a death rate such as few cities in Christendom can equal. The appalling mortality of the past summer, especially among children, has given rise to a great amount of sorrow and indignation, and not a little criticism of the medical and police authorities. That much might have been done to improve the health of the city by sanitary measures there is no doubt; the great source of disease and death in the city, however, is the tenement-house system, whereby families are massed by the hundreds in huge barracks, destitute of light, ventilation and the means of keeping clean. Only by the dispersion of the tenement-house population can the death rate be reduced to reasonable limits. There is no way by which such a desirable result can be effected humanely save by providing means for carrying the poorer working people to and from country homes more rapidly and cheaply than is possible with surface roads."

"The spectroscope has explored the far-off space of heaven. The light of hundreds of stars has been analyzed, and nebulae, scarcely visible, have had the quality of their radiations revealed by its aid. The light, in some cases very feeble, with which a number of stars shine gives a spectrum with dark lines like the solar spectrum. This fact proves that the constitution of these stars is like that of our sun. Aldebaran sends us records not only of hydrogen, magnesium and calcium, which abound in solar light, but also of metals such as tellurium, antimony and mercury. Matter is everywhere the same. We meet the hydrogen of water in the sun, in Sirius and in the nebulae. Everywhere matter moves, everywhere it vibrates, and these movements appear to be inseparable from the atoms."

# THE AUTHORS

ALVA MYRDAL ("The International Control of Disarmament") is a Swedish expert on disarmament who has served as the cabinet minister in charge of disarmament affairs and as her country's chief delegate at the international disarmament conference in Geneva. She is the wife of the well-known author and scholar Gunnar Myrdal. In addition to her official work on disarmament she is a member of the Swedish parliament and was Swedish ambassador to India. She was graduated from the University of Stockholm in 1924, obtaining her master's degree at the University of Uppsala in 1934 and her Ph.D. from the University of Leeds in 1962. Among her many activities in Sweden and the U.S. was the Swedish Training College for Preschool Teachers, which she founded in 1936 and served as director from 1936 to 1948.

JOSEPH S. MILLER ("The Structure of Emission Nebulas") is associate professor of astronomy at the University of California at Santa Cruz and associate astronomer at the Lick Observatory. He did his undergraduate work at the University of California at Los Angeles and received his Ph.D. from the University of Wisconsin. "I find that astronomy consumes most of my waking hours," he writes, "and a good many generally reserved for sleeping, but I have managed to pursue a few outside interests. One of them is wine: for the past four years I have taught a course in enology at Crown College—a course that stresses the scientific aspects of viniculture as well as appreciation of the end product. I also enjoy bicycling and have spent many weekends exploring back roads in the Santa Cruz Mountains."

PETER SATIR ("How Cilia Move") is professor in the department of physiology-anatomy at the University of California at Berkeley and director of the Electron Microscope Laboratory. "I was graduated from Rockefeller University in 1961," he writes, "shortly after my first article appeared in *Scientific American* ["Cilia," February, 1961]. I spent most of the next six years at the University of Chicago in zoology and in biology at the college, rising through the junior ranks of the academic ladder. In 1967 I received my first appointment at Berkeley. I have had the good fortune to spend several periods in various laboratories abroad in Denmark, Switzerland and, most recently, at the Zoological Institute at the University of Tokyo, as a Guggenheim Fellow. During my stay in Denmark I met, wooed and won my wife, who is a scientist in her own right, pursuing a full-time career. With two biologists in the family and two active sons, we keep rather busy. I find time for unstrenuous walks in the Berkeley hills and for reasonably frequent trips to theaters, restaurants, shops and museums of all kinds. These interests in city amenities are perhaps what one would expect of a native New Yorker, but I also appreciate the beautiful country that surrounds the Bay area and the open attitudes that one finds here."

DAVID R. SAFRANY ("Nitrogen Fixation") is president and research director of the Aerokinetics Corporation, which he founded; its business is energy conversion. He was graduated from Rensselaer Polytechnic Institute in 1959 and obtained his Ph.D. (in physical chemistry) there in 1964. After an additional year there as a research associate he did research work with various corporations until he set up his own firm in 1972. Safrany describes himself as a conservationist and an admirer of the California coast, which he rides along on a motorcycle and flies over in an airplane that he pilots himself.

JAMES MARSTON FITCH, JOHN TEMPLER and PAUL CORCORAN ("The Dimensions of Stairs") are respectively professor of architecture at Columbia University, associate professor of architecture at the Georgia Institute of Technology and associate professor of rehabilitation medicine at the Boston University School of Medicine. Fitch is at Columbia's Graduate School of Architecture and Planning, where among other things he directs what he describes as "a program in the retrieval and recycling of historic sites and structures." Templer, who was born in England, has a bachelor's degree in architecture from the University of Pretoria and the graduate diploma in town planning from the University of the Witwatersrand. His master's degree and Ph.D. are from Columbia. Corcoran was graduated from Georgetown University and received his M.D. there in 1959.

JÜRGEN NICOLAI ("Mimicry in Parasitic Birds") is at the Max Planck Institute for Behavioral Physiology in Seewiesen, where for many years he worked with Konrad Lorenz. He is now in the department of ethology and serves

# NOW $69.95

# The Sinclair Scientific just solved your biggest problem. Price.

## (Logs, trig, arithmetic. Scientific notation, too.)

The Sinclair Scientific is the only pocket calculator that offers scientific capacity at a truly affordable price. And look how much you get:

**log and anti-log (base 10)**
**sin and arcsin**
**cos and arccos**
**tan and arctan**
**automatic squaring**
**automatic doubling**
**$x^y$, including square and other roots**
**four basic arithmetic functions**
**plus scientific notation ($10^{-99}$ to $10^{+99}$).**

All for an incredible $69.95.

### What makes a scientific calculator scientific?

To be a really valuable tool for engineers, scientists, technicians and students, a calculator must provide all of the above.

Clearly, a scientific calculator without scientific notation severely limits the size of numbers with which you can work easily.

And scientific notation without transcendental functions is little more than window dressing on an arithmetic calculator.

Granted, there are companies other than Sinclair offering excellent units with all the essential ingredients.

But they sell at much higher prices.

What's more, the Sinclair Scientific isn't just portable. It's pocketable.

Less than ¾-inch thin. And 3¾-ounces light. It's the world's thinnest, lightest scientific calculator.

actual size

### What makes the Sinclair Scientific so inexpensive?

Two important technological breakthroughs.

First, the British-built Sinclair Scientific has a single integrated circuit. Engineered by Sinclair.

Second, Sinclair's exclusive keyboard has only four function keys. All of which provide "triple-action" by changing from standard to upper or lower case mode.

### Old hands at small miracles.

Sinclair has been an innovator in calculator miniaturization right from the start.

In the last two years Sinclair brought to America the Sinclair Executive and Executive Memory – the world's thinnest, lightest calculators – as well as the Cambridge.

The Sinclair Scientific is backed by an unconditional one-year replacement warranty.

Sinclair also maintains a service-by-return mail operation in the U.S. (and everywhere else in the world) to handle any product problems.

### How to get your Sinclair Scientific.

The Sinclair Scientific may be ordered by phone or mail. And you may charge it to your American Express, Diners Club, Master Charge or BankAmericard account.

If you're not completely satisfied, you may return the unit within 2 weeks of receipt for prompt refund.

Just fill in the coupon and mail it to: Sinclair Radionics, Inc., 375 Park Avenue, New York 10022. Or call (800) 223-5764, toll free.

sinclair™
SCIENTIFIC
The logical choice.

### Specifications

**Functions:**
4 arithmetic
2 logarithmic
6 trigonometric

**Keyboard:**
18 key format with 4 "triple-action" function keys using standard, upper and lower case operation.

**Display:**
5-digit mantissa
2-digit exponent
(both signable)

**Exponent:**
200-decade range, from $10^{-99}$ to $10^{+99}$

**Logic:**
Reverse Polish, with post-fixed operators for full flow chain calculations.

**Power Source:**
Battery operated with 4 inexpensive AAA penlight batteries, providing over 25 hours of use.

**Size:**
4⅜" high;
2" wide; 11/16" thick.

**Weight:**
3¾ oz.

**Warranty:**
1 year.

## ORDER FORM

For credit card phone orders call (800) 223-5764, toll free.
(New York residents call (212) 688-6623.)

**To: Sinclair Radionics, Inc. 375 Park Avenue, New York, New York 10022**

Please send me _____ Sinclair Scientific(s) at $69.95 (plus $1.50 per unit, shipping and handling) including batteries, carrying case, instruction booklet and warranty. New York residents add sales tax.)

☐ Enclosed is my check for _____.

☐ Please charge my credit card account.

Credit Card _____

Account No. _____

Expires _____

Master Charge Code # (4 digits) _____

Signature _____

Name _____

Company _____

Address _____

City _____

State _____ Zip Code _____

16 C

In the United Kingdom: Sinclair Radionics, Ltd., London Road, St. Ives, Huntingdonshire.
In Germany: Sinclair Elektronik GmbH., 8012 Ottobrunn (Munchen), Rosenheimer Landstrasse 39.

also as a university lecturer at the University of Munich. Nicolai studied zoology, botany and anthropology at the University of Mainz from 1949 to 1954, receiving his doctorate there in 1954. In pursuit of his interest in birds he has made five research trips to Africa, visiting Tanzania, Kenya, Uganda, Cameroons and Nigeria.

EMILIO BIZZI ("The Coordination of Eye-Head Movements") is professor of neurophysiology in the department of psychology at the Massachusetts Institute of Technology. "I was born in Rome," he writes, "where, after laboring through the required years of Latin and Ancient Greek, I moved on to the University of Rome's Medical School, graduating in 1958. Ironically, while interning at the medieval hospital of the city of Siena I became attracted to modern neurophysiology. This interest led me to the Institute of Physiology at the University of Pisa, where I received my neurophysiological training. Thereafter I pursued my career in the U.S. at Washington University, the National Institutes of Health and M.I.T. I have no major interests outside my work, although I play tennis as often as possible and follow closely contemporary literature and music."

MICHAEL H. JAMESON ("The Excavation of a Drowned Greek Temple") is professor of classical studies at the University of Pennsylvania. Born of American parents in London, he spent much of his childhood in Peking, where his father was a college teacher. He traces his interest in archaeology and antiquity to that time. He acquired his bachelor's degree and his Ph.D. (both in Greek) at the University of Chicago, spent a year at the American School of Classical Studies in Athens and then taught at the University of Missouri for four years before joining the Pennsylvania faculty in 1954. From 1966 to 1968 he was dean of the Graduate School of Arts and Sciences at Pennsylvania. Since 1968 he has been director of the university's Center for Ancient History. The work Jameson describes has been conducted for the American School of Classical Studies in Athens by the University of Pennsylvania and Indiana University with support from the National Endowment for the Humanities, the Ford Foundation and the National Geographic Society. He is interested in "an anthropological approach to history and in what can be called historical ecology—the relation of man to environment throughout history."

Now! Preview the great new

Deutsche Grammophon

recording of the complete

# FOUR SYMPHONIES OF BRAHMS

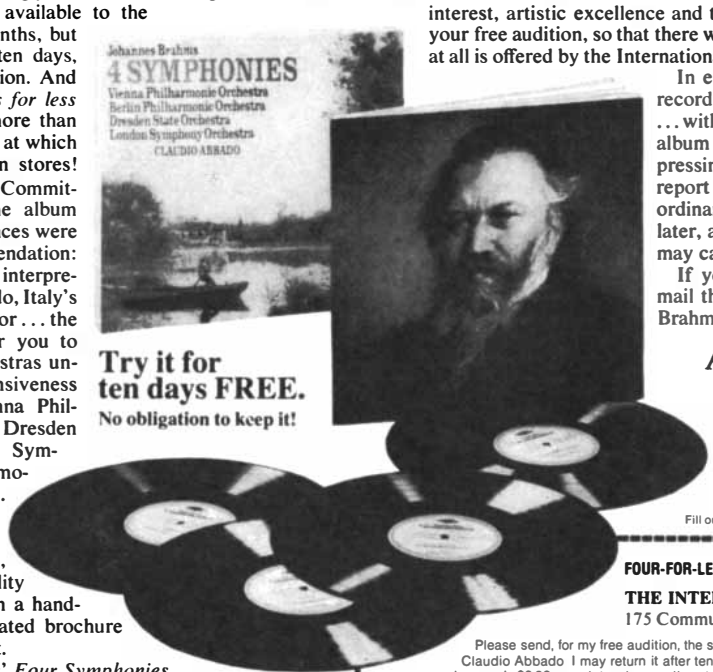
superbly conducted by CLAUDIO ABBADO

# ALL 4 RECORDS FOR LESS THAN THE PRICE OF ONE

plus free auditions of important new recordings
months before they are available to the public!

BRAHMS' *Four Symphonies*—most magnificent since Beethoven's! The extraordinary new recording of them from Deutsche Grammophon offers Claudio Abbado conducting *four* of the world's greatest orchestras. This album will not be available to the American public for many months, but you may preview it now for ten days, without charge, without obligation. And you may keep all *four records for less than the price of one*—yes, more than 78% off the suggested list price at which they will eventually be sold in stores!

The Carnegie Hall Selection Committee's report accompanying the album points out why these performances were singled out for their recommendation: freshly conceived, deeply felt interpretations by young Claudio Abbado, Italy's foremost present-day conductor... the unprecedented opportunity for you to enjoy and compare four orchestras unsurpassed in precision, responsiveness and tonal splendor—the Vienna Philharmonic, Berlin Philharmonic, Dresden State Orchestra and London Symphony... plus Deutsche Grammophon's incredibly lifelike sound. The four stereo records themselves are superior, silent-surfaced European pressings, manufactured to highest quality control standards, and come in a handsome hinged case with illustrated brochure and *exclusive* Committee report.




Our preview offer of Brahms' *Four Symphonies* is limited! So please mail card or the coupon on this page today. Listen for ten days. Then either return the album and pay nothing, or keep it for only $6.98 plus small postage/handling charge (and sales tax in New York). This is a savings of *$24.94 off* the suggested retail price when the regular edition is finally released.

**How you can be first to enjoy the greatest new recordings!**

By replying now, you assure yourself of the privilege of receiving, for future ten-day free previews, special *advance editions* of major new albums... chosen *exclusively* for the International Preview Society by the Carnegie Hall Selection Committee. Since the Society is *not* a record "club", you will never be required to buy any album, nor will you receive a "record-of-the-month" like clockwork. Only when the Committee discovers a new album of the most extraordinary musical interest, artistic excellence and technical quality will it be sent for your free audition, so that there will often be months when no album at all is offered by the International Preview Society.

In every case, you will preview these records *long before* they reach the stores ...with the option of purchasing any album you wish, in a superb imported pressing, including exclusive Committee report... never at a higher price than the ordinary edition will retail for months later, and frequently at *big savings!* You may cancel this arrangement at any time.

If you truly love great music, please mail the coupon today for your exciting Brahms' *Four Symphonies* preview.

All 4 records only

**$6.98**

(plus small postage/handling charge) if you decide to keep them after your free audition!

Fill out and mail this coupon if card is missing.

Q52

**FOUR-FOR-LESS-THAN-THE-PRICE-OF-ONE PRIVILEGE**

**THE INTERNATIONAL PREVIEW SOCIETY**
175 Community Drive, Great Neck, N.Y. 11025

Please send, for my free audition, the superb recording of Brahms' *Four Symphonies* by Claudio Abbado. I may return it after ten days and owe nothing, or keep all four records and pay only $6.98 complete, plus small postage/handling charge (and sales tax in New York). This is $24.94 *off* suggested list price of the regular edition, when it becomes available months from now. I will also receive, at intervals of one or more months, free ten-day previews of albums chosen by the Carnegie Hall Selection Committee, far in advance of general release. I may keep any album for no more, and often for far less, than suggested list price. *I am not obligated to buy any minimum number of albums and I may cancel this arrangement at any time*

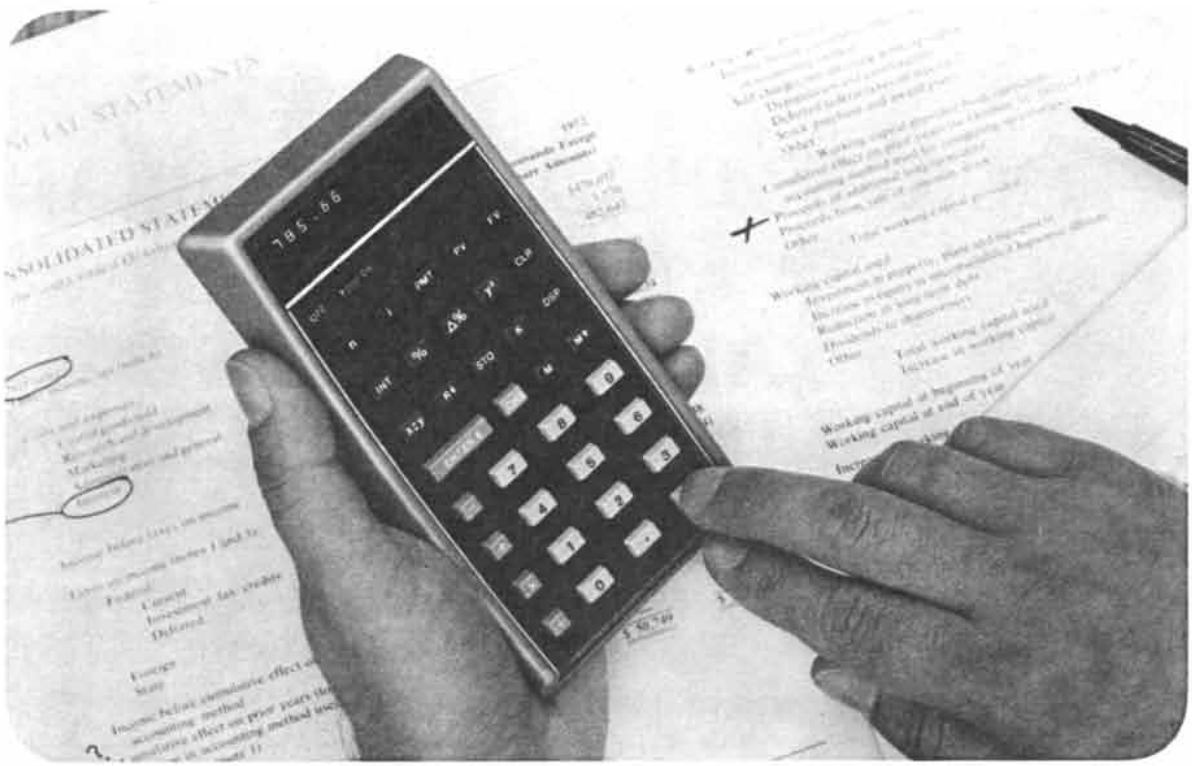
NAME _____
(please print)

ADDRESS _____

CITY _____ STATE _____ ZIP _____

Offer limited to the U.S. and its possessions. Offer expires 1/31/75
Limit: One membership per household—Only new members eligible

**19237**/Q52



## The HP 70: a new business calculator you won't outgrow.

A little over a year ago, we introduced the HP-80, the first pocket-sized calculator for time-and-money management. Although we designed it expressly for the financial specialist, it wasn't long before non-specialists were buying it by the thousands, simply because the HP-80 served the needs of business far better than any existing pocket-sized calculator. Out of this realization came the new HP-70.

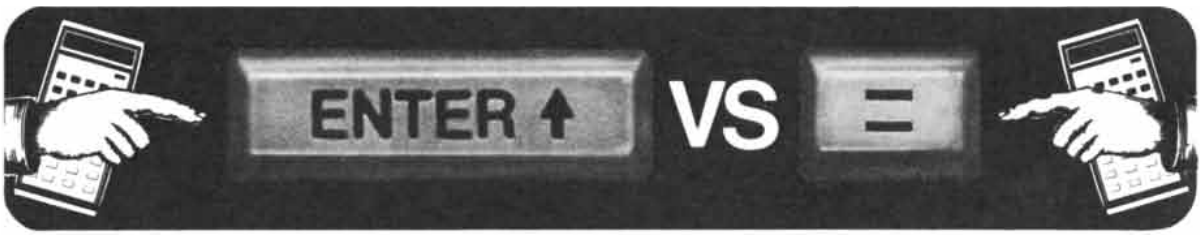
Designed to solve the *general* time-and-money problems of the businessman rather than the *special* calculations of the financial professional, the HP-70 is easier to use than its financial counterpart. Its keyboard is simpler. Yet the HP-70 retains the row of five financial keys found on the HP-80, pre-programmed to solve more than 100 time-and-money problems. Problems such as compound interest, effective return on investment, declining balance depreciation, mortgage payments, finance charges, future value of an annuity, etc. All you do is press the keys. The ponderous equations and cumbersome tables are stored electronically in the HP-70.

To make these powerful financial keys really easy to use, we gave each its own memory bank. Now you can enter the numbers of a financial problem in any order and change any number in the problem without having to re-enter the other numbers—a nice convenience.

The HP-70 also has two addressable memories and a four-memory operational stack. This stack, plus the non-algebraic computer logic we use in all our pocket calculators, lets you solve all problems in the same consistent way, even when they include serial, chain, or mixed calculations. (You'll find more information on our non-algebraic approach across the page.) You never have to jot down and re-enter intermediate results obtained during chain calculations because the operational stack stores them for you and brings them back to the working register—automatically.

Each of the HP-70's two addressable memories (the K and M keys) lets you store any number that you might want to use later in a lengthy calculation or repeatedly in a series of calculations. Since you can use one of these registers (M) as an accumulator, you can keep two running totals at the same time—one in the stack and one with M+.

As you discover more and more ways to put your HP-70 to work, its exceptional calculating power will give you capability you won't outgrow. And the quality and care of its manufacture mean that it will be fully operational when you need it. The $275* price includes rechargeable batteries, AC adapter and battery charger, soft case, owner's handbook, and reference guide.




## HP pocket-sized calculators give you answers you can trust. (Why we chose to be different.)

Most pocket-sized calculators have an = key, and you can pick up one you've never seen before and solve a simple equation by entering values the way they're written and pressing = to see your answer.

Pick up one of our pocket-sized machines and you'll hunt in vain for the = key. Instead you'll see one that says ENTER and you'll wonder what to do with it.

To find out, you'll have to look in the Operating Manual. When you do, you may think, "Why did they go this route when they could have kept it simple?"

But as you work the examples in the manual and practice with your own problems, you'll find yourself thinking, "This is very easy, and I'm getting answers I can really trust."

Easy because you always enter your data the same way—from left to right, the way you read an equation, but without having to keep track of parentheses and brackets.

Trust because you see the result of each intermediate calculation displayed as you perform it, and you can check your progress and correct errors as you go. You can also go back and review stored numbers at any time.

When we started out to design a small, powerful scientific calculator several years ago, we assumed it would have an = key. It did, in the breadboard stage.

But along the way from breadboard to prototype, our designers realized that there was a better method which got around some of the severe limitations inherent in the "algebraic" approach.

These limitations start showing up as problems get more complicated and you have to spend a lot of time figuring out how to put them into an algebraic calculator. "A plus B times C minus D divided by E equals" won't do it.

You find you must write down intermediate answers to re-enter later. You may work your way through a long equation and see an answer you know is wrong. Where did you goof?

Or you may get a wrong answer and think it's right. Disaster.

Our non-algebraic approach not only makes it unnecessary for you to write and re-enter intermediate results, but gives you the continuous feedback you need to avoid ending up with a wrong answer.

It's an approach we first adapted and applied some years ago, using an unambiguous, parentheses-free language developed by the Polish mathematician Jan Lukasiewicz.

Before you invest in a pocket-sized calculator, you should look beyond its ability to solve simple problems in the most obvious way. To understand more fully the difference between our approach and the algebraic one, return the coupon for a copy of the booklet ENTER VS. =. You owe it to yourself to read it. We will also send you information on our pocket-sized calculators: the HP-35 electronic slide rule ($225*), the HP-45 advanced scientific calculator ($325*), the HP-65 programmable computer calculator ($795*), the HP-70 business calculator ($275*), and the HP-80 financial calculator ($395*).

For more information on these products write to us. Hewlett-Packard, 1505 Page Mill Road, Palo Alto, California 94304.

*Domestic USA prices only. 00434

Mail to: Hewlett-Packard, 1505 Page Mill Road, Palo Alto, CA 94304

Please send me the *ENTER VS* = booklet, together with information on HP pocket-sized calculators.

Name _____ Title _____

Company _____

Address _____

City _____ State _____ Zip _____

HEWLETT hp PACKARD

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


**OFFSHORE DRILLING: BLIGHT OR BONANZA?** Off our shores may be enough oil and gas to take care of an important part of America's petroleum needs for many years to come.

**Conoco believes we should be developing that resource full speed.**

But our governments—federal and state—are holding up exploration in some areas because of concern for the environment. And where governments are ready to move they are frequently stopped by lawsuits brought by environmental groups.

**Here are the facts:**

Of 18,000 wells drilled off our shores, only four resulted in oil spills of any consequence.

Technology has improved, and the risks of spills are less than ever.

Oil companies currently hold leases on about three percent of America's offshore continental shelf, much of it in the Gulf of Mexico. From these offshore areas we are now getting about 18 percent of domestic oil production.

It usually takes millions of dollars to explore one offshore site; and, if successful, from two to seven years to bring an established field into full production.

Conoco believes that the environmental risks are now so well in hand that the industry should be allowed to step up its search for offshore oil. This is one of the best ways we have to help reach America's goal of near self-sufficiency in energy.

If you'd like to be more informed on both the pros and the cons, write Conoco and we'll send you the following material: a booklet explaining and supporting offshore drilling published by the American Petroleum Institute; and a report of an offshore drilling conference debating both sides of the issue, sponsored by the New York Ocean Science Laboratory.

Write Conoco, Dept. B54, GPO Box 29, New York, New York 10001.

**Continental Oil Company** conoco

Established 1845 SCIENTIFIC AMERICAN October 1974 Volume 231 Number 4

# The International Control of Disarmament

*The widely acknowledged need for a separate United Nations agency responsible for verifying and controlling disarmament agreements is intensified by the current impasse in bilateral "summit" talks*

by Alva Myrdal

Two recent developments—the failure of the latest round of bilateral "summit" meetings between the U.S. and the U.S.S.R. to make any significant progress toward the immediate goal of strategic-arms limitation (let alone toward the ultimate goal of nuclear disarmament) and the announcement by India that it had become the sixth nation to successfully test a nuclear explosive—serve to emphasize the urgent need for renewed international efforts to end the continuing waste and insecurity of the arms race. As Sweden's delegate to the Political Committee of the United Nations and to the ongoing UN disarmament conference at Geneva from its beginning in 1962 until my retirement last year, I have been in a position to witness the ebb and flow of the arms-regulation proposals that have been put forward at various times by various nations. Out of this experience I have distilled a few general observations about how (and how not) to negotiate an effective arms agreement.

First and foremost it should not be tolerated that the two superpowers exercise a world hegemony based largely on their incessant arms race and at the same time play an insincere game of disarmament at the negotiating tables. Moreover, in view of the fact that objections have so often been raised against agreements on arms limitation or genuine disarmament on grounds of the difficulty or the alleged impossibility of controlling their implementation, the problem of control and verification must be tackled in a new and dynamic way.

I have therefore come to the conclusion that there is one practical measure that could serve to bind together and facilitate all disarmament efforts. That is to create a new UN agency charged with the collection and dissemination of information regarding the fulfillment by the nations of the obligations they incur under disarmament agreements and regarding ongoing changes in national armaments. Such an agency would begin its work modestly, depending on information available from national sources, but ultimately it would stand ready to accept control functions. It might even become a repository and publisher of pertinent satellite-surveillance data. I am convinced that this approach to the internationalization of knowledge about arms and disarmament would secure the foundations for the mutual confidence among nations on which the negotiation and observance of significant disarmament agreements must in the first and last instance rest. Here I shall set forth some of the reasoning behind this proposal.

The value of any agreement on disarmament or the regulation of armaments depends not only on how large a part of the community of nations subscribes to it but also on mutual confidence that it will be duly upheld. Ultimately such confidence depends on the trustworthiness of the nations that are parties to the agreement. All properly share in the concern that they should be able to rely on the other parties to fulfill their obligations. The overriding assumption must be that any government that has negotiated a disarmament (or nonarmament) agreement, and that during the preparatory period has worked to get the agreement tailored as much to its wishes as mutuality allows, will enter as a party to the agreement with no intention of breaking it or of cheating.

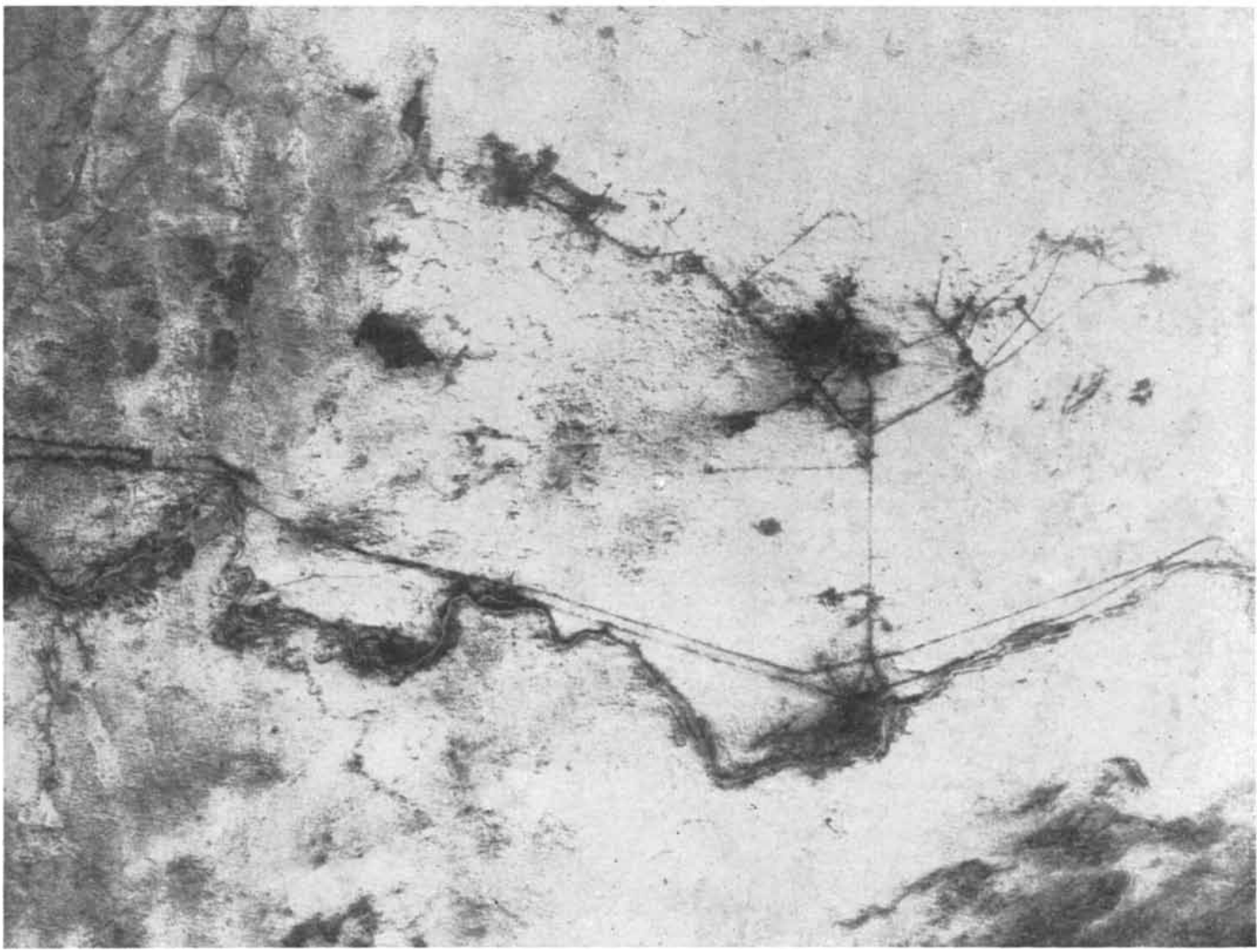
The historical record speaks for the validity of this assumption. It is doubtful, in fact, that there has ever been an instance of a clandestine violation in the arms field. Some Americans have charged that the U.S.S.R. violated the nuclear-test moratorium that prevailed at the end of the 1950's. In the first place there could not have been any such violation because there had been no formal agreement; moreover, the U.S. had already

indicated its intention of resuming testing. In the second place the Russian test explosion was not secret. Even less was India's recent nuclear explosion a violation, inasmuch as India had for 20 years worked assiduously for a ban on all nuclear-weapons testing, and when it did not sign the Non-Proliferation Treaty, it led those nations that openly expressed reservations about the treaty's discriminatory character.

In fact, open abrogations of treaties are rather more likely than clandestine ones. It is of course true, as some commentators have pointed out, that certain ultramodern weapons (for example binary chemical weapons) might be easier to conceal than nuclear weapons. Nonetheless, the main assumption retains its strength, namely that the political commitment made when entering an arms agreement is the most reliable guarantee, whether or not it is supported by technical devices for detection.

If willful violations of an arms agreement should occur, they would raise problems that would be essentially political and not technical. Collective sanctions could not be enforced against such breaches; penalties could not be meted out by an international organization, nor could punitive expeditions be put into the field. The risk is the political one that other parties will find reasons to abrogate the treaty. In arms-regulation treaties there are often provisions for such withdrawal, according to the formula that a party may decide "that extraordinary events ... have jeopardized the supreme interests of its country."

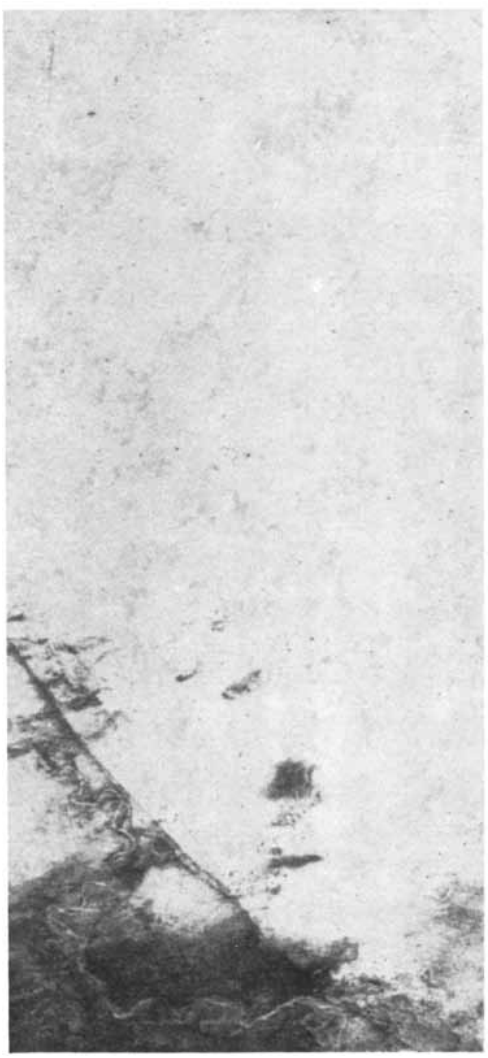
Concern about loyal adherence to agreements has been surprisingly and, in my view, irrationally vociferous during disarmament negotiations. It has caused considerable attention to be paid to questions of control and verification. The fact that the matter of control is held to be important, however, sets out the practical limits of the area of agreement. There is a plethora of differing demands for control and even different interpretations of



**RUSSIAN SPACE CENTER near Tyuratam in southern Kazakhstan appears in this U.S. satellite photograph. The photograph, which was made from an altitude of about 570 miles by a camera aboard the first Earth Resources Technology Satellite (ERTS 1), was released by the National Aeronautics and Space Administration in 1972 following a decision to make all such nonmilitary-satellite photographs available to the public on request, regardless of the possible strategic significance of the area photographed. The author argues that the same policy of openness should be extended to include the currently secret military-satellite programs of both sides, on the grounds that such an approach would encourage the mutual confidence among nations on which the verification and control of disarmament agreements depend. The ultimate repository for all such arms-surveillance data would be a new United Nations agency to be called the International Disarmament Control Organization. The military satellites of both the U.S. and the U.S.S.R. are of course capable of obtaining photographs with much greater resolution. The comparatively poor resolution of this**

what control clauses in already signed and ratified treaties really mean.

I believe two major unresolved issues are at the bottom of the confusion. One is the remaining uncertainty about what the real purpose of a control system is to be in relation to an arms-regulation agreement. I shall return to this issue. The other issue is the continued dominance of a long-standing tendency to view the problem in terms of the control deemed to be both necessary and acceptable in a totally disarmed world. It was in this context that the control problem was originally raised and discussed. Questions about what control measures are necessary and applicable to partial or collateral measures should be related to the specific conditions, which differ from case to case, if they are to be relevant to the pragmatic attitude that is now the prevailing approach to disarmament negotiations and debates.



**ERTS photograph is enhanced somewhat by the fact that it was made in winter; roads, railroads and other installations show up clearly as black lines and spots against the light-colored background of the surrounding snow-covered desert. The main "cosmodrome," or rocket-launching complex (*top center*), is approximately 15 miles north of the railway town Tyuratam (*bottom center*).**

It was on the submerged reef of obfuscations concerning the scope of control that the first major postwar arms-limitation effort on the part of the U.S. and the U.S.S.R. foundered. Since the debate over the Baruch Plan right after World War II the accusations have become stereotyped: the U.S.S.R. wants to confine control to what is being disarmed, whereas the U.S. wants to control what exists of armaments.

Although this is a rather quixotic feud, it regrettably still turns up and causes vexations at point after point. I am convinced, however, that a meeting of minds must and will occur. In their abstract form both positions are impossible. In concrete cases both positions may well be abandoned, as they were in the first strategic-arms-limitation talks (SALT I), or considerably modified, as they were in the nuclear-test-ban negotiations.

In a strict sense "control of disarmament" could be interpreted as being limited to monitoring the destruction of armaments. Indeed, in the early days of the Geneva disarmament conference there was often talk of such "bonfires" of bombers and other weapons. In the convention prohibiting the production of biological weapons the destruction of the existing stock is explicitly required. President Nixon in 1969 made the magnanimous gesture of promising such destruction of U.S. stockpiles. The fact remains that international observation teams have never been invited to witness the destruction of any weapons. The logical flaw in the demand that control be limited to the implementation of disarmament is, of course, that in order to appreciate how much has been destroyed or eliminated one has to know what share of the total it represents.

"Control of armaments retained," on the other hand, is a logically more defensible proposition. In its most literal sense—corresponding to the "bonfire" in the preceding category—it would require proof in the form of physical inspection of all facilities of any party to a treaty in order to ascertain that elimination or reduction of armaments has been effected so as to correspond to the levels agreed on. So far the Russians have been categorically opposed to incorporating the right to obligatory inspections in any of the treaties that have been negotiated. Objections might well be raised by other countries with regard to possible future arms agreements.

What is confusing the issue is to a large extent the fact that the postures taken by various nations are not clearly addressed either to a situation of general and complete disarmament or to one where disarmament is less perfect. One reason is probably, as I have mentioned, the lingering dominance of the idea of a foolproof control system. This criterion has always been associated with a state of world affairs where nations were totally disarmed and the UN had the duty of policing the world society. Such an ambitious scheme was embedded, for example, in the basic guidelines for disarmament negotiations that were submitted to the UN General Assembly in 1961 and go under the name of the McCloy-Zorin Agreed Principles. The agreement states as the sixth principle:

"All disarmament measures should be implemented from beginning to end under such strict and effective international control as would provide firm assurance that all parties are honoring their obligations. During and after the implementation of general and complete disarmament, the most thorough control should be exercised, the nature and extent of such control depending on the requirements for verification of the disarmament measures being carried out in each stage. To implement control over the inspection of disarmament, an international disarmament organization including all parties to the agreement should be created within the framework of the United Nations. This international disarmament organization and its inspectors should be assured unrestricted access without veto to all places as necessary for the purpose of effective verification."

The two contrary positions held by the superpowers can, I surmise, be disregarded in their extreme form, when (as is now the rule) all relevant negotiations have moved on to the practical consideration of varied and limited disarmament measures. Problems of verification and control are considerably easier to solve in practice than in theory—either on the political ground that enough confidence prevails or on the technological ground that sufficiently reliable means of monitoring events in the field of interest are becoming available. Even so there will often remain margins of uncertainty that may require a mustering of political will to compromise if they are not to be used as excuses for disagreement.

The futility of demanding control pro-

visions that would give 100 percent assurance that no violation of an arms regulation agreement would ever go undetected becomes much more indisputable when one considers the other continuing source of confusion, namely that there has not as yet been any clear analysis and unambiguous acceptance of what is to be the function of control and verification. In general, interest has been focused on only one function: the ability to check whether or not a violation of treaty obligations has occurred. I would call this an after-the-fact approach. The main interest here has been the legalistic one of keeping the parties accountable. The evidence sought should be so strong that it would hold up before a judiciary authority. Thus the accused party should preferably have been caught red-handed in the act of violation or should have left indisputable material proof—hence the insistence on obligatory inspections. Only if guarantees were given that such a practically foolproof system for after-the-fact verification existed, the reasoning goes, would enough trust be engendered to allow an agreement to be signed. That is still the U.S. position in regard to some important treaties, including a total ban on nuclear-weapons tests. As a result negotiations on this issue have for decades just been treading water.

Control can be regarded from a theoretically different angle, however, insofar as it has another important function, namely to prevent violations. This might be called the before-the-fact approach. The role of control and verification would then be to establish sufficient risks of disclosure that the risks as such act as a deterrent against any nation's attempting to violate a treaty obligation.

In relation to a treaty banning nuclear tests, for example, the before-the-fact approach could be effective if the parties were willing to accept the now very low margin of uncertainty that some of the lowest-yield tests might go undetected and to apply to those uncertainties a strategy of nonresponse. The risk would

| SIGNED | IN FORCE | | | NUMBER OF PARTIES |
|---|---|---|---|---|
| 1959 | 1961 | ANTARCTIC TREATY | PROHIBITS ALL MILITARY ACTIVITY IN ANTARCTIC AREA | 17 |
| 1963 | 1963 | PARTIAL-NUCLEAR-TEST-BAN TREATY | PROHIBITS NUCLEAR EXPLOSIONS IN THE ATMOSPHERE, IN OUTER SPACE AND UNDER WATER | 106 |
| 1967 | 1967 | OUTER-SPACE TREATY | PROHIBITS ALL MILITARY ACTIVITY IN OUTER SPACE, INCLUDING THE MOON AND OTHER CELESTIAL BODIES | 71 |
| 1967 | 1967 | TREATY OF TLATELOLCO | PROHIBITS NUCLEAR WEAPONS IN LATIN AMERICA | 18 |
| 1968 | 1970 | NON-PROLIFERATION TREATY | PROHIBITS ACQUISITION OF NUCLEAR WEAPONS BY NON-NUCLEAR NATIONS | 82 |
| 1971 | 1972 | SEA-BED TREATY | PROHIBITS EMPLACEMENT OF NUCLEAR WEAPONS AND OTHER WEAPONS OF MASS DESTRUCTION ON OCEAN FLOOR OR SUBSOIL THEREOF | 52 |
| 1972 | | BIOLOGICAL-WEAPONS CONVENTION | PROHIBITS DEVELOPMENT, PRODUCTION AND STOCKPILING OF BACTERIOLOGICAL AND TOXIN WEAPONS AND REQUIRES DESTRUCTION OF EXISTING BIOLOGICAL WEAPONS. | 31 |

**MULTILATERAL ARMS AGREEMENTS negotiated in recent years are listed here in chronological order, together with short statements of their major provisions. The column at extreme right gives the total number of parties to each treaty as of December 31, 1973. The convention on biological weapons had been ratified by 31 nations but was still not in force at the beginning of this year.**

strategy of nonresponse. The risk would be minimized by the factor of deterrence: the calculable probability that violations will be detected. That is obviously a far more effective method than to search for 100-percent-positive proof in every single suspected event whether or not the event is a violation.

In practical terms solutions to these theoretical riddles do not present great difficulties. The degree of trust is always of the essence when an agreement is contemplated. The before-the-fact deterrence approach presupposes a climate of confidence, but it has the comparative merit of motivating a willingness to take a first step on somewhat shakier ground, since it then allows confidence to grow gradually as it is tested by experience. To demand foolproof methods for after-the-fact verification is to impose an impossible requirement.

It can be safely stated that for both after-the-fact verification and before-the-fact deterrence a main interest is to increase the scope of knowledge about what happens to armaments as well as disarmament. It would be particularly true, however, that the more pertinent knowledge was available, the more reliance could be placed on the before-the-fact approach.

Speaking generally, the imperative task for practical policy should be to reduce the margins of uncertainty in order to change the demand for control from an impossible absolute requirement to a manageable relative requirement. Increased knowledge is really the link between the two extreme requirements concerning what to control: armaments sacrificed or armaments retained. In addition increased knowledge serves the purpose both of deterrence (the before-the-fact risk of disclosure) and of monitoring (the after-the-fact search for evidence of violations).

In this light the prospects for progress should become brighter. Where a fair degree of knowledge prevails, even without any specific verification apparatus, the margin of uncertainty correspondingly decreases, making it less likely that a would-be violator might be tempted to play a hazardous game.

How to expand the store of knowledge in fields that become sensitive when disarmament measures are being negotiated is a practical question having to do with the availability of techniques. Nonetheless, even this technological problem has a political aspect. It involves the question of what kind of world society one wants to foster for the future: one based on openness or one based on secretiveness. With respect to disarmament there is much to be said for the view that maximizing the common fund of knowledge through a total abandonment of secrecy would best serve all interests. Confidence would be encouraged and deterrence would become particularly strong if all the nuclear powers were to participate. Niels Bohr suggested toward the end of World War II that the only way to deal with the problem of nuclear arms was to totally abandon the secrecy surrounding nuclear explosions.

Even Edward Teller, a man not given to accepting such a thorough change of the ways of the world, has said that a possible path is "a gradual and well-planned abandonment of all secrecy concerning technical and scientific facts." The spread of knowledge is unavoidable anyway, and the lack of knowledge is not the main factor limiting nuclear-weapons production. The policy of openness must be actively pursued. As Teller argues: "We should at the same time exert as much pressure as we possibly can on every nation in the world that they likewise permit complete freedom for the flow of information. At the present time technical facts are subject to secrecy in many nations. We should try by every means to reverse this trend toward secrecy. Every additional secret is an obstacle to the free collaboration and the eventual union of nations. A strong and widespread condemnation of all practices of secrecy may in the long run have a strong effect even on those countries which value this form of security most."

Secrecy has many negative ramifications. It cannot long preserve a purely technical military monopoly. On the other hand, spreading exaggerated notions about an adversary's superiority or one's own insufficiency tends to cause overreaction. The notorious "missile gap" of the early 1960's was a case in point, but the phenomenon has a timeless ability to open up terrifying perspectives.

Among the important positive aspects of openness is that it both increases knowledge about ongoing developments that might threaten to step up the arms race and multiplies opportunities for verification. The most advanced means of observation have been by-products of the arms race (satellites, sonar and so forth). Regrettably these devices are now monopolized by a few nations. Ideally they belong to mankind's accumulated fund of knowledge and should be allowed to serve the interests of everyone's security.

The one dictum that is generally valid is that openness promotes all knowledge. Knowledge does not thrive in the dark. This wider aspect of the ethics of access to knowledge is beyond the scope of this article. Here I wish only to emphasize that military secrecy is the most obnoxious of all attempts to conceal truth.

Verification is no substitute for trust. The question of trust in a disarmament agreement remains preeminently a political problem, but verification can obviously contribute to trust and thus to a buttressing of international agreements. Technical means can facilitate trust even outside the domain of verification, as is evidenced by the "hot line" set up between the U.S. and the U.S.S.R. in 1963.

It would nonetheless be a mistake to rely too much on technical controls. In the early days of disarmament negotiations great store was set by "strict and effective international control" (as had been stated in the McCloy-Zorin principles), undoubtedly because it was then hoped that general and complete disarmament would emerge out of one set of decisions, in spite of the fact that the political realities were far from ripe for it. Particularly as new methods of verification with technical aids emerged, "control" sometimes was in danger of becoming a technical pipe dream. An example was the plan to build hundreds of seismological stations around the world for monitoring a test ban.

Overreliance on the inspection approach was pervasive throughout the period from 1958 to 1965, even in academic studies of arms control and disarmament. The search went on for perfect methods of control, until it looked as though what disarmament would lead to was a world police state. The consoling thought is that this approach is now virtually a thing of the past. It was based on a mixing of mental images: a scheme for general and complete disarmament introduced into a political situation that was totally inhospitable to such brave idealism.

The task as seen from the perspective of today is rather how to obtain increasingly wide and reliable knowledge of armaments and potential changes in their quantity and quality throughout the world. A reassuring openness, I would hold, is the avenue to disarmament control. Even if the goal of total openness is distant, to take steps in that direction is clearly in line with the trend of our time, which is toward increasing communication of all kinds.

In the field of interest to arms-limitation endeavors, however, the problems of verification and control have also suf-

fered from being treated too much as generalities. The search for verification methods must span a great diversity of problems. The choice of method must be specific, depending on the type of armament proscribed and the developmental stage to be verified, as well as on the technical means at one's disposal.

This requirement for deciding on specific verification methods should not exclude the usefulness of cross-checking events by employing several different approaches. Hence efforts to observe changes in hardware can be augmented not only by the examination of budget figures but also by the analysis of trade statistics. For example, trade in phosphorus can be analyzed for the purpose of monitoring the possibility of chemical-weapons production.

To date no systematic analysis has been attempted to determine what different disarmament measures would optimally require of verification methods and whether some measures might not make use of a common operative handling. Instead disparate approaches to control have been taken in an unsystematic way in the treaties and recommendations negotiated so far. The picture of

| SIGNED | IN FORCE | | |
|---|---|---|---|
| 1963 | 1963 | "HOT LINE" AGREEMENT | ESTABLISHES DIRECT RADIO AND TELEGRAPH COMMUNICATIONS BETWEEN U.S. AND U.S.S.R. FOR USE IN EMERGENCY |
| 1971 | 1971 | "HOT LINE" MODERNIZATION AGREEMENT | INCREASES RELIABILITY OF ORIGINAL "HOT LINE" SYSTEM BY ADDING TWO SATELLITE-COMMUNICATIONS CIRCUITS |
| 1971 | 1971 | NUCLEAR-ACCIDENTS AGREEMENT | INSTITUTES VARIOUS MEASURES TO REDUCE RISK OF ACCIDENTAL NUCLEAR WAR BETWEEN U.S. AND U.S.S.R. |
| 1972 | 1972 | HIGH-SEAS AGREEMENT | PROVIDES FOR MEASURES TO HELP PREVENT DANGEROUS INCIDENTS ON OR OVER THE HIGH SEAS INVOLVING SHIPS AND AIRCRAFT OF BOTH PARTIES |
| 1972 | 1972 | SALT I ABM TREATY | LIMITS DEPLOYMENT OF ANTI-BALLISTIC-MISSILE SYSTEMS TO TWO SITES IN EACH COUNTRY |
| 1972 | 1972 | SALT I INTERIM OFFENSIVE-ARMS AGREEMENT | PROVIDES FOR FIVE-YEAR FREEZE ON AGGREGATE NUMBER OF FIXED LAND-BASED INTERCONTINENTAL BALLISTIC MISSILES (ICBM'S) AND SUBMARINE-LAUNCHED BALLISTIC MISSILES (SLBM'S) ON EACH SIDE |
| 1973 | 1973 | PROTOCOL TO HIGH-SEAS AGREEMENT | PROHIBITS SIMULATED ATTACKS BY SHIPS AND AIRCRAFT OF EACH PARTY AIMED AT NONMILITARY SHIPS OF OTHER PARTY |
| 1973 | 1973 | NUCLEAR-WAR-PREVENTION AGREEMENT | INSTITUTES VARIOUS MEASURES TO HELP AVERT OUTBREAK OF NUCLEAR WAR IN CRISIS SITUATIONS |
| 1974 | | SALT II ABM TREATY | LIMITS DEPLOYMENT OF ANTI- BALLISTIC-MISSILE SYSTEMS TO ONE SITE IN EACH COUNTRY |
| 1974 | | SALT II THRESHOLD NUCLEAR-TEST-BAN TREATY | PROHIBITS UNDERGROUND TESTS OF NUCLEAR WEAPONS WITH EXPLOSIVE YIELDS GREATER THAN 150 KILOTONS |
| 1974 | | SALT II INTERIM OFFENSIVE-ARMS AGREEMENT | COMMITS BOTH PARTIES TO NEGOTIATE EXTENSION OF SALT I INTERIM OFFENSIVE-ARMS AGREEMENT THROUGH 1985 |

**BILATERAL ARMS AGREEMENTS between the U.S. and the U.S.S.R. cover mostly peripheral matters of immediate concern only to the two "superpowers." Critics of the agreements reached in the strategic-arms-limitation talks (SALT I and II) contend that the bilateral approach has succeeded only in outlawing weapons systems that neither side wanted anyway, with the net result that the arms race has merely been redirected. The SALT II treaties, which were signed in Moscow on July 3, have not yet been ratified by the U.S. Congress and hence are not yet in force; the SALT II "threshold" nuclear-test ban is not intended to go into effect until 1976.**

the controls attempted is therefore one of great disarray.

Thus the problems of verification are to a large extent concerned with the right choice of techniques and procedures. Such choices should, however, be guided by the overarching aims of maximizing trust and minimizing policing on the one hand and gradually widening the appropriate knowledge through international cooperation on the other.

It is of prime importance to resort as little as possible to intrusive methods of investigation, be they conducted by spies or through on-site inspections. Modern technology is very helpful in this regard, since it provides ever more efficient means for unobtrusively checking events from a distance. Radar and sonar, the analysis of radioactive emissions and the monitoring of telecommunication frequency patterns are just a few of the less intrusive methods of surveillance made available by modern technology. One such method has had special prominence in the recently concluded agreements, namely observations from satellites. The satellite surveillance of launchers for land-based strategic missiles, for example, made it unnecessary for the parties in the SALT I agreement to seek the right of direct observation of the missiles.

The use of sensors on satellites for making observations raises in the most acute form problems of ethics that are inherent in all attempts at control and verification. Control by whom? Because satellite observations are now virtually the monopoly of the superpowers, there must be an international demand that the observations of military satellites be made openly available, at least to the countries being observed. Only a certain sharing of data of relevance to environmental matters is in the offing so far. At a recent meeting of the UN Working Group on Remote Sensing of the Earth by Satellites the representative of the U.S. promised that data from his country's satellites would be released regularly for international use. The same demand for the internationalization of knowledge must be made in regard to observations of relevance to armaments and disarmament. The most desirable development would be direct international management of a satellite system.

Internationalization of at least the pertinent data is a *sine qua non*. It must be realized how great the inequality between nations is with respect to information that may be highly relevant for fostering mutual trust. Satellites are the most prominent example of the advanced techniques for observation that only certain nations possess. More generally, international equity demands that information should not be secretly held by the strongest powers but should be placed under international control. Great emphasis must be put on both increasing and internationalizing knowledge.

To cement the basis for mutual trust, cooperation should be directed to a gradual increase of knowledge. A wide, nonpolitical dissemination of information will serve this purpose in a general way, but there would also be occasion to build up a body of reliable knowledge in specific cases where doubt has arisen about the fulfillment of pledges.

Even in these specific cases the type of verification and control that relies on police surveillance and aims at a verdict by judicial authority is as far as possible to be avoided. Instead a gradual process of mutual contributions by the parties concerned to clarify matters should be tried. The Swedish delegation on international disarmament negotiations has sketched such a system of a largely voluntary control procedure, called "verification by challenge." The official presentation of it in brief form stipulates that under the proposed arrangement "a party suspected of having conducted an underground test, in violation of the treaty, would be expected voluntarily to offer clarifying information to allay suspicion, the assumption being that the suspected party would itself be vitally interested in establishing its innocence. An 'invitation to inspection' might be forthcoming spontaneously in some instance and under pressure in more severe cases of doubt. If such a challenge went unheeded on several occasions, other parties to the treaty would acquire the right to withdraw from it.... The threat of withdrawal might induce the accused party to offer clarification of the suspected event, or if the accusation persisted, to invite inspection. The system of 'verification by challenge' would be useful whether or not obligatory inspections were envisaged in the treaty. If obligatory inspections were envisaged, 'verification by challenge' would help reduce the size of the unresolved problem, and if inspection were not envisaged, it would help resolve suspicions."

In more concrete terms it can be said that in this scheme of verification by challenge the sanction of publicity in effect replaces the sanction of on-site inspection. Thus the method of deterrence is expected to be effective. The key to the entire problem of disarmament control is not reliance on specific formulas in a treaty but the construction of a firm basis of universal confidence through a cumulative process of fully shared factual information.

All the arguments advanced in the foregoing point to the need for some international agency to which the functions of verification and control can be entrusted. That some such body is necessary has also been acknowledged in all the more general plans for disarmament.

Sometimes these plans have been ambitious enough to aim at a kind of "managing agency for disarmament." That was the idea indicated in the McCloy-Zorin Agreed Principles, which called for an international disarmament organization within the UN. No date was set for its creation, but the demand that "the most thorough control should be exercised... during and after the implementation of General and Complete Disarmament" suggests an early inception of the control machinery.

The two draft treaties on general and complete disarmament submitted in 1962 by the U.S. and the U.S.S.R. respectively differed on several points having to do with inspection and control but not with respect to the establishing of an international disarmament organization, which both regarded in principle as being fundamental. They did not differ much on the timetable either. The U.S.S.R. wanted such an organization to begin operating "as soon as disarmament measures are initiated," which is not much different from the U.S. formulation to establish it "upon the entry into force of the Treaty." In the painfully protracted but unproductive negotiations that have followed, not much attention has been given to this organization, which in its draft shape was quite a full-fledged one with a complex organizational pattern.

Several tentative proposals along less ambitious lines have been made, however. Thus an early initiative was taken in 1962 by the nonaligned delegations to the Geneva Disarmament Committee to set up an international commission of scientists for the specific purpose of verifying a test-ban agreement. Soon thereafter the Swedish delegation pleaded that such a scientific commission be set up on an interim basis even before an agreement was reached. In 1972 the delegations of the Netherlands, Sweden and Yugoslavia made separate proposals for a similar institution, this time in relation to the verification of a prohibition on the production of chemical weapons.

In 1973 I widened the suggestion to


encompass all aspects of disarmament control and monitoring. A similar position was endorsed in a statement by P. H. Kooijmans of the Netherlands. My proposal that a central disarmament-control organization be set up immediately was motivated by two main concerns. The primary justification for the establishment of some organizational framework on a preparatory basis was that we risk, through *ad hoc* methods of work, heaping on one another control arrangements that remain unintegrated and even unrelated to one another. The second thesis was that we need such an organ as soon as possible because, in addition to following the implementation of control arrangements that are already presumed in various treaties, a constant watch should be kept on progress in the general direction of disarmament.

In the wake of the laborious search for some international agreement incorporating at least a semblance of disarmament, signs have become visible that international cooperation on control issues might become feasible. The nonaligned nations, supported in general terms but not strongly by the Western side, have struggled in vain to get such operational cooperation under the aegis of the UN Secretary General incorporated in some treaties. In the current discussions of a comprehensive test-ban agreement the socialist countries also seem to have moved somewhat closer to acceptance of organized international cooperation with respect to the exchange of data. An interesting signal was detected in relation to a prospective treaty on chemical weapons in a working paper by seven socialist states. That side has always been the one upholding the view that "national means" were sufficient for verification. In this paper a quite comprehensive system of internal control was outlined, based on national committees of control, but also

**VERIFICATION PROCEDURES relied on to ensure compliance with the recent multilateral arms agreements have varied widely, as can be seen in this tabular presentation prepared as part of a working paper submitted last year by the Swedish delegation to the UN disarmament conference in Geneva. In the author's view the task of negotiating specific verification and control measures for specific disarmament agreements is much easier in practice than in theory, "either on the political ground that enough confidence prevails, or on the technological ground that sufficiently reliable means of monitoring events ... are becoming available."**

with a "possibility for publication of reports for general information," the system also in certain cases to be accompanied by the voluntary exchange of information among states.

A significantly more active role for this international data-exchange system, presupposing also an international committee to furnish guidelines for making such reporting comparable and standardized, was indicated by a group of Russian scientists in a paper on chemical disarmament published by the Stockholm International Peace Research Institute (SIPRI) in 1973.

My proposal is nominally attached to the idea of an international disarmament organization, since such an organization was in principle accepted by the UN as early as 1961, but the plan as I now want to present it envisions a system much more than an agency. Such a comprehensive system must be organically and hierarchically built up from the national level to various international levels, of which the one with judiciary functions should be not only the highest but also the last one resorted to. The design of the proposed system corresponds to a structure with four levels.

The broad base of such a structure must be the national means of detection and verification that are organized within countries not only for internal purposes (for the control of poisonous chemical agents, weapons production and the like) but also for monitoring events abroad by remote observation (by means of seismological stations, analysis of radioactive particles in the air and so forth). All such data should be published.

Questions referring to the machinery needed for control at the national level will obviously be handled by each state according to its own traditions. In all countries with aspirations to democratic government there is already one parliamentary standing committee or more scrutinizing the activities of the military, with decision-making power closely tied to the sanctioning of allocations.

Parliamentary action, however, is rarely enough to keep an issue under vital civic control. Thus in the working paper I have mentioned the seven socialist states call for the establishment of national committees comprised of "representatives of governmental and public organizations" including "specialists in chemistry and economics" with broad responsibilities in relation to the production of chemical means of warfare. In Sweden a citizens' committee has been appointed to scrutinize the acquisition of weapons from the viewpoint of compliance with the international humanitarian rules of war.

The ideal of openness should be pursued first and foremost at the national level. Citizens should be counted on to serve as the watchdogs ensuring that disarmament agreements are respected and that any moves in the direction of militarization are pushed back. When information on all matters relating to arms and arms limitation is spread openly within a nation, the information usually becomes available to other nations as well. It would obviously be more of a building block for a world system of verification if such nationally assembled information could be regularly, systematically and as a matter of routine addressed to an international center for further dissemination instead of finding its way sporadically to more than 100 countries. A broad flow of information would enhance everyone's sense of security—and security, after all, is the main objective of all verification.

The most general feature of such an approach would include greater publicity about arms production, arms deployment and arms trade. A kind of unobtrusive, broadside monitoring of disarmament dynamics would be gained by the more detailed publication of defense budgets. More specific information can be requested by separate agreements, including information obtained by methods of distant detection.

The crucial question then arises of whether this requirement calls for the creation of an international disarmament organization for centralizing all pertinent information emanating from national sources, or whether it is sufficient that such information be relayed to one of the likewise international but delimited organs in the appropriate specialized field. There are, however, few such bodies already in existence that have been given the function of monitoring matters of relevance to disarmament. Only the International Atomic Energy Agency has a surveillance function so far, and that is limited to surveillance of the production of nuclear weapons. For chemical weapons, where the need for an international control body is currently being discussed, no suitable existing agency has been indicated. Other agencies that exist or might be created are likely to have constitutions and constituencies of their own, often outside the framework of the UN, and they might not want to add the monitoring of disarmament to their functions.

Moreover, the areas to be covered for civilian benefits and for arms control are often not at all congruent. This would be the case with an International Ocean Regime, if it is established after the Law of the Sea Conference concludes its work. It might, of course, be of assistance in monitoring the deployment of unwanted marine installations (for example platforms for military refueling that may be forbidden). Such a regime, however, must be limited to surveillance in its own environment; it cannot serve as a guard against the land-based production and development of vessels or devices that may have to be ruled out for the sake of disarmament. Marine development, environmental protection and disarmament nonetheless have a joint stake in the internationalization of knowledge.

There are large areas of interest to disarmers that would remain outside the purview of even the most ambitious scheme for a world system of transnational agencies. Here a two-pronged approach is imperative. On the one hand, the access to technologies, particularly advanced ones requiring instrumentation beyond the national means of every country, must be internationalized together with the institutionalizing of sectorial networks for the wide dissemination of scientific and technical data. This would greatly benefit the accumulation of knowledge on developments related to armaments and disarmament. On the other hand, the specialized need for keeping disarmament matters in focus makes it inescapable that an international disarmament control organization must be established, possessing from the outset enough independent status and access to scientific expertise to command respect for objectivity and impartiality.

My proposal is therefore that a new body should be set up to receive such notifications and to perform all kinds of services for the cause of disarmament. It would be the second level in a control system, but it would be the first international level. I suggest that it be called the International Disarmament Control Organization. Its immediate function should be to act as an intermediary, or a clearinghouse, for providing knowledge about the implementation of disarmament agreements and also about spontaneous developments in the direction of disarmament. Such knowledge could be provided by member states using their national means of detection, verification and control.

As a supplemental activity the center should collate knowledge derived from various open sources: from scientific publications and statistics on production,

trade and the transfer of arms and sensitive raw materials or commodities. It should profit by close cooperation with international agencies established for specialized purposes. It could also provide guidelines and even establish a more systematic international program for collecting national information, including the setting of rules and procedures for making the data of verification comparable. The process of verification itself, however, is conceived of as being conducted by national authorities within individual states.

An International Disarmament Control Organization might, particularly at the start, function as an open clearinghouse at the service of all nations by relaying information on disarmament matters to them in the form of bulletins. If its competence were broad enough to include its serving as a channel for an obligatory input of data from member states, it would have considerably more impact than the few existing international institutions, such as SIPRI, that have been set up independently to serve a similar purpose.

So far no mandatory verification by special investigations has been made a part of the proposal for an International Disarmament Control Organization. There will, however, be a need for extending the function of the organization to take on the responsibility as coordinator and referee for collecting evidence of a more active, investigatory character. Even for the future, with the possibility of enhanced world cooperation, it would seem necessary to have an International Disarmament Control Organization functioning for the purpose of directing specific inquiries to existing specialized bodies in cases of disagreement on questions of compliance with disarmament treaties.

The most critical question is how such a disarmament monitoring service could move toward a more inquisitive process of verification in cases where doubts have been expressed. Such activity would correspond to a third level in the control system. A difficulty is introduced by the fact that different disarmament treaties have, and will continue to have, quite differently composed constituencies or parties. Here the solution must be that the same secretariat serve a circle of "collateral constituencies," consisting of the full membership of parties concerned in each treaty. Their constituencies would assemble in regular review conferences, which several treaties already prescribe but which have so far not been given any organizational basis.

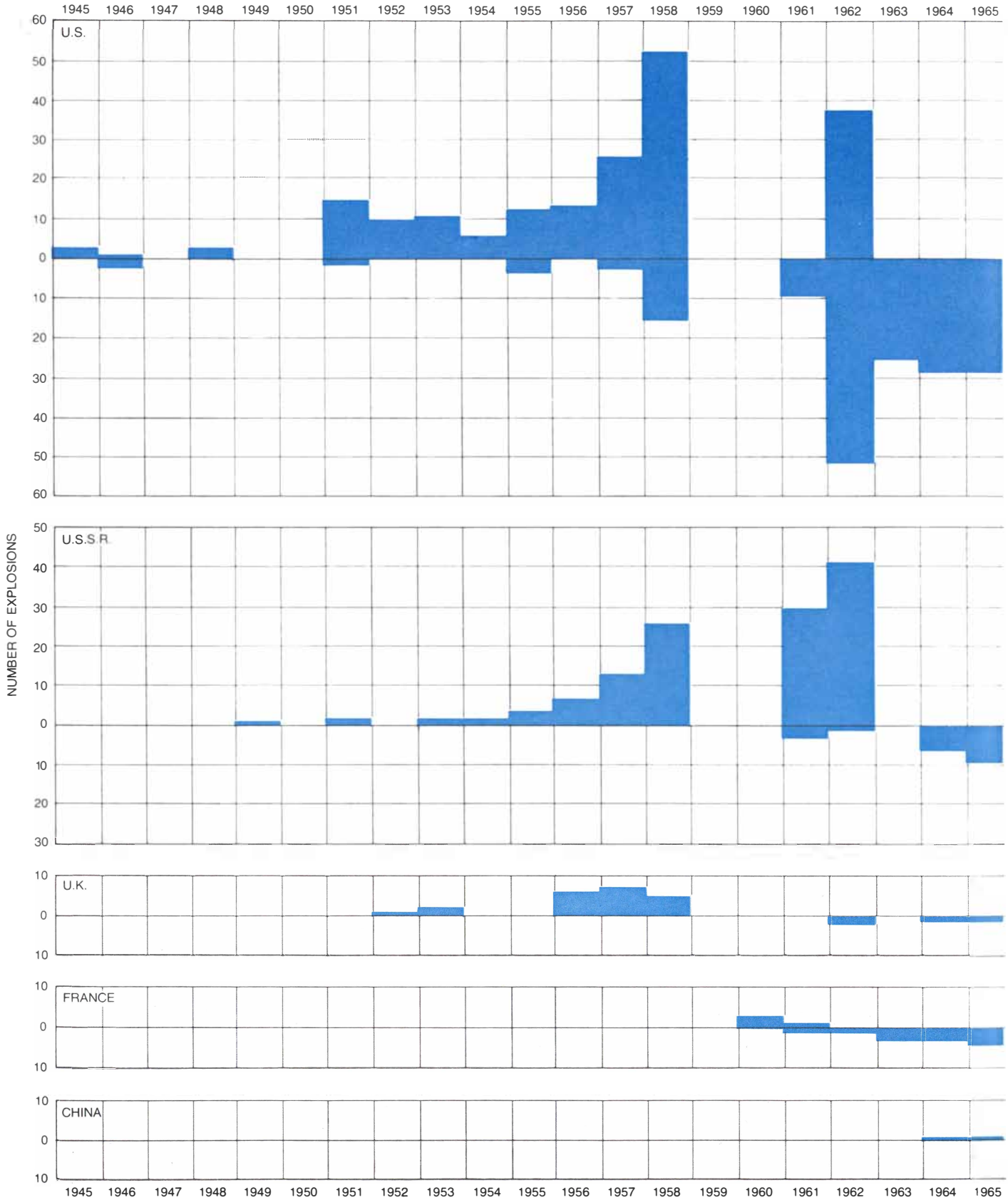
Such a separation of constituencies becomes particularly necessary when there is dispute about whether or not a specific treaty has been violated. The "verification by challenge" procedure would then begin by the complaining party's directing a query to another party (or other parties) through the good offices of the International Disarmament Control Organization. If the series of challenges and responses does not lead to satisfactory results, an investigation might be appropriate. If it were, the international control body would be the correct organ for handling it. For the actual "field investigations," if such a procedure is in order, it may well be that the international body should entrust them to various existing expert bodies or to *ad hoc* expert groups. The material collated and

**SIGNERS AND NONSIGNERS of the Non-Proliferation Treaty are identified on this world map, the signers in gray and the nonsigners in color. The principal nonsigners are China,**

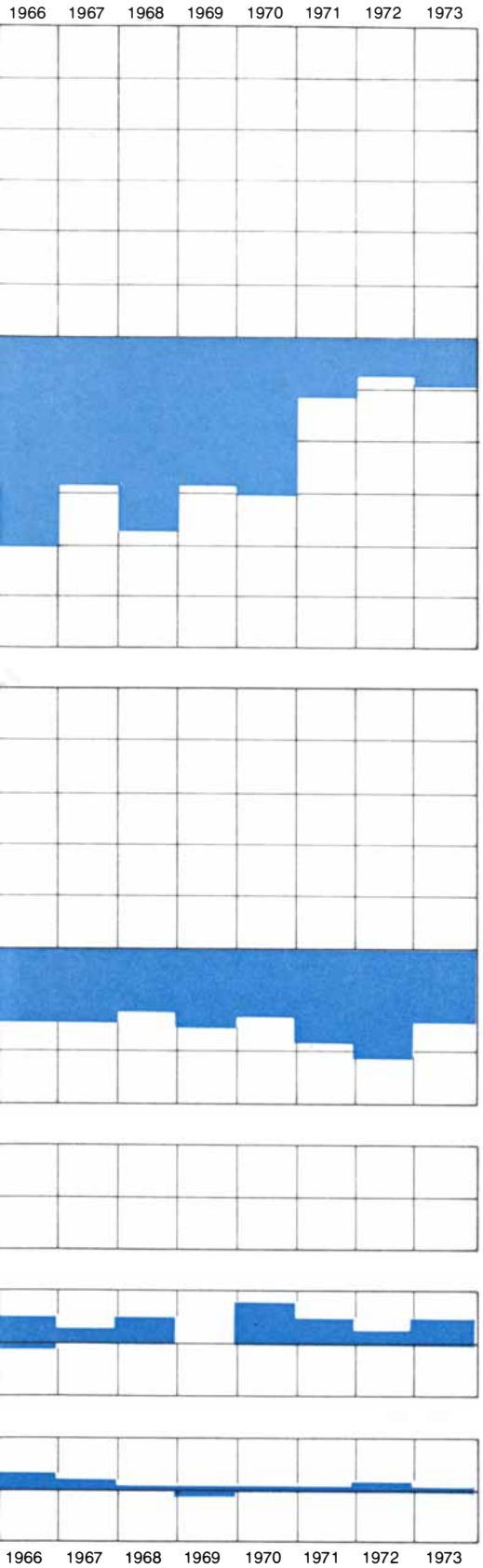
accumulated by the international organization could then, if one of the parties wanted to lodge a complaint about treaty violation, accompany such an accusation to the appropriate judiciary organ. Presumably that would be the UN Security Council, which can thus be conceived of as being the fourth tier in the control system. It would be important, however, to maintain a strict separation of powers: the International Disarmament Control Organization should never itself pronounce verdicts. It should only assemble, collate, coordinate and transmit data.

The separation of the investigative and jurisdictional functions, referring them to different organs, must be made clear and explicit. The Security Council should not be asked to take on this dual responsibility. Only the creation of an International Disarmament Control Organization, independent of the Security Council with its veto powers, can guarantee a fair, equitable and expert collation of evidence. The lack of such an institution is a fundamental deficiency at this time, giving arguments to those who are reluctant to enter disarmament agreements. During the debate in the

**France, India, Israel and South Africa. Several of the signers have delayed ratification, and some (for example Egypt) may not at present have any intention of ratifying. There are now 84 full parties to the treaty together with 22 signers that have not (yet) ratified.**




**KNOWN NUCLEAR EXPLOSIONS** from 1945 through 1973 are tallied by nation in this bar chart according to whether they took place in the atmosphere (*bars above base lines*) or underground (*bars below base lines*); the latter category includes a few cases of underwater explosions. The totals, which were compiled by the Stockholm International Peace Research Institute, are based not only on official announcements by the nations responsible but also on seismic monitoring and other data-gathering techniques employed by various nations, including Sweden. The totals do not include 33 nuclear tests known to have been conducted by the U.S.S.R. before 1958, for which exact dates are unavailable, nor do they include 23 unspecified U.S. explosions that are known to have




**occurred between 1961 and 1963. The abrupt change from atmospheric testing to underground testing by the U.S., the U.S.S.R. and the U.K. coincides with their signing of the Partial-Nuclear-Test-Ban Treaty in 1963. The 1974 totals would of course include India.**

UN on the convention prohibiting the production of biological weapons strong objections were raised against the handing over of decisions concerning investigations directly to the Security Council, with its veto power to protect allies and friends.

There are convincing arguments for not postponing action on establishing such an organizational nucleus. There is, in fact, no reason to wait for agreements in treaty form before starting to monitor what happens in one field or another of interest with regard to disarmament. The sequence can just as well start with the monitoring. Indeed, monitoring activities for the verification and control of arms-regulation measures might well stimulate the emergence of formal agreements. If a target date should be set for launching such an organizational innovation for the sake of disarmament, I would suggest the Review Conference for Non-Proliferation of Nuclear Weapons in May, 1975, when the need will make itself felt strongly and urgently.

In regard to the form of organizational structure, it would be of value to examine whether one might follow fairly closely the pattern set for continuous international work on problems of the environment. The general constituency must be the same worldwide one in both cases, encompassing the full range of the UN membership.

An interim International Disarmament Control Organization need not be large. Costs should be kept to the minimum of administrative necessities. Unlike the UN Environment Program, the International Disarmament Control Organization would hardly need a fund for operations. Like the Environment Program, however, it should have a semi-independent status within the UN, its budget allocation being underwritten by the UN. Such a relative degree of independence is particularly needed for an agency whose entire usefulness hinges on its objectivity and freedom from political shackles.

In sum, it appears highly desirable that a new international body should be set up for detecting possible transgressions of international rules in the field of arms and for following changes in the quality and quantity of armaments in various parts of the world. Such a body would be responsible for collating the relevant data, initiating specific inquiries, disseminating its findings and acting as a referee for those who are parties to various treaties. Such a body might also have to order the installation of specific equipment as well as produce manuals and guidelines for the art of monitoring. An agency of this type can be expected to grow on the basis of its own merits as a service organ.

As for deterrence—the threat of disclosure that would discourage clandestine violations—the process could take a more general, "passive" course. The new agency could then link up with existing or prospective sectorial international agencies, collating and distributing data from many sources. The main requirement for deterrence is that knowledge be widened and internationalized; the more information that is amassed and the more transparent all fields of human activity become, the more repellent becomes the risk of disclosure of any activities that are formally forbidden or even those that just go against the sense of what is commonly accepted as being internationally right—of being in the true interests of mankind. The denser the data exchange in a field is, the more it can serve as an alarm system, awakening some nation, some organization, some individuals to take action to get more of the unwanted developments prohibited.

Here is a promising connection between the suggested more informal approach to the control problem and the formalization of prohibitions by treaties with their specific control obligations. My contention is that such formalization will be eased when knowledge becomes more reliable, and when a consensus on desirable disarmament measures becomes firmer. These trends grow together, mutually reinforcing each other.

Moreover, it should be emphasized that any treaty prohibition extends beyond the parties who have gone through the formal process of ratification; the treaty tends to take on a character of customary law and can often be argued to be binding on all. At least the psychological and "ethical" impact is not to be minimized. There is a true bond of universal interest in the verifiability of how loyal and widespread is compliance not only with prohibitions but also with rules for good conduct among nations.

Initiatives allowing the UN to study the most appropriate forms and functions for fulfilling the dual task of spreading knowledge about armaments and of verifying disarmaments are the most urgent ones that I should have wanted to press if I were still a disarmament delegate. A proposal for an International Disarmament Control Organization ought to be forthcoming from some farsighted member nation, such as the nonaligned countries have often proved to be.

# The Structure of Emission Nebulas

*An emission nebula is a cloud of interstellar gas that shines by its own light. Each kind of ion in it has a characteristic color, and three-color photography reveals how the ions are distributed*

by Joseph S. Miller

More than 90 percent of the matter in our galaxy is in the form of stars, but the small fraction represented by the tenuous gaseous medium in the space between the stars has an importance for astronomy that is far out of proportion to its relative abundance. This interstellar medium is the raw material out of which new stars are being formed. It consists not only of matter left over from the formation of stars but also of matter that was once in the interior of stars and was then returned to interstellar space. The medium is therefore part of a cosmic recycling process, and it is closely studied for clues to what happens in the interior of stars over the course of stellar evolution.

The interstellar medium is mostly invisible, but here and there it is illuminated by the light of stars or it shines with its own light. In the latter case the atoms of the interstellar gas have been ionized—stripped of one electron or more—by the radiation of hot stars nearby. These luminous patches are the emission nebulas.

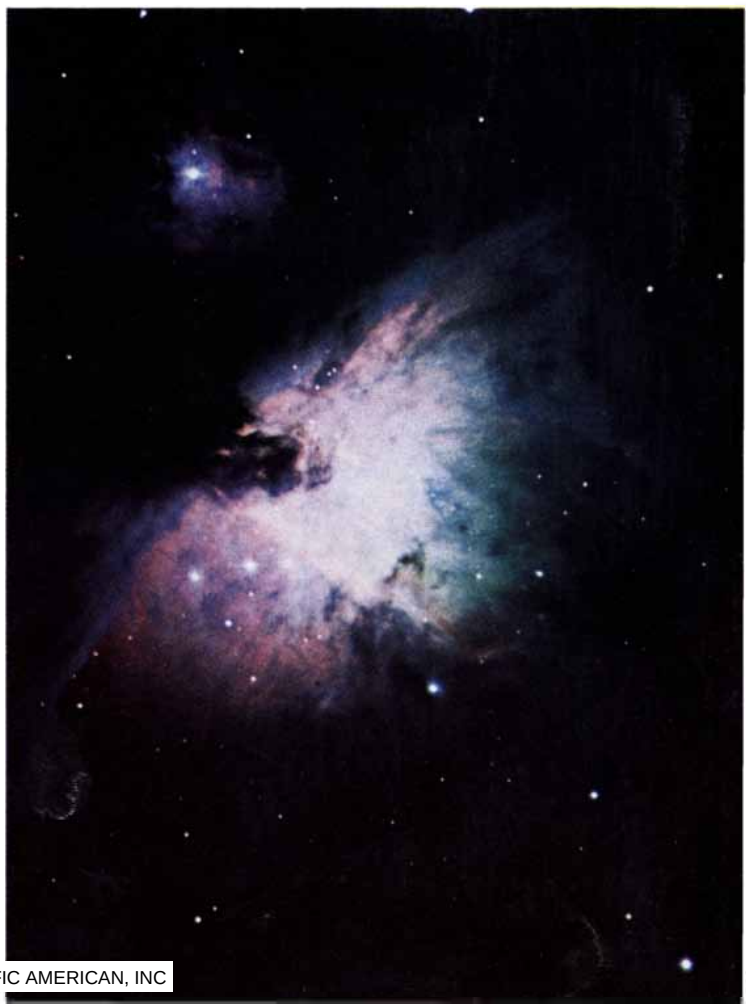
The emission nebulas can be investigated in a variety of ways: by direct photography, by spectroscopy and by observations in the radio region of the electromagnetic spectrum. Direct photographs provide information on the gross morphology of the nebula but typically yield little knowledge of the physical conditions within the gas. Spectroscopic and radio observations reveal the physical conditions but usually only in a small number of places within the nebula or as an average for the nebula as a whole, so that the relation between the physical conditions and the morphology is seldom well defined. My colleagues and I at the Lick Observatory have been taking another approach: color photography utilizing the dye-transfer process rather than standard color emulsions. The result more closely approximates the color response of the eye than that of the standard emulsions; moreover, the balance of colors reveals the ionization structure of the nebula.

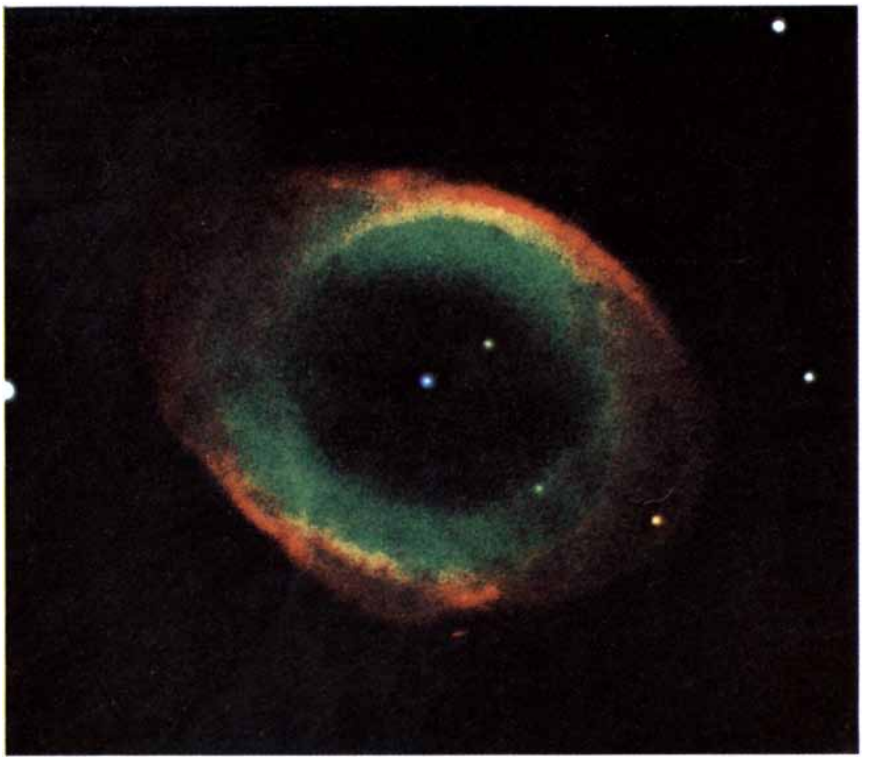
The existence of interstellar matter was not generally recognized until the latter half of the 19th century. It had been known for some time that certain areas of the sky, largely confined to the Milky Way, were strikingly deficient in stars. William Herschel and other astronomers of the late 18th century and the early 19th interpreted such areas on the basis that the stars were simply not there, that these were relatively empty tunnels through the general distribution of stars. It was also known that there were nebulous objects in space that could not be resolved into stars by telescopes.

It was not until 1874, when the English astronomer William Huggins observed that a nebulosity in the constellation Draco had a spectrum characteristic of a rarefied gas, that it was established that some of the nebulous regions were definitely not composed of unresolved stars. Fifty years later the work of Robert J. Trumpler at the Lick Observatory convincingly demonstrated that there was also particulate matter in space capable of extinguishing the light of stars, and that most of the apparently empty places along the Milky Way were in fact regions with a high concentration of such particles. The clouds of interstellar dust grains normally reveal themselves by blocking the light of the stars behind them, but a few shine by the reflected light of nearby stars. These reflection nebulas are useful for studying the properties of the interstellar dust grains, but they are distinguished from the true emission nebulas, which shine as a result of the atomic processes going on within them.

There are three main classes of emission nebulas, although the atomic processes that give rise to their radiation are basically the same. The first class is made up of the H II regions, clouds of ionized hydrogen associated with places in the galaxy where stars have recently formed and where their radiation ionizes the gas near them. H II is the spectroscopic designation for the ionized hydro-

**DYE-TRANSFER PROCESS was used to make a carefully controlled color photograph of the Great Nebula in Orion. The Orion nebula is an H II region, a region where stars have just formed and where their ultraviolet radiation gives rise to ionized hydrogen (H II). Three separate black-and-white photographs were made of the nebula with the 36-inch Crossley reflecting telescope at the Lick Observatory. Each photograph was exposed through a different color filter and then made into a mat that can absorb a dye. Each mat was dipped into the dye that is the complementary color of the filter used for the photograph: a cyan dye for the red exposure (*top left on the opposite page*), a magenta dye for the green exposure (*top right*) and a yellow dye for the blue exposure (*bottom left*). Then the mats are placed one at a time in careful registration on a paper that absorbs the dye. The end product (*bottom right*) is a photograph that not only shows how the nebula would appear to the unaided eye but also reveals the distribution of its light-emitting ions. For example, the green color is radiated by doubly ionized oxygen (O III), that is, oxygen atoms stripped of two electrons. The reddish color is largely due to hydrogen-alpha emission, a wavelength radiated by the hydrogen atom when its electron drops from the third energy level above the ground state, or state of lowest energy, to the second energy level. The blue-gray color at the eastern (*left*) edge of the nebula is starlight reflected from grains of interstellar dust.**





**RING NEBULA IN LYRA, photographed with the 120-inch reflector at the Lick Observatory, is an example of a second type of emission nebula. It is a planetary nebula, which results from an old star ejecting its outer layers. The blue star in the center of the nebula is responsible for ionizing oxygen to form O III out to a certain radius (*green regions*), at which point hydrogen-alpha emission and emission from N II dominate picture (*red regions*).**

**DUMBBELL NEBULA IN VULPECULA, photographed with the Crossley reflector, is another planetary nebula whose central blue star is responsible for ionizing the surrounding gas. Between the central star and the red regions there is a substantial amount of hydrogen, so that the energetic photons capable of doubly ionizing oxygen are depleted at a point well inside the edge of the nebula. Between the central star and the green regions, however, there is less hydrogen, and the oxygen is doubly ionized out to the edge of the nebula.**

gen atom; H I is neutral, or un-ionized, hydrogen.

The second class of emission nebulas consists of the planetary nebulas. Such a nebula is a sphere or shell of gas that has been expelled by a star in a late stage of its evolution; the gas is ionized by the small, hot star that remains in the center. The term planetary nebula comes from the fact that early observers using small telescopes perceived these round objects as resembling planets rather than the normally pointlike stars. The third class of emission nebulas consists of the remnants of supernovas, massive stars that have violently exploded near the end of their existence. Supernova remnants shine by at least two different ionization mechanisms, to which I shall return.

The three classes of emission nebulas present us with objects that can be studied from several viewpoints. First, they are objects of extremely low density and furnish an opportunity to investigate how low-density gases are ionized and excited, and how they interact with radiation. Second, detailed observations of their spectra can yield information on their chemical composition. Third, since the various classes of nebulas are associated with different stages of stellar evolution, they can provide information about important events in the life cycle of stars.

The source of the energy that heats and ionizes the H II regions and the planetary nebulas is the same: ultraviolet radiation from the hot stars within them. Since hydrogen is by far the most abundant element in interstellar space, an understanding of its ionization is central to investigating the physical processes in the nebulas.

It takes 13.6 electron volts (eV) of energy to completely strip the single electron of the hydrogen atom from the atom when the atom is in its ground state, or lowest energy level. Since this energy is high compared with the energy that is typical of either photons (quanta of radiation) or subatomic particles in interstellar space, the majority of the interstellar hydrogen in our galaxy is in the neutral un-ionized form. Stars with a surface temperature of more than 25,000 degrees Kelvin, however, emit a substantial amount of ultraviolet radiation, the photons of which have energies above 13.6 eV. When a hydrogen atom absorbs a photon with an energy greater than 13.6 eV, the electron is ejected with a kinetic energy equal to the difference between the energy of the impinging photon and 13.6 eV. The inverse of this photoionization process, recombination, is also possible: an ionized hydrogen

CRAB NEBULA, photographed with the Crossley reflector, is an example of a third type of emission nebula: the remnant of a supernova. The central star of the Crab Nebula exploded in A.D. 1054, throwing off some of its mass. The bluish central portion is an area where electrons are being accelerated in a strong magnetic field and are emitting synchrotron radiation. The most energetic photons of the synchrotron radiation are doubly ionizing oxygen atoms and are causing the filaments closer to the nebula's center to glow green.

CYGNUS LOOP, also called the Veil Nebula, is the remnant of an old supernova that exploded some 50,000 years ago. This photograph, made with the Crossley reflector, shows only a fifth of the entire remnant, which covers an area in the sky three degrees across (six times the diameter of the full moon). Its light originates by a mechanism different from the synchrotron process. As the remnant expands its atoms collide with the atoms of the interstellar medium, forming a shock front. The material of the interstellar medium is swept up, ionized, compressed and heated by the interaction. It then slowly cools as the shock front moves on. As it cools, electrons recombine with the ions of the interstellar medium and emit radiation in the visible region of the electromagnetic spectrum.

atom (a proton) captures an electron, emitting a photon as it does so. For a region in interstellar space to be in equilibrium the rate of photoionization must equal the rate of recombination.

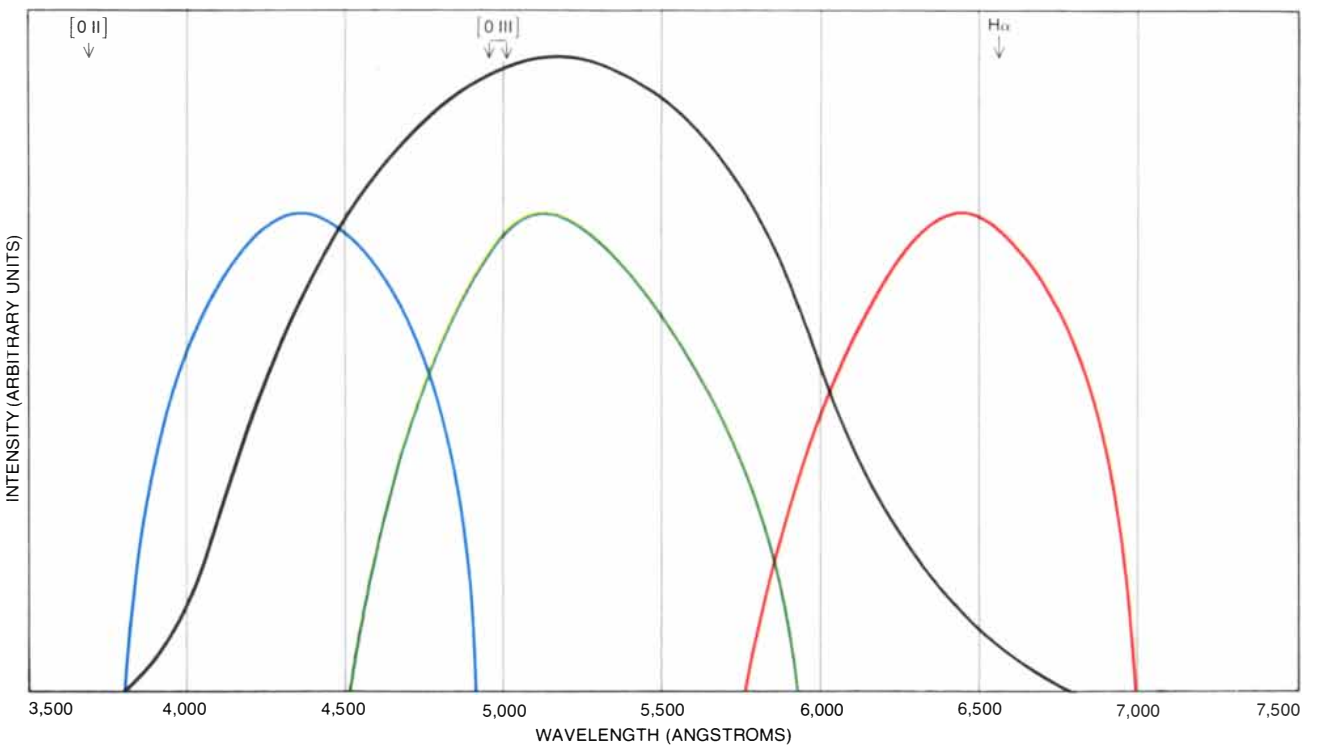
The processes of photoionization and recombination can be depicted graphically in energy-level diagrams [*see illustrations on pages 40 and 41*]. The electron can be captured at any energy level of the atom and then cascades rapidly to the ground state. It can follow any one of many downward energetic paths, but it is possible to accurately calculate what will happen statistically in a large number of recombining atoms. The results of these calculations, which are in excellent agreement with observations, indicate that the transition from the hydrogen atom's third energy level to the second will produce the most intense visible radiation. That radiation is in the red region of the visible spectrum at a wavelength of 6,563 angstroms and is designated hydrogen alpha.

The pioneering work of Bengt Strömgren in the 1930's on the theory of the photoionization of hydrogen demonstrated that if the star were embedded in a uniformly distributed gas of low density, the hydrogen would be almost completely ionized out to some particular distance from the star, at which point there would be a thin transition zone between the ionized gas and the neutral gas. This distance, the Strömgren radius, is determined chiefly by the density of the gas and by the total number of ionizing photons emitted by the star. For the star and the gas to be in equilibrium the rate at which atoms are ionized in the entire volume must equal the rate at which the star emits ionizing photons; the rate of recombination of the ions and electrons depends on the square of the density of the gas and to some extent on its temperature.

Since the relative number of photons emitted whose energy is greater than 13.6 eV is quite sensitive to the star's temperature, there is a wide range of possible Strömgren radii for the hydrogen in emission nebulas. If the amount of gas around a hot star is small enough for the star to have at least enough photons to ionize the entire cloud, the nebula is said to be gas-bounded. Alternatively, if the star does not generate enough radiation to thoroughly ionize the cloud, the nebula is said to be radiation-bounded. Most H II regions are believed to be radiation-bounded; planetary nebulas can fall into either category.

Elements other than hydrogen are ionized in an analogous way. These elements can be divided into two types: those that have an ionization energy that is less than that of hydrogen (13.6 eV) and those that have an ionization energy that is greater. The elements with an ionization energy less than that of hydrogen are ionized throughout most of interstellar space because they do not have to compete with hydrogen for high-energy photons. Those with an ionization energy greater than 13.6 eV, however, must compete with hydrogen, and their Strömgren radii in a particular nebula are generally equal to or less than the Strömgren radius for hydrogen.

For example, helium requires photons with an energy of at least 24.5 eV to remove its first electron to form singly ionized helium (He II) and 55 eV to remove its second electron to form doubly ionized helium (He III). Most H II regions and planetary nebulas contain enough high-energy photons to ensure that helium will be at least singly ionized throughout most of the volume in which hydrogen is ionized, but only the hottest stars found in the center of planetary




COLOR RESPONSE OF THE EYE (*black*) is compared with the color response of the three separate blue, green and red photographs used in the dye-transfer process for constructing the color photographs of the gaseous nebulas. Also indicated are the emission lines of singly ionized oxygen (O II) at the wavelength of 3,727 angstroms, the two lines of doubly ionized oxygen (O III) at the wavelengths of 4,959 and 5,007 angstroms and the line of hydrogen alpha ($H\alpha$) at the wavelength of 6,562 angstroms. The brackets indicate that the two lines of oxygen are "forbidden" transitions, that is, transitions that are not normally observed in laboratory spectroscopy. Standard color emulsions tend to be insensitive in the green region of the spectrum and thus do not record emission of O III. Neither the eye nor the blue exposure is sensitive to the emission line of O II in the near-ultraviolet region of the spectrum.

nebulas are capable of producing large regions of He III. Light of a wavelength that is characteristic of an electron recombining with He III to form He II is never seen in the H II regions of our galaxy. Even in the planetary nebulas with the hottest central stars the Strömgren radius of the He III is always smaller than the radius of H II, a direct result of the relatively few photons with an energy above 55 eV emitted by any star [*see illustration on page 42*].

For every 10 atoms of hydrogen in interstellar space there is one atom of helium, but the next most abundant elements (carbon, nitrogen and oxygen) are only a thousandth as abundant as hydrogen. They too are subject to the photoionization-recombination process, but because of their low abundance their ability to compete with hydrogen and helium for high-energy photons is small. The atoms that do manage to become ionized have a characteristic Strömgren radius for each state of ionization. The zone of highest ionization is usually closest to the central star. The ionization structure of an emission nebula is therefore the result of a fairly complex interplay between the radiation emitted by the central star and the depletion of this radiation as it proceeds outward and ionizes the gas.

What factors determine the temperature of the gas at a particular point in an emission nebula? Again the individual processes are fairly simple and the total situation is complex because a large number of processes are at work simultaneously. Basically the temperature at any point is determined by the balance of heating processes and cooling processes. Heating results directly from photoionization: the electron ejected when an atom is ionized interacts in a matter of seconds with other electrons and ions in the gas, contributing its energy to the overall kinetic energy of the gas. Recombination, however, favors the lower-energy, slower-moving electrons. The radiation given off by the electron as it recombines generally has no effect on the energy balance of the gas because it ultimately escapes from the nebula. The net result is that the gas as a whole gains kinetic energy, since on the average the electrons being recaptured from the gas by recombination are less energetic than the electrons being added to the gas by photoionization.

If photoionization and recombination were the only processes inside a nebula, the gas could heat up to a temperature of more than 100,000 degrees K. There are, however, other heat-loss mechanisms that keep the temperature in the range between 5,000 and 20,000 degrees. The most important of these mechanisms is the inelastic collision of electrons with partly ionized atoms of elements heavier than hydrogen and helium, that is, collisions where an electron gives up some of its kinetic energy to excite the electrons still bound within the ion to higher energy levels. The bound electrons then drop back to the ground state and emit radiation that leaves the nebula.

Since a typical electron in the gas has a kinetic energy of only a few electron volts, only those atoms that have energy levels within a few electron volts of the ground state are readily excited by inelastic collisions. Therefore the small amounts of neutral hydrogen and helium present in the nebula are of no importance for cooling, since their lowest excited states are more than 10 eV above the ground state. Ions of oxygen, nitrogen, neon and other trace elements, however, do have low-lying excited states. And even though their abundances are small compared with the abundances of hydrogen and helium, they are quite efficient in reducing the energy content of the gas. It happens that the radiation they emit results from transitions of the excited electrons that violate the simplest forms of the rules of quantum mechanics and are therefore called forbidden transitions. Although the forbidden transitions are not normally important in the laboratory, they are of paramount significance in the low-density gases of interstellar space.

One of the most important nebular coolants is oxygen. In the singly ionized form O II, radiation is produced in the ultraviolet region of the spectrum as a result of the collisional excitation of the lowest pair of energy levels. O II can be responsible for a large percentage of the energy lost in nebulas of relatively low ionization. In nebulas of higher ionization the doubly ionized form O III is often a strong coolant through transitions from its first excited level to the ground state, which emit green light at the wavelengths of 5,007 and 4,959 angstroms. These two emission lines are often among the strongest radiated by nebulas in the visible region of the spectrum.

Even with the trace elements acting as coolants to balance the heat generated by photoionization and recombination the nebula is not in thermodynamic equilibrium. For a gas to be in such equilibrium there must be a single temperature that characterizes the level of excitation, the degree of ionization, the kinetic energy of the particles in the gas and the amount of radiation present. With gaseous nebulas different temperatures are generally needed to describe these parameters. Therefore the powerful general equations that can be applied to describe a gas in thermodynamic equilibrium cannot be employed with gaseous nebulas, and one must treat each process individually. Nonetheless, the availability of large computers and increasing knowledge of atomic processes have made possible considerable progress in theoretical studies of the physical structure of gaseous nebulas. Not only are the general features of that structure well understood but also the theories bear up well when their predictions are compared with the observations.

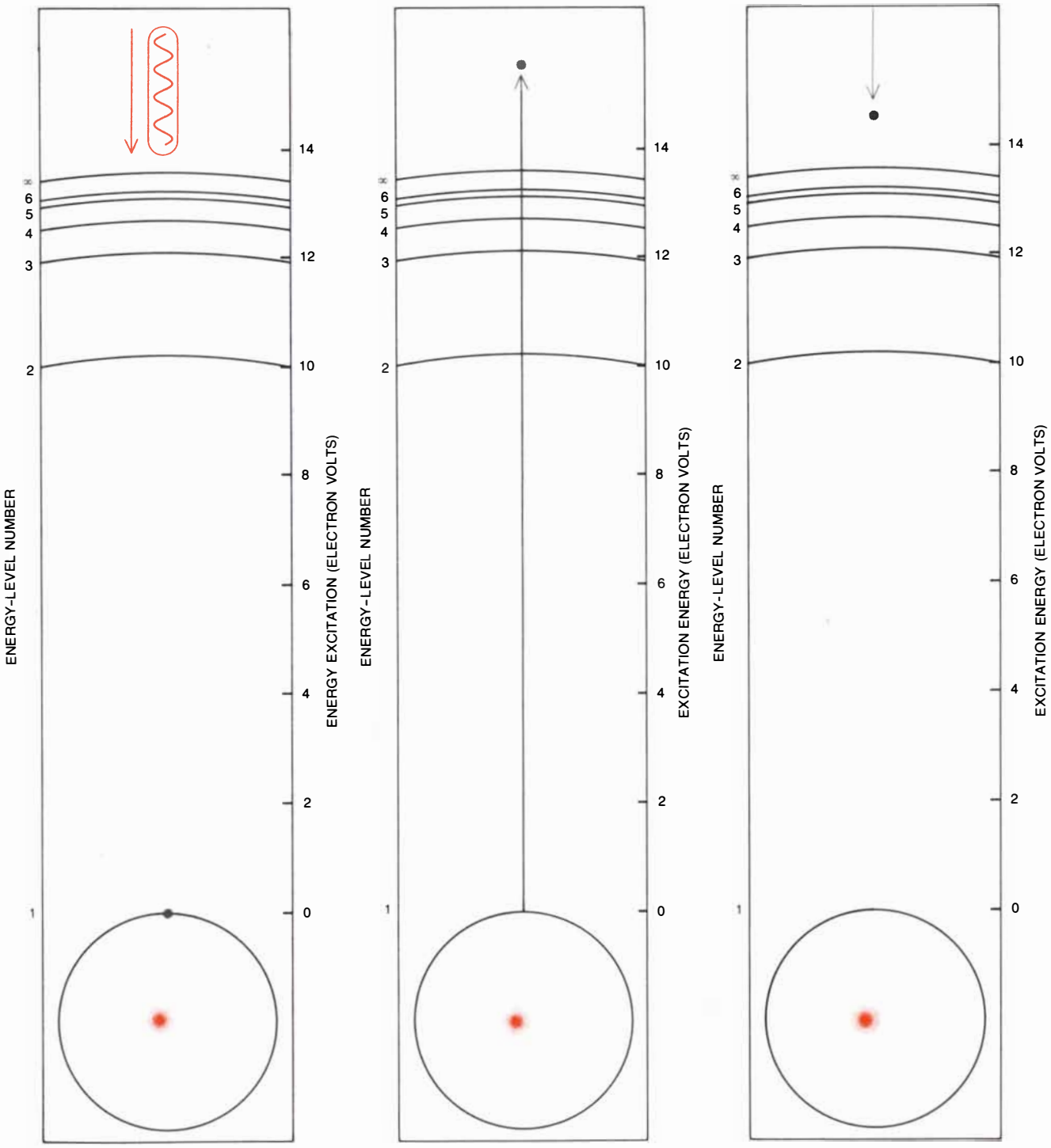
Since the radiation from the ions that are indicators of the degree of ionization in a nebula is emitted at widely separated wavelengths in the visible region of the spectrum, gradations in the ionization structure of a nebula are correlated with variations in color that can be photographed. Standard color films, however, suffer from a color response that not only is uneven over different parts of the spectrum but also depends on the length of the exposure time. Therefore such films are not suitable for photographing astronomical objects in a controlled and repeatable way. Their limitations are particularly serious for the color photography of nebulas because the two important green lines of O III fall in a region of the spectrum where standard color emulsions are quite insensitive.

At the Lick Observatory, George H. Herbig and Merle F. Walker, aided by the advice and technical skill of Eugene A. Harlan and the late Rulon Watson, have resurrected the dye-transfer process for studying the emission nebulas. In this process, which was once the principal means of making high-quality color photographic prints, three separate black-and-white pictures are made of an object through three different color filters. The individual photographs are then made into mats that can absorb dyes, the amount of the dye absorbed being related to the intensity of the original exposure. Each mat is then dipped into a different dye, and the mats are successively placed in careful registration on a special paper that absorbs the dyes from them. If the dyes and filters are carefully chosen and the photographic exposures are carefully balanced, the dye-transfer process is capable of producing controlled and repeatable color pictures.

The human eye is unable to perceive colors very well at the low light levels

characteristic of most diffuse astronomical objects, and the dye-transfer pictures were intended to match the color response of the eye as closely as possible in order to show how the objects would actually look if the colors were seen directly. The matching was done by calibrating the emulsions of the photographic film to the wavelengths of light admitted by the filters. Standard color charts were used for laboratory tests and selected astronomical objects were chosen for field tests. The final product is, of course, only an approximation of the color response of the eye, but the results are far superior to what can be obtained with standard color film.

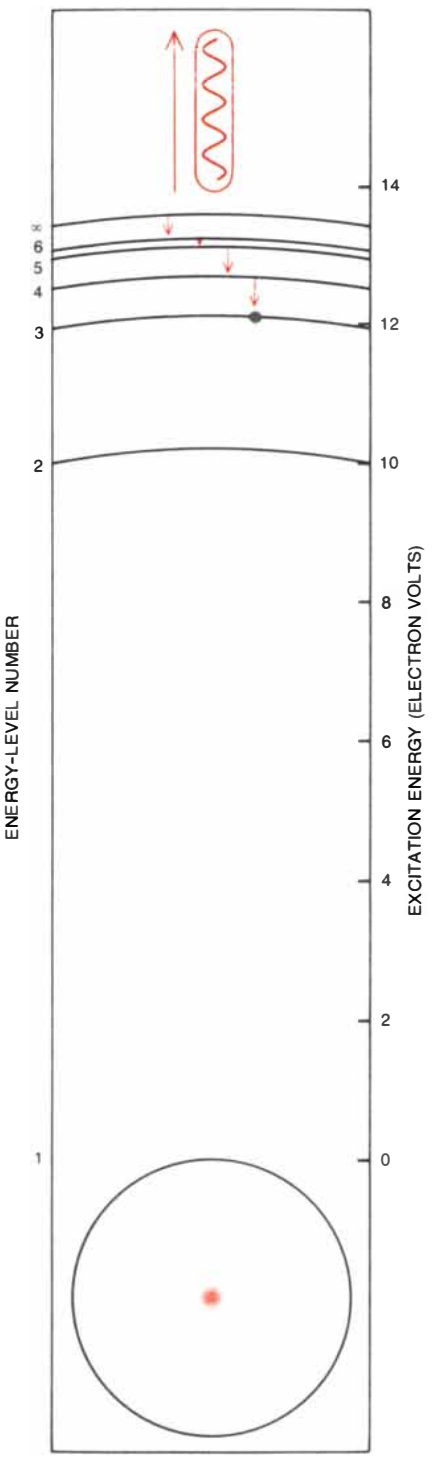
It turned out that, although the original intent was to match the response of the eye, the color balance exceeded our expectations in revealing the ionization




**PHOTOIONIZATION** of hydrogen is illustrated in this energy-level diagram. The neutral, or un-ionized, hydrogen atom is in its lowest energy state when its single electron (*gray*) is in the ground state (*Level 1 at left*) with respect to the single proton (*color*) of the nucleus. For an atom of hydrogen to be ionized a quantum of incoming radiation must have an energy of at least 13.6 electron volts. Atom absorbs radiation and ejects electron with a kinetic energy equal to the difference in energy of quantum and 13.6 electron volts (*right*).

**RECOMBINATION** is the inverse of photoionization. Here the single proton of a hydrogen ion captures an electron (*left*), which cascades down to lower energy levels in the atom, emitting radiation as it does so. The energy of the typical recombining elec-

structure of the emission nebulas, particularly the structure of the ionized oxygen, which is an excellent tracer of the overall ionization. A spectacular example of the wealth of information that can be obtained from such photographs are the pictures of the Great Nebula in Orion.




**tron is less than that of a typical electron ejected during photoionization. In the illustration the energy is captured to the third energy level (*right*); if electron now drops just one more energy level, then it would emit red radiation of hydrogen-alpha line.**

The Orion nebula is probably the most notable of all the H II regions both because of its closeness to the earth (1,500 light-years) and because of its brightness, although the total amount of matter in it is not overly large compared with the average H II region. In the photograph of the nebula's center on the cover of this issue of *Scientific American* the four bright stars known as the Trapezium (for their geometric arrangement) are clearly visible. The brightest of the four is a very hot star with a temperature of about 40,000 degrees K. It is the one chiefly responsible for ionizing the nebular gas. The intense green color of the gas is due to the radiation of the two forbidden lines of O III. To the southeast there is a reddish border around the nebula where there apparently is not enough O III to generate the green radiation, so that the red radiation from hydrogen dominates. There is also a contribution from the red lines of singly ionized nitrogen (N II), an ion found in regions of relatively low ionization.

The transition from the predominance of O III to the predominance of hydrogen and nitrogen shows up more clearly on the longer-exposure photograph of the Orion nebula at bottom right on page 35, which shows the tenuous outlying portions of the nebula as well as the central regions. Although the brilliant central features are considerably overexposed and the green from O III is not faithfully reproduced, to the west one can see that the radiation extends with decreasing intensity as far as the main body of the nebula can be detected. In that direction it seems there is enough radiation to ionize all the gas, so that O III exists throughout the entire volume. In other words, in its western region the nebula runs out of gas before it runs out of ionizing photons, and that region is gas-bounded.

To the north and southeast, however, photons capable of ionizing oxygen to form O III are depleted, and there are large regions where the overall level of ionization is lower. In those regions the color of the nebula is dominated by the red from hydrogen and nitrogen. Finally, in the extension of the nebula to the southeast it can be seen that the material is divided into two regions, one primarily red and the other blue-gray. Again the red is emitted by hydrogen and nitrogen; the blue-gray, however, is starlight reflected by interstellar dust. In that region there is relatively little ionized gas, but the geometry of the gas and the dust is complex enough to make it somewhat unclear how the dusty reflection nebulosity is related to the rest of the nebula and the stars.

Like the Orion nebula, most other H II regions show a wealth of detail that complicates efforts to study their physical structure. Their complexity reflects the rather chaotic distribution of the interstellar medium in regions where stars have recently formed. Many of the planetary nebulas, on the other hand, are quite regular and symmetrical and must have originated with the gentle, smooth ejection of gas by the parent star in the interior. Exactly how a star ejects its outer envelope, however, is still entirely unknown.

The prototypical planetary nebula is the Ring Nebula in Lyra, which appears in the top photograph on page 36. A hot blue star is located at the center of the "doughnut hole" in a region of gas that is particularly low in density. The parts of the nebula closest to the star show the green of the radiation from O III, and there is a relatively clear, symmetrical transition to the red of the hydrogen and nitrogen at the outer edge. Spectroscopic observations show that at this reddish outer edge there is also an increase in the strength of the emission lines of O II in the ultraviolet region of the spectrum, but since the color photographs are not sensitive to ultraviolet radiation the emission from hydrogen and nitrogen dominates the picture. Nevertheless, the color transition corresponds to a distinct place where the energetic photons capable of ionizing O III are exhausted. Another example of such a region of transition from O III to O II in a nebula is the Dumbbell Nebula in Vulpecula, which appears in the bottom photograph on page 36.

The color photographs thus illustrate in a beautiful and clear way the operation of a rather complex set of phenomena associated with the photoionization of a diffuse gas by a star or stars. The photographs have accurately guided detailed spectroscopic observations of the Ring Nebula and the Dumbbell Nebula with the 120-inch reflecting telescope at the Lick Observatory. Moreover, they convincingly verify the results of theoretical calculations.

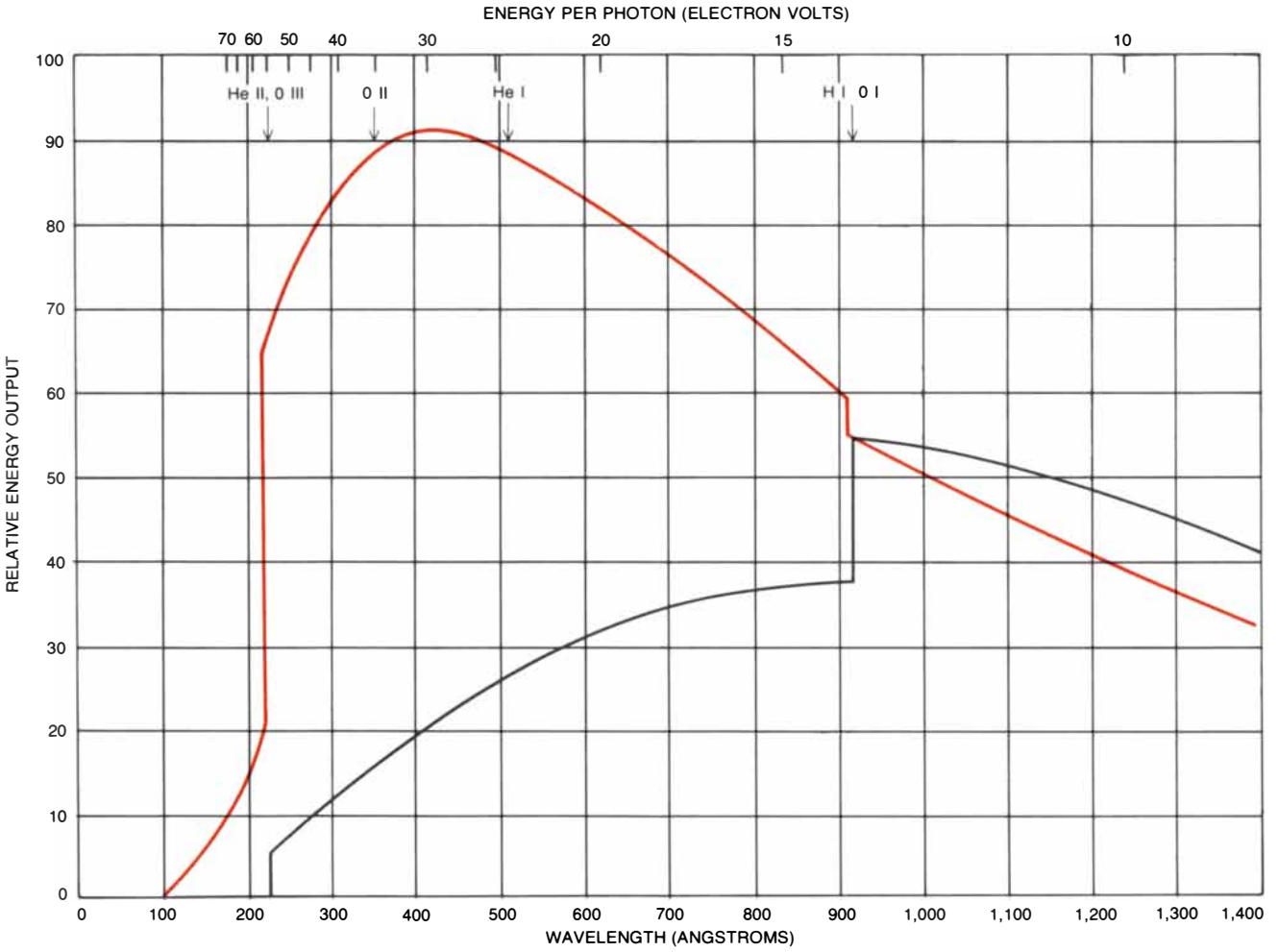
As I have mentioned, remnants from supernovas represent a separate class of gaseous nebulas whose source of energy is not necessarily the photons from a central star. The best-known remnant, the Crab Nebula in Taurus, presents an example of gas ionized by the synchrotron radiation emitted by electrons spiraling at high velocities in a magnetic field. The spectrum of the radi-

ation from a star approximates the spectrum of a black body, or ideal radiator, and the distribution of the energy at different wavelengths falls off rapidly in the far-ultraviolet region of the spectrum. Synchrotron radiation, unlike the blackbody radiation, can show a flatter energy distribution extending far into the ultraviolet or beyond it. The Crab Nebula emits some of its energy in the X-ray region of the spectrum. Thus synchrotron radiation is composed of many more high-energy photons than the radiation from stellar sources, and the resulting ionization structure of a supernova remnant can be quite different from the structure of an H II region or a planetary nebula.

The Crab Nebula is the product of a supernova that exploded in the year 1054 and is therefore an example of a relatively young remnant that has not yet begun to interact to any large degree with the interstellar gas of the galaxy around it. Most supernova remnants we observe are much older and derive their luminosity from the fact that the expanding shell of the remnant collides with the surrounding interstellar gas. The collision gives rise to a shock front in which the gas swept up by the expanding shell is highly ionized; the light we observe is generated behind the shock wave, where the ions and electrons are recombining and cooling. The Cygnus Loop, also known as the Veil Nebula, is a good example of this phenomenon.

Interpreting the color photographs of either the Crab Nebula or the Cygnus Loop is not as clear-cut as interpreting those of either the H II regions or the planetary nebulas. The Crab Nebula has two dissimilar components: a relatively smooth bluish region and many multicolored filaments and pieces of filaments [*see upper photograph on page 37*]. The blue region is the place where light is generated by the synchrotron process. It is believed the strong magnetic field of the pulsar at the center of the nebula is responsible for the acceleration of the electrons to the high speeds required to emit the synchrotron radiation [see "Supernova Remnants," by Paul Gorenstein and Wallace Tucker; SCIENTIFIC AMERICAN, July, 1971].

Since the synchrotron radiation extends into the far ultraviolet, it ionizes the gas in the filaments, which then glow with the colors characteristic of their chemical composition and ionization. As can be seen in the photograph, the distinction between the red filamentary re-




**CURVES OF THE DISTRIBUTION OF ENERGY over the far-ultraviolet region of the spectrum are shown for two stars at different temperatures. The distributions are adjusted so that both stars emit the same amount of energy at a wavelength a little longer than 900 angstroms. The color curve is for a star with a temperature of 100,000 degrees Kelvin, typical of the central star of a planetary nebula; the black curve is for a star with a temperature of 50,000 degrees K., representative of the hottest stars found inside H II regions. The relative energy output indicates how much energy per wavelength one star puts out with respect to the other. For example, at a wavelength of 250 angstroms the hotter star is emitting some nine times as much energy as the cooler star. Energy per photon at different wavelengths is indicated across top. Also shown is minimum energy of ionization for singly ionized helium (He II), for singly and doubly ionized oxygen (O II and O III), for neutral helium (He I), for neutral hydrogen (H I) and for neutral oxygen (O I).**

gions and the green ones is not at all symmetrical or orderly. It is possible, however, to work out the three-dimensional distribution of the filaments in space. Correlating their spatial arrangement with the colors in the photograph and with measurements of the intensity of their various spectral lines indicates that the regions of higher ionization (the green filaments) tend to be located closer to the center of the nebula than regions of lower ionization (the red filaments) and may even be embedded in the portion of the nebula emitting the synchrotron radiation. Such an arrangement would imply that the higher ionization is principally due to the fact that the higher intensity of the ionizing photons in the nebula's center more effectively overcomes the balancing process of recombination.

At greater distances from the Crab Nebula's center the gas seems to be subjected to radiation of lesser intensity, and thus even though the energy distribution over the spectrum of the radiation may be the same, the equilibrium between photoionization and recombination will be established at an overall lower state of ionization. Complicating the situation even more is the fact that in the Crab Nebula several states of ionization of a given element coexist in one place instead of being more simply stratified into regions of different Strömgren radii, as they are in the H II regions and the planetary nebulas.

Whereas the Crab Nebula is now a little more than 900 years old, the Cygnus Loop originated with a supernova that exploded perhaps 50,000 years ago. The rapidly expanding debris of the star has collided with a large amount of interstellar gas during this interval, gradually slowing in its expansion from an initial rate of approximately 10,000 kilometers per second to its present rate of some 100 kilometers per second. The visible light from the nebula is the direct result of this continuing deceleration as the expanding remnant ionizes the gas of the interstellar medium. Although detailed theoretical models of this process have been devised and the calculations are in excellent agreement with the observed emission spectrum of the remnant, the physical phenomena responsible for the gross color variations from place to place are still unknown.

We have seen that in color photographs three intrinsically different kinds of objects--H II regions, planetary nebulas and the Crab Nebula—display an intricate structure that can be understood in a rather straightforward way in terms of the theory of ionization by radiation of a low-density gas. I believe our theoretical understanding of the details of this complex process is quite good. It is now possible to extend the study of ionized interstellar gases and consider them in the context of other astronomical problems.

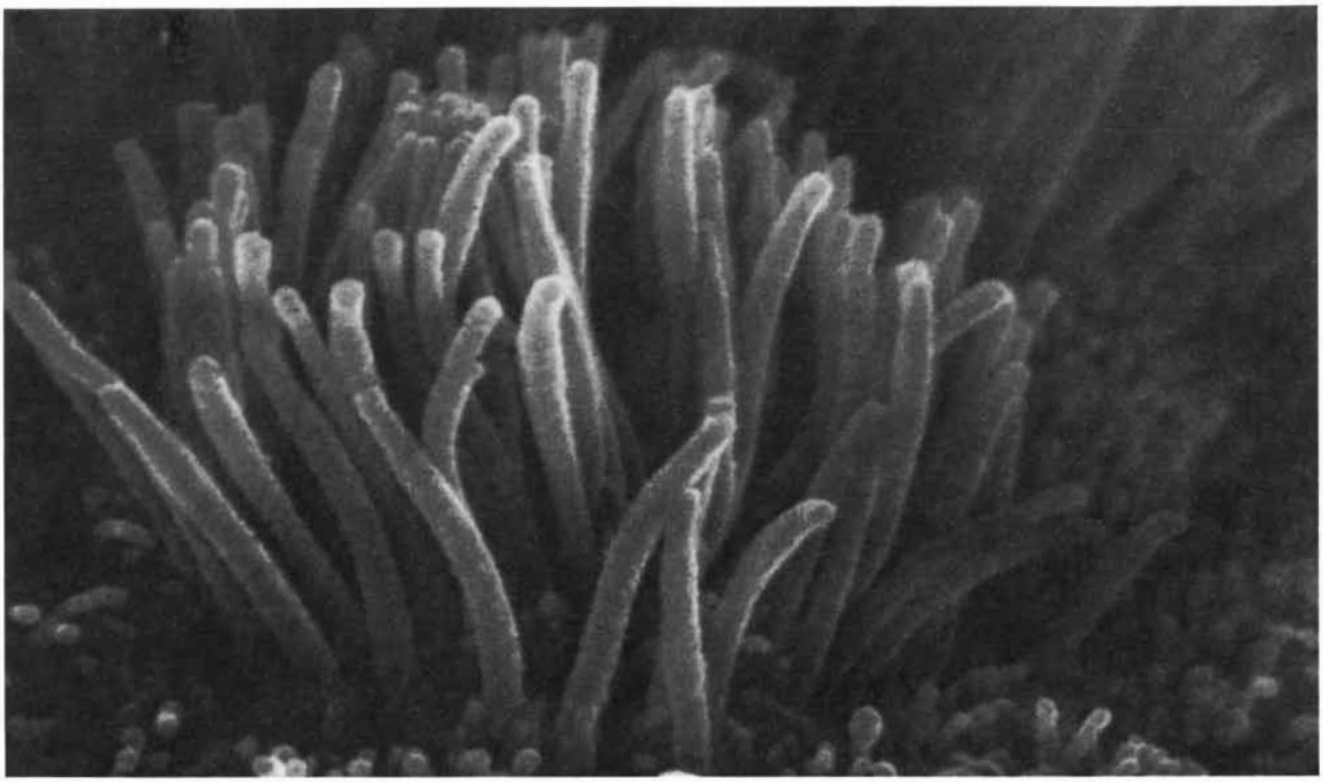
One important problem for which emission nebulas can supply some valuable answers is the question of the chemical composition of the matter in our galaxy, both in the past and at present. Armed with the knowledge of the physical structure of nebulas one can interpret the strengths of the various emission lines from different elements in terms of the relative abundances of the same elements in the nebulas. The results are quite interesting. It turns out that H II regions like the Orion nebula, which are young objects, and the planetary nebulas, which are made by stars that are on the average several billion years old, have abundances of helium, neon, oxygen, argon, nitrogen and sulfur that are very similar. It appears, then, that planetary nebulas are the ejected outer layers of stars whose chemical composition was not much changed by nuclear reactions earlier in the life of the star. Such evidence must be kept in mind during future detailed investigations of how planetary nebulas are ejected from their parent stars.

Furthermore, the composition of the material out of which the old stars in the planetary nebulas formed differs little from the composition of the present-day interstellar medium. Thus we are actually looking far back into the past by studying, for example, the planetary nebula located in the globular cluster M15 in the constellation Pegasus. The star responsible for this object is believed to be at least 10 billion years old and was formed at a very early stage in the evolution of our galaxy. Studies of other stars in the cluster indicate that in them the abundance of metals such as iron is deficient with respect to the abundance of hydrogen by a factor of more than 100 compared with those abundances in the sun, which is a much younger object. The planetary nebula, however, shows about the same ratio of helium to hydrogen as the Orion nebula, indicating that the abundance of helium has not changed much over most of the history of the galaxy. The abundance ratios of oxygen and neon with respect to hydrogen in the planetary nebula are about a tenth as great as the ratios in the Orion nebula.

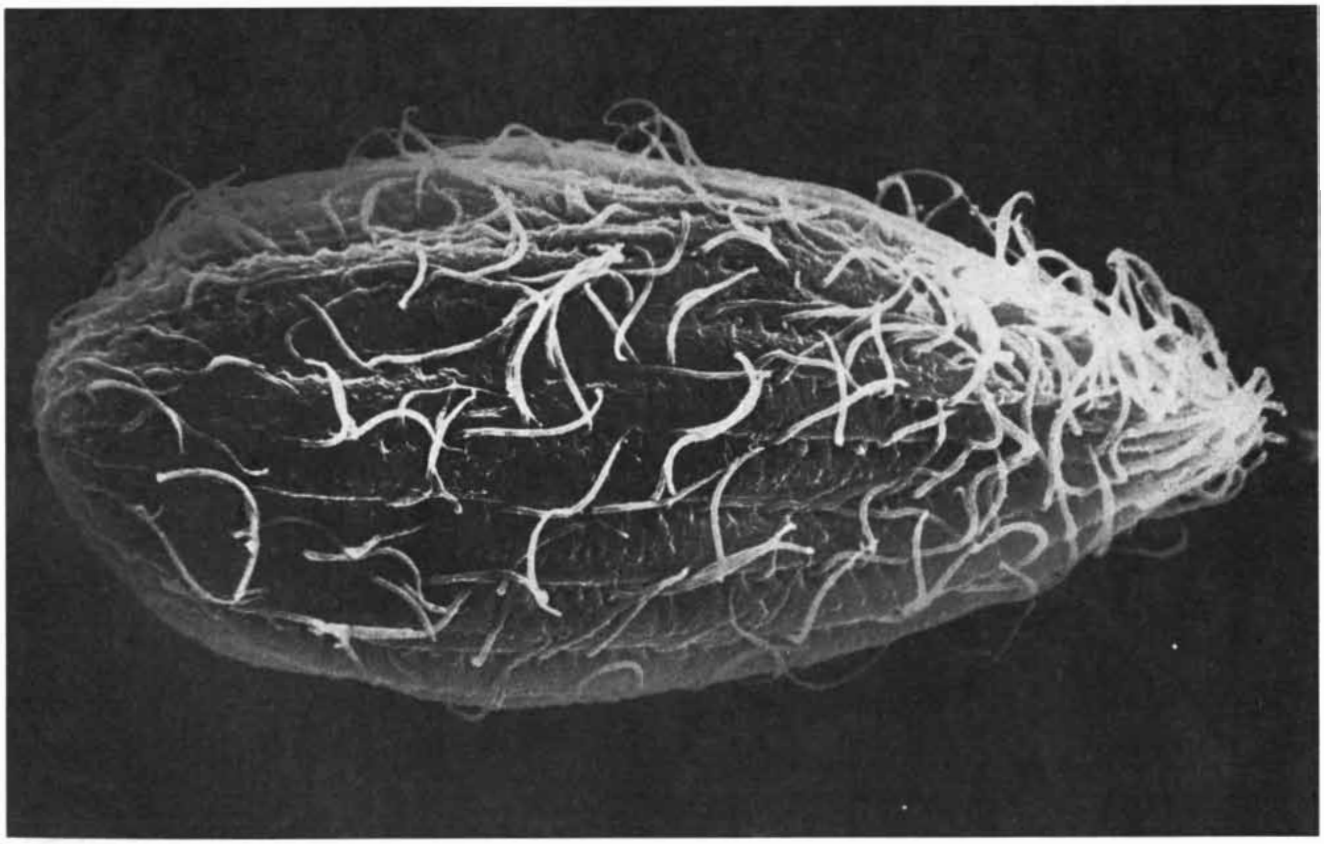
Our present view of the chemical evolution of the galaxy is that it very likely started out with material composed entirely or almost entirely of hydrogen and helium. As a result of the nuclear synthesis of heavier elements in the interior of stars and the recycling of matter by supernova explosions the galaxy has gradually been enriched in the heavier elements. The composition of the planetary nebula in M15 suggests that oxygen and neon were built up more rapidly than the metals, since even in this very old object these lighter elements have abundances that are within 10 percent of their abundances in the interstellar medium as it is currently constituted. Many details have yet to be worked out, but the overall picture seems to be soundly sketched.

We might then ask if the abundances of the elements in the filaments of the Crab Nebula confirm this picture, since they represent material being returned to the interstellar medium at the present time. Unfortunately for astronomers the Crab Nebula generates more questions than answers. A recent analysis of the abundance ratios of the elements conducted at the Lick Observatory by measurement of the intensities of the spectral lines of a few filaments selected by consulting the color photographs showed that the ratio of helium to hydrogen was much larger than normal: between six and 10 times larger. It is not possible at present to determine the abundance of metals, but at least it appears that the ratios of oxygen, neon, nitrogen and sulfur with respect to hydrogen are close to those found in the H II regions. These results vary from what we had expected for a supernova remnant, but it is quite likely that the Crab Nebula originated in a very unusual supernova explosion. As far as we know it is unique in our galaxy. In any case there is still much left to be learned about the formation of the elements during the history of the galaxy.

The study of emission nebulas has reached maturity. The basic physical mechanisms operating appear to be well known, so that we can understand the rich detail of the color photographs in terms of processes at the atomic level. With this knowledge we have been able to address questions concerning the chemical evolution of the galaxy. Some astronomers are beginning to turn their attention to the ionized gases in more distant objects in the universe. We can now hope to considerably increase our knowledge of these objects as a result of what we have been able to learn by studying the emission nebulas closer at hand.



CILIA IN MOUSE OVIDUCT move eggs along the tube from the ovary toward the uterus. The cilia, enlarged some 23,000 diameters in this scanning electron micrograph made by Ellen R. Dirksen of the University of California at San Francisco, are about .25 micrometer in diameter and project a few micrometers above the surfaces of the cells lining the oviduct, which are covered by microvilli.



CILIA ON TETRAHYMENA, a protozoon, propel the one-celled animal through water. The cilia cover the entire surface of the organism, enlarged 3,500 diameters in this scanning electron micrograph made by Birgit Satir of the University of California at Berkeley and Jean-Paul Revel of the California Institute of Technology. Cilia are same diameter as oviduct cilia but about twice as long.

# HOW CILIA MOVE

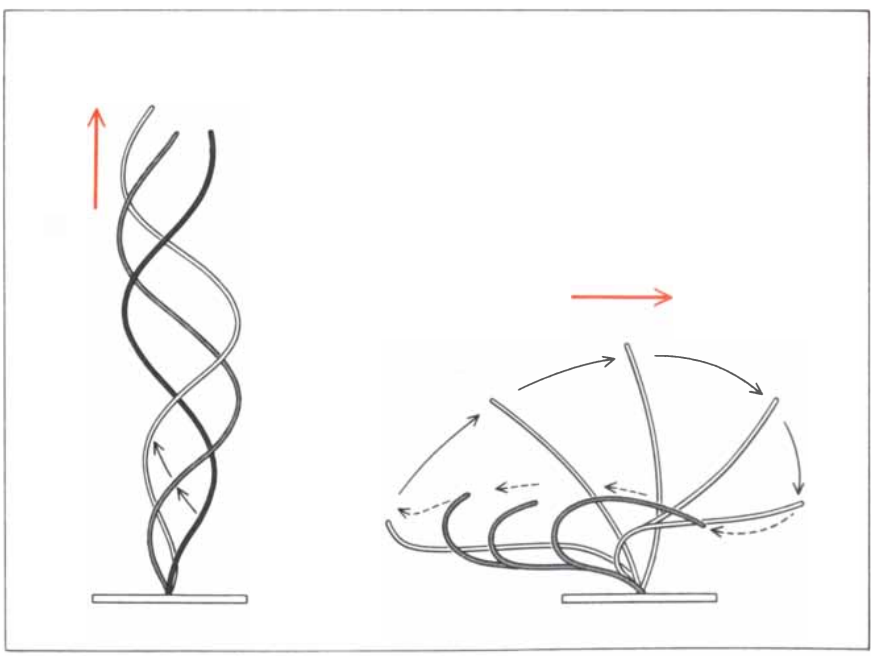
## The hairlike organelles that propel swimming cells or move liquids over fixed cells are composed of sheaves of microtubules. The cilia beat when the microtubules, powered by ATP, slide past one another

by Peter Satir

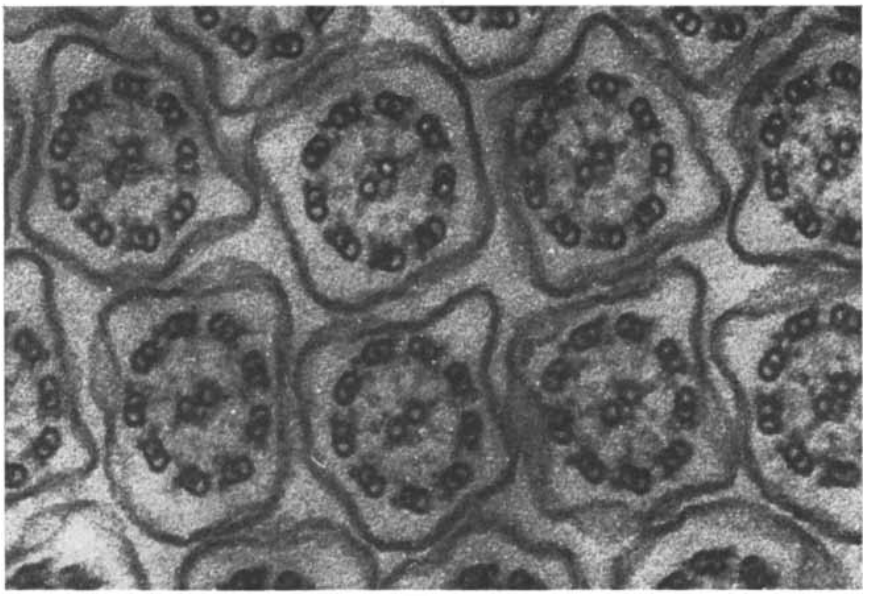
Cilia are "eyelashes" that are found on the free surfaces of a wide variety of cells, ranging from single-celled organisms such as protozoa to the epithelial cells of human mucous membranes. Some cells have dense fringes or mats of cilia and some have only one or a few of them, in which case they are called flagella. (Bacteria also have flagella, but these are an entirely different organelle.) Whatever their superficial appearance, all motile cilia have essentially the same structure and the same basic function: to move a stream of liquid. We are beginning to understand how they move.

Ciliary motion can have either of two effects, depending obviously on whether the cell is a free-swimming one or is fixed in place. Cilia attached to a moving cell, such as a paramecium or a spermatozoon, propel it through a liquid; cilia attached to fixed cells, such as those that line the gill of a mussel or the trachea or reproductive tract of a human being, move a liquid over the surface of the cells. Most cilia beat between 10 and 40 times a second, but the form of the beat is quite variable. The flagellum constituting the tail of a sperm propagates a wave that undulates as it moves from the base of the tail to the tip and drives the sperm head forward; in the case of a sea-urchin sperm that wave moves in a single plane at velocities of about half a millimeter per second. The cilia on the gill cells of the mussel *Mytilus* beat in a more complex way: an oarlike, planar effective stroke drives water over the gill and is followed by a whiplike return stroke in which the cilium swings out to one side as it regains its original position. Given the size of cilia or flagella, the hydrodynamics are suggestive of an oar working in molasses rather than in water. The viscous forces produced in the medium in which a mat of cilia functions cause the individual cilia to interact in a special way, producing a metachronal, or out-of-phase, wave that propagates through the mat.

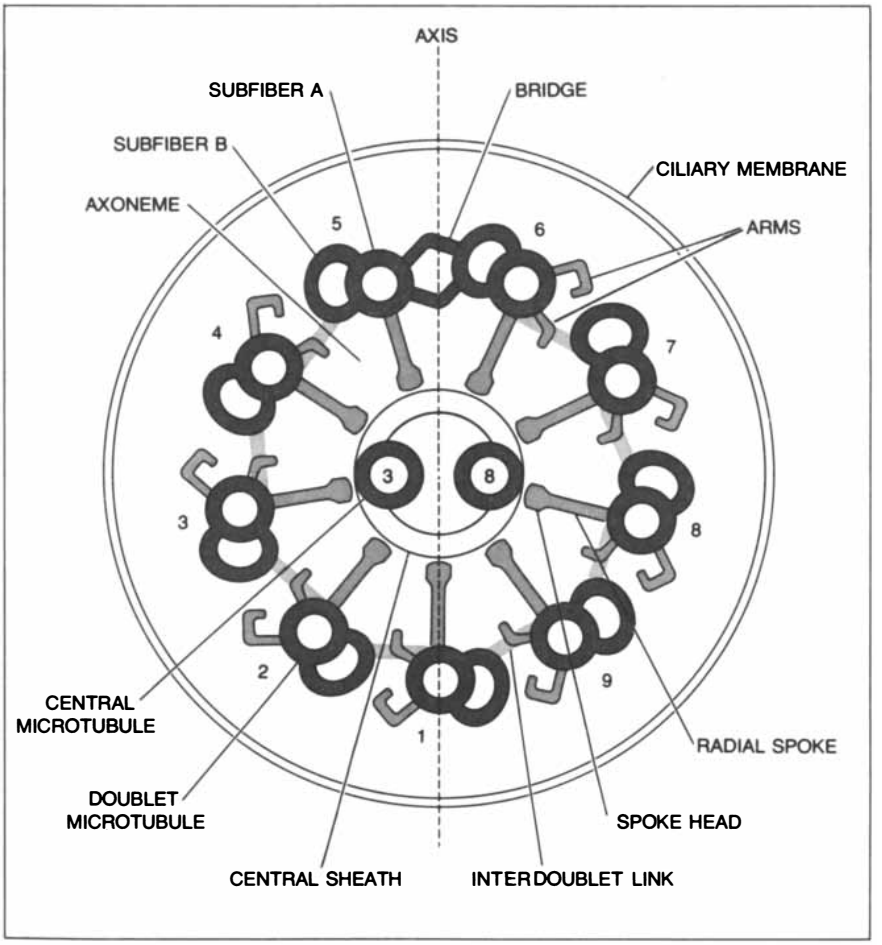
In 1961, when the subject was last reviewed in *Scientific American,* not much more than that was known about ciliary motion and its mechanism [see "Cilia," by Peter Satir; SCIENTIFIC AMERICAN, February, 1961]. The structure of the organelle was beginning to be uncovered, on the other hand, largely as the result of work by Irene Manton of the University of Leeds and by Don W. Fawcett and Keith R. Porter, who were then working at the Rockefeller Institute for Medical Research. Their pioneering electron-microscope studies and those of others had established the remarkable identity of the ultrastructure in a wide range of cilia in different organisms. Each cilium could be seen to consist primarily of a sheaf of filaments, or tubular elements, arranged in what came to be called the 9 + 2 pattern, with nine doublet elements surrounding a central pair of singlets. The sheaf of filaments, called an axoneme, was surrounded by and enclosed in an extension of the cell membrane, which meant that it was essentially a protruding portion of the cytoplasm



**CILIARY MOTION takes various forms; these are two typical patterns. Successive waves propagated toward the tip of a flagellum (*left*) move water (*colored arrow*) and thus propel the sperm head in the opposite direction. The beat of a typical cilium (*right*) consists of a straight-armed effective stroke (*white cilia*) that moves a liquid (*colored arrow*), followed by a curling return stroke (*gray cilia*), often in a third dimension (out of plane of paper).**



**CROSS SECTIONS OF CILIA from the gill cell of a mussel are enlarged about 100,000 diameters in this transmission electron micrograph. Each consists of a cylindrical axoneme enveloped by an extension of the cell membrane. The axoneme, .2 micrometer in diameter, includes a ring of nine doublet microtubules surrounding two single central microtubules.**



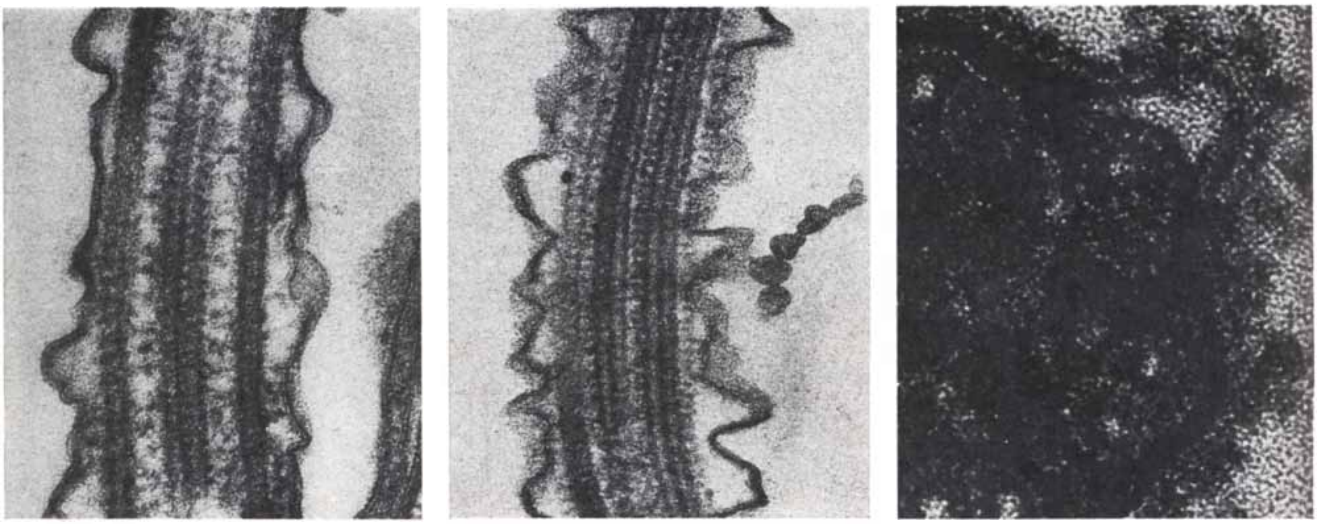

**INDIVIDUAL CILIUM CROSS SECTION like those in the micrograph at the top of the page is diagrammed. Each doublet microtubule has two subfibers, *A* and *B*; arms project from subfiber *A* toward the next microtubule's subfiber *B*. The axis is an imaginary line defining the plane of the cilium's effective stroke, bisecting the bridge, the central microtubule pair and the doublet designated No. 1. Cilium is viewed here from base toward tip.**

of the cell. Below the axoneme could be discerned a basal body: a short cylinder identical in structure with a centriole, one of the organelles that constitute the poles of the mitotic spindle during cell division in most animal cells. Electron micrographs showed that the outer doublets of the axoneme were more complicated in structure than the central singlets; successive cross sections made it clear that the two kinds of filaments and the individual elements of the doublets were of different lengths. In 1961 the substructure of the individual tubular elements had not yet been revealed by high-resolution micrography and the biochemistry of the cilium was still rudimentary. There was speculation about how the filaments might interact to make the cilium move, but the experimental information was lacking. One could not even be sure that the basis of ciliary motion lay solely within the axoneme, or indeed that the cilium itself beat actively.

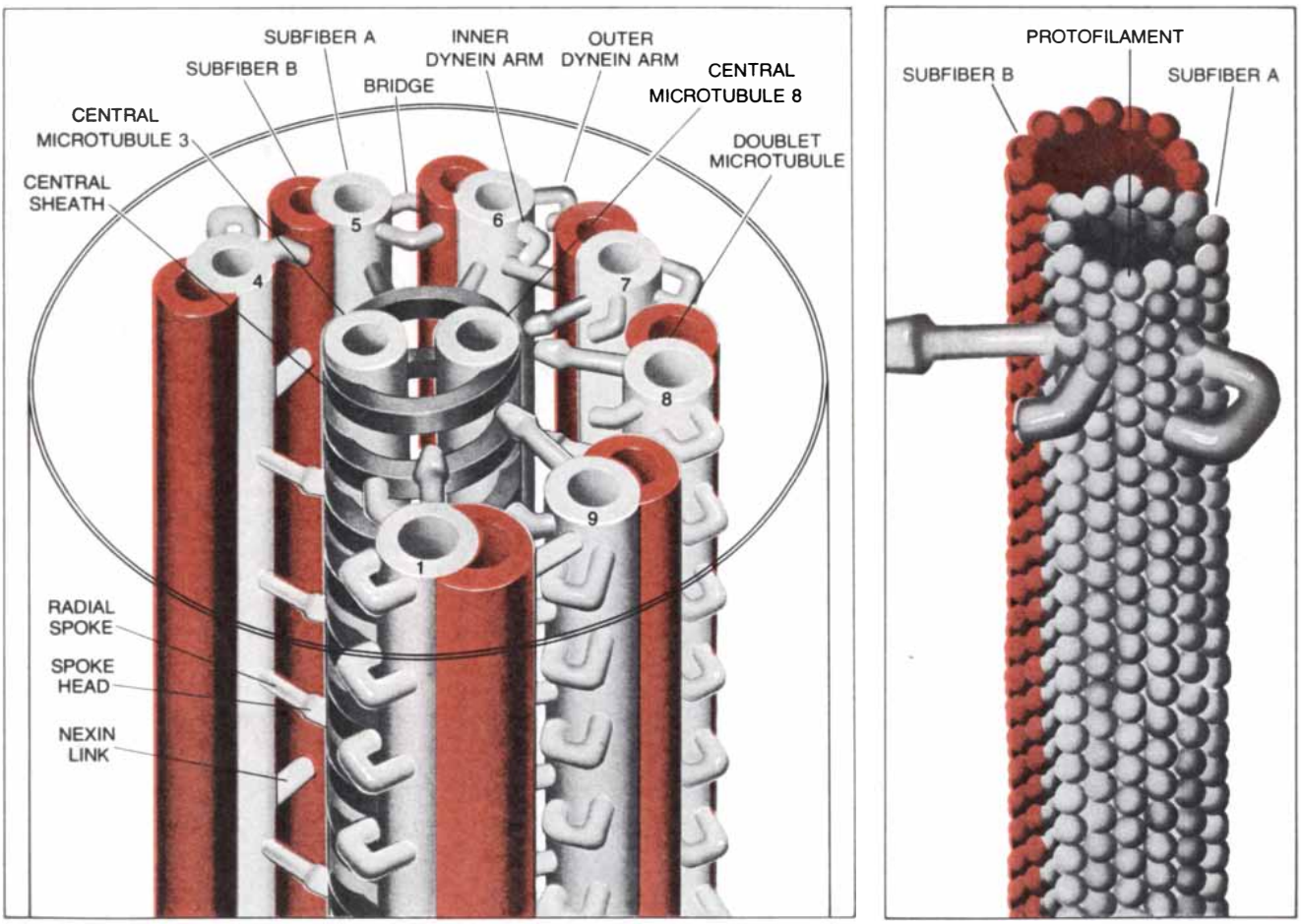
Current understanding of the mechanism of ciliary motion is the result of observations and experiments in many laboratories, which I shall present here somewhat out of chronological order. One of the basic findings was the demonstration that the cilium is not simply a waving appendage powered by contractile elements elsewhere in the cell. That was established in an elegant experiment performed by Stuart Goldstein, Michael Holwill and Norman Silvester of the University of London. They aimed a laser microbeam at the base of the cilium of a living protozoon, *Crithidia,* and thus severed the cilium without damaging it. The severed cilium swam off on its own. Repeated experiments indicate that the isolated cilium keeps going as long as its energy reserves hold out.

To study the biochemistry and kinetics of cilia it was essential to have a "model system" that could be subjected to experimentation. Following the lead of muscle-cell investigators, in the early 1950's Hartmut Hoffmann-Berling of the Max Planck Institute for Medical Research in Heidelberg added glycerin to ciliated cells and tissues. The result was exciting: once the glycerin was removed and the cells were washed, the addition of adenosine triphosphate (ATP) to the preparation caused the cilia to beat. This indicated, and further experiments have confirmed, that ATP is the energy source for motility in cilia, as it is for most cellular activity.

The precise effect of the glycerin became clear about 10 years later when Frank Child and I, working at the Uni-



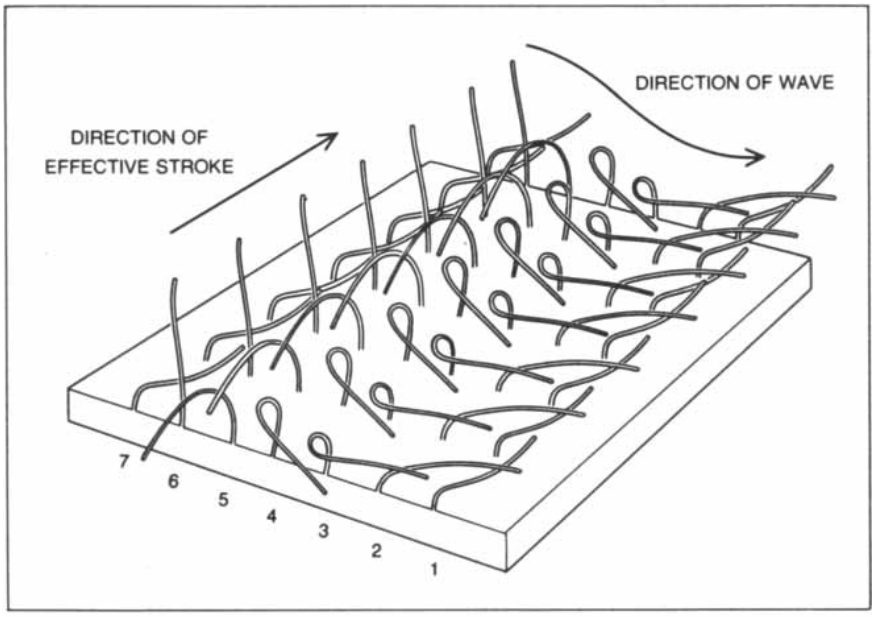
**DETAILS OF CILIARY STRUCTURE** are revealed in two micrographs of longitudinal sections of *Tetrahymena* cilia, made by Birgit Satir, and a high-power micrograph of part of a cross section of a mussel-gill cilium, made by Fred D. Warner of Syracuse University and the author. The longitudinal sections show the axoneme and the wrinkled ciliary membrane. A section through the middle of the cilium (*left*) shows two doublets attached by radial spokes to the sheath surrounding the central microtubules. A section nearer the surface (*center*) shows two of the outer doublets, with arms projecting from one toward the other, forming a ladderlike pattern. The cross section (*right*), which enlarges the cilium some 550,000 diameters, resolves the microtubules' subunits: the protofilaments.



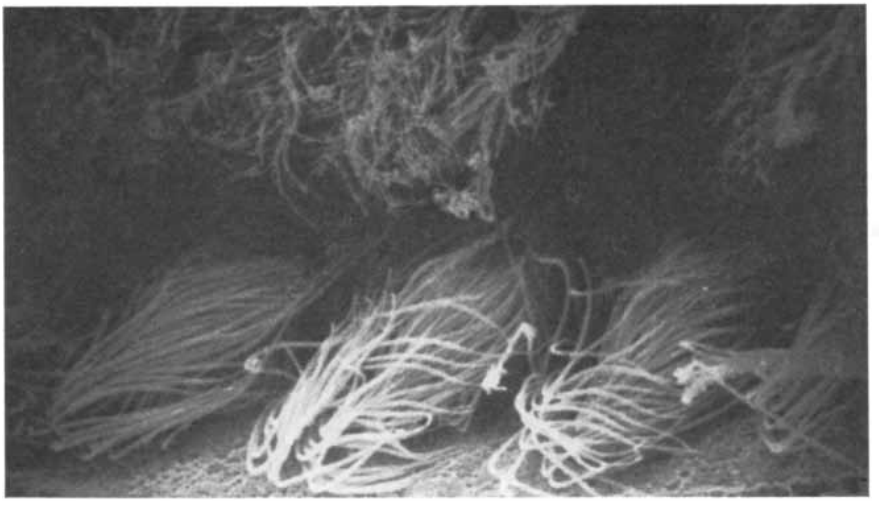

**ULTRASTRUCTURE OF CILIUM** is diagrammed schematically (*left*), largely on the basis of the interpretation of electron micrographs. The central sheath apparently consists of a series of tilted hoops, to which the doublets are connected by the heads of radial spokes. The doublets are linked by molecules of a protein, nexin. The two sets of arms, inner and outer, are molecules of the enzyme dynein. The construction of an individual doublet is shown in more detail (*right*). The cylinders are composed of tubulin molecules arrayed in rows (13 in subfiber *A*, 10 or 11 in subfiber *B*) that constitute protofilaments. As in the drawing on the opposite page and in the cross-section micrographs, the view is from base toward tip, that is, the tip of the cilium is toward the bottom of the page.

versity of Chicago, examined glycerinated cilia in the electron microscope. We could see that the glycerin had destroyed the membranes of the cilia and the cell but had left the 9 + 2 axonemes virtually intact, and the naked axoneme was capable of ciliary motion. In the living cell ATP, produced by mitochondria, which usually lie near the base of the cilium, diffuses out toward the tip of the axoneme. An essential function of the ciliary membrane is to maintain around the axoneme the proper concentration of ATP and also of certain essential ions such as magnesium. After the membrane is destroyed by glycerin it therefore becomes necessary to supply the ATP and the necessary ions from the outside, and by the same token it becomes possible to vary the concentration of ATP and other substances for experimental purposes.

This investigative technique was subsequently refined by Barbara and Ian R. Gibbons and their colleagues at the University of Hawaii. They removed the heads from sea-urchin sperm and treated the tails with a detergent, Triton, instead of glycerin. The resulting model tails, which are simple membraneless 9 + 2 axonemes, swim just like living sperm. Moreover, the precise dimensions of their beat can be controlled by alterations in the medium in which they are reactivated. In particular the concentration of ATP determines the frequency of the beat.




**BEATING CILIA on the surface of a cell interact so that in one direction, usually perpendicular to the direction of the effective stroke, they are out of phase. The result is a metachronal, or out-of-phase, wave that appears to move over the mat of cilia. In this diagram the cilia in row *1* are at the end of the planar effective stroke. Those in rows *2*, *3* and *4* are in successive stages of the curling return stroke. Those in row *5* have ended the return stroke and are beginning the effective stroke, which has ended in row *7* (the same as row *1*).**

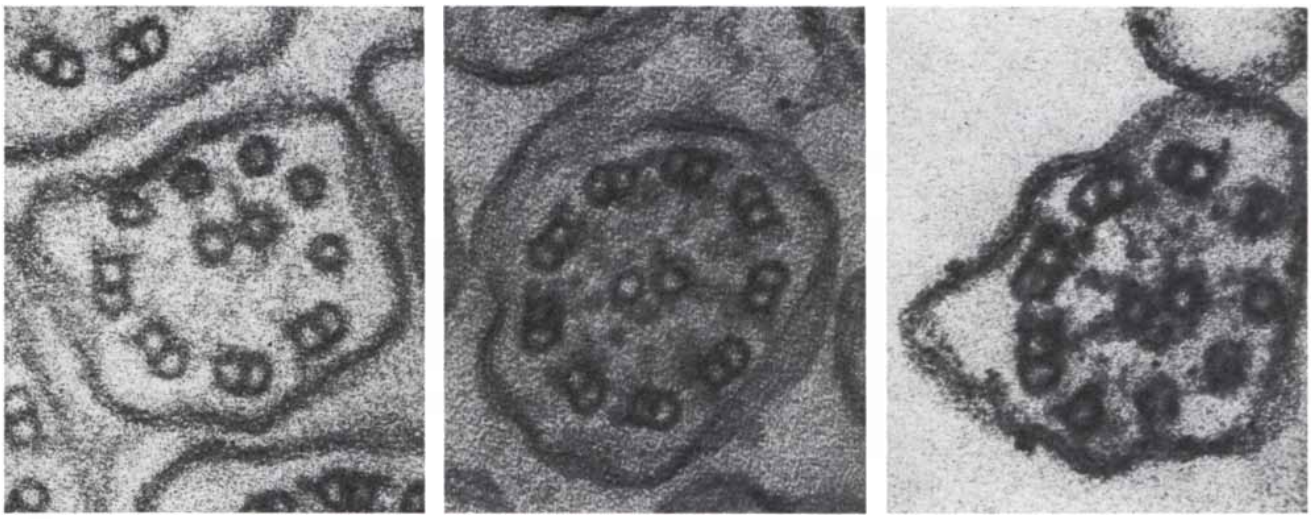


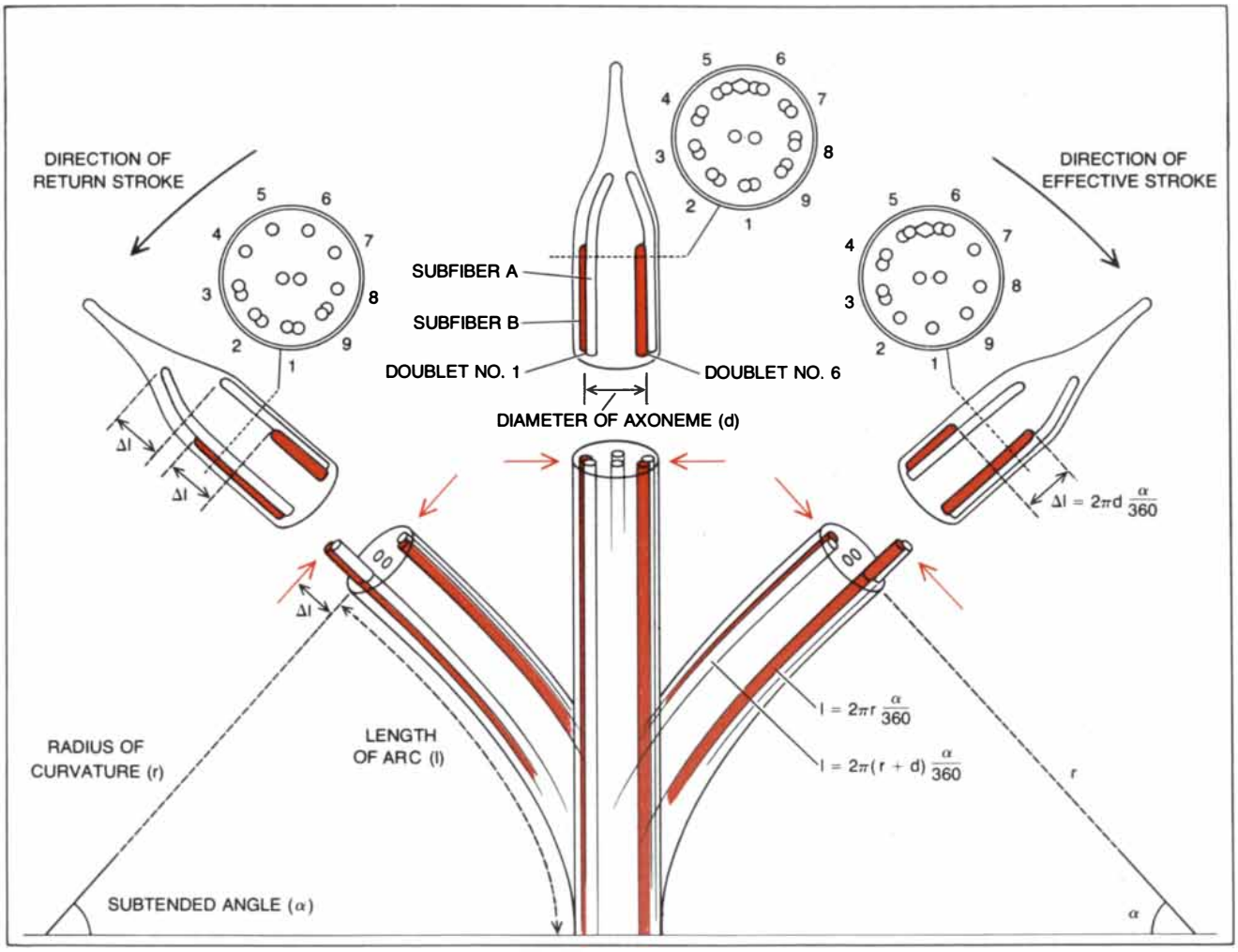
**METACHRONAL WAVE in the lateral cilia of mussel-gill cells (*foreground*) was preserved by instantaneous fixation with osmium tetroxide. The scanning electron micrograph was made by the author and Dirksen. The directions of the effective stroke and the metachronal wave are about as they are in the diagram at the top. Cilia in different stroke positions in preparations like this one are examined to determine the pattern of subfibers at their tips.**

The energy to move the cilium is supplied, as in other cellular work processes, by the breakdown of ATP through hydrolysis into adenosine diphosphate (ADP) and phosphoric acid. The enzyme that accomplishes this hydrolysis in cilia was identified by Ian Gibbons and A. J. Rowe, working at Harvard University, as an ATPase, which they named dynein. Gibbons found that he could extract dynein from axonemes if he put them in a solution containing a chelating agent, which in effect sequesters the magnesium ions. When he examined the dyneinless axonemes in the electron microscope, he found they lacked a specific structure characteristic of normal axonemes: a series of projections called arms that protrude from one side of each of the nine outer doublets. If he then took a solution containing the dyneinless axonemes and added dynein and the right amount of magnesium ion to it, the dynein was reincorporated into the axonemes. And at the same time that the return of dynein was signaled by the reappearance of ATPase activity, the arms also reappeared. In other words, the arms are dynein molecules.

Many additional details of the construction of the doublets and singlets of the axoneme are now clear. Each tubular element is actually a microtubule, a structure found in the cytoplasm of many kinds of cells, including the axons of nerve cells. Microtubules are composed of small protein molecules called tubulins. There are at least two kinds of tubulin molecule in each microtubule. They are strung together like Poppit beads in a row to form protofilaments; the protofilaments are arrayed alongside one another, something like individual columns of building stones, to make up the wall of the microtubule.

Each axonemal doublet is composed of two subfibers, *A* and *B*. Subfiber *A* is a complete microtubule with 13 protofilaments. Subfiber *B* is shorter in length and has fewer protofilaments—only 10 or

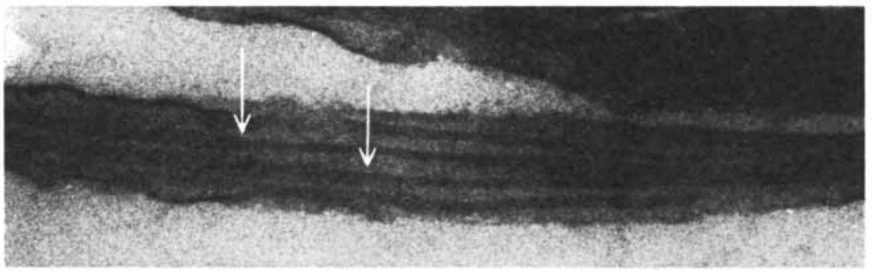




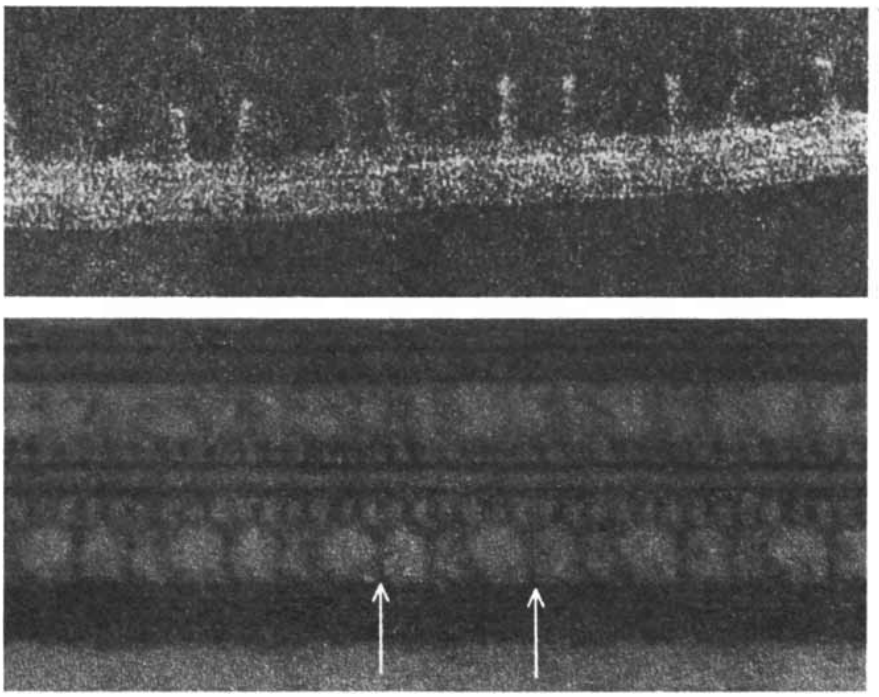

**SLIDING OF MICROTUBULES** is manifested by differences in the pattern of subfibers near the tip of the cilium. Here the relations among sliding, bending and tip pattern are shown for two doublets, Nos. 1 and 6, in three stroke positions: straight up, bent in the effective-stroke direction (*right*) and bent in the return-stroke direction (*left*). The colored arrows mark equal distances along the microtubules from the beginning of the bend. Because the microtubules are flexible but of constant length the doublet on the inner, concave side of a bend must slide tipward. The resulting displacement ($\Delta$) can be measured at the tip, at the end of subfiber *B* or at the end of the bent region. The length ($l$) of the arc of the bend is given by the formulas. Because the length of the inner arc is proportional to the radius ($r$) and the length of the outer arc is proportional to that radius plus the diameter of the axoneme ($d$), the difference in length ($\Delta$) is proportional to the axoneme diameter ($d$). As the three base-to-tip cross-section diagrams of the indicated part of each cilium show, in the straight position all subfiber *B*'s should be present but in the bent positions subfiber *B* of the doublets on the outer side of the bend will be missing; the expectation is confirmed by the three electron micrographs (*top*). When the dynein arms cause microtubules to slide past one another, shear resistance in the cilium changes sliding to bending.

11—so that it sits alongside subfiber *A* like an intersecting, incomplete cylinder. The dynein molecules, it became clear, are spaced along subfiber *A* in two differently oriented rows; they form two sets of arms (inner and outer) that point toward subfiber *B* of the adjacent doublet and are long enough to bridge the space between microtubules. This is reminiscent of the structure of a muscle fibril, in which the ATPase ends of myosin molecules on the thick filament point toward and, under the right conditions, connect with thin filaments so as to slide sets of filaments past one another and thus cause the muscle to contract [see "The Cooperative Action of Muscle Proteins," by John M. Murray and Annemarie Weber; SCIENTIFIC AMERICAN, February]. If the ciliary microtubules could "walk along" one another by means of their dynein arms, one could envisage that, given the right sequence of activity and appropriate shear resistance within the axoneme, bends would form and would propagate in such a way as to produce the movements characteristic of cilia. The hypothesis that this is the mechanism of ciliary motion is called the sliding-microtubule hypothesis.

The first experimental evidence that microtubules slide when cilia move was obtained in my laboratory at the University of Chicago in 1964. I was trying to determine the sequence in which the nine doublets terminate at the tip of a cilium. I did that by examining the cilia of cells that had been fixed quickly in osmium tetroxide, thus "freezing" the cilia in the middle of a metachronal wave and making it possible to look at the tips of cilia in different positions in the electron microscope. I found that whereas different cilia that were in the identical stroke position appeared to have identical tips, the tip pattern in cilia in one stroke position was not the same as it was in cilia in another position. A special feature in the gill-cell cilia of the freshwater mussel I was examining enables one to number each microtubule unequivocally: there is a diamond-shaped bridge between the two microtubules that are at the leading edge of the cilium when the cilium is in its effective stroke, that is, when it is actively moving liquid. If an imaginary line connecting the center of the bridge with the middle of the central pair of single microtubules is extended across the cilium, it transects a doublet on the opposite side. This doublet, which is at the leading edge of the cilium during its recovery stroke, is designated No. 1. That makes the doublets that are connected by the bridge Nos. 5 and 6.



**ADJACENT DOUBLETS in a longitudinal section near the tip (*right*) of a mussel-gill cilium are enlarged about 60,000 diameters in this micrograph. As the arrows indicate, subfibers *B* of the two doublets end at different points; the distance between the ends is the displacement caused by the sliding of the doublets. Both subfibers *A* continue toward the tip.**



**RADIAL SPOKES are seen in electron micrographs made by Warner. A flattened-out subfiber *A* from a blowfly sperm tail is enlarged some 250,000 diameters and negatively stained (*top*), revealing the protofilaments and the projecting spokes. The periodicity of the spokes is accentuated in a micrograph of a section of a mussel-gill axoneme (*bottom*). On the lower side of the axoneme, where the spokes are directly in the plane of the section, they are seen to be arranged in triplet groups with a major repeat (*arrows*) of .086 micrometer. The print was made by multiple exposure of the negative, which was advanced .086 micrometer between successive exposures. The image remained sharp because the periodicity is invariant.**

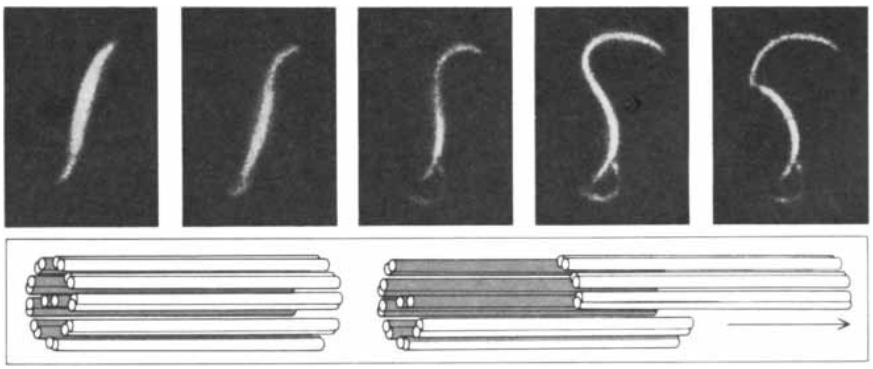
What I found was that when the cilium was bent so that doublets Nos. 5 and 6 were at the leading edge, No. 6 appeared to be longer at the tip than No. 1. When doublet No. 1 was at the leading edge, on the other hand, it appeared to be the longer. I concluded that doublets No. 1 and No. 6 must actually be about the same length and that they must be displaced as the cilium bent in a manner consistent with the sliding of the microtubules past one another [*see illustration on preceding page*]. Tip differences can be measured quite accurately for every microtubule doublet in the axoneme, and the amount of displacement can also be measured at several points along the axoneme. The displacement is strictly proportional to the amount of bend, as would be predicted by the rigorous application of simple geometric relations if the microtubules do not actually change length during the beat. The constant of proportionality depends on the diameter of the axoneme, since that is the amount by which the radii of curvature of the inner and the outer microtubules differ. Whereas cilia vary enormously in length, from about a micrometer (a thousandth of a millimeter) to two millimeters, the diameter of the axoneme is invariably about .2 micrometer. The constancy of diameter suggests that the sliding mechanism must be quite general for these

organelles in many different species and tissues.

At the University of California at Berkeley, together with Fred D. Warner of Syracuse University, I have been able to carry the geometric analysis a step further. We have found that there is a built-in ruler along the microtubules in the form of a repeat pattern of the projections known as radial spokes, which appear to connect the doublets to the central sheath. The repeats mark off stretches exactly .086 micrometer long in all the cilia we have studied. The periodicity is invariant in both straight and bent regions of the axoneme, effectively demonstrating that the microtubules do not contract at all.

This interpretation was amply and strikingly confirmed by Keith E. Summers and Ian Gibbons at the University of Hawaii. They took detergent-treated sperm tails and exposed them briefly to the protein-digesting enzyme trypsin, so that the trypsin broke up some of the more complicated linkages in the axoneme but left the microtubules and the dynein arms more or less intact. When they then added ATP, the axonemes no longer swam; instead they fell apart! Summers and Gibbons made dark-field photomicrographs of the axonemes at the moment the ATP reached them. At that instant each axoneme grows longer, and correspondingly thinner, as groups of doublets manage in effect to walk along their neighbors and then on beyond them [*see top illustration on this page*]. Clearly the trypsin uncouples the sliding of microtubules from the bending of the axoneme.

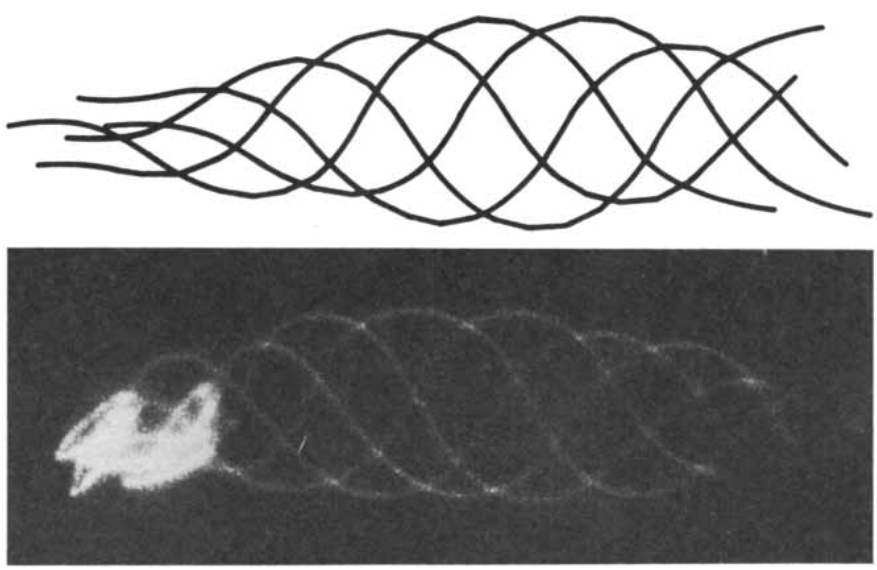
The radial spokes are among the structures that are digested by trypsin. So are certain interdoublet links composed of a protein called nexin. It seems likely that one of these structures normally provides the shear resistance within the axoneme that converts sliding into bending. The best candidate for this function, in $9 + 2$ axonemes at least, seems to be the radial spokes. For one thing, they are aligned in the direction in which the bending takes place. For another, Warner and I have found that the configuration of the spokes is different in straight regions of the axoneme from what it is in bent regions. In straight regions most of the spokes stand perpendicular to the microtubule doublet at which they originate; in bent regions many of the spokes are tilted and perhaps stretched, as would be expected if they had been firmly attached at both the doublet and the central sheath when the doublet slid forward.



**SLIDING AND BENDING are uncoupled when axonemes are subjected to trypsin, which breaks down some of their protein structures, notably the spokes. Keith E. Summers of the National Heart and Lung Institute and Ian R. Gibbons of the University of Hawaii treated sea-urchin sperm tails with detergent and trypsin and added ATP. The sequence of dark-field micrographs (*top*) shows how the axoneme elongates and becomes thinner as the ATP causes the microtubules to slide along one another. The diagram (*bottom*) indicates schematically how the sliding of spokeless microtubules could have this effect.**

Such attachment would provide effective resistance in places where bends develop, and thus bring about bending, when the doublets move past one another. If that is the case, then in axonemes in which the attachment is defective the cilia should bend with difficulty or even be nonmotile. Sir John Randall and his colleagues at Kings College London discovered mutants of the single-celled green alga *Chlamydomonas* in which the nine doublet microtubules and the arms look structurally normal but in which the central sheath and the central pair are improperly formed. Such mutants are indeed nonmotile, even though the structural basis for microtubule sliding is presumably present. The biochemistry of these cilia is not well established, however, and there may be deficiencies other than spoke attachment that explain the mutant behavior. (A further difficulty is that there are unusual sperm tails that have axonemes with microtubules without the $9 + 2$ arrangement and possibly without spokes, but that nevertheless generate bends.)

The assumptions of the sliding-microtubule hypothesis have been further tested by Charles J. Brokaw of the California Institute of Technology, who developed a computer program based on



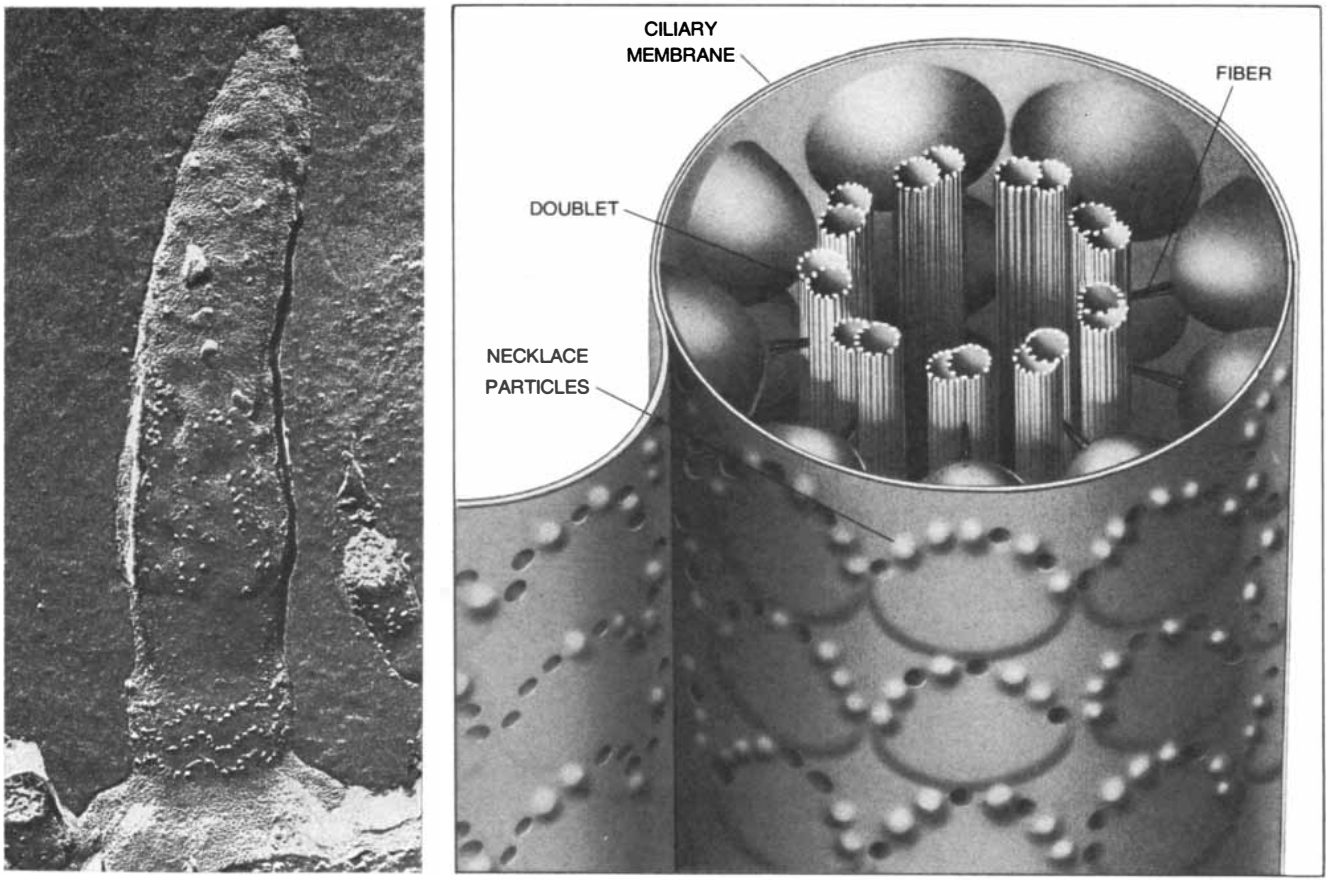
**COMPUTER SIMULATION of a beating sperm tail (*top*) approximates the real thing (*bottom*) when the data supplied to the computer reflect the assumptions of the sliding-microtubule hypothesis. Charles J. Brokaw of Cal Tech devised a computer program that in effect generates a working model of a sperm flagellum. Each line in the computer printout represents the sperm tail in a successive beat position. The superposed flash micrographs, which were made at the rate of 120 flashes per second, are of a swimming sea-urchin sperm. The white oblong shapes (*left*) are successive images of the sperm head.**

microtubule sliding that can simulate the movement of a flagellum. The program generates values for bending rates along the flagellum, which are then used to print out successive portraits of the swimming organelle. The computer-simulated flagella develop and propagate bends that have wave variables identical with those of living or detergent-treated 9 + 2 axonemes [*see bottom illustration on preceding page*]. The computer printouts are much less like actual sperm-tail patterns when equations representing contraction of the microtubules are substituted for sliding.

If the sequential bends that arise along a relatively long axoneme are opposite and symmetrical, the undulating axoneme swims forward in a straight line. Many sperm tails do just that. Some cilia and flagella are capable of behavior much more complex than simple, continuous forward swimming, however. For example, when a paramecium encounters an obstacle, it can back away and then go ahead in a slightly different direction. It does so by altering and in some cases completely reversing the direction of the stroke of some of its cilia. In such cases the fundamental sliding-microtubule mechanism must be affected by specific cellular controls. Roger O. Eckert, Yutaka Naitoh and their collaborators at the University of California at Los Angeles have found that ciliary reversal is caused by an influx of calcium ion into the cell. The same ion is known to be the signal for muscle contraction. The point at which the calcium acts on the sliding mechanism of the axoneme is still not clear, and yet the U.C.L.A. observation is very important in that it suggests how changes in the concentration of ions within a cell may be directly related to the behavior of the cell and its cilia. Investigation of processes that alter ciliary behavior has considerable significance for human health, since the arrest of ciliary motion by various agents (such as tobacco smoke) can lead to pathological changes in the underlying cells.

One possible mechanism for relating microtubule sliding and ion concentrations is a "microtubule-membrane complex" we have noted at the base of cilia that are capable of complex behavior. Fibers extend from each axonemal doublet to particles in the adjacent ciliary membrane, which are arranged in a "ciliary necklace" with several strands [*see illustration below*]. The sliding of microtubules when cilia move could displace parts of the necklace and thereby alter the permeability of part of the membrane to certain ions. The alteration might in turn provide a feedback pathway that could regulate the magnitude or direction of sliding. Such a mechanism may function in cells other than ciliated ones, where microtubules may also be connected to intramembrane particles.

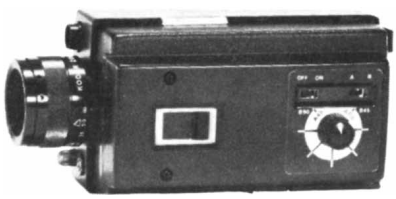
Microtubules, as I have indicated, are present in many cells and they may be capable of sliding past one another. In doing so they may serve as general organelles of transport, not only moving membrane particles but also moving chromosomes during cell division or moving enzymes and other molecules along the axons of nerve cells. All such ideas are quite speculative at this stage of investigation, but it is clear that the verification of the sliding-microtubule mechanism of ciliary motility has important implications for many unsolved questions in cell biology.




CILIARY NECKLACE is seen in a "freeze fracture" image of a mussel-gill cilium (*left*). The preparation was frozen and then fractured, cleaving the ciliary membrane; a cast of the interior of the membrane was made, so that the electron micrograph of the cast is a replica of the membrane interior, showing three strands of particles near the base of the cilium. The diagram (*right*) is a reconstruction of the microtubule-membrane complex that apparently gives rise to the necklace. The membrane has been split open, revealing the necklace particles, which are connected to the middle of the doublet microtubules by wineglass-shaped structures.

**Fear of crime can help see growth of**
**tissue**
**tides**
**traffic jams**
**corrosion**
**construction projects**
**storms**
**birds' nests**

## Pick up one of these



Widespread worry over suspicious activities brought forth the KODAK ANALYST Super 8 Camera. The quantity produced has been sufficient to keep the price less than $245, complete with mounting bracket, sound-absorbent box, sighting mirror.

If interested in acquiring one or more, write Kodak, Graphic Data, Department 661B, Rochester, N.Y. 14650. (Hold your payment. Just write.)

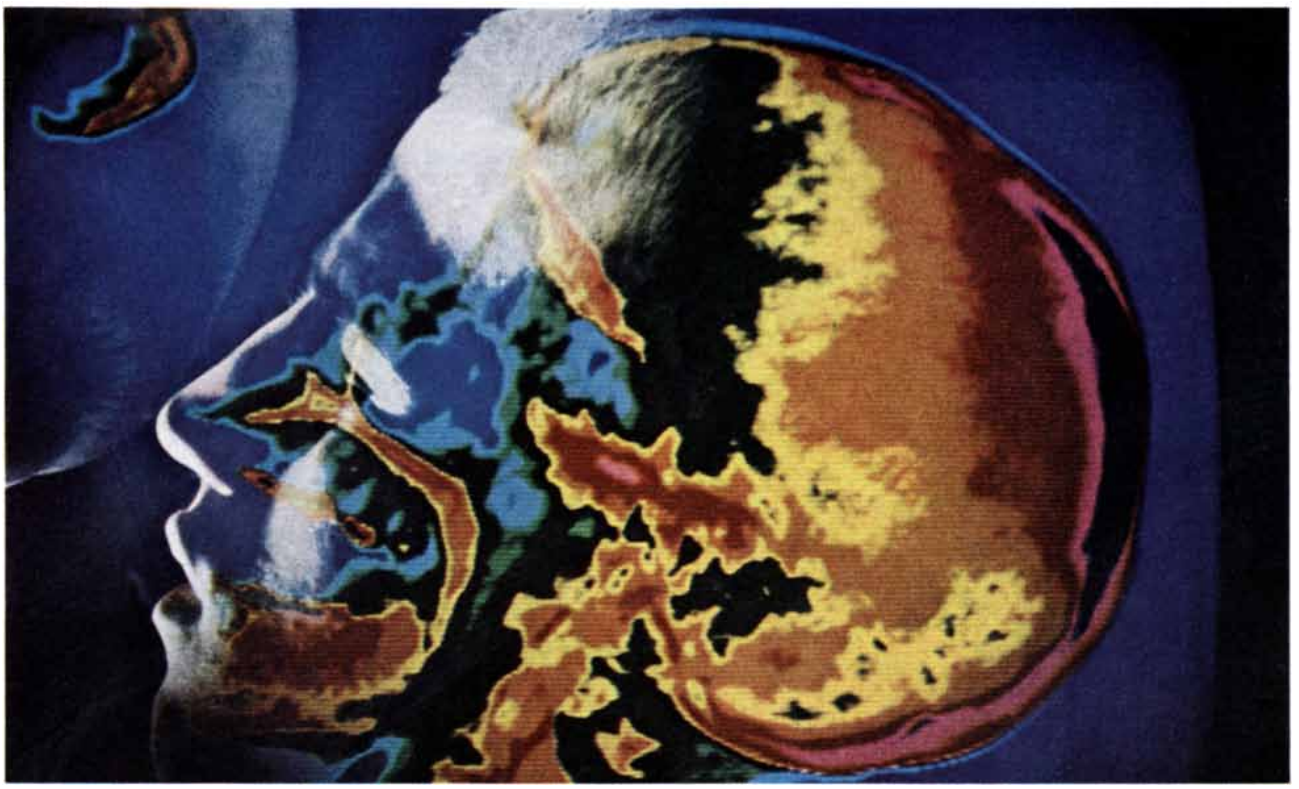
The camera takes 3600 pictures at a loading, 7200 if you can forego showing them on a movie screen. Choose your interval anywhere from 1¼ seconds to 90 seconds. For longer intervals a suitable timing device can be plugged in. Shutter speed is 1/100 sec. Operates as cold as −18 C. Powered by 4 AA-size 1.5 V alkaline batteries or by converter from 117 V, 60 Hz outlet. Aperture adjusts itself to the light conditions up to *f*/1.9. Lens has focusing scale and zooms to permit choice of field coverage from 10 inches x 14 inches at 6 feet to 45 feet x 61 feet at 150 feet.

Takes a variety of films that come in 50-ft super 8 cartridges, including KODAK 4-X Reversal Film 7277 at 400 daylight speed, color films at 160 tungsten speed, black-and-white films that can be processed to a negative or reversal image either commercially or with your own simple equipment. Also takes 100-ft cartridges of KODAK MFX Film that can be processed to a negative or positive for study on a film editor or viewer. Interesting frames can be blown up to paper prints.

Kodak

*Price subject to change without notice.*

# Why jump into a specialized field until you've studied all the possibilities?



If you're a new physician, you might not have decided on a field of specialization yet. That's easy to understand. But while you're deciding, you can be getting valuable experience and expert guidance towards that decision as a physician in the U.S. Air Force. Air Force physicians are exposed to many fields of medicine. They practice with highly professional people in modern facilities. And all Air Force physicians have seemingly unlimited opportunities for furthering their education and professional growth. Air Force graduate medical education programs are available. And because Air Force physicians participate in Group Practice (a collection of specialists in different fields complementing each other by expanding total capability) they all expand their individual ability and horizons with each new patient.

As an Air Force physician (and officer) you can apply for worldwide work assignments. The salary is excellent. And you get a 30-day paid vacation each and every year.

If you're still deciding on a medical specialty...or a future...look into the Air Force before you make a final decision. We think you'll like what you see. For more detailed information about Air Force opportunities in medicine please fill out and mail the coupon.

C-SA-104

Air Force Opportunities
P.O. Box AF
Peoria, IL 61614

Please send me more information on the Air Force Physician Program. I understand there is no obligation.

Name _____ Sex ☐ M ☐ F
(Please Print)

Address _____

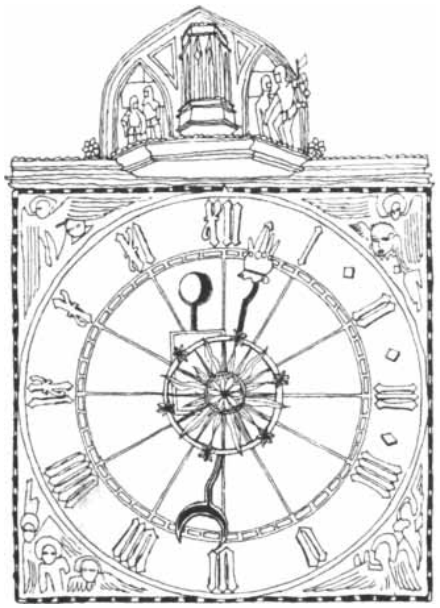
City _____

State _____ Zip _____ Phone _____

Soc. Sec. # _____ Date of Birth _____

**Health Care at its best.**
**Air Force.**

# SCIENCE AND THE CITIZEN



### *Binary Subtraction*

In a little-noted action, coinciding closely with the resignation of President Nixon, the House of Representatives approved by a vote of 315 to 70 a resolution urging the Senate to ratify the 1925 Geneva Protocol banning the first use of chemical weapons in war. Although the U.S. was one of the original signers of the protocol, it is the only major power that has not formally ratified it. Five years ago, when President Nixon sent the protocol to the Senate for ratification, he declared the U.S. would never be the first to use lethal or incapacitating chemical weapons, but he insisted that the protocol did not cover tear gas and herbicides, both of which were then being used in Vietnam. At about the same time the United Nations General Assembly was supporting by a vote of 58 to three a resolution declaring that tear gas and herbicides were covered by the protocol.

As part of its recent action the House also deleted from the Department of Defense budget a request for $5.8 million to start production of a new generation of binary nerve-gas weapons. These are devices that contain two relatively nontoxic ingredients in separate containers that are mixed and react to form a potent nerve gas only after the device is on its way to the target. Evidently the Department of Defense would like to replace its present enormous stockpile of nerve-gas weapons (estimated to contain 40 million pounds of gas) with binary weapons, in part because state governments object to the transportation of nerve-gas munitions across their borders and in part because European nations are uneasy about the large quantities of nerve-gas weapons in the stockpiles of the North Atlantic Treaty Organization (NATO). Representative Wayne Owens, Democrat of Utah, has estimated that to replace the existing stock of nerve-gas weapons with binary ones would cost between $1 billion and $2 billion.

Meanwhile in Geneva the Conference of the Committee on Disarmament (CCD) has been meeting five months a year for two years. Twice a week delegates of 25 nations listen to one another discuss the elimination of chemical weapons from the arsenals of the world. Under the 1972 Biological Weapons Convention the U.S. and other powers committed themselves to negotiate in good faith toward an early agreement to ban such weapons.

At hearings held by a new subcommittee on arms control and disarmament of the House Armed Services Committee, Fred C. Iklé, director of the Arms Control and Disarmament Agency, testified: "If we start on a new type of production program [of binary nerve gases], it becomes even harder to envisage constructive arms-control agreements limiting competition in chemical weapons." The House evidently heeded the testimony of Iklé and others who spoke against the new program. In transmitting the resolution urging the Senate to ratify the Geneva Protocol the House subcommittee suggested that the Senate has the power to ratify the protocol without excluding herbicides and tear gas.

### *Arms and the Congress*

Congress should have a more effective voice in making decisions on national security, particularly with regard to strategic-nuclear-weapons systems, and to that end organizational and procedural changes are necessary. Planning and the authorization of funds should be projected over a longer period. Funding should be better coordinated with the requirements of arms-control negotiations. And Congress should have available to it professional resources for the evaluation of proposed military programs; that will require, among other things, improved access to secret military information for members of Congress, their staffs and their outside advisers. These are the conclusions of a statement issued last month by the research and policy committee of the Committee for Economic Development (CED). The statement was drawn up by a subcommittee on decision making for national security chaired by Franklin A. Lindsay, president of the Itek Corporation; the project director was Thomas C. Schelling of Harvard University.

The statement points out that the useful life of a major new weapons system "typically does not begin until nearly a decade after production is authorized; its lifetime then extends through that second decade and beyond." Congressional committees reviewing strategic weapons should therefore focus on their long-term justification and should insist that requests for authorization be made "in the context of projected second-decade needs." Moreover, components of the annual defense budget should be projected for five years, and "items requiring long lead time should be the subject of actual five-year authorizations." Each year Congress would revise the authorization for the next four years and add a new fifth-year authorization.

Decisions on U.S. weapons systems are strongly influenced by what the U.S.S.R. is doing and are also intended in part to influence Russian decision making, according to the statement. Congress can properly participate in the process of bargaining on arms limitation by authorizing or failing to authorize a specific program, and even by indicating that it might authorize one. There is danger, however, in adopting programs simply as bargaining chips, the CED subcommittee warned. "There are certain diplomatic tactics for which the legislative branch of government in a democratic society is just not suited." The bargaining-chip tactic encourages indecisiveness and either can saddle the country with weapons that were meant to be merely threats or can be exposed as dangerous bluffing. The CED subcommittee believes "Congress has a positive role to play in the process of strategic arms limitation bargaining" but urges Congress "to be doubly cautious about authorizing any system that is justified principally in terms of its bargaining value."

Government programs have in general

become "increasingly technical, complex and esoteric," and moreover military programs are veiled in secrecy. It is therefore difficult for Congress to evaluate military requests. The CED subcommittee recommends the establishment of a new professional research organization: "an office, institute or institution for research and evaluation" that could help Congress to play "a stronger role in scrutinizing and reviewing the defense budget and weapon programs." Congress should also encourage the existence of "diverse centers of analytic strength" by regularly calling for independent advice and testimony from organizations and individuals outside the Government.

Central to proper Congressional participation is "the need for security clearance, for access to classified information and the right to have custody of it, and for participation in classified discussion and examination as well as analysis and research." Security clearance is now the responsibility of the executive agencies. Members of Congress, sitting in executive session, are sometimes given classified information, and a few Congressional staff members have top security clearance, but the arrangements are "awkward and discriminatory." The CED subcommittee believes Congress must become independent of the Department of Defense in this area by enacting "legislation establishing its own bipartisan procedure for certifying individuals and organizations cleared for access to security information."

### *Weight-watching*

Does the force of gravity change over a period of time? Observations of the motions of the sun and the moon have yielded evidence that the universal gravitational constant $G$ may indeed be decreasing at the small but distinct rate of one part in $10^{10}$ per year. If these preliminary results are confirmed by experiments currently in progress, they will profoundly affect modern theories of the physical world.

It has long been known that the length of the terrestrial day is increasing by .002 second per century as a result of the fact that the rate of the earth's spin is slowly decreasing. Most of the retardation is caused by the friction of the tides. The angular momentum lost by the rotating earth is transferred to the moon, which absorbs it by receding from the earth at a rate of three centimeters per year. There is, however, a puzzle: the rate at which the earth has been slowing down is only about half the rate that would result from tidal friction alone. The discrepancy can be accounted for if the value of $G$ is decreasing. Indeed, such a decrease was predicted as early as 1937 by P. A. M. Dirac, and it is also incorporated in the cosmological theory put forward by Fred Hoyle and J. V. Narlikar and the gravitational theory developed by Carl H. Brans and Robert H. Dicke.

Thomas C. Van Flandern of the U.S. Naval Observatory and his colleagues have been examining the motion of the moon by timing occultations, or eclipses, of stars as the moon passes in front of them. The observations are made with respect to two independent standards of time. One standard is atomic time, which is measured with an atomic clock whose rate is presumed to be uniform; the other is ephemeris time, which is based on the orbital motion of the earth around the sun and is thus subject to any changes resulting from a change in $G$. Observations made with respect to both standards show that the times at which the moon occults specific stars in its path get later as the moon recedes from the earth and hence takes longer to complete each orbit. Observations made with respect to atomic time measure the total change in the moon's orbital motion from all causes. Observations made with respect to ephemeris time, however, exclude any contribution from a changing value in $G$ because the earth is receding from the sun at a rate proportional to the rate at which the moon is receding from the earth, and the two effects exactly cancel each other. Therefore any differences between the measurements made by one time standard and those made by the other would reveal the effect due exclusively to a decrease in the value of $G$. It is such differences that Van Flandern and his colleagues have detected.

There are three other experiments in progress whose results promise either to verify or to negate the results from the observations of lunar occultations within about a year. One is the laser-ranging experiment utilizing the three reflectors placed on the moon by U.S. astronauts; it will precisely measure the rate at which the moon is receding from the earth. The second is a radar-ranging experiment that will try to determine the rate at which planets other than the earth (notably Mercury) are receding from the sun. The third is a sensitive laboratory experiment to detect whether or not the rate of spin of two small masses rotating between two large ones decreases with time as a result of a decreasing gravitational attraction between the two sets of masses.

If the occultation results are correct in indicating that the force of gravity is

slowly weakening, the puzzle about the discrepancy in the rate at which the earth's rotation is decreasing would be solved. Moreover, theories that predict that gravity is weakening also predict that the solid earth is expanding at the rate of about a centimeter per century, which could have something to do with large-scale geological processes such as continental drift. And if the value of $G$ is indeed decreasing, it would mean that the general theory of relativity is incomplete, since it does not take such a phenomenon into account.

### *Enrichment by Laser*

Laser methods have been successfully employed by several laboratories to enrich natural uranium in its content of the fissionable isotope U-235. Moreover, it is reported that high levels of enrichment can be achieved in a single pass through a laser device. The technique consists in selectively exciting atoms of U-235 with laser light tuned to a wavelength that is not absorbed by the more abundant atoms of U-238, then ionizing the excited U-235 atoms and collecting them electrostatically. Although the technique has clear military implications, its chief attractiveness lies in its potential for producing cheaper enriched uranium fuel for nuclear power plants.

In the U.S. uranium is currently enriched by the gaseous-diffusion method developed in the 1940's. The enrichment step accounts for about a third of the total fuel-cycle cost of a nuclear power plant. By 1983 the demand for enriched uranium for such plants is expected to exceed the capacity of the present diffusion facilities, and new facilities will have to be built. Enrichment by high-speed gas centrifugation has been widely discussed, but the feasibility of the process on a large scale has yet to be demonstrated. According to a study by David F. Edwards of Colorado State University, the laser technique opens up the possibility of an isotope-separation process with the effectiveness of the gaseous-diffusion and gas-centrifuge methods at a small fraction of their cost.

The first report of uranium enrichment with a laser was released last year by the Exxon Nuclear Company and the Avco Everett Research Laboratories. Since that time similar successes have been reported by groups in Israel and at the Lawrence Livermore Laboratory. In the Livermore experiment vaporized uranium metal was illuminated by laser light at a wavelength of 5,915.4 angstroms and by ultraviolet radiation from a mercury-vapor lamp. The laser light

was absorbed by U-235 but not by U-238. (The absorption peak for U-238 is .1 angstrom lower.) The ultraviolet radiation then kicked some of the electrons out of the excited U-235 atoms, and the positively charged ions were collected in a negatively charged cup. The flow of U-235 was too small for significant accumulation in the cup, but the effect could be measured in a mass spectrometer.

The laser-excitation method can be used to separate isotopes of other elements. Experiments in the U.S.S.R. have demonstrated that nitrogen 15 can be separated from nitrogen 14 by the photodissociation of ammonia. Workers in the U.S. have reported the laser separation of hydrogen, boron and bromine isotopes. According to Edwards, isotope separation by laser excitation promises to provide many of the natural isotopes at a cost many times lower than the costs prevailing.

### *Poor Man's Satellite*

High-altitude balloons have been used for astronomical purposes since the mid-1950's. They have attained altitudes up to 155,000 feet, and the flights have lasted as long as several days. Currently the National Scientific Balloon Facility at Palestine, Tex., a division of the National Center for Atmospheric Research, is conducting experiments with a type of balloon that is capable of attaining similar altitudes, of carrying payloads of some 500 pounds and of remaining aloft for many months and repeatedly circumnavigating the earth.

The balloon being used for such long-duration flights is a superpressure balloon: its internal pressure is slightly higher than the external pressure of the atmosphere when it reaches the desired altitude. Most standard balloons are zero-pressure balloons: their internal pressure is the same as the external pressure. At night the gas cools and diminishes in volume until it is denser than the surrounding air, with the result that the balloon sinks unless ballast is periodically released. When the balloon runs out of ballast, its flight comes to an end. In a superpressure balloon the gas remains at a pressure higher than that of the surrounding air even after the balloon cools, so that there is no need for ballast.

Since the superpressure balloon is flying at an altitude of some 30 miles, it is above 99 percent of the earth's atmosphere. It is therefore a good observation platform for long-term astronomical investigations that cannot be pursued from the ground. Such investigations include

# Wayne Sutterfield takes a lot of garbage

Look, we've all got to take a little sometime...but this guy goes looking for it.

Wayne Sutterfield is Refuse Commissioner for the City of St. Louis and he has 900 tons of garbage on his hands each day. So Wayne Sutterfield believes in new approaches to solid waste management.

The city's garbage is continually combed for steel cans. Sutterfield's powerful magnetic conveyor belts literally suck them right out of the garbage at the rate of about 78,000,000 a year. Steel's unique magnetic property makes it possible.

The reclaimed cans are marketed to a steel producer who recycles them during the production of new steel products. The economics help Sutterfield and the people of St. Louis 'clean up' in the bargain. The same thing is happening today in 20 cities across the country.

But steel can recovery isn't the only exciting thing that happens to the city's garbage. St. Louis burns some 300 tons of its combustible solid waste each day (mixed with coal) to generate electric power in a conventional boiler at the Union Electric Company. That power is enough to serve 25,000 homes. The program reduces air pollution, saves landfill and conserves coal.

St. Louis was the first American city to try it, but it won't be the last.

Sometimes it pays to take a little garbage.

For more information write, **Tinplate Producers, American Iron and Steel Institute, 1000 16th Street, N.W., Washington, D.C. 20036.**

# ...and cleans up in the bargain.

Steel
the recycled material

**Tinplate Producers**
**American Iron**
**and Steel Institute**

# He wants oil without oil spills. So do we.

Meet Walt Welles, conservationist, former navigator on an oceanographic schooner. Loves the sea. Worries over what man might do to it. Hates oil spills.

But he also likes what petroleum does for his enjoyment of nature. Quick-drying sails made from petrochemical feedstocks. A maintenance-free fiberglass hull, also petroleum-derived. An auxiliary engine that couldn't go anywhere without diesel fuel.

So he wants conservation—plus adequate supplies of petroleum. Which, to Walt Welles, is simply a matter of realistic trade-offs on conservation issues involving oil.

Offshore drilling? "Yes, within reason," he says. "I certainly don't want to look at nothing but oil wells when I put to sea. But I'm told a single production platform today does the job of many earlier ones." We could also tell him some of our plans for the future, like producing oil on the bottom of the sea. In many cases with no visible platforms at all. This new system is being tested right now.

As to building deepwater ports to handle very large tankers: "I'm excited, as a conservationist, that crude oil could be unloaded at a buoy 10 or 20 miles out to sea, then carried by submarine pipeline to storage tanks ashore. That means fewer vessels in our ports. Fewer risks of collision and oil spills. And no need to dredge our harbors, damaging the eco-system."

It also means that we'd be able to handle our imported crude oil in the most economical way.

"The trick," Welles says, "is to make the ocean work for all of us. Yield her fish *and* her oil. Without messing up the critical balance of nature. Or the beauty we all love."

Well said, Walt. We couldn't agree more.

Mobil®

# More car in a small car. That's par for Duster.

Like all the best-selling compacts from Plymouth, Duster's game is strong and complete.

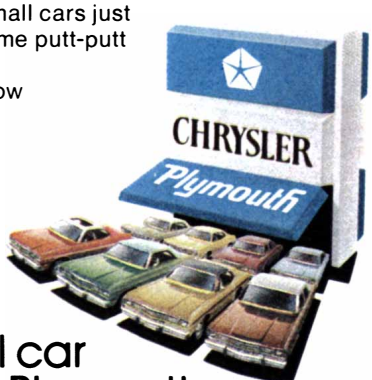
**You get great distance from a Duster.** You'll get a good long roll from a gallon of gas. And speaking of distance, with unleaded gas, Duster's new spark plugs can go as much as 30,000 miles without replacement. Which fits in just fine with Duster's money-saving Electronic Ignition System (no points or condenser to replace).

CHRYSLER MOTORS CORPORATION

**Check the form on this superstar.** Inside, there's comfortable seating for a fivesome. On the small car circuit, five is one better than par. And Duster's trunk takes honors among small cars, because it's big enough for the longest of drives.

**There aren't any holes in Duster's game.** It's priced lower and has more room than some very popular imports.

So, you see, some small cars just drive for show. And some putt-putt to save you dough. But Duster does it *all*. So now you know.




America's **No.1** selling small car comes from Plymouth.

the recording of cosmic rays, the collection of interplanetary dust and micrometeorites and the study of many celestial phenomena at wavelengths in the electromagnetic spectrum that are blocked by the atmosphere.

The cost of flying an experiment aboard a superpressure balloon is only a few percent of the cost of placing it in orbit around the earth. Accordingly the superpressure balloons have been called "poor man's satellites." While the balloon is aloft investigators can remain in radio communication with it in order to issue commands to their apparatus and retrieve data. At the end of the flight the payload can be recovered. Thus the balloon also provides a relatively inexpensive way for an investigator to make a trial run of an experiment that may later be put into space.

### *Lineups*

Managers and other baseball strategists give much thought to the batting order, following such precepts as putting a good runner first and a big hitter in the "cleanup," or fourth, spot. Does it make any difference? R. Allan Freeze of the University of British Columbia, writing in *Operations Research*, says that "the effect of using the best batting order rather than the worst is less than three extra wins per 162-game season."

Freeze conducted a computer simulation of some 200,000 baseball games, programming the computer with the main features of the *Sports Illustrated* Baseball Game, which is normally played by the repetitive throwing of three dice. (With his program he could simulate a set of 10,000 games in less than three minutes.) For his data he chose two groups of players from the New York Yankees. One group consisted of nine of the batters and four of the pitchers on the 1970 team and therefore, according to Freeze, represented "an average grouping of major-league players." The other group was "a composite lineup of all-time Yankee greats" and so was "of Hall of Fame caliber." Freeze simulated games with the teams in the traditional lineup, a second lineup putting the hitters in descending order of productivity and a third putting them in ascending order of productivity. He reached the "hardly surprising" conclusion that the traditional lineup is helpful. "What may be surprising is the very small influence of what most baseball connoisseurs would consider to be a monumentally absurd strategy," putting the big hitter (say Babe Ruth) last in a weak lineup.

# NITROGEN FIXATION

Life requires nitrogen that has been "fixed" through combination with other elements. Thermodynamics imposes a constraint on all possible methods of fixation: invariably energy must be supplied

by David R. Safrany

Nitrogen is one of the major constituent elements of life; it is an essential material in the construction of proteins and other biological molecules. Nitrogen is also the major component of the atmosphere: dry air consists of about 79 percent nitrogen. The abundance of the element does not ensure its biological availability, however, because nitrogen can be incorporated into most living systems only in "fixed" form, that is, only when it is combined with certain other elements.

The natural supply of fixed nitrogen is limited, and it imposes a limit on the capacity of world agriculture. Supplemental nitrogen must be provided by chemical fertilizers containing nitrogen fixed by industrial methods, chiefly the Haber process, in which nitrogen is combined with hydrogen to form ammonia. In 1972 more than 14 million tons of ammonia were manufactured in the U.S., and production is doubling about every five years. The demand will continue to grow as the world population increases, and in particular as the techniques of modern agriculture, which include the intensive application of nitrogenous fertilizers, are introduced in the underdeveloped regions of the world.

Among industrial chemical processes the Haber process is remarkably efficient, and its product is comparatively inexpensive. Until the "energy crisis" of the past year the manufacturing cost of ammonia was stable at slightly more than one cent per pound. Because of the enormous quantity of fixed nitrogen consumed, however, and because most of it contributes to the production of a basic commodity (food) a price reduction would have important and widespread consequences. During the past 60 years numerous alternatives to the Haber process have been proposed. The commercial success of such proposals is determined less by the ingenuity of the chemist in discovering novel reactions than by thermodynamic constraints that apply to all possible methods of nitrogen fixation.

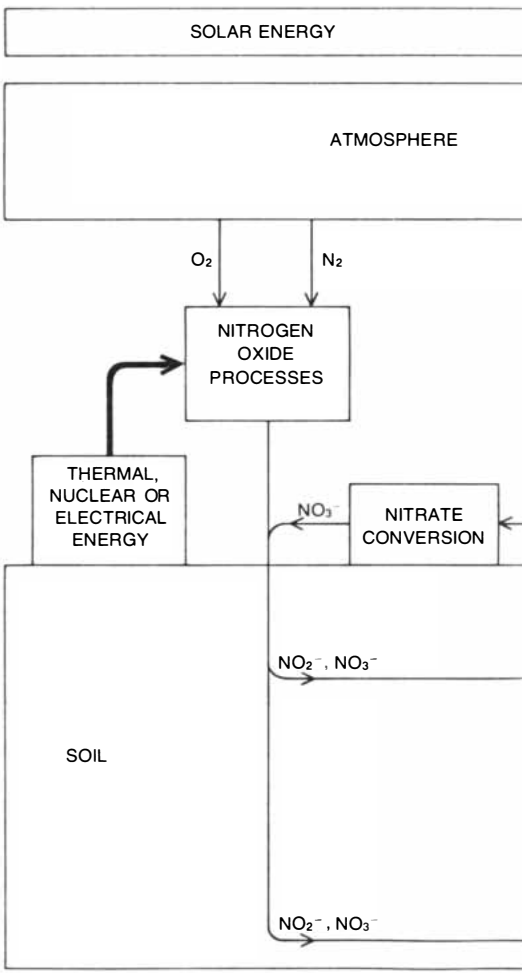
Nitrogen is useful to life only in fixed form because nitrogen as it is found in the air, as a diatomic gas ($N_2$), is nearly inert. In order to make it combine with other substances an irreducible quantity of energy must be added to the system. This energy cost can be paid in any of several currencies, through an energy-rich raw material, for example, or directly as thermal or electrical energy, but it must be paid. For this reason any large reduction in the cost of fixed nitrogen is likely to result not from improvements in the technology of fixation itself, which is already highly developed, but from advances in the technologies that supply the energy or raw materials needed for fixation.

## Free Energy

Thermodynamics is the science that enables one to calculate the energy absorbed or dissipated by the reactants during a chemical process, and thereby to determine whether or not a process is possible under given conditions. A fundamental concept in thermodynamics is the concept of "free energy," an intrinsic property of substances expressed in kilocalories per mole. (A mole is the equivalent in grams of the molecular weight of a substance.) In a chemical reaction the change in free energy, denoted $\Delta F$, can be calculated simply by subtracting the sum of the free energies of the reactants from that of the products.

An analogy can be made between free energy and gravitational potential energy, as exemplified by the physics of a roller coaster. The roller coaster car at the top of the highest hill has maximum potential energy; as it rolls downhill its potential energy decreases and is converted to kinetic energy, or motion. When the car reaches ground level, its potential energy is exhausted.

Similarly, in a chemical reaction free energy decreases and is converted to kinetic energy, which is manifested as




**NITROGEN CYCLE describes a continuous interchange of the element between the atmosphere and living organisms. In broad outline, the molecular gas ($N_2$) is converted**

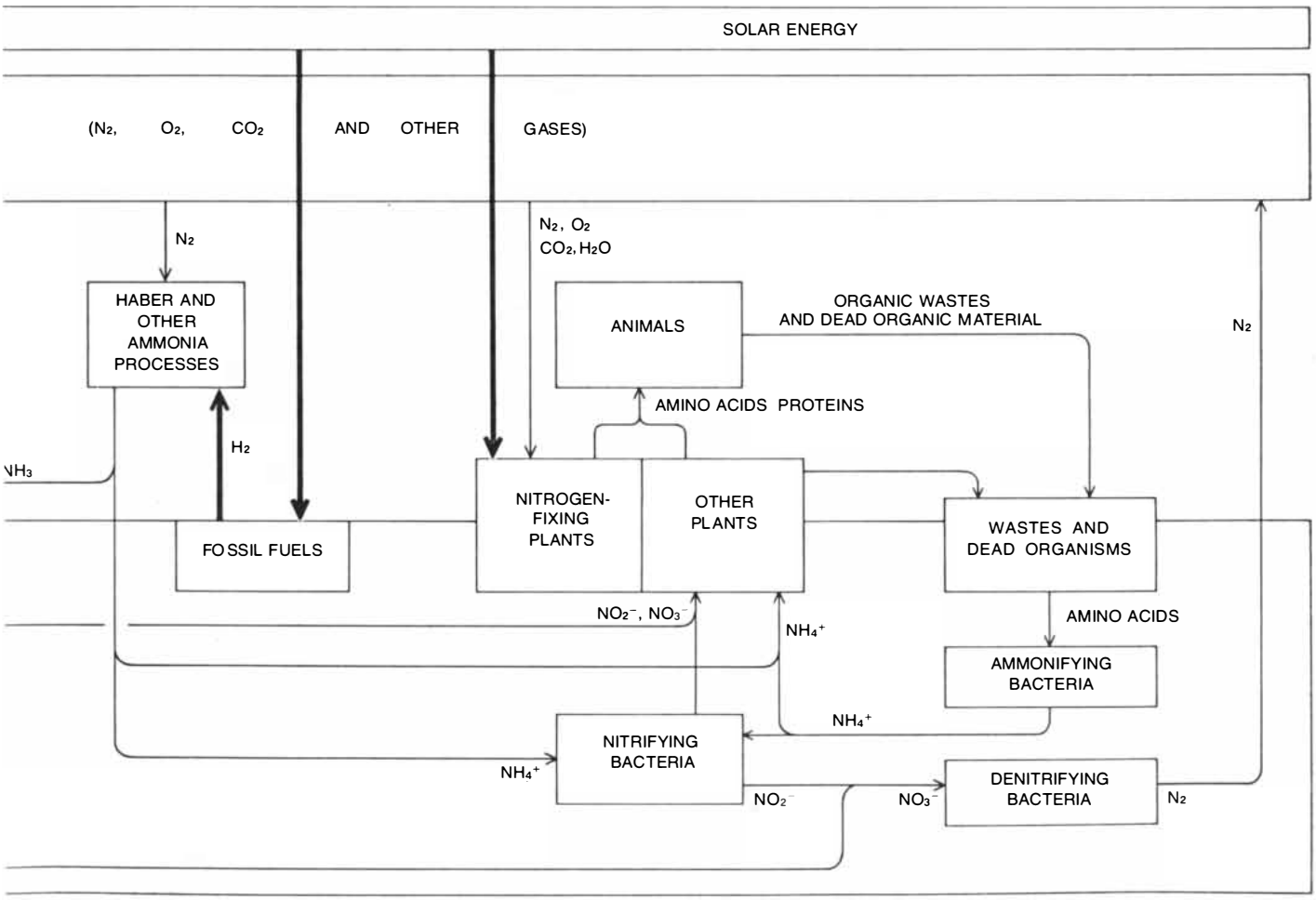
heat, until the system reaches thermodynamic equilibrium, the point of lowest free energy for that system. The reference level from which free energy is measured—the equivalent of ground level for the roller coaster—is generally set arbitrarily at the free-energy level of the stable forms of the elements involved in a reaction at some specified temperature, usually 25 degrees Celsius. In other words, the free energy of a substance in its equilibrium form at room temperature is set equal to zero. Calculated values of $\Delta F$ therefore indicate the tendency of the reactants to revert to their stable form.

Reactions that proceed with a decrease in free energy (that is, where $\Delta F$ is negative) are called exoergic reactions. Because all chemical processes are reversible there is for every exoergic reaction a corresponding endoergic one, in which free energy must be supplied. Hydrogen and oxygen, for example, combine exoergically to form water; the decrease in free energy is 57 kilocalories per mole of hydrogen consumed or water produced. Thermodynamics predicts that the reverse reaction, in which water breaks down into hydrogen and oxygen, can occur only when 57 kilocalories per mole is supplied to the system. The prediction is confirmed by the commonplace observation that water never explodes into its component gases.

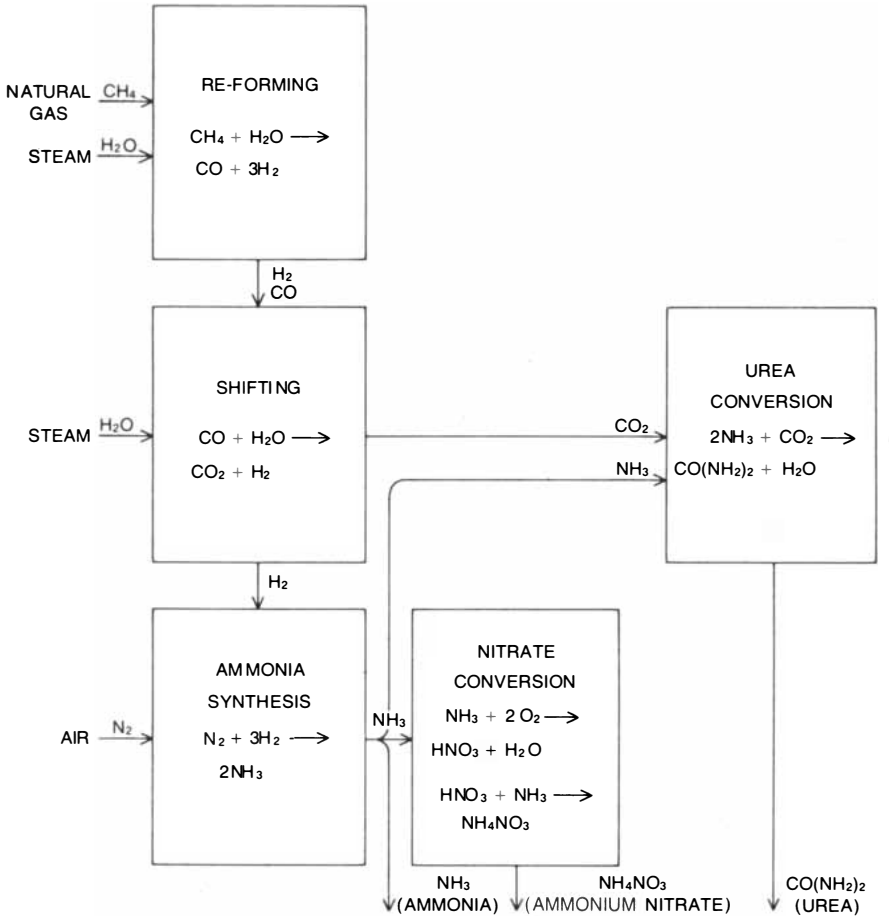
Although thermodynamics is a powerful instrument for analysis, it cannot tell us all we might want to know about the behavior of chemical systems. Thermodynamics predicts with certainty, for example, that without an external source of energy an endoergic reaction will not take place; predictions about exoergic reactions cannot be made with the same certainty. A calculated decrease in free energy indicates only that a postulated reaction is possible, not that it will actually occur. As we have seen, in the reaction of gaseous hydrogen and oxygen $\Delta F$ is $-57$ kilocalories, indicating that the reaction is strongly exoergic. Nevertheless, a mixture of hydrogen and oxygen is stable and, if it is left alone, the gases will not combine; they explode to form water only when the reaction is initiated by a small input of energy, such as that provided by an electric spark. The energy required to initiate an exoergic reaction is called the energy of activation.

Just as thermodynamics cannot predict whether or not an exoergic reaction will proceed, it cannot predict what path the reaction will take if it does proceed. Change in free energy is calculated from the properties of the reactants and products alone; there may be any number of intermediate products and dozens of possible mechanisms relating them, but thermodynamics cannot tell us anything about them. Activation energy and the paths of chemical reactions are outside the province of thermodynamics; they are the subject of the field of study called chemical kinetics.

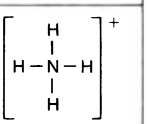
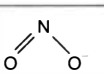
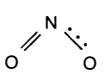
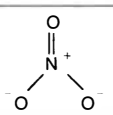
In spite of these limitations, thermodynamics provides a valuable framework for evaluating possible chemical proc-




**into fixed forms ($NH_4^+$, $NO_2^-$, $NO_3^-$); these are then degraded and returned to the atmosphere by the concerted action of several kinds of bacteria. All methods of fixation require a source of energy (*heavy lines*). Nitrogen-fixing plants (such as the legumes) obtain energy directly from sunlight; solar energy, accumulated in fossil fuels, also drives the synthesis of ammonia by the Haber process. The nitrogen oxide processes, which are not now commercially competitive, require thermal, electrical or nuclear energy.**




**HABER PROCESS** is the catalytic synthesis of ammonia. Hydrogen is derived from methane by the re-forming and shifting reactions and then combined with nitrogen at high temperature and pressure in the presence of a catalyst. The ammonia can be utilized directly or it can be converted to nitrate or urea, two common fertilizers. The Haber process, developed 60 years ago, today accounts for virtually all industrial fixed-nitrogen production.

| SUBSTANCE | MOLECULE | | ION IN AQUEOUS SOLUTION | | |
|---|---|---|---|---|---|
| | FORMULA | STRUCTURE | NAME | FORMULA | STRUCTURE |
| MOLECULAR NITROGEN | $N_2$ | $N \equiv N$ | —— | —— | —— |
| AMMONIA | $NH_3$ |  | AMMONIUM ION | $NH_4^+$ |  |
| NITRIC OXIDE | NO | $N \cdot\cdot O$ | NITRITE ION | $NO_2^-$ |  |
| NITROGEN DIOXIDE | $NO_2$ |  | NITRATE ION | $NO_3^-$ |  |

**MOLECULAR SPECIES** important in the nitrogen cycle are compounds of nitrogen with hydrogen or oxygen. The triple bond in molecular nitrogen is in part responsible for the stability of the molecule. Ammonia is the product of the Haber process, of some other potential industrial methods and apparently of biological fixation as well; the nitrogen oxides would be formed in other proposed processes. The three dots in the NO and $NO_2$ structural diagrams represent three-electron bonds. Fixed nitrogen is absorbed by plants only in ionized form; one ion can be converted to another by industrial methods and by bacteria.

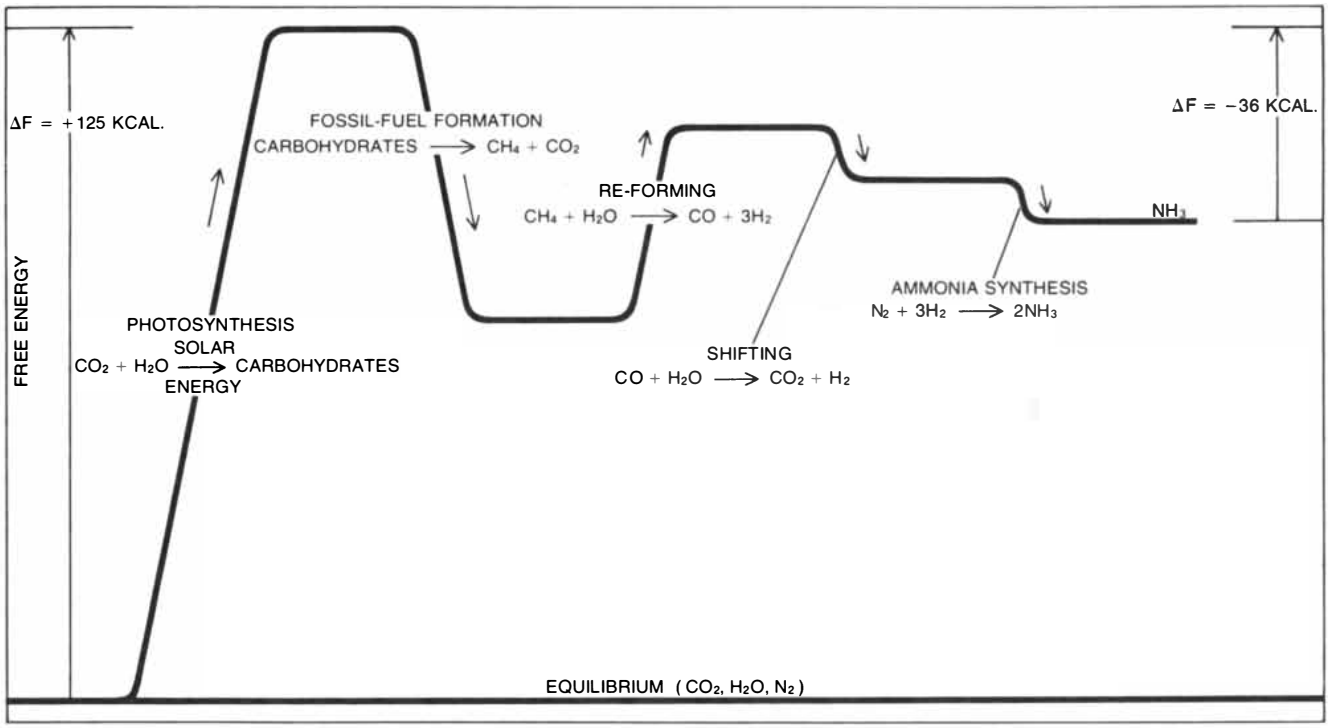
esses, such as those employed in or proposed for nitrogen fixation. Although some processes that initially appear promising from thermodynamic considerations may prove unattractive for kinetic reasons, any process that is forbidden by thermodynamics must be immediately excluded from consideration, since no manipulation of chemical kinetics can alter the fundamental free-energy transactions.

## The Haber Process

The ultimate use of most industrial fixed nitrogen today is fertilizer, but a major part of the production of the earliest Haber fixation plants was diverted to a much different purpose: the manufacture of munitions, most of which are based on nitrate compounds. Earlier supplies of fixed nitrogen were obtained largely from Chilean deposits of saltpeter (sodium nitrate) and from various chemical methods, all of them inefficient, that produce carbon-nitrogen compounds (cyanides). The introduction of synthetic ammonia just before World War I is considered to have had a significant effect on the prosecution of that war. In the long run, however, the agricultural uses of fixed nitrogen have proved much more important than munitions and all other uses.

Plant growth demands large quantities of fixed nitrogen, most of which is incorporated into proteins and other biological molecules and becomes available to animals in the form of food. Subsequently the nitrogen is returned to the soil in organic wastes and dead organisms, and eventually it is returned to the atmosphere, thus forming a closed cycle [*see illustration on preceding two pages*].

The high-yield crops of modern agriculture can quickly deplete the soil of fixed nitrogen. The supply can be replenished by applying organic fertilizers, such as mulch or manure, but the nitrogen content of these substances is not high. Another method is the growing of legumes such as beans, peas, alfalfa and clover, which are exempt from the general dependency of plants on fixed forms of nitrogen because they are able to utilize atmospheric nitrogen directly. The legumes fix nitrogen with the aid of bacteria that live symbiotically in nodules in their roots. Although the magnitude of biological fixation is difficult to measure, it appears likely that the activities of the legumes and their symbionts, and of certain algae and free-living bacteria, make the largest contribution to the world ni-




**THERMODYNAMIC ANALYSIS of the Haber process relies on the concept of "free energy," an intrinsic property of substances that determines their potential for combining with other elements. When free energy increases (the change in free energy is denoted ΔF), energy must be supplied to the system and the reaction is called an endoergic one; when a reaction is exoergic, on the other hand, free energy decreases and is dissipated as heat. The ammonia synthesis that is the ultimate event in the Haber process proceeds with a modest decrease in free energy, but the overall process is strongly endoergic. It begins with the photosynthesis of carbohydrates in green plants, promoted by light quanta from the sun. The carbohydrates decay exoergically to form fossil fuels such as methane ($CH_4$), which can then be re-formed endoergically to yield hydrogen gas. Additional hydrogen is liberated in the shifting reaction. Much of the free energy of the products of photosynthesis thus resides in hydrogen and is available to promote ammonia synthesis.**

trogen cycle [see "The Nitrogen Cycle," by C. C. Delwiche; SCIENTIFIC AMERICAN, September, 1970].

The most efficient way to replenish soil nitrogen is through the application of chemical fertilizers, in which nitrogen is a major component. The nitrogen can be supplied as anhydrous ammonia, as the ammonium ion ($NH_4^+$) or as the nitrate ion ($NO_3^-$). In some cases both ions are supplied, in the form of ammonium nitrate ($NH_4NO_3$). Another common vehicle for fixed nitrogen, urea, has the formula $CO(NH_2)_2$ [*see bottom illustration on opposite page*]. All these products can be manufactured from ammonia, and ammonia is therefore the primary raw material of the nitrogenous fertilizer industry.

The first efficient synthesis of ammonia directly from its component elements was achieved by Fritz Haber in Germany in 1909; the first commercial plant, designed by Haber and Karl Bosch, began operating in 1913. In the Haber process one molecule of nitrogen ($N_2$) and three molecules of hydrogen ($H_2$) combine at elevated temperature and pressure and in the presence of a catalyst to form two molecules of ammonia ($NH_3$). Modern catalysts are sophisticated and highly efficient, and their use has become routine. Historically, however, the development of catalysts was a formidable task. Haber's chief contribution to the technology of nitrogen fixation was the discovery of a catalyst suitable for commercial production, and it was for this achievement that he received a Nobel prize in 1918.

The Haber process is exoergic: $\Delta F$ is $-9$ kilocalories. In other words, the sum of the free energies of $N_2$ and $H_2$ exceeds the free energy of $NH_3$ by nine kilocalories, which must be dissipated during the reaction as heat. This simple energy equation misrepresents the thermodynamics of nitrogen fixation, however, because it embraces only one part of the overall chemical system that culminates in the synthesis of ammonia. In order to grasp the extent of the misrepresentation one must consider the source of the raw materials consumed in the process, and in particular the contribution each makes to the sum of the free energies of the reactants.

The diatomic gas is the equilibrium state of nitrogen under terrestrial conditions; it is for this reason, in fact, that nitrogen is most abundant on the earth as a molecular gas rather than in some other form, such as ammonia. (Nitric acid could be considered an alternative equilibrium state, but it can be ignored because kinetic considerations inhibit its formation.) Nitrogen gas is thus very near a condition of minimum free energy and has almost no potential for combining with other elements. In this fact is the crux of the nitrogen-fixation problem: because nitrogen gas is at equilibrium or near it free energy must be supplied in order to form nitrogen compounds.

If nitrogen is at equilibrium, then in the Haber process free energy must be brought to the reaction by the other raw material: hydrogen. Hydrogen gas is in fact rich in free energy; its equilibrium state under terrestrial conditions is not the molecular gas but water. The source of this free energy is apparent from the origin of commercial hydrogen; it is derived from fossil fuels.

Coal, oil and natural gas were formed from dead organic matter; in the context of this discussion it is sufficient to assume that they were formed from one class of organic substances, the carbohydrates, compounds whose composition is

some multiple of the empirical formula $CH_2O$. Carbohydrates are the direct product of the photosynthetic activity of green plants, in which carbon dioxide ($CO_2$) reacts (indirectly) with water. Just as water is the equilibrium form of hydrogen, carbon dioxide is the equilibrium form of carbon under terrestrial conditions, and both reactants have minimal free energy. The overall photosynthetic process must therefore be highly endoergic: $\triangle F$ is $+125$ kilocalories per mole. As is well known, the energy needed to promote this reaction is provided by light quanta from the sun. About three-fourths of the energy supplied, 93 kilocalories, resides in the carbohydrate product and is available as free energy to drive subsequent exoergic reactions.

One of these subsequent reactions is the degradation of carbohydrates to hydrocarbons (compounds composed of carbon and hydrogen only) during the formation of natural gas and petroleum. One such reaction is the transformation of carbohydrates into methane ($CH_4$), the simplest hydrocarbon and the principal constituent of natural gas. The decomposition, which evolves carbon dioxide as well as methane, is exoergic, but $\Delta F$ is only $-54$ kilocalories, so that not all the initial free energy of the carbohydrates is released. Neither is any substantial portion carried off by the $CO_2$; therefore of 93 kilocalories per mole of free energy embodied in plant carbohydrates 39 kilocalories are retained by the methane of natural gas.

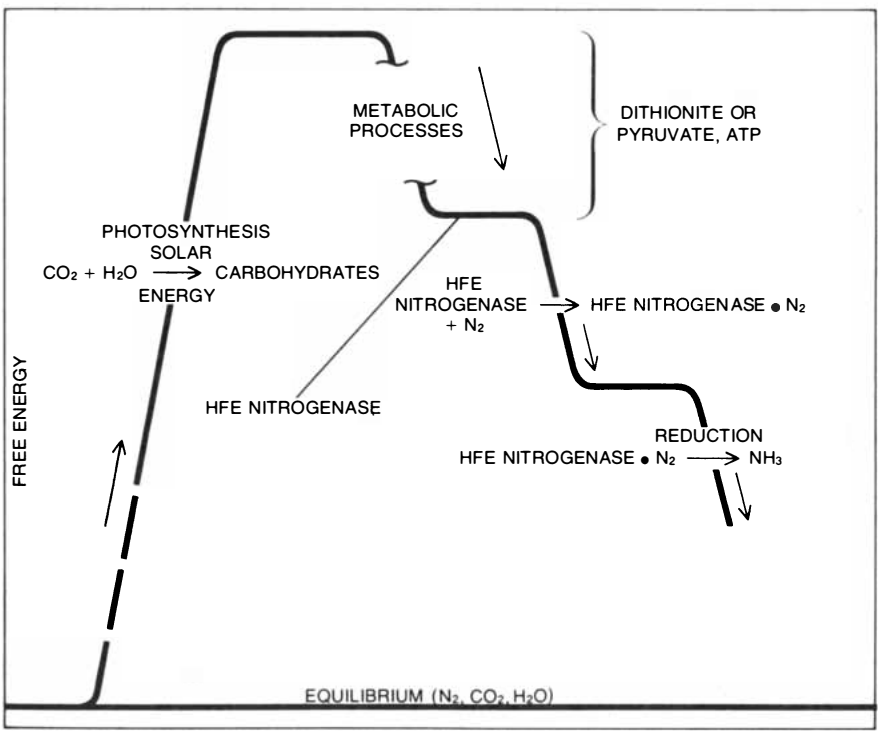
In the U.S. methane is the customary commercial source of hydrogen. In the process called re-forming it is treated with high-temperature steam, producing carbon monoxide (CO) and molecular hydrogen. The reaction is endoergic ($\Delta F = +34$ kilocalories), the required free energy being provided in thermal form by the steam. In a subsequent process, called shifting, the mixture of carbon monoxide and hydrogen is reacted with additional steam to generate carbon dioxide and more hydrogen. This reaction is exoergic, with a decrease in free energy of seven kilocalories [*see top illustration on page 66*].

As in the decomposition of carbohydrates, carbon dioxide removes only negligible free energy from the system, because it is near equilibrium. Since an additional 34 kilocalories of free energy is supplied to the methane during reforming, and since shifting dissipates only seven kilocalories, the hydrogen produced by these reactions must retain a substantial part of the free energy initially present in the methane of natural gas.

The free energy inherent in hydrogen gas is the driving force of the Haber process; indeed, the series of reactions that liberates hydrogen is an essential complement to the reaction that takes place during ammonia synthesis. It is the endoergic photosynthesis of carbohydrates on which the exoergic reaction of nitrogen and hydrogen ultimately depends. The sum of all these reactions is included in what can be called the overall Haber process [*see illustration on preceding page*].

Hydrogen not only provides the energy for the Haber process; it also determines its cost. In fact, it is indicative of the efficiency of the Haber process that the cost of ammonia is almost entirely the cost of the hydrogen consumed in making it. Until recent months the price of hydrogen had been stable for decades at from six to seven cents per pound; ammonia is about a sixth hydrogen by weight, and until recently it could be made for a little more than one cent per pound.

Even at that cost the volume of ammonia production has provided strong motivation to those seeking cheaper methods of nitrogen fixation. As the prices of petroleum and natural gas increase, the endeavor is certain to become more urgent, since the prices of hydrogen and ammonia must rise along with them. Several of these proposed methods are intended to replace hydrogen with some less expensive substance that might react conveniently with atmospheric nitrogen; others would retain hydrogen as a reactant in the hope that the development of a much improved catalyst would eliminate the need for high temperature and pressure. In the latter category are possible methods fashioned by analogy to the mechanisms presumed to operate in those living organisms capable of fixing atmospheric nitrogen.




**EXTRACTS FROM BACTERIA of the genera *Azotobacter* and *Clostridium* fix nitrogen in vitro. One possible thermodynamic model of the mechanism of this fixation process is diagrammed. A series of metabolic reactions, whose nature is unknown, generates a nitrogenase, an enzyme that promotes reactions of nitrogen high in free energy (HFE nitrogenase). The HFE nitrogenase combines with atmospheric nitrogen, and the resulting complex is reduced to yield ammonia. Fixation takes place only in the presence of a strong reductant, such as dithionite or pyruvate; adenosine triphosphate (ATP) is also required.**

## Biological Fixation

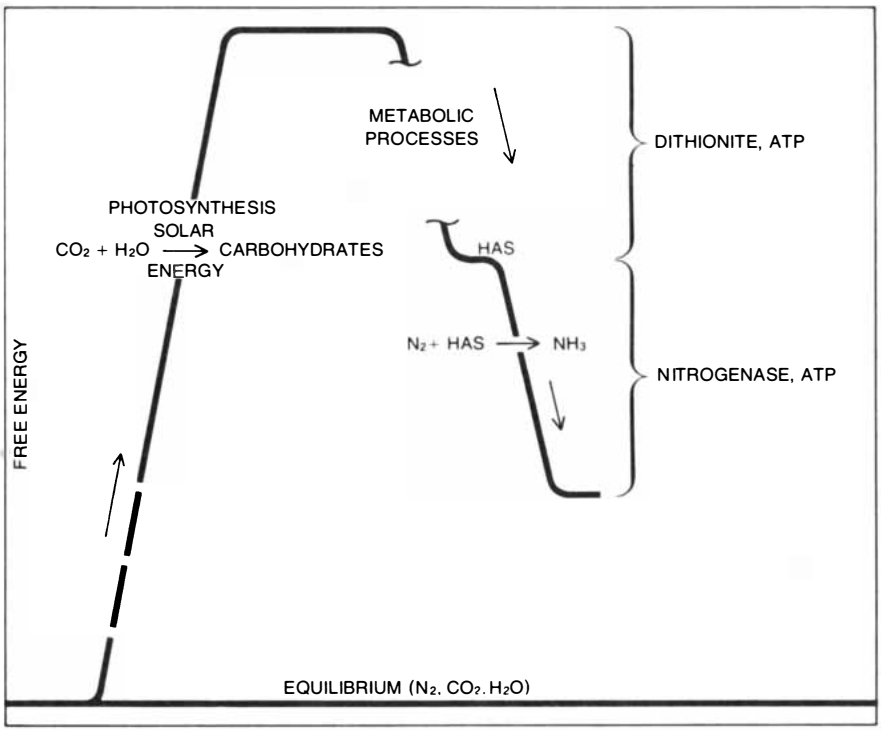
The activities of nitrogen-fixing organisms have been known for more than 100 years and have been intensively studied for decades. The details of the biochemical mechanism of fixation nevertheless remain obscure. It has been known for some time that trace amounts of iron and calcium, and somewhat more molybdenum, are required by all bio-

logical fixation systems and that large amounts of carbohydrates are invariably consumed during the synthesis of a relatively small amount of fixed nitrogen. Furthermore, two kinds of enzyme have been found to be present in all nitrogen-fixing organisms: a nitrogenase and a hydrogenase, which respectively promote reactions of nitrogen and hydrogen. The presence of a hydrogenase is particularly intriguing in that it suggests that ammonia or some other compound involving nitrogen-hydrogen bonds is the product of the synthesis.

Until 1960 all studies of biological fixation had to be made in the organisms themselves in vivo. In that year J. E. Carnahan, John E. Castle, Leonard E. Mortenson and Howard F. Mower of E. I. du Pont de Nemours & Co. isolated an extract from the anaerobic bacterium *Clostridium pasteurianum* that fixes nitrogen in vitro. The extract was found to be active in the presence of sodium pyruvate, a substance closely related to pyruvic acid, which participates in certain metabolic cycles. As in the case of the carbohydrate consumed during fixation in vivo, large quantities of pyruvate were required to fix only a little nitrogen.

Two possible functions were suggested for the pyruvate: it might serve as a source of hydrogen and it might effect fixation through the involvement of adenosine triphosphate (ATP), the substance that participates in almost all biological energy transactions. It was subsequently discovered that the reaction proceeds in the absence of pyruvate if molecular hydrogen is supplied and if ATP is present in small amounts. The *Clostridium* extract could be separated into two fractions, one that metabolized pyruvate and one that participated in the actual fixation process. The latter fraction could be further separated into two constituents, one containing iron and the other iron and molybdenum.

Another experimental system, an extract of the aerobic bacterium *Azotobacter vinelandii*, was later studied by William A. Bulen and Richard C. Burns of the Charles F. Kettering Research Laboratory. In this case pyruvate could be replaced by sodium dithionite ($Na_2S_2O_4$), a powerful reducing agent, or donor of electrons. Dithionite was soon found to be active with the *Clostridium* extract as well, and reduction by dithionite has since become a standard test for nitrogen-fixing activity. The *Azotobacter* extract could also be divided into two fractions, one containing iron and the other iron and molybdenum. In addition, as in the earlier experiments small quantities




**ALTERNATIVE MODEL of nitrogen fixation by extracts from bacteria postulates a "hydrogen-active species" (HAS) capable of reacting directly with molecular nitrogen. A reducing agent and ATP are still required, but in this interpretation of the experimental results the free energy that must be supplied to the system comes not from the nitrogenase but from the hydrogen-active species. The nitrogenase serves merely as a catalyst, promoting the reaction at room temperature; the identity of the hypothetical HAS is unknown.**

of ATP were required for fixation [*see illustrations on these two pages*].

These investigations provide only an inadequate description of the mechanism of biological fixation, but they do enable us to construct possible thermodynamic models. The ultimate reactants must be carbon dioxide, water and atmospheric nitrogen, and the primary product is apparently ammonia. The thermodynamic similarity of the system to the overall Haber process is unmistakable. In both processes the first step is the endoergic photosynthesis of carbohydrates. In the Haber process various transformations, mainly exoergic, convert carbohydrates to hydrogen that is rich in free energy, which then reacts exoergically with nitrogen. In the biological system, although the reaction path is unknown, it is certain that exoergic processes ultimately derived from carbohydrates must produce one or more substances of high free energy capable of combining with molecular nitrogen.

The salient mystery in biological fixation is the identity of these high-free-energy substances. It is a mystery that thermodynamics cannot solve. The substances may be the iron-containing or the iron-and-molybdenum-containing components of the nitrogenases, or they may be other substances as yet undiscovered. It is equally plausible from the point of view of thermodynamics that the active agent is molecular hydrogen itself, or perhaps some postulated "hydrogen-active species" whose reaction with nitrogen is catalyzed by a nitrogenase and promoted by ATP. The fact that pyruvate and dithionite reduce nitrogen to ammonia in laboratory extracts by no means proves that similar strong reducing agents participate in the natural process. Thus the behavior of the system as it is now known can be described by more than one thermodynamic model, and additional information from kinetic studies is required in order to choose between alternative explanations.

One other puzzle of biological fixation is at least approachable through thermodynamics, even if any proposed answers must for now remain somewhat speculative. The puzzle is why the process seems to be so inefficient, why so much carbohydrate must be consumed to produce a little fixed nitrogen. A reasonable explanation is that not all the free energy of the carbohydrates is directed to the synthesis of ammonia. In biological systems competing reactions are common, and in

this case there is reason to believe the hydrogenase reactions that lead to nitrogen fixation would show a greater tendency to convert carbohydrates to carbon dioxide and water than to convert them to the precursors of ammonia. Fixed nitrogen might even be a by-product of a process whose primary function is unrelated to nitrogen metabolism.

## Metallo-Organic Complexes

An interesting approach to understanding biological nitrogen fixation is through constructing nonbiological systems that seem to imitate the performance of the living cell. A catalyst capable of promoting the reaction of nitrogen and hydrogen at room temperature and pressure, for example, might help to illuminate the nature and mechanism of the enzymes found in *Clostridium* and *Azotobacter*. There is another motive for developing such catalysts: they might make it possible to synthesize ammonia under milder conditions than the Haber process requires. Several groups of investigators in the U.S., Canada and the U.S.S.R. have recently reported the discovery of reactions in which this goal seems to have been achieved.

Nitrogen fixation under ambient conditions is not in itself novel. In one long-recognized mechanism metallic lithium combines with atmospheric nitrogen at room temperature to form lithium nitride ($Li_3N$). Ammonia can be produced by reacting lithium nitride with water, yielding as a by-product lithium hydroxide (LiOH). The lithium hydroxide will react further with hydrochloric acid to form lithium chloride. Each of these reactions is exoergic, but they proceed only at the cost of dissipating the high free energy of metallic lithium. If ammonia is to be produced continuously by this method, lithium must be regenerated from the chloride by electrolysis, an endoergic process that consumes as much energy as the preceding reactions liberate [*see illustration below*].

The recently reported reactions employ not pure metals but metallo-organic compounds containing titanium or one of a few other metals. An important point is that titanium is one of the group of elements called the transition metals, and so are the iron and molybdenum found in nitrogenases.

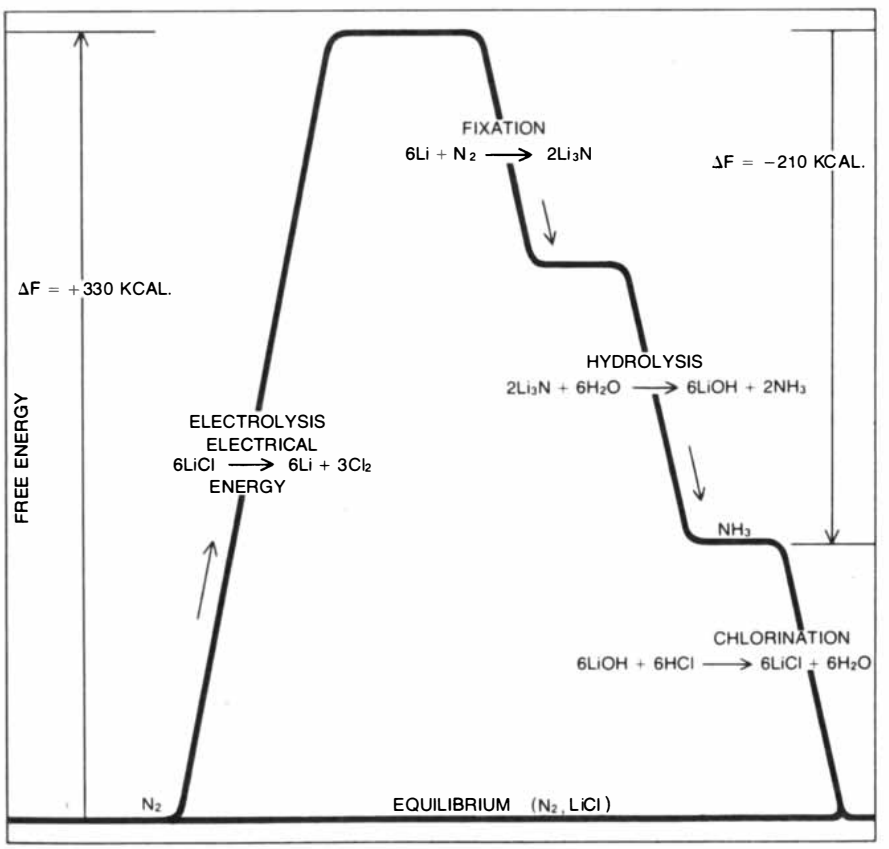
The titanium compounds were first synthesized by Eugene E. van Tamelen and his colleagues at Stanford University. One of the compounds, titanium diisopropoxide, consists of two propyl groups—chains of three carbon atoms—bonded to a titanium atom through oxygen atoms connected with the center carbon of each propyl chain. For brevity the molecule can be designated $Ti(OR)_2$, where OR represents the isopropoxide radical.

$Ti(OR)_2$ combines with molecular nitrogen at room temperature to form the complex $Ti(OR)_2 \cdot N_2$. Sodium naphthalide, a strong reducing agent denoted NaNp, reduces the complex to a nitride whose character has not been determined and that is therefore denoted simply [reduced $Ti(OR)_2 \cdot N_2$]. Ammonia is evolved when [reduced $Ti(OR)_2 \cdot N_2$] is reacted with isopropyl alcohol, and simultaneously the $Ti(OR)_2$ is converted to $Ti(OR)_4$. Ammonia continues to be formed as long as NaNp is added to the system; alternatively metallic sodium can be added, since it combines exoergically with naphthalene liberated in the last stage of the reaction to regenerate NaNp. The NaNp is needed to restore $Ti(OR)_4$ to the form of $Ti(OR)_2$.

This series of reactions is much more complex than that employed in fixation with metallic lithium, but its principle of operation is identical. All the reactions described are exoergic, but if the process is to operate continuously, metallic sodium must be regenerated by electrolysis, an endoergic step. In both methods it is electrical energy from outside the chemical system, and not energy inherent in the reactants, that provides the free energy for fixation [*see illustration on page 75*]. Thus it does not appear that a new class of catalysts has actually been discovered.

The metallo-organic process is at least superficially similar to biological fixation. In both cases the reactions proceed at room temperature, and both employ compounds containing transition metals. In addition both the extracts from biological systems and the metallo-organic complexes can operate with strong reducing agents.

On the other hand, biological systems in vitro are capable of fixing nitrogen with molecular hydrogen as well as




**LITHIUM AND NITROGEN combine to form lithium nitride ($Li_3N$) in a reaction that proceeds exoergically at room temperature and pressure. From this primary fixation product ammonia can be generated by reacting the nitride with water, producing as a by-product lithium hydroxide (LiOH). Fixation by this method is thermodynamically permissible because lithium, like hydrogen, is high in free energy; if ammonia is to be produced continuously, however, the high-free-energy lithium must be regenerated. This is accomplished by converting the hydroxide to the chloride, then decomposing the chloride by electrolysis. The regeneration is endoergic and requires 330 kilocalories per mole.**

# For less than $100, even engineers and scientists can use the mathematical genius of a Casio FX-10.


**Solve problems like these in seconds:**

- sin 12° 45′ 9″
- $e^{\frac{\pi}{2}}$
- $4.56^{1.23}$
- log 2.3
- ln 0.31
- $2.3^7$
- $2 \times \sqrt{2}$
- $a = r \cdot \cos\theta$
- $s = \pi r^2 \rightarrow r = \sqrt{\frac{s}{\pi}}$

Now you can retire your faithful slide rule. Because Casio's new FX-10 scientific calculator answers practically every need of engineers, scientists, architects or anyone who works with complex formulas.

The FX-10 features Casio's unique independent key system. One touch of a key gives you trigonometric functions, common and natural logarithms, exponential function, square and powers, square root, reciprocal calculation and calculation involving $\pi$. In all, the 10 most-used scientific math functions at your fingertip.

Other advantages include constants in all four functions, automatic floating decimal, true credit balance readings and overflow check.

What's more, you can take this battery-powered, hand-sized calculator with you for office, laboratory or on-site figurework. Or operate it on AC, with the adaptor.

All this at Casio's low suggested retail price of only $99.95—a tremendous bargain, as you'll discover if you compare the FX-10 with other models. You can check it out today at your Casio Dealer's, or the coupon will bring you detailed information.

Casio Electronic Calculators, Consumer Products Div.,
Attn: National Sales Manager, Suite 4011,
One World Trade Center, New York, N.Y. 10048

☐ Your new FX-10 sounds great. Tell me more.

Name _____

Street _____

City _____ State _____ Zip _____

From the world's leading manufacturer of electronic calculators.

Casio®

NEW YORK (212) 432-9230 CHICAGO (312) 885-1800 LOS ANGELES (213) 923-4564 ATLANTA (404) 434-5451

# There's nothing like a
# But there's a lot more to a


With a Polaroid portrait identification card, a glance tells you whether the person and the picture are the same. But an ID card should do more than match faces; it should suit the conditions of use.

For that reason we've developed all sorts of ways to make our ID cards more convenient for both the user and the issuer.

In addition to our standard cards (A and B)—carried in pocket, purse, or wallet—we have others with more sophisticated functions.

For instance, there's a card (C) which can be punch-coded. A computer reads it just as accurately as you match a picture to a face.

We have badges (D) with a clip for attachment to outerwear. They have regular or extra-large portrait areas.

We have a card (E) which can be encoded to act like a key. Slipped into an electronic device, it allows access to secure areas, parking lots, etc.

We have a new credit card (F) combining a magnetic stripe with a portrait and an embossed name and number. It's another way of encoding information securely for automatic banking transactions.

Chances are we have an ID card that wil exactly fill your needs.

**Security starts with the card itself.**

The Polaroid ID card is highly secure. It's lam nated with a clear, tough plastic which become integral with the card.

The card works so well that more than 6000 or ganizations all over the world use it. You'll find our ID cards in industry, universities, research and computer centers. They're used as credit cards and drivers' licenses.

People like to carry them. They like having their own color portrait, and they like being able to

Polaroid ®

# ortrait for an ID card.
# ID card than a portrait.


rove who they are.
They carry them with confidence.
Credit cards, for example. Before the portrait, any people were afraid to carry them around. ut now, with the portrait, a lost or stolen credit rd is no longer a serious problem.
You can't forge a face.

**A handy system for making cards.**

Any organization can produce its own ID cards sily and quickly with a low-cost Polaroid ID-3 nd Identification System.
It can be carried around like a suitcase, so you n take it from one branch of your organization another. An operator can be trained to run it in few minutes.
With an ID-3, you can immediately issue permanent ID cards to all your employees. If anybody blinks, you can make a new card on the spot.

You can use different background colors to color-code the employee's function, work area or security clearance.

Security starts with the ID-3. It's the basis of entire security systems.

We'd like to demonstrate it in your own office or factory.

We'll put your portrait on a card and let it speak for itself.

For more information or a demonstration, write to Polaroid Corporation, Dept. 113-255, 549 Technology Square, Cambridge, Mass. 02139. Or you can call collect (617) 547-5176.

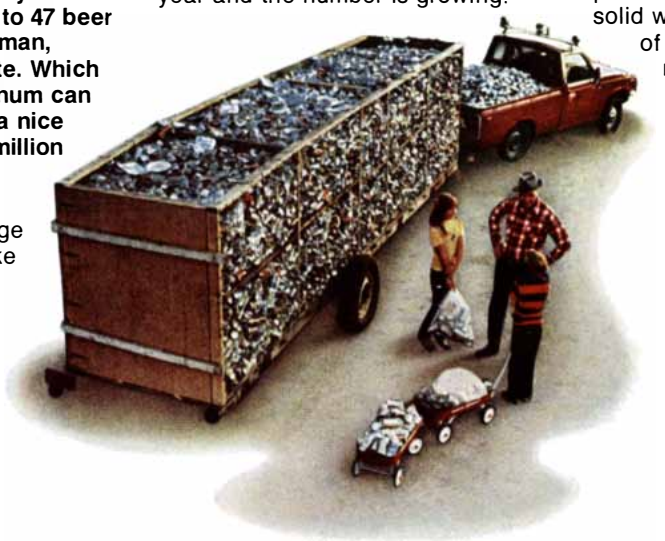
## The Polaroid 2-Minute ID System.

# There's a new kind of Arizona prospector, searching for a new kind of gold...aluminum.

**The people in Arizona have discovered that if all that glitters isn't gold, what's left can still be pretty valuable. Like aluminum. In 1973, over 94 million cans, or about 36 percent of those used, were collected and recycled in Arizona. That comes out to 47 beer or soft drink cans for every man, woman and child in the state. Which means that Arizona's aluminum can prospectors have received a nice chunk of the estimated 15 million dollars paid to the nation's collectors since 1970.**

There's not another beverage packaging material quite like aluminum. It's lightweight, chills quickly, keeps things fresh, opens easily, has high scrap value and can be recycled repeatedly.

The collection and recycling of aluminum cans is working. More than 1.6 billion were collected last year and the number is growing.



Soon, new methods now under development will make possible the collection of even greater numbers. Ultimately, we can expect that improved technology will make it possible to recycle most municipal solid waste, and the high value of aluminum scrap will help make it economically attractive to do so.

For more information on Alcoa's "Yes We Can" recycling program and a list of collection centers, write to us—Aluminum Company of America, 352-K Alcoa Building, Pittsburgh, PA 15219.

**The reasons for using aluminum are found in aluminum itself.**

ALCOA

with dithionite and other strong reducing agents, whereas the metallo-organic processes have not been operated with hydrogen. Moreover, there is reason to doubt that reagents as powerful as dithionite or sodium naphthalide could be tolerated inside the cell. Finally, evidence from the metallo-organic process suggests that in such systems nitrogen fixation proceeds only with a very large change in free energy, necessarily including a large input of free energy to overcome an endoergic step. It seems doubtful that carbohydrate metabolism could provide the required free energy.

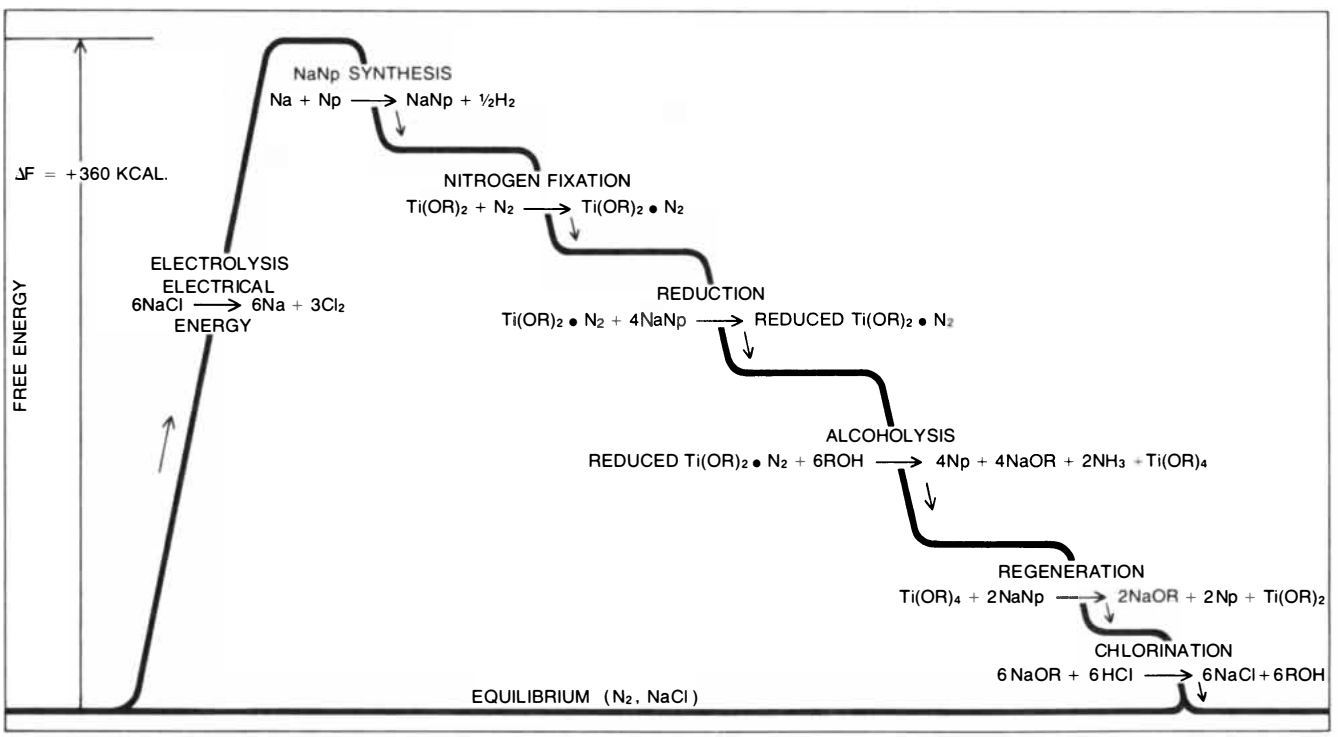
Any analogy between the metallo-organic process and biological fixation implies that the final reactant in the biological system, the substance that actually combines with nitrogen, is a metal-containing component of a nitrogenase. The high energy cost of regenerating such molecules suggests instead that the nitrogenase may be a more typical enzyme that does not require an external source of energy for regeneration and that the ultimate reactant is a hydrogen-active species. In that case the reactant would not have to be regenerated; fixation itself would proceed through a relatively simple mechanism analogous to the Haber process [*see illustrations on next two pages*]. There is an obvious objection to this hypothesis too, however: no reactions at low temperature between nitrogen and an appropriate hydrogen-active species have yet been observed. The resolution of the issue will have to await further study of biological fixation systems.

## Oxidative Methods

The fixation techniques discussed above have all involved the reduction of nitrogen to form ammonia. There are also oxidative processes in which the products are nitrogen oxides. These include several compounds, such as nitric oxide (NO), nitrogen dioxide ($NO_2$), nitrogen trioxide ($N_2O_3$) and nitrogen tetroxide ($N_2O_4$), that tend to be formed together. It is convenient to speak of the mixture as if it consisted entirely of nitric oxide.

The potential attraction of such compounds should be immediately apparent: both raw materials can be conveniently obtained directly from the atmosphere without cost. Unfortunately they cannot be induced to combine without the expenditure of a large amount of energy. Unlike hydrogen, oxygen has insufficient free energy to react with nitrogen under ordinary conditions; neither element is sufficiently far above its equilibrium state. The formation of nitric oxide is endoergic; $\Delta F$ is +41 kilocalories, and this quantity of energy must be supplied from an external source. The methods by which the required free energy can be supplied fall into two categories: activation by heating and low-temperature activation by ionization.

Thermal activation is accomplished by heating air to a temperature of from 2,000 to 3,000 degrees C. in a furnace. It is often achieved unintentionally in fossil-fuel power plants and internal-combustion engines, producing air pollution in the form of a brown haze of nitrogen dioxide. Thermal activation was also the basis of two industrial methods of fixation, both of which are now obsolete. In the Wisconsin process the combustion of fossil fuels provided the necessary heat, whereas electrical energy was employed




**METALLO-ORGANIC COMPLEXES, such as titanium diisopropoxide [$Ti(OR)_2$], also fix nitrogen under ambient conditions. In a procedure developed by Eugene E. van Tamelen, $Ti(OR)_2$, which is rich in free energy, combines with nitrogen and is then reduced by sodium naphthalide (NaNp) to a compound whose nature is unknown and that is therefore designated [reduced $Ti(OR)_2 \cdot N_2$]. Treatment with isopropyl alcohol (ROH) liberates ammonia as well as free naphthalene and sodium propoxide (NaOR); at the same time $Ti(OR)_2$ is converted to $Ti(OR)_4$, which has little free energy and cannot combine with nitrogen. $Ti(OR)_2$ is regenerated by reaction with additional sodium naphthalide. If ammonia synthesis is to be continuous, sodium naphthalide must be regenerated from metallic sodium, produced by the electrolysis of sodium chloride at an energy cost of 360 kilocalories per mole. A similarity has been suggested between the metallo-organic complexes and the nitrogenases of biological systems, in part because both contain transition metals. An equally cogent analogy can be made between fixation by the metallo-organic compounds and fixation by lithium.**

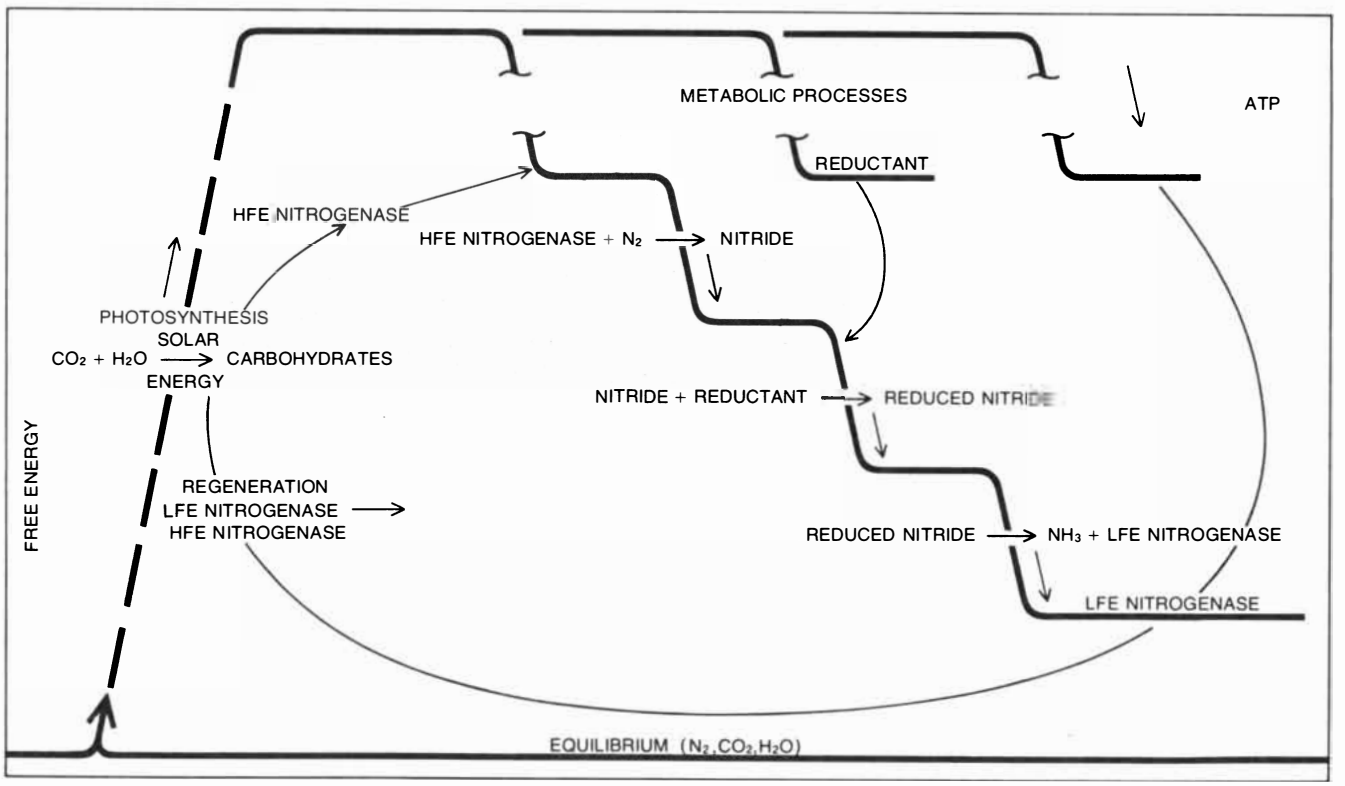
in the Birkeland-Eyde "flaming arc" process.

At high temperature the average velocity of gas molecules increases, and so do the frequency and violence of collisions between molecules. When the velocity becomes high enough, some molecules no longer rebound from one another elastically but convert some of their kinetic energy into the free energy of excitation or ionization. If the stable configuration of the electrons in the molecule is disrupted, or if the normal molecular vibrations and rotations are intensified, the molecule enters an excited state; if an electron is dislodged altogether, the molecule becomes an ion, bearing a positive charge. Excited molecules and ions possess very high free energy; the $N_2^+$ ion, for example, has a free energy of 358 kilocalories. As a consequence they readily enter into many chemical combinations, and the products of these combinations often retain sufficient free energy to support further reactions [*see illustration on page 78*]. At high temperature nitric oxide represents a state of lower free energy than the excited and ionized component elements do, and as a result there is a strong tendency for nitrogen and oxygen to associate in NO molecules.

The efficiency of thermal activation is low, however. The collision processes that lead to activation are reversible, and excited molecules can relinquish their free energy through collisions with molecules that have not been excited; at temperatures that are commercially practical the latter are always in the majority and excited molecules rapidly lose their energy. A dynamic balance is reached between activation and deactivation, with the result that at 2,200 degrees C. only about 2 percent of the mixture is converted to nitric oxide. Because all the air must be heated, and not just the small fraction that contributes to the formation of the product, the process requires about 20 times the minimum theoretical energy requirement of 41 kilocalories per mole. Most of the energy goes into heating the bulk gas instead of into promoting the desired reactions.

Low-temperature ionization has the significant advantage that in principle all the molecules of the gas can be ionized or excited. The activation can be accomplished by subjecting the air to an electric potential of a few thousand volts, so that a low-temperature discharge is initiated, or by exposing air to an intense flux of ionizing radiation inside a nuclear reactor. In either case the gas molecules are bombarded by fast-moving ions and the collisions are inelastic. The resulting cascade of reactions can produce a substantial yield of nitrogen oxides.

As in all endoergic processes, the telling factor in calculating the feasibility of low-temperature activation is the cost of energy. If the source of energy is electricity, its cost would seem to forbid fixation by air activation as a commercial enterprise. The "chemonuclear" technique, on the other hand, utilizes a remarkably cheap form of energy: the kinetic energy of the nuclear fragments produced by the fission of a uranium nucleus. The pioneering work in this field has been done by Paul Harteck and his colleagues at Rensselaer Polytechnic Institute and the Brookhaven National Laboratory. They have found that special reactor designs are required to ensure that a large proportion of the fission fragments escape into the gas; their many elegant experiments have shown, however, that with the proper design




**BIOLOGICAL NITROGEN FIXATION can be described by either of two thermodynamic models; studies of fixation by bacterial extracts in vitro and by metallo-organic complexes have not yet provided the means to discriminate between them. In one model (*left*), constructed by analogy to the metallo-organic systems, nitrogen reacts with an enzyme of high free energy (HFE nitrogenase). The resulting nitride is reduced and exoergically yields ammonia, the enzyme being converted to a low-free-energy (LFE) form. Three series of couple reactions are required by the model: one to produce the HFE nitrogenase, one to produce the reductant and one to convert**

more than half of the fission energy can be made available to promote chemical reactions. The chemonuclear process thus has the potential of high energy-conversion efficiency. Whether or not it can become a practical method of nitrogen fixation depends on economic and technological considerations.

## The Economics of Fixation

The several methods of fixation I have been discussing employ a variety of raw materials, sources of energy and reaction processes; superficially they seem to be hardly related at all. They can nevertheless be conveniently and accurately compared, since the same thermodynamic laws govern them all. In fact, from the point of view of thermodynamics all the processes have a similar overall form. In each case energy is expended in preliminary endoergic reactions that yield substances high in free energy; these then react exoergically in the fixation process itself. It can even be predicted with certainty that any as yet undiscovered fixation systems will conform to the same rule.

The economic evaluation of alternative fixation methods therefore depends ultimately on how energy is supplied in the endoergic step. The solar energy latent in carbohydrates and other organic matter can be employed, as it is in the Haber process, but only at a high cost in raw materials. Alternatively the free resources of the air can be tapped, but only when an external source of thermal, electrical or nuclear energy is provided.

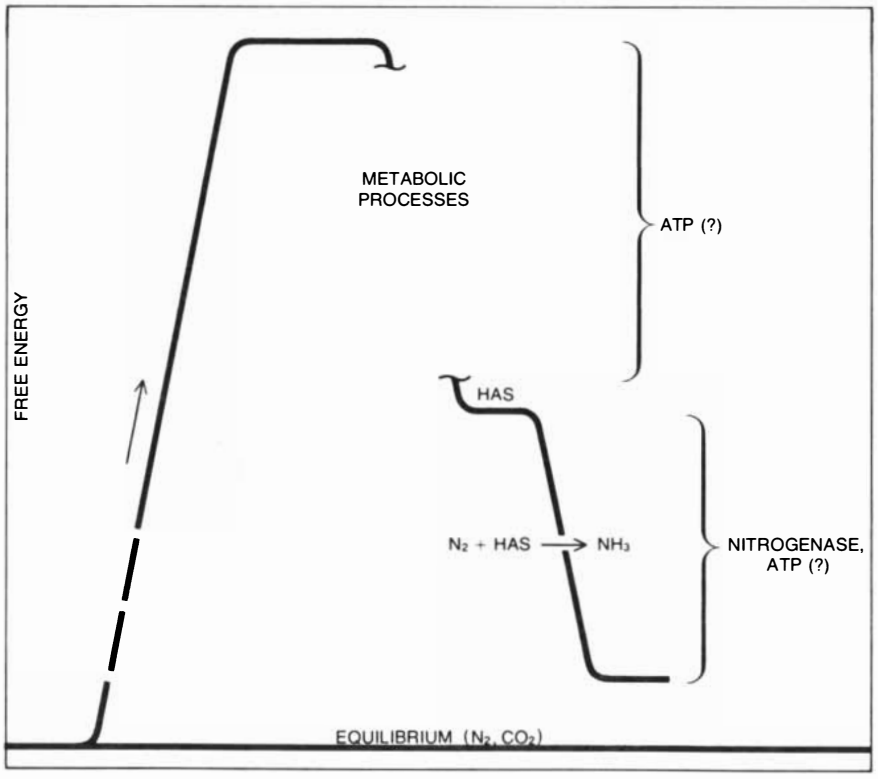
For any fixation process the theoretical minimum energy requirement can be calculated directly from thermodynamic information. Estimating the cost of this energy, however, has recently become a difficult exercise in prognostication. The prices of petroleum and natural gas are influenced by many political, economic and technological factors, and even for the immediate future they cannot be reliably predicted. The most that can be said is that the price will eventually be stabilized again, since continued price increases will make practical the recovery of otherwise unprofitable reserves, such as shale oil. The level at which stability will be achieved is for now unknown.

Even if the future prices of oil and gas remain uncertain, however, the relative costs of the various methods of nitrogen fixation can be calculated. Hydrogen for the Haber process and thermal and electrical energy for other methods derive for the most part from the same source—fossil fuels—and any increase in fuel costs will raise the cost of each method proportionately. For this reason it will suffice to discuss the costs of fixation in terms of the prices that prevailed before the "energy crisis" began.

Economic analysis and the history of the nitrogen-fixation industry both suggest that if fossil fuels are to be employed in fixation, then the Haber process is the most efficient way to utilize them. When hydrogen cost six to seven cents per pound (and ammonia could therefore be made by the Haber process for slightly more than one cent per pound), thermal energy cost about .1 cent per kilowatt-hour and electrical energy, also largely derived from fossil fuels, cost about .8 cent per kilowatt-hour. At these prices the metallo-organic system, which requires 5.7 kilowatt-hours of electricity for each pound of ammonia produced, would yield a minimum theoretical price of 4.6 cents per pound of ammonia; the similar lithium nitride method would cost 4.3 cents per pound. Clearly neither could compete with the Haber process; moreover, these prices are based on thermodynamic requirements only and thus assume 100 percent efficiency, which could not be attained.

The development of a "supercatalyst" for ammonia synthesis, perhaps from the study of enzymes mediating biological fixation, might make it possible to combine nitrogen and hydrogen at room temperature. Although it may seem surprising, such a catalyst could have only a marginal effect on the cost of ammonia. From the point of view of thermodynamics a supercatalyst system would be only a variant on the Haber process, since the raw materials and the product would be the same and the change in free energy would be unaffected. Because virtually all the cost of Haber ammonia is the cost of hydrogen the price of the product would be little changed.

Among the oxidative processes low-temperature ionization requires 3.3 kilowatt-hours of energy per pound of nitric oxide produced. If this requirement were met with electrical energy (at former prices), the cost of fixation would be 2.6 cents per pound. Thermal activation requires 18 kilowatt-hours per pound. Employing electrical energy for that purpose is hopelessly expensive: at least 14 cents per pound of product. The direct application of fossil fuels is considerably

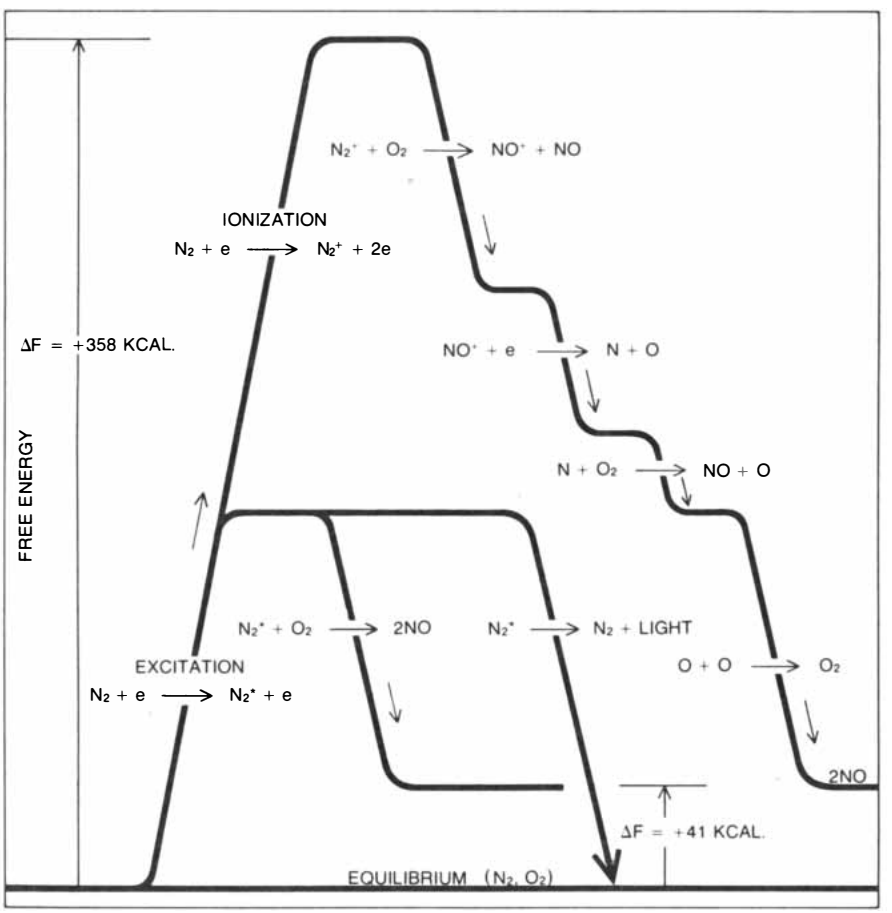



**the LFE nitrogenase back to the high-free-energy form. An equally plausible model (*right*) requires only one series of coupled reactions, the product of which is a high-free-energy substance other than the nitrogenase, probably a hydrogen-active species (HAS). Nitrogen combines with the HAS in a reaction catalyzed by the nitrogenase. The observed requirement for adenosine triphosphate could be accounted for by either of the models.**

cheaper, but not cheap enough. The theoretical minimum cost of 1.8 cents per pound already exceeds the cost of Haber ammonia, and inefficiency has not been allowed for. In addition, because $NH_3$ is lighter than NO, .6 pound of ammonia can be oxidized to yield one pound of nitric oxide, so that to be competitive the oxidative processes must manufacture nitric oxide not at one cent per pound but at .6 cent per pound.

The most promising competitor of the Haber process is the chemonuclear ionization of air. As in low-temperature electrical activation, 3.3 kilowatt-hours per pound of nitric oxide is required, but the cost of raw fission energy is much lower than that of electricity. The uranium-235 fuel employed in a nuclear reactor costs about $20 per gram, but the energy released by fission is so great that the cost of fission-fragment energy is only from .05 to .1 cent per kilowatt-hour. At .05 cent per kilowatt-hour the theoretical cost of producing nitric oxide is about .17 cent per pound. At that price chemonuclear fixation would definitely be competitive with the Haber process, but in practice the price cannot be achieved. As we have seen, only about half of the fission fragments give up their energy by exciting or ionizing gas molecules; moreover, only about half of the energy so transferred is available to promote fixation. (The rest is dissipated as heat.) The overall efficiency is therefore about 25 percent, and a more realistic manufacturing cost for the product is .7 cent per pound. Since the product is nitric oxide, the equivalent cost for ammonia is .6 cent per pound.

In addition to the problem of costs there is a technological impediment to the adoption of chemonuclear fixation. The process depends on the release of fission fragments, which are highly radioactive, into the working medium, and these fragments would have to be removed before the product could be marketed. Decontamination to acceptable levels might prove impossible, and even if it were found to be feasible, it would almost certainly add to the cost.




**ACTIVATION OF AIR is the basis of several processes in which the free energy required for nitrogen fixation is supplied directly to nitrogen itself. Activation can be accomplished by heating, by an electrical discharge or by ionizing radiation. In each case several oxides of nitrogen are formed; here only nitric oxide (NO) is considered. Excitation occurs when a nitrogen molecule collides at sufficient energy with another molecule or with an electron or other particle. The excited molecule ($N_2$*) can combine with oxygen to form nitric oxide, or it may decay to the equilibrium state by emitting a photon. Collisions at higher energy produce nitrogen ions ($N_2^+$), which participate in a cascade of reactions with oxygen atoms and molecules. The result is the synthesis of two molecules of nitric oxide.**

## Future Fixation Methods

If the problem of decontamination could be overcome, and if a relatively modest improvement in efficiency could be attained, chemonuclear fixation might well become competitive with the Haber process. Even considering the steadily increasing cost of hydrogen, however, it does not appear that the chemonuclear technique offers the prospect of a large reduction in the price of fixed nitrogen, partly because the price of uranium 235 is also expected to rise. The improvement, if any, would be modest, and although it would certainly transform the nitrogen-fixation industry, it would probably have only minimal impact on agricultural prices or practices.

Extraordinary cost reductions are far more likely to result from the application of new techniques to more conventional processes, and in particular from potential methods of obtaining less expensive hydrogen for ammonia synthesis. Inexpensive hydrogen is important in itself, since many industries other than nitrogen fixation require it; moreover, it has recently been suggested that hydrogen gas could serve as a nonpolluting alternative to fossil fuels [see "The Hydrogen Economy," by Derek P. Gregory; SCIENTIFIC AMERICAN, January, 1973].

Besides fossil fuels the most obvious source of hydrogen is water, from which it can be extracted by electrolysis. The process is of course endoergic, theoretically requiring 57 kilocalories per mole, or 15.4 kilowatt-hours of energy per pound of hydrogen, and the economic possibilities are not encouraging. The most advanced electrolytic cells now being developed promise an efficiency of 50 percent, so that about 31 kilowatt-hours per pound would be required. At the energy cost prevailing before the price rises of the past year began, hydrogen manufactured by electrolysis would cost about 25 cents per pound. Even if the cost of electricity could be reduced to .2 cent per kilowatt-hour, which is the transmission cost, electrolytically made hydrogen would just maintain the former price of fossil-fuel hydrogen: about six cents per pound.

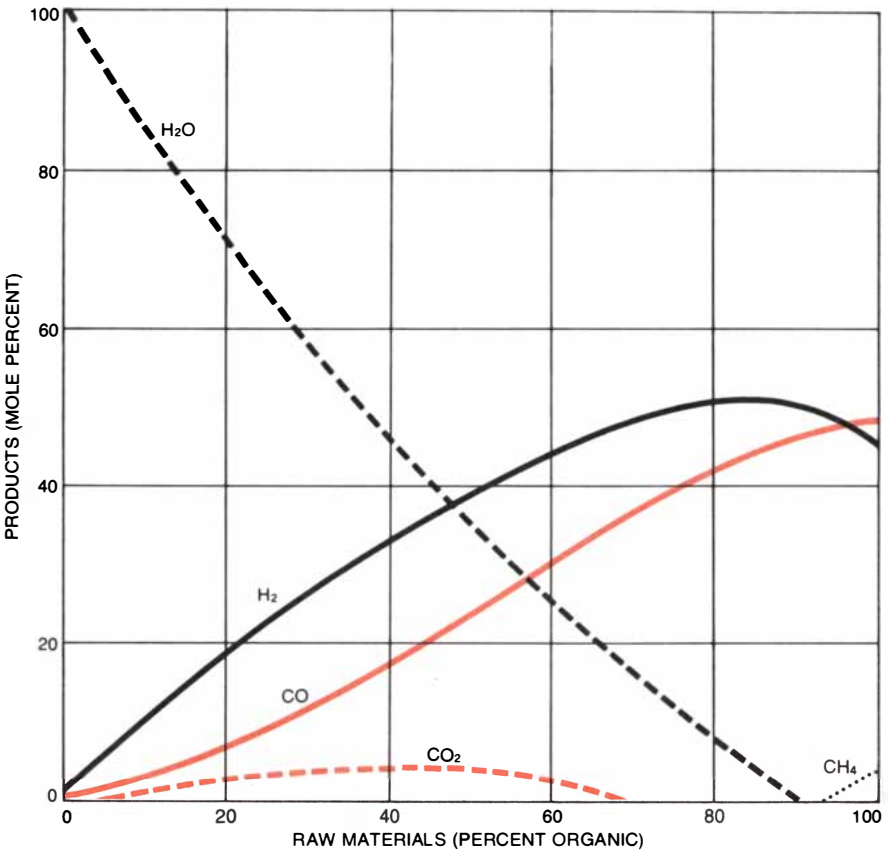
Certain more innovative methods might substantially reduce the cost of

hydrogen—in theory even eliminate it entirely—by employing as a raw material the carbohydrates in waste organic materials. In one such proposal the wastes would be fermented by the action of microorganisms to yield methane, which could then be converted to hydrogen and carbon dioxide by the conventional re-forming and shifting methods. In another the Haber reaction would be bypassed entirely by employing a catalyst or a system of catalysts capable of combining nitrogen directly with the hydrogen in carbohydrates. The latter process must remain speculative for the time being, since the necessary catalysts have not been developed.

The raw materials for these processes can be obtained virtually without cost. Since the reactions involved are exoergic and energy costs are therefore also zero, either one suggests the prospect of free fixed nitrogen. Actually the cost would probably be high. Both processes would be slow, and enormous quantities of bulk solids would be required, only a small portion of which would be converted to useful products. If a catalyst active directly on carbohydrates and molecular nitrogen could be developed, it might be best applied directly to the soil, where it could generate plant nutrients *in situ*.

In the fermentative processes, which exist today and can be investigated, most of the material is converted not to methane but to other products of little commercial value or else is left as unreacted residue. Furthermore, even if homogeneous organic wastes could be obtained as raw material, the biochemistry of fermentation is so complex that the outcome of the reactions cannot be predicted or controlled well enough to achieve commercially competitive efficiency.

A chemically less complicated but nevertheless more grandiose scheme has been proposed by the author. A water slurry of organic wastes would be vaporized in an underground cavity by the heat of a thermonuclear explosion. All the substances would be dissociated into their component atoms, which would recombine, if the proper conditions were established, to yield hydrogen and carbon monoxide almost exclusively [*see illustration on this page*]. The process would take advantage of the extremely low cost of raw fusion energy, about .001 cent per kilowatt-hour, or about a fiftieth the cost of raw fission energy. By vaporizing a few hundred thousand tons of material at a time, hydrogen might be manufactured for less than one cent per pound, and ammonia could therefore be produced for less than one-sixth cent per pound.




**THERMONUCLEAR EXPLOSION could be employed to produce hydrogen for the Haber process from organic wastes. The products of such an explosion were calculated theoretically, assuming a temperature of 2,000 degrees C. and a pressure of 500 atmospheres. A few hundred thousand tons of waste, in a water slurry, would be vaporized in an underground cavity. Under optimum conditions virtually all the material would be converted to hydrogen and carbon monoxide; the carbon monoxide could be utilized to generate additional hydrogen. Hydrogen manufactured in this way could cost as little as one cent per pound.**

As in the chemonuclear process, the products of thermonuclear conversion would be contaminated with radioactive substances; in this case, however, the contaminants might be readily eliminated. The most important one would be the hydrogen isotope tritium. It would be reacted with nitrogen, by the Haber process, to produce tritiated ammonia. The ammonia would not be sold, however, but oxidized to nitric oxide and water; the tritium would be confined to the water (which could eventually be sold as an isotope), and the nitric oxide, the primary product, would be free of contamination.

Another method for utilizing the heat of nuclear reactions has been suggested by G. De Beni and C. Marchetti of Euratom. They propose to dissociate water by direct thermal activation inside a nuclear reactor. As in electrolysis, the minimum energy requirement is 15.4 kilowatt-hours per pound of hydrogen. That is a large sum, but raw fission energy is sufficiently inexpensive, so that the overall cost of the product might be competitive with that of hydrogen derived from fossil fuels.

Because the heat of fission rather than the ionizing radiation would be utilized, contamination of the product would not become a problem. Another technological difficulty has not been solved, however. Direct thermal dissociation requires a temperature of from 2,500 to 3,000 degrees C., which is far beyond the range of practicality for industrial reactors available today; De Beni and Marchetti consider 750 degrees the maximum allowed. By substituting a series of coupled reactions (involving iron oxides and chlorides) for direct dissociation, they propose to reduce the temperature required in the process to 1,000 degrees, and it may be possible to reduce it further.

A more distant possibility is the utili-

zation of controlled nuclear fusion for nitrogen fixation. Bernard J. Eastlund and William C. Gough of the Atomic Energy Commission have pointed out that the exhaust gases of a fusion reactor would be so hot—at a temperature of millions of degrees C.—they could break down any material into its component atoms and thus make pure elements available for recycling. Such a "fusion torch" could be useful to nitrogen fixation in any number of ways, for example by dissociating water or decomposing carbohydrates. Organic wastes could be heated and then quenched to produce hydrogen and carbon monoxide under controlled conditions by the same mechanism operating without control in an underground thermonuclear explosion. Finally, nitric oxide could be manufactured by thermal activation of air. As in any other method of thermal activation the efficiency would be quite low, and only about 2 percent of the reactants would combine to form nitric oxide. Nevertheless, the cost of energy would in this case be so low, about .001 cent per kilowatt-hour, that nitric oxide could be produced for .018 cent per pound.

The fusion torch is an attractive concept, and it would undoubtedly transform many technologies in addition to nitrogen fixation; it appears that it will not be available, however, for many years to come. In the interim nitrogen will almost certainly continue to be fixed by the Haber process, which is now more than 60 years old but which remains the most efficient of the methods available to us today.

A reduction by an order of magnitude in the price of fixed nitrogen would have extraordinary worldwide consequences for agriculture. It is not our dependence on the Haber process itself that precludes such a price reduction but the dependence of the Haber process on fossil fuels for hydrogen. As the prices of petroleum and natural gas continue to rise, and as abundance becomes apparent shortage, this dependence could become alarming. Consequently the most promising fields of investigation for the immediate future are those that could lead to potential sources of less expensive hydrogen.

| PROCESS | PRODUCT | RAW MATERIAL | RAW MATERIAL COST (CENTS PER POUND) | ENERGY REQUIRED (KCAL. PER MOLE) | ENERGY REQUIRED (KILOWATT-HOURS PER POUND) | ENERGY COST (CENTS PER KILOWATT-HOUR) | TOTAL COST (CENTS PER POUND) |
|---|---|---|---|---|---|---|---|
| HABER | $NH_3$ | $H_2$ | 1 | 0 | .04 | — | 1.04 |
| CARBOHYDRATE-CATALYST | $NH_3$ | CARBOHYDRATES | 0 | 0 | — | — | 0 |
| SUPERCATALYST | $NH_3$ | $H_2$ | 1 | 0 | — | — | 1 |
| METALLO-ORGANIC | $NH_3$ | — | — | 179 | 5.7 | .8 | 4.6 |
| LITHIUM NITRIDE | $NH_3$ | — | — | 165 | 5.3 | .8 | 4.3 |
| LOW-TEMPERATURE IONIZATION | | | | | | | |
| ELECTRICAL | NO | AIR | 0 | 180 | 3.3 | .8 | 2.6 |
| CHEMONUCLEAR | NO | AIR | 0 | 180 | 3.3 | .05 | 0.17 |
| THERMAL ACTIVATION | | | | | | | |
| ELECTRICAL | NO | AIR | 0 | 1,000 | 18 | .8 | 14 |
| FOSSIL FUEL | NO | AIR | 0 | 1,000 | 18 | .1 | 1.8 |
| FUSION TORCH | NO | AIR | 0 | 1,000 | 18 | .001 | 0.018 |

**ECONOMICS OF NITROGEN FIXATION depends on the balance between the cost of raw materials and the cost of energy. Almost the entire cost of ammonia made by the Haber process is the cost of raw material (hydrogen); because the process is exoergic thermodynamically no external source of energy is necessary. Methods that eliminate the need for hydrogen require that energy be supplied. Costs given are those that prevailed before the price of fossil fuels began to rise about a year ago. For the Haber process actual costs are given; for all others the costs are theoretical minimums calculated from thermodynamic requirements. The biological process employing carbohydrates has a theoretical cost of zero but would probably be too slow and have too low a yield to be practical. Chemonuclear activation of air has a low theoretical cost but because of its inefficiency would be only marginally competitive.**


Coins shown ¾ actual size

# *To mark the first issue of its new national coinage... The Government of Belize announces the special minting of an all silver Collector's Proof Set*

THE GOVERNMENT OF BELIZE (formerly British Honduras) is proud to announce the first minting of its all-new national coinage in 1974. To honor this occasion the Government has announced the issue of a Special Collector's Proof Set—unlike any issued by any other country in the world today.

Every one of the eight coins in this special Collector's Proof Set—from the ten-dollar piece to the one-cent piece—will be minted *in solid sterling silver.*

This set has been authorized by the Government of Belize to provide collectors with particularly beautiful and valuable examples of the nation's new coinage. The face of each coin will bear an exquisitely sculptured portrait of one of the fascinating tropical birds of Belize. The reverse will carry the National Seal of the Government of Belize. And every coin in this unique set will be a beautiful gem Proof—minted from special dies so that the finely detailed design stands out in frosted relief against a brilliant, mirror-like background.

**All orders for Proof Sets must be postmarked by October 31, 1974.**

Those who wish to acquire this historic Collector's Proof Set are advised that there is a limit of one such set per order. The official issue price is $100. the set, and only those orders postmarked by October 31, 1974, will be accepted.

The circulating coinage of Belize will bear the same designs as the coins in the Collector's Proof Set, but will not be minted in silver. Six of the circulating coins will be struck in cupro-nickel, one in nickel-brass, and one in bronze. Proof and Specimen Sets of the circulating coinage are also being made available to collectors.

TRADITIONAL PROOF SETS, containing flawless Proofs of the circulating coins are available at the official issue price of $35. per set. The ordering deadline is October 31.

SPECIMEN SETS, containing brilliant non-proof specimens of the coins, may be ordered at the official issue price of $20. per set. These sets will be available until authorized limits have been reached.

The total number of Collector's Proof Sets and Traditional Proof Sets to be issued will exactly equal the number ordered by October 31. Thereafter, the special proof dies will be destroyed, so that no other Proof Sets of the 1974 coinage of Belize can ever be minted again.

To place your order, complete the form below and mail it directly to the official minter of the coins: The Franklin Mint, Franklin Center, Pennsylvania 19091.

ORDER FORM

**THE NEW NATIONAL COINAGE OF BELIZE**

The Franklin Mint
Franklin Center, Pennsylvania 19091

Please accept my order for the following:

☐ One COLLECTOR'S PROOF SET, minted in solid sterling silver, at the official issue price of $100. the set . . . $_____

_____ (QUANTITY) TRADITIONAL PROOF SET(S), minted in the same metals as the circulating coinage, at the official issue price of $35. per set . . . . . $_____

_____ (QUANTITY) SPECIMEN SET(S), in the circulation metals, at the official issue price of $20. per set . . . . . $_____

Total $_____

Add $1. per set for postage and handling $_____

REMITTANCE ENCLOSED $_____

Mr.
Mrs.
Miss_____
PLEASE PRINT CLEARLY

Address_____

City, State, Zip_____

**ORDERS FOR PROOF SETS MUST BE POSTMARKED BY OCTOBER 31, 1974**

All orders subject to acceptance.
Please allow 8 weeks for delivery. 36-65

# THE DIMENSIONS OF STAIRS

From antiquity to modern times stairways have been built by rules of thumb. A study of how people move and expend energy in going up and down stairs suggests a more rational approach to their design

by James Marston Fitch, John Templer and Paul Corcoran

Sir Henry Wotton, a 17th-century English diplomat and writer, included architecture among his interests, and he once offered certain strictures on the construction of stairs. "To make a compleate Staire-case," he wrote, "is a curious peece of Architecture: The vulgar Cautions are these.

"That it have a very liberall Light, against all Casualtie of Slippes, and Falles.

"That the space above the Head, bee large and Airy... because a man doth spend much breath in mounting.

"That the Halfe-paces bee well distributed, at competent distances, for reposing on the way.

"That to avoyd Encounters, and besides to gratifie the beholder, the whole Staire-case have no nigard Latitude, that is, for the principall Ascent, at least ten foot in Royall Buildings.

"That the breadth of every single Step or Staire bee never lesse then one foote, nor more then eighteen inches.

"That they exceede by no meanes halfe a foot in their height or thicknesse; for our Legges doe labour more in Elevation, then in Distention...."

The simple good sense underlying Wotton's "vulgar Cautions" is quite often neglected or forgotten in buildings today. The pedestrian encounters stairs that are distinctly uncomfortable if not dangerous—not because they are inadequately maintained or badly eroded but because the geometry of their design does not match the natural human gait. On such a stairway one risks an accident, either from not being aware of irregularity or from not paying attention to it.

The authors of this article have been looking into the design of stairs over a period of years. In that time important contributions to the work have been made by José R. Bernardo, an architect in New York, and Neville O'Reilly of Rutgers University. Our attention was attracted to the subject of stairs by certain general considerations of an architectural nature and also by reports of a high accident rate on a monumental stairway in New York: the four broad steps leading up from the plaza at Lincoln Center to the walkway in front of the new Metropolitan Opera House. The general considerations had to do with the deployment of people in space, particularly in buildings.

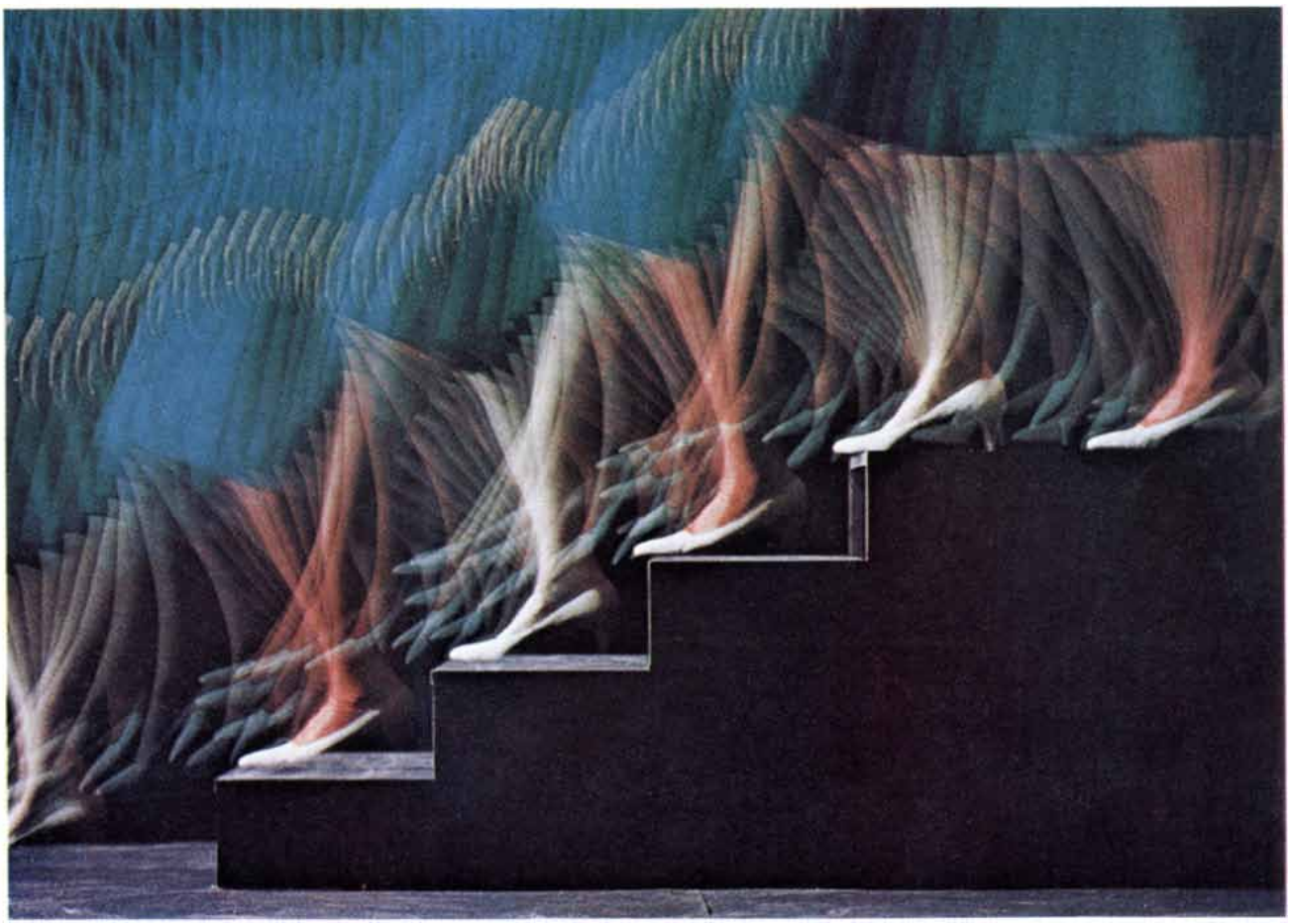
The way an individual acts in space is the result of all the forces that play on him at any given moment. The forces are both external and internal, objective and subjective. The values of all of them are in constant flux. Much energy is required merely to occupy space—to overcome such forces as friction, gravity and atmospheric pressure—let alone to accomplish socially productive work. Hence the conservation of energy is a matter of considerable consequence to the individual.

People conserve energy by a complex feedback system in which sensory input continuously modifies spatial orientation. Information the senses provide on environmental events is converted into decisions on the movements to be made. This marvelously sensitive and precise cost accounting of the input and output of energy is largely carried on unconsciously, or at least below the level of consciousness. It is therefore difficult to observe and quantify.

Nonetheless, if the overall effectiveness of architecture is to be raised above its present unsatisfactory level, architects must evolve a holistic theory of human deployment in architectural space. Although much has been written about the dimensions of the body and the expenditure of energy in various tasks, little of the material has any direct relevance to architectural problems. Most studies of human movement, beginning with the historic work of Eadweard Muybridge nearly a century ago, seem to have been largely descriptive. Moreover, they have usually focused on the motion of the individual subjects and neglected to provide much information on the circumstances in which the tests were made. Muybridge, for example, photographed nude men and women running up and down stairs but did not provide data on the dimensions of the steps or show how motion would have been modified by stairs of other proportions. Similarly, members of the remarkable Ducroquet family of French orthopedic physicians have carried on for decades a study of normal and pathological walking. All of it, however, was on level places. How would it be altered by stairways?

The need for holistic data is nowhere more apparent than in the design of stairs. Here is a device almost as old as architecture itself (the ramp may be older), and yet little has been written on the subject. When one examines the material, as we have had occasion to do, it turns out to consist largely of empirical rules of thumb. No one seems to have examined closely the sense of comfort or strain, safety or danger that the user ex-

**WOMAN ON STEPS** appears in the photographs on the opposite page, which were made with a stroboscopic flash lamp as she went up and down a short flight in a laboratory. It was one of four flights of differing design built to test how the various designs affected the human gait. The woman wears a red stocking and a white one to assist the observations.

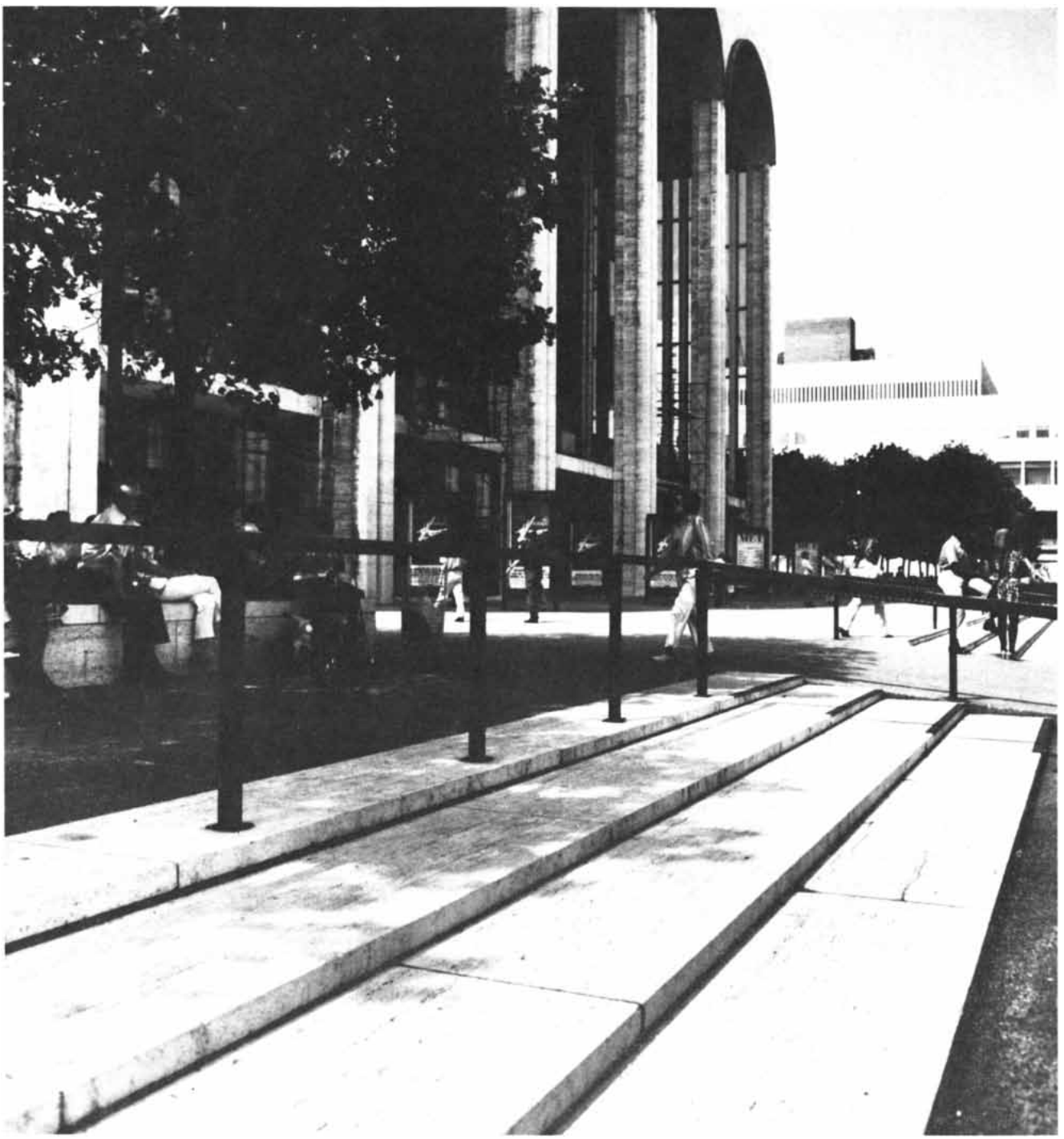


periences in climbing or descending a particular kind of stairway.

The restraints on all human movement, including travel on stairs, are both physiological and cultural. The physiological restraint is a function of the energy required to lift or lower the body from one point in space to another. The cultural restraint is a function of the gait and posture dictated by the customs and costumes of the culture.

The physiological factor is more or less constant. The metabolic rate of a man climbing a "normal" set of stairs at a normal speed is two to five times that of a man walking on a horizontal surface at three miles per hour. The cultural factor is variable. Naked young men led the Panathenaic procession up the ramp of the Acropolis, whereas the pace and posture of Empress Eugénie on the stairs at the palace of Versailles would have been established by heavy hoopskirts.

All stairs impose a certain gait and hence a definite psychic orientation on the user. This constraint has always been well understood by the architects of monumental structures such as temples, palaces and opera houses. In those places the stairway plays an important role in the required ceremonial behavior of both climbers and beholders. No comparable attention, however, has been paid to the



LINCOLN CENTER STAIRWAY in New York is of unusual configuration, having exceptionally shallow risers and deep treads. In addition it leads in the downward direction from a broad walkway in front of the Metropolitan Opera House to a busy plaza, so that people approaching it from the direction of the opera house may not see it. The stairway has been the scene of numerous stumbles.

design of stairways for mundane utilitarian purposes.

The stairway at Lincoln Center appears to impose an awkward gait on its users because of its shallow risers (3.375 inches) and deep treads (25 inches). As originally installed the steps were hard to see because they were made of a light travertine similar in color to the rest of the plaza. Moreover, they are a considerable distance away from the front of the opera house, and a pedestrian approaching the busy plaza from the direction of the opera house may simply not notice them—a surmise reinforced by the fact that most of the people who fell or stumbled were descending the steps.

The full extent of the trouble with the stairway is not known, since both the architects and the management of the center have declined to discuss the matter. It is known, however, that the accidents have been frequent enough and serious enough to have led to a number of lawsuits and to a series of physical modifications of the stairway. Indeed, at the places where pedestrian traffic is the heaviest the steps have been replaced with ramps. In much of the remaining space the steps have been closed off with bronze railings.

Since the general configuration of the steps satisfies the conventional standard for the design of exterior stairs, we were intrigued by the evidence that something was wrong with them. We were also motivated by the scantiness of the architectural literature on stairways. We decided to inquire into the matter in the hope of providing a better basis for a theory of design of stairs.

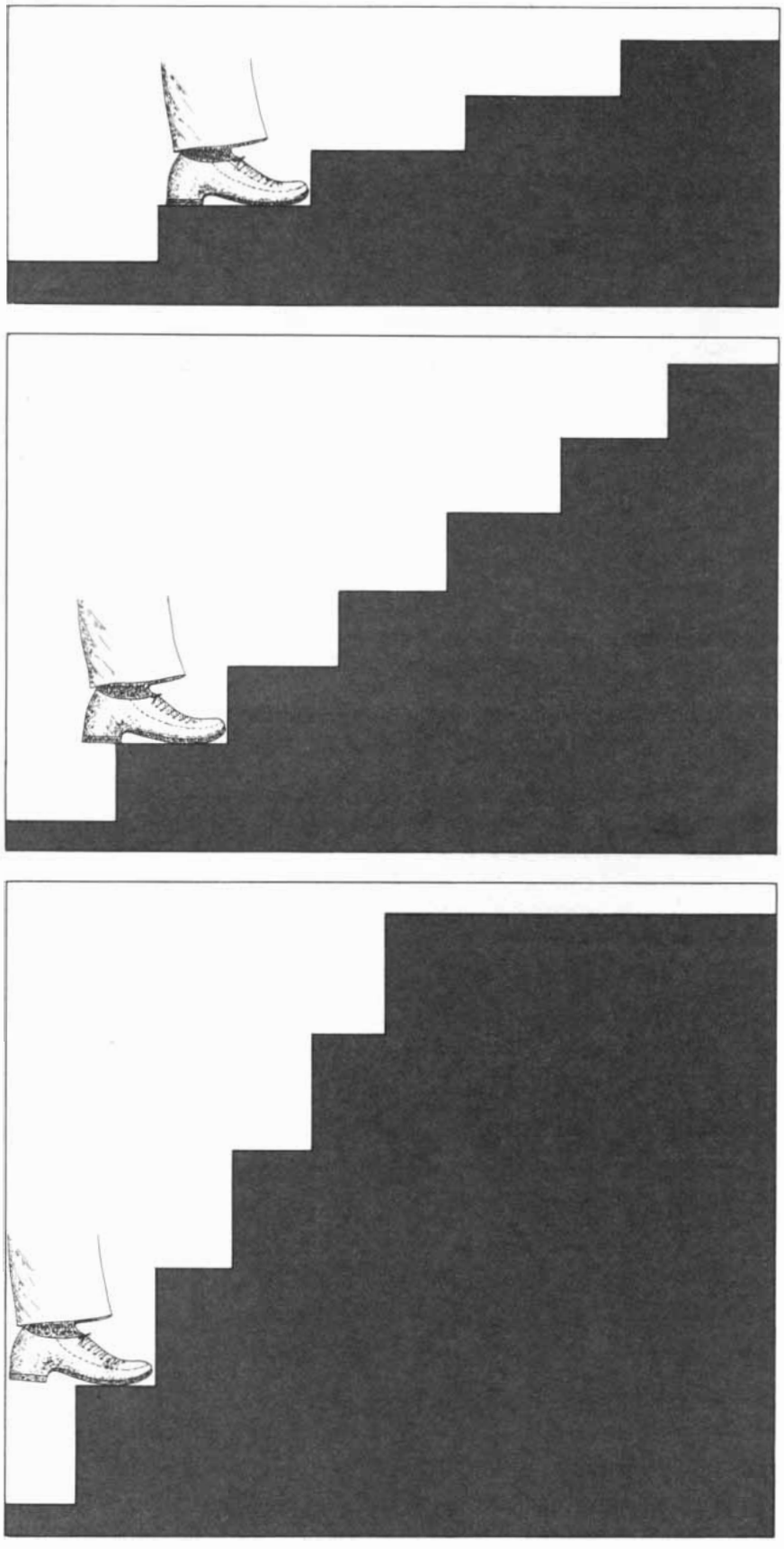
**HELICAL STAIRS are in City Hall in New York. They carry traffic between the entrance hall and the rotunda. A study of how people traveled up and down the stairs showed that in ascent 65 percent of the users kept to the right on the flight at the right, whereas no one kept to the right on the other flight. In descent the patterns of use were essentially reversed.**

The technical literature on stairway accidents proved to be somewhat limited in that the few reported studies had focused on accidents in the home. Residential stairways are usually steeper than those in public places and more deliberately utilitarian. Nonetheless, the studies indicate that most stairway accidents involving injuries severe enough to require medical attention occur during descent and that in a high percentage of the cases the stairways lack handrails. Slipping is the primary cause of stairway accidents.

The steepness of a staircase appears to be unrelated to the accident rate. Dimensional irregularities, caused by poor design rather than wear, are strongly correlated with the accident rate. None of the studies examined a subject of particular interest to us: the size of risers and treads and the various ways in which the height of the riser and the depth of the tread can be satisfactorily combined.

We therefore undertook a series of investigations directed at identifying riser and tread sizes and combinations that seemed to be most congruent with the human gait and hence felt safe and comfortable to users. We were by no means the first in the field. Almost every architectural writer since Vitruvius has expressed views on suitable combinations. It was not until about 1672, however, that François Blondel, director of the Royal Academy of Architecture in Paris, concluded from personal observation that the normal pace in level walking (24 inches, he said) must be decreased by a regular and fixed amount to allow the foot to be raised in climbing stairs. Blondel derived a formula stating that the pace must be decreased by two inches for every inch of riser. Expressed mathematically, the formula specified that the depth of the tread should be 24 inches minus two times the height of the riser, or that the total of the tread and two times the height of the riser should be 24 inches.

The apparent anthropometric logic behind Blondel's thesis recommended his formula to generations of architects in spite of the fact that, in stairways where the risers are to be made higher or lower



**RISER AND TREAD COMBINATIONS are portrayed according to a formula derived some 300 years ago by François Blondel, director of the Royal Academy of Architecture in Paris. He held that the depth of the tread should be 24 inches minus two times the height of riser. Application of formula is shown here for risers of five (*top*), seven and nine inches.**

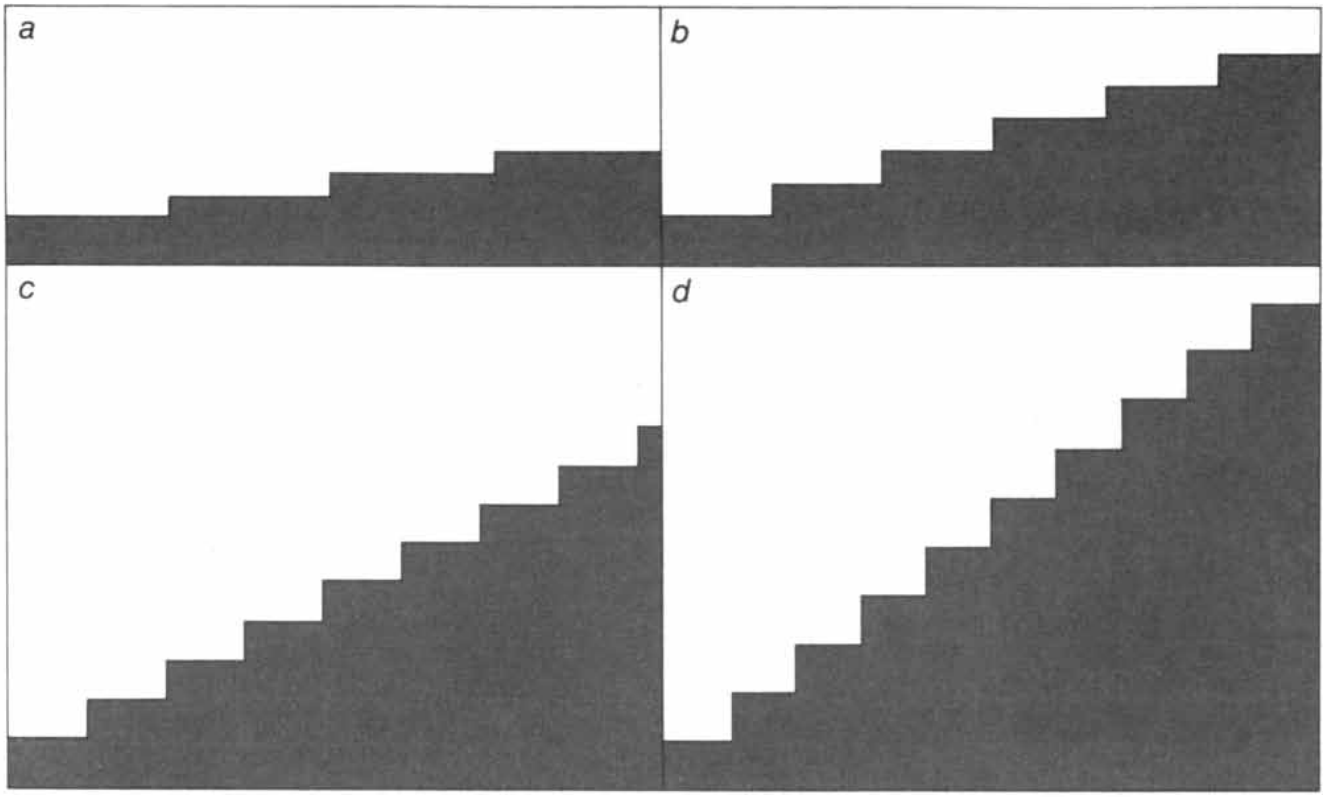
than usual, the rule produces either extremely narrow or extremely wide treads. Moreover, the average human being is bigger now than most people were in the 17th century, and today's inch is shorter than the one employed by Blondel in his calculations. It is therefore somewhat surprising to find Blondel's rule or variations of it still entrenched in many building and fire codes. The National Fire Code, for example, specifies that the depth of the tread plus two times the height of the riser should total between 24 and 25 inches.

Our study also led us to a number of apparently quite arbitrary formulas that govern only utilitarian interior stairs; exterior stairs are either not mentioned or are explicitly exempted. The formulas are notable for their lack of unanimity. For a four-inch riser, as a case in point, treads ranging in depth from 13 to 28.3 inches are acceptable.

The formulas are particularly equivocal when the angle of ascent is less than 20 degrees. Here the architect is expected to design stairways for which the formulas offer no consensus. Faced with such a situation he is apt to fall back on intuition, which may be what happened in the design of the unfortunate stairway at Lincoln Center.

To test the effect of various stairways on the user's gait and sense of comfort we built replicas of four stairways and photographed people walking up and down them. Although each one was only three feet wide, it reproduced accurately the configuration of the real steps: the ones at Lincoln Center, a monumental flight in the central plaza at Columbia University and two ordinary stairways meeting the standards of the New York City building code. Each replica consisted of four steps so as to make it comparable with the Lincoln Center stairway. The angle of ascent ranged from 7.7 degrees (Lincoln Center) to 36.8 degrees (city code). Men and women wearing ordinary street clothes and their own choice of shoes were photographed in the light of a high-speed repetitive flash lamp as they went up and down the steps.

These exploratory tests confirmed our suspicion that certain step configurations, such as the ones in the Lincoln Center stairway, impose a curiously irregular gait on both men and women in both ascent and descent. We decided that our next move should be to examine the relation between gait and the configuration of steps in a methodical way, testing the entire range of riser-




**REPLICA DESIGNS embodied a variety of riser and tread configurations and pitches. The Lincoln Center configuration (*a*) had 3.375-inch risers, 25-inch treads and a pitch of 7.7 degrees. A replica of a monumental stairway on the campus of Columbia University (*b*) had five-inch risers, 17-inch treads and a pitch of 16.4 degrees. The other designs are found in many public places. One (*c*) had six-inch risers, 12-inch treads and a pitch of 26.6 degrees. The other (*d*) had 7.5-inch risers, 10-inch treads and a 36.8-degree pitch.**

tread combinations normally found or recommended. For this purpose we employed a mechanical stairway that operated as a treadmill.

The stairway had been developed originally by the U.S. Department of Agriculture for monitoring the energy expenditure of women on stairways. It can be adjusted for a range of pitch from 22 to 45 degrees, riser height from four to 11 inches, tread depth from seven to 15 inches and speed of ambulation from about 20 feet of vertical rise per minute to about 100 feet. Our tests were carried out at the metabolic research laboratories of the Helen Hayes Hospital, a New York State physical rehabilitation facility. Sixteen subjects participated. They wore special rubber overshoes fitted with pressure-sensitive switches connected to a polygraph. In each trial the polygraph recorded the duration of the contact of the subject's toes and heels with the treads and also the rhythm of his gait.

The tests were run at three angles of pitch (25, 35 and 45 degrees), with three different riser-tread combinations for each angle. We therefore had nine configurations, each of which was tested at three different speeds of ambulation. The first speed forced the subject into a rhythm bordering on a run; the second was the normal pace that people have been observed to use on stairs, and the third was a speed close to slow motion.

On the basis of 1,291 successful tests we developed equations that enable us to predict with confidence the number of missteps that can be anticipated with any of the riser-tread combinations. (A misstep would not necessarily lead to a stumble or a fall. As defined for the tests, a misstep is a step that is not completed within certain predetermined tolerances.)

The results show that fewer missteps are likely in ascent than in descent at all speeds and for all riser-tread combinations. Indeed, all the test stairs can be climbed with little likelihood of a mishap, at least at one of the three speeds. Nonetheless, the results indicate that a range of riser-tread combinations, having risers of from 6.3 to 8.9 inches and treads of from 7.7 to 14.2 inches, results in stairs that are likely to be safer to climb, as reflected by the low number of missteps, than other configurations of the test group.

In descent the fewest missteps were recorded when the subjects were walking at the slowest pace. This finding held for every configuration of steps. It bears out what we found in the accident studies, which was that many of the falls occurred when people were hurrying down the stairs.

It appears from our results that the safest stairs for descent are not the same as the ones for ascent. In descent treads that are at least 12.3 inches deep will result in the fewest missteps at all speeds. The number increases with each reduction in tread size. Treads less than nine inches deep will be associated with the greatest number of missteps regardless of the height of the riser and the speed of ambulation. Risers in the range from 4.6 to 7.2 inches are, however, likely to cause fewer missteps in descent than lower or higher ones. Our findings suggest that in the rare situations where a stairway is to be designed for unidirectional movement the configuration might differ according to whether the movement is to be up or down.

A study made by Joan S. Ward and William Beadling of the Loughborough University of Technology in England suggests why smaller treads raise difficulties in descent. Ward and Beadling noted that, as the size of the tread

is progressively decreased, a point is reached where the shod foot can no longer fit on the step without being twisted sideways. This awkward movement induces a crablike gait and a poorly balanced posture.

The crablike gait will begin to appear whenever the tread is too small for the foot. Therefore if stairs are to be adequate for the largest people, the tread will need to be at least 11.8 inches deep, allowing for a nominal quarter-inch between the heel and the riser and for the toe of the shoe to overhang the step by about 1¾ inches. A more conservative design, which would accommodate 95 percent of the population, would make the smallest allowable tread 11.1 inches.

These data suggest that 11-inch treads should be a reasonable minimum, at least for stairs in public places. The minimum allowed by most codes now is 9½ or 10 inches.

Having identified certain riser-tread combinations that are least likely to cause falls, we turned to the question of what combinations might be more comfortable than others in terms of the energy demand they impose on the user. Put another way, the question we sought to answer was: Are some stairs more tiring than others?

Again we found the published material scanty. A good deal has been written about the energy cost of level walking and walking on ramps, but the technical difficulty of measuring the metabolic cost of going up or down stairs has limited the number and quality of the studies published on that subject. The problem is to separate with precision the energy expended on the basic functions of life from the additional energy spent climbing stairs. The separation is particularly difficult when the stairway is typical of the ones generally encountered.

To simplify the technical procedures for measuring energy expenditure on stairs one needs a long flight of stairs, which is relatively uncommon. One team of investigators carried out its tests on a flight of stairs that went up a mountain. Other investigators chose to have their subjects work on shorter flights but make several ascents and descents on each of the tests. This procedure effectively prevented the investigators from separating the work of descent from the work of ascent.

Once again the mechanical stairway came to our rescue, since it enabled us to simulate stairways of any length. To carry out the metabolic studies we employed 19 different riser-tread combinations and three different rates of vertical ascent. We measured a subject's energy expenditure by recording the amount of oxygen he consumed while walking on the treadmill and subtracting from that total the energy cost of the basic metabolic functions. With 639 observations we formulated an equation that enables us to predict the energy demand of any of the stair combinations in terms of the speed of ascent, the pitch of the stairway, the height of the riser, the depth of the tread and the weight of the subject.

**MECHANICAL STAIRWAY enabled the experimenters to test a number of different riser and tread combinations, pitches and speeds of ambulation for both their effect on the human gait and the amount of energy required to climb under various conditions. Here the machine appears in a double exposure that portrays how it can be set at different angles of pitch and how it serves for both ascent and descent. The apparatus is at the Helen Hayes Hospital in West Haverstraw, N.Y., a state facility where the experimental work was done.**

The experiments showed that walking on stairs produces a pattern of energy expenditure that differs notably from the patterns exhibited in walking on the level or on ramps. In level walking and on ramps people generally choose a speed of ambulation at which the total energy cost of walking a given distance will be

# Conversation Pieces

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

### *The Great Communications Traffic Jam*

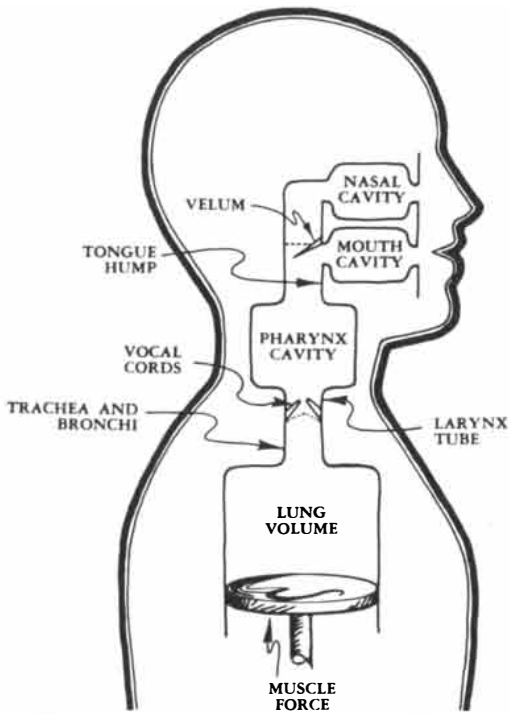
If you've ever listened to a citizen's band radio, you know about crowded airwaves. With two or three people trying to talk on each channel, there's often more frustration than communication.

The situation is more controlled for transcontinental and international traffic but the jams are building up there, too, particularly at the Christmas and Mother's Day peaks in the United States.

One solution is to use higher frequencies, even laser communications, and take advantage of the extra bandwidth. But the technology is expensive and, so far, it's developing gradually rather than explosively.

A possible interim solution is voice compression. It won't do a thing for CB fans yet, but it looks very promising for the telephone companies and government communication networks. Here's how it works.

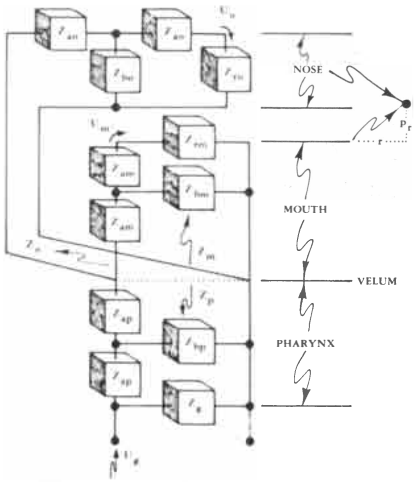
You first put the voice signals from a microphone through an electronic box that digitizes them. This simply means that each up or down step in the range of the speaking voice is turned into a set of ones and zeros, or data bits.




*The human speech generation system consists of a bellows (the lungs) that pumps air past the vocal cords and through a series of resonating cavities (the pharynx, mouth, and nose). Nerve impulses regulate frequency and volume.*

If you use a high data rate (i.e. divide the voice range into many, *small* up-and-down steps and sample them frequently) you get enough digital data at the receiving end to reconvert the digital signals into natural voice. If you compress the voice by using fewer, and therefore bigger, up-and-down steps (or by sampling the steps less frequently), you may get the *words* out at the other end but the speech has a flat quality.

The crucial problem, then, is how to conserve bandwidth by compressing the voice but still maintain good quality. The people in TRW's Voice Processing Lab have been remarkably successful in doing just that. With an unusual combination of linguistic and mathematical modeling skills, they're beginning to develop systems that will do the whole job automatically. With large-scale integrated circuit technology, plus volume production, costs may eventually be quite low, too.

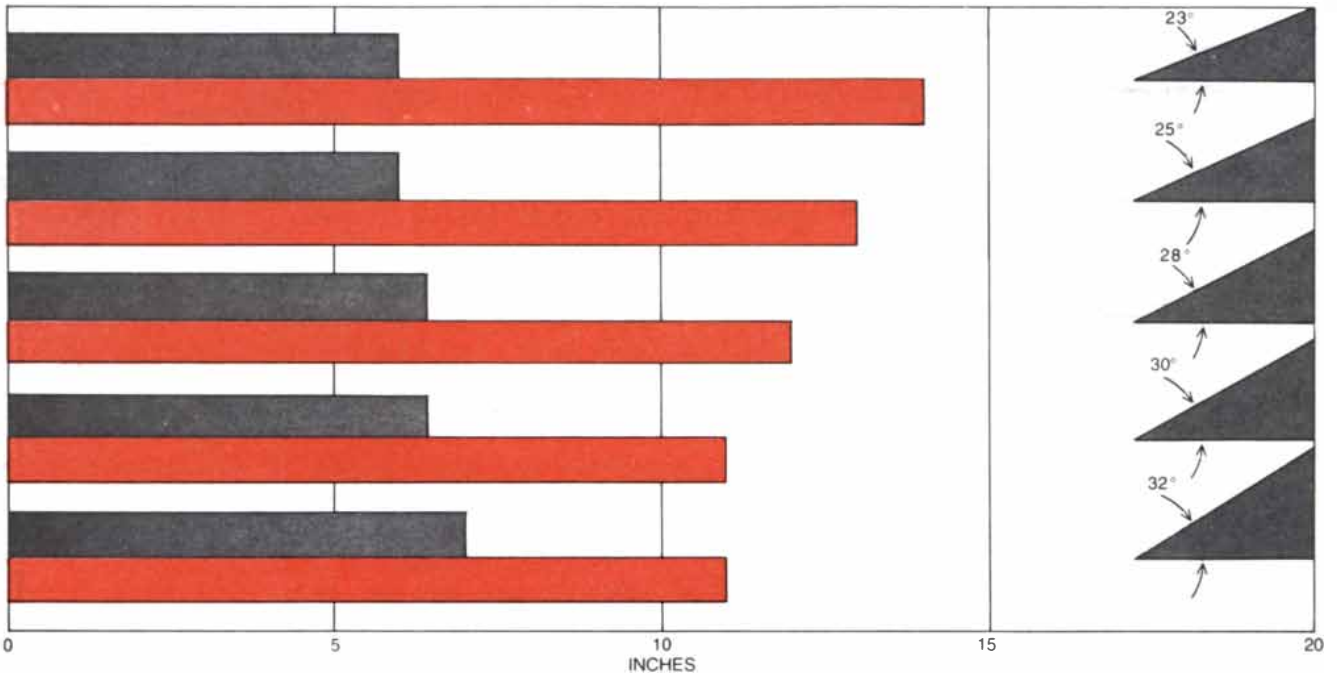



***Careful evaluation of the properties of human speech forms the basis for refinement of the basic algorithms (mathematical descriptions of the human speech system and its mechanical and electronic analogs) that enable programmers to compress the digital speech signals. TRW's multidisciplinary team is working with specialists in electronics to develop techniques that use relatively few data bits per unit of speech, yet preserve the natural quality of the voice.***

For further information, write on your company letterhead to:

**TRW®**
**SYSTEMS GROUP**

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




**PROPORTIONING OF STAIRS to conform with the natural, informal human gait is depicted. The chart shows riser height (*gray*), depth of tread (*color*) and angle of steepness for five designs. These designs and several others were found suitable by the authors.**

at a minimum. The total energy cost of walking a mile on the level is much higher at one mile per hour or six miles per hour than it is at three miles per hour. People therefore usually choose speeds close to three miles per hour for walking. A similar phenomenon of optimal speed is also exhibited by people on ramps.

This finding does not hold good for walking on stairs. People seldom climb stairs at a speed that is most economical in terms of total energy expenditure. Apparently stairs are barriers to which the human being is not well adapted physically. Climbing stairs is perhaps more like running, which obviously does not conserve energy.

One can consider ramps and steps simply as prosthetic devices for assisting the human organism in climbing from floor to floor and can compare them in various geometric combinations in terms of the energy expended in using them. Basing our observations on the usual rate of climb, we found that within the range of the observations the total energy cost declined as the stairway became steeper. On ramps also the total energy cost per vertical meter decreases as the pitch is increased until a pitch of about 17 degrees is reached; from there on the energy cost rises. Ramps, however, demand a higher expenditure of energy per meter of rise than the usual range of stairs.

Our findings have implications for design in structures where the sole concern is the total energy cost of vertical movement. When one must consider the energy cost of both horizontal and vertical movement, one finds that a ramp with a gradient of less than about eight degrees is more economical than any stairway that is likely to be encountered in normal activity.

On the basis of these considerations one can conclude that in a conventional two-story house a stairway is more economical than a ramp in terms of the total energy expenditure. (The stairway, of course, is also more economical in terms of the amount of space it takes up.) In a transit terminal, however, where one often needs to travel considerable distances horizontally as well as vertically, a ramp within the normal range is more economical than any stairway in terms of the total energy demand.

It is not the total energy cost of an action that one perceives as fatigue. Fatigue is experienced after expending energy at a high rate over a short period of time. Clearly it takes about the same amount of energy to run up a stairway as to walk up, but running is more tiring. It is therefore worth considering the rates of energy expenditure imposed by various riser-tread combinations, particularly since the studies of step geometry by Ward and Beadling indicated that the stairways least preferred by a group of test subjects were the ones demanding high rates of energy expenditure.

We found that the rate of energy expenditure on stairs is generally greater for high risers than for low ones. Steep stairs with high risers and small treads, however, demand no more effort than shallower stairs with much lower risers and large treads. The findings have implications for the design of stairways.

Clearly the findings support the general validity of Blondel's 300-year-old thesis. For comfort, in terms of low rates of energy expenditure, a stairway should be designed to provide a congruence between its riser-tread combination and the human gait. To keep the rate of energy expenditure low, the designer must either combine low risers with deep treads or reduce the size of the treads as the height of the risers increases. It must be said, however, that the formulas of Blondel and his successors are too restrictive, too narrowly idealistic and too unresponsive to human adaptability. The range of options for combining risers and treads at any level of energy expenditure cannot be expressed satisfactorily by a linear equation.

Let us now assemble the conclusions from the analysis of gait and the analysis of energy expenditure and apply them to the design of stairways. Assuming that the stairway is to have people going both up and down, the designer can choose from a broad range of riser-tread combinations. The stairs will have risers of from four to seven inches coupled with treads of from 11 to 14 inches. A stairway designed within this range will result in fairly low rates of energy expenditure (at normal speeds) and a low rate of missteps.

# So simple. So reliable. So effective. So successful. Sony.

The simplicity, reliability, and effectiveness of the Sony U-Matic Videocassette System are what have made it so successful.

So successful, in fact, that an independent survey* shows that *3 out of every 4 purchasers of video recording/playback equipment in 1974 will buy the ¾″ U format Videocassette System.*

So successful in use that the U. S. Army has publicly stated that their ¾″ U-matic equipment is "99.9% reliable."

And the ¾″ U format is Sony's format. Originated and engineered by Sony. Perfected by Sony. And proven effective by thousands of users in Business and Industry: from the Fortune 500 to smaller regional companies. In Education: from Universities to Public School Systems. In Medicine, in Cable TV, Government. Wherever people must interchange information.

It's the new concept in communication. With all the impact that you know television has. Simple to use, ideal for training, teaching, product demonstrations, messages from management.

Here are some special features of the Sony ¾″ U-matic format:

Delivers a picture that's superior in resolution, color fidelity, and stability.

2-track sound for bilingual or technical and non-technical presentations.

Playing time of up to 60 minutes, twice that of some other systems.

Push-button controls. Pause, automatic rewind and shut-off.

A book-sized cassette that can be mailed inexpensively anywhere in the U.S. or the world.

And everything the Sony name stands for in quality, reliability, and nationwide service.

It's Sony's easy and effective way for you to move into a new world of communications. Find out how easy. Write today.

## Sony. The proven one!

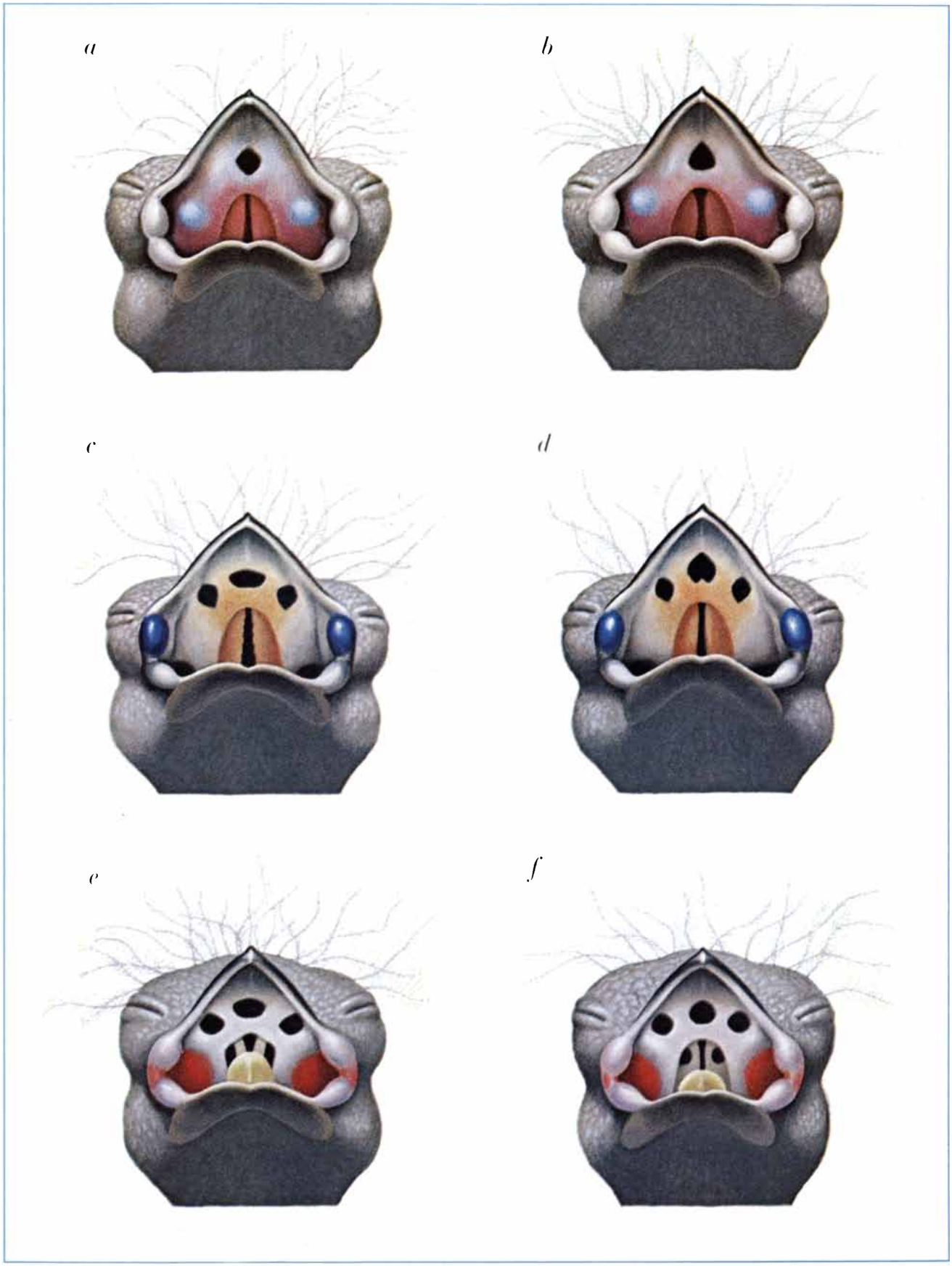
Sony Corporation of America SN-182
Video Products Dept. SA-104
9 West 57th Street, N. Y., N. Y. 10019

☐ Send me more information about the Sony Videocassette System.
☐ Have a representative contact me for a demonstration.

NAME
TITLE
ORGANIZATION
ADDRESS
CITY
STATE ZIP
PHONE (AREA CODE)
TYPE OF BUSINESS

**Sony® U-matic® Color Videocassette System**

* *Educational & Industrial Television*, February 1974
All TV reception simulated




MOUTH MARKINGS of host nestlings, which stimulate the host parents' feeding response, are mimicked in detail by all the parasitic species of African widow birds. Shown here are the markings of a host-species nestling, the melba finch (*a*), and its parasitic sibling, the paradise widow bird (*b*), of a second host species, the purple grenadier (*c*), and its parasite, the straw-tailed widow bird (*d*), and of a third host-parasite pair, Jameson's fire finch (*e*) and the purple combassou (*f*). Many other mimicries of the type are known.

# MIMICRY IN PARASITIC BIRDS

## Various species of birds lay their eggs in the nest of another bird, which then incubates the eggs and feeds the young. The widow birds of Africa achieve this result by some remarkable feats of mimicry

by Jürgen Nicolai

The number of eggs a female animal produces is inversely proportional to the probability that the egg will give rise to a female that produces more eggs. All invertebrate animals and most lower vertebrates need to broadcast eggs by the hundreds and thousands in order to ensure the existence of subsequent generations. The number of eggs laid by a bird, although much smaller, also reflects the probability of reproductive success. Consider the greatest of ocean birds, the wandering albatross. It occupies breeding grounds on isolated islands of the Southern Hemisphere where it is completely undisturbed, and the female albatross lays a single egg every other year. In contrast, most species of European tits and both species of goldcrests respond to the stresses of severe winters and numerous predators by laying two yearly clutches of eggs, each numbering between eight and 12.

Any bird that is subject to severe environmental stress or predation pressure is at a reproductive disadvantage. It must incubate its eggs with the heat of its own body, and so the eggs must be kept all in one place, usually in a nest. A predator therefore needs to search out only a single target. One group of birds, the turkeylike megapods of Australasia and the Pacific, has surmounted this handicap by an ingenious stratagem. Instead of building nests they pile up mounds of plant material where they bury their eggs to be incubated by the warmth of plant decay [see "Incubator Birds," by H. J. Frith; SCIENTIFIC AMERICAN, August, 1959]. Still other bird species have discovered another solution to the problem. They deposit their eggs, one at a time, in the nest of another species; the eggs are then incubated and the hatchlings raised by the host.

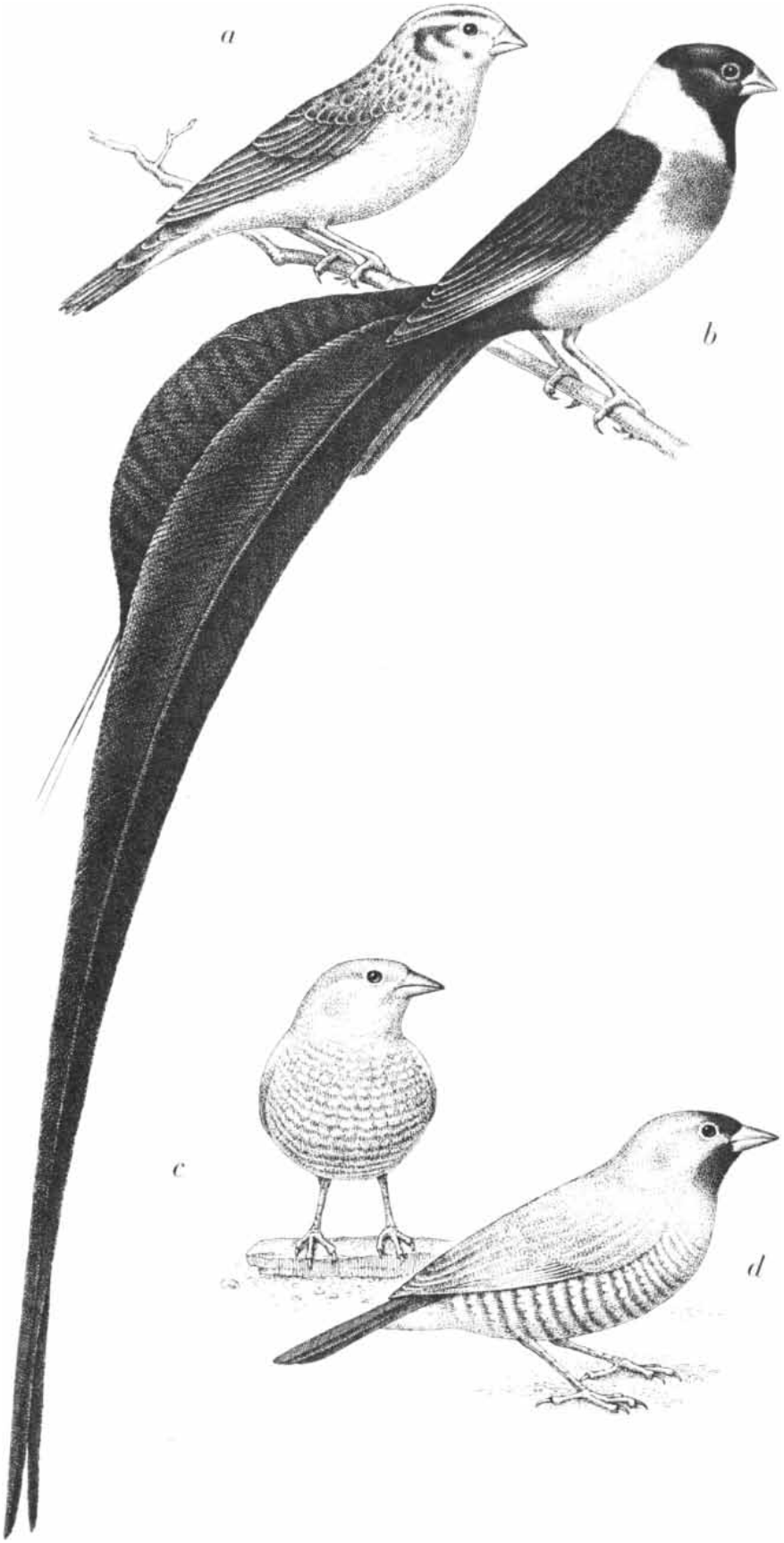
Birds that practice this kind of parasitic parenthood are found around the world. Among them are various members of the cosmopolitan family of cuckoos, some species of American cowbirds, a black-headed duck (*Heteronetta atricapilla*) in South America, all species of honey guides and certain weaverbirds in Africa. I have investigated the behavior of this last group, observing in particular the several species of a subfamily, the Viduinae, within the family of weaverbirds. The birds in this subfamily are commonly called widow birds. My work has been conducted both in the field and in our laboratory at the Max Planck Institute for Behavioral Physiology in Seewiesen.

Although a parasitic bird has overcome the disadvantage of putting all its eggs in one nest, the very nature of its solution to the problem introduces certain other difficulties. For one thing, if parasitic parenthood is to be successful, the parasite's reproductive cycle must be synchronized with the cycle of the host species. For another, it would be fatal if the host rejected either the parasite's egg or the hatchling that emerged from it. With respect to the egg a strategy of mimicry has evolved. For example, parasitic cuckoos lay eggs with markings that closely resemble those of the host's eggs. Widow-bird eggs are unmarked, as are their host's; they differ from the eggs of the host species only slightly in size and shape.

With respect to the acceptance of the hatchling the parasite strategies that have evolved are various. Several species of cowbirds punch tiny holes in the host's eggs before depositing their own, thus ensuring that no host egg ever hatches. An African honey-guide hatchling emerges from its egg armed with formidable hooks at the tip of its beak; its host's offspring are all soon fatally wounded, leaving the parasite to consume all the food the host provides. Some cuckoo nestlings simply nudge their foster siblings out of the nest during the first few days of life; others mature so much more rapidly than their nestmates that the foster siblings' development is inhibited and they die prematurely.

In the long run, of course, strategies that result in the death of the host's young are unproductive. The parasitic parent needs to use the nests of several host pairs each year and a dwindling host population means fewer nests. The widow birds have evolved a strategy that circumvents this problem; like the egg strategy, it is imitative. Each widow-bird nestling is indistinguishable from its host's nestlings in size, in color and markings, in gesture and in call. Thus the intruders can grow up among their alien nestmates with no risk of being rejected by their foster parents. Such a strategy ensures that the numbers of the host population do not diminish.

It seems surprising that the widow birds should have evolved this particular strategy. Their chosen hosts—various species of finches, the entire family Estrildidae—are noted for having mouths that are colored and marked in a complex and conspicuous manner [*see illustration on opposite page*]. No two of the 125 species of estrildid finches have identical mouth markings. The color of the palate may be whitish, red, yellow or bluish. The palate markings may form a three-spot pattern or a five-spot pattern or may be only a fine horseshoe-shaped line; the spots themselves may be black or violet. Moreover, the fledglings' gape papillae vary in shape and color. They may be ivory white, cornflower




**PARASITIC ADULTS, the female (*a*) and male (*b*) paradise widow bird, do not in any way mimic the appearance of the foster parents of their young, the female (*c*) and male (*d*) melba finch. The male paradise widow bird is seen in its bright breeding plumage.**

blue, yellow or faintly violet, and they may appear as simple thickenings or resemble small pearl-shaped warts.

In parallel with these elaborate mouth markings there has evolved among the estrildid finches a precise, genetically based "knowledge" of species-specific patterns and stimulus-receiving mechanisms that are correspondingly selective. By the simple experimental method of placing nestlings of one finch species in the nest of other species my colleagues and I have demonstrated that the parent birds will feed only the nestlings that display the appropriate species-specific markings. Even minor deviations in pattern are noted by the parent birds, and the deviants are ruthlessly weeded out by starvation.

A parasite nestling constantly interacts with its foster parents from the time of its emergence until it leaves the nest. The host birds are thus an essential part of the parasite's environment. The host sets a series of examples that instructs the parasite with respect to such matters as environmental standards and feeding habits. One instance of this process of parallel adaptation is the synchronization of the parasite's and the host's reproductive cycle.

Most species of estrildid finches breed during the rainy season. The finches' gonads begin to swell when, after months of drought, a series of showers brings the vegetation to renewed life. The timing is appropriate because the increase in the food supply that comes with the rainy season makes it easier to feed nestlings. Several finch species begin their breeding activities when the rains start; others do not begin until the middle of the rainy season or toward the end of it. A few species, such as the yellow-winged pytilia and the aurora finch of West Africa, delay their breeding until the dry season, a time when they find feeding conditions appropriate.

In each of these instances the parasitic widow birds synchronize their reproductive cycle with that of their host. One element in the cycle is that when courtship begins, the male widow bird displays a splendid and conspicuous plumage. The molt into breeding plumage occupies a period of four to six weeks. This means that the maturation of the widow bird's gonads, which initiates the molt, actually occurs sometime before the host's gonads mature.

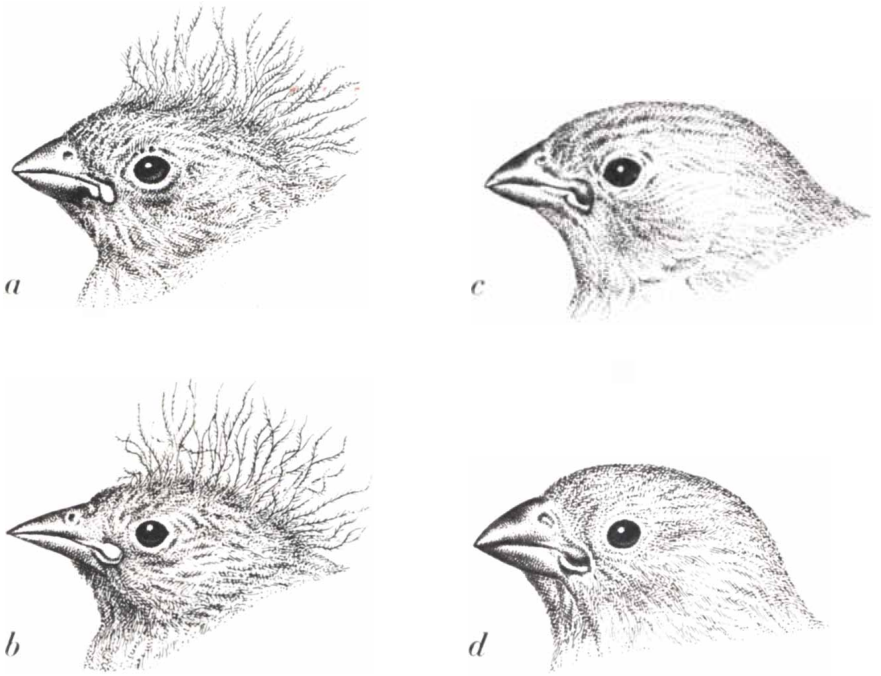
To cite some examples, the paradise widow bird and the straw-tailed widow bird have hosts that breed during the rainy season. That requires the male

widow birds to develop their breeding plumage before the end of the dry season. Similarly, two species of widow bird, the Togo paradise widow bird and the Kongo paradise widow bird, have hosts that breed during the dry season. The male parasites' molt must therefore take place toward the end of the rainy season. Under these circumstances it is clear that the onset of the parasites' reproductive activity cannot simply be triggered by the onset of the hosts'; it is evidently set in motion by some kind of seasonal rhythm. The specific rhythms remain unidentified but one is justified in assuming that, because the rainy and dry seasons follow each other with great regularity in most of Africa, the gonads of rainy-season breeders begin to be stimulated after the passage of a certain length of dry-season time, and vice versa.

Now, in any fine-tuned system of parasitic adaptation one of the greatest potentials for disruption is accidental hybridization between species of parasites that are closely related and coexist in the same area. As far as the widow birds are concerned, any hybrid offspring would display intermediate mouth markings rather than markings that match those of their foster siblings. That in turn means they would be refused food by the host species of both parents. This hazard, with its long-term potential for the extinction of all widow birds, has been avoided by the evolution of a kind of behavior that is unknown among other parasitic birds. In brief, when the male widow bird sings, it includes in its song certain unique finch-species phrases it has learned from its foster parents.

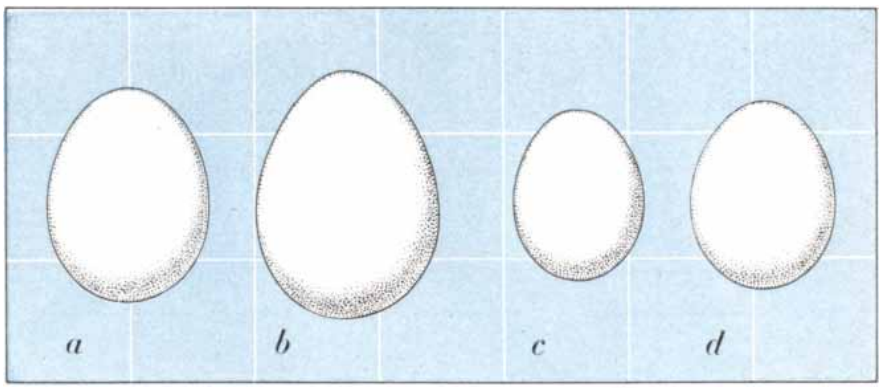
Widow birds are polygamous. At the start of the breeding season a male stakes out a large territory that it defends against potential rivals. Throughout the breeding season the male perches on certain selected trees and bushes in the territory and sings its unique melody. If one compares the songs of various widow-bird species, either by listening or by analyzing sound spectrograms, it soon becomes apparent that the songs of all species have a few phrases in common. For example, they all include harsh chattering sounds; the motif is evidently related to the vocalizations produced by the widow birds' closest relatives, the bishop birds of the subfamily Euplectinae. Since these chatterings appear to represent an ancient genetic heritage, we call them widow-bird phrases.

The major portion of any male widow bird's song consists of quite different vocalizations. Its motifs vary from one species to another, and each species-specific repertory perfectly imitates the motifs characteristic of the widow bird's finch host. We therefore call such motifs host phrases. They include the finch's long-distance call, its contact call, its distress and anger calls, its greeting phrases and, of course, its routine song. The parasite's mimicry of this repertory is so exact that it sounds like the finch's to the human ear and looks like it in a spectrogram [*see illustration on page 97*]. Furthermore, the host finch itself cannot distinguish between the widow bird's imitation and the song of its own species.

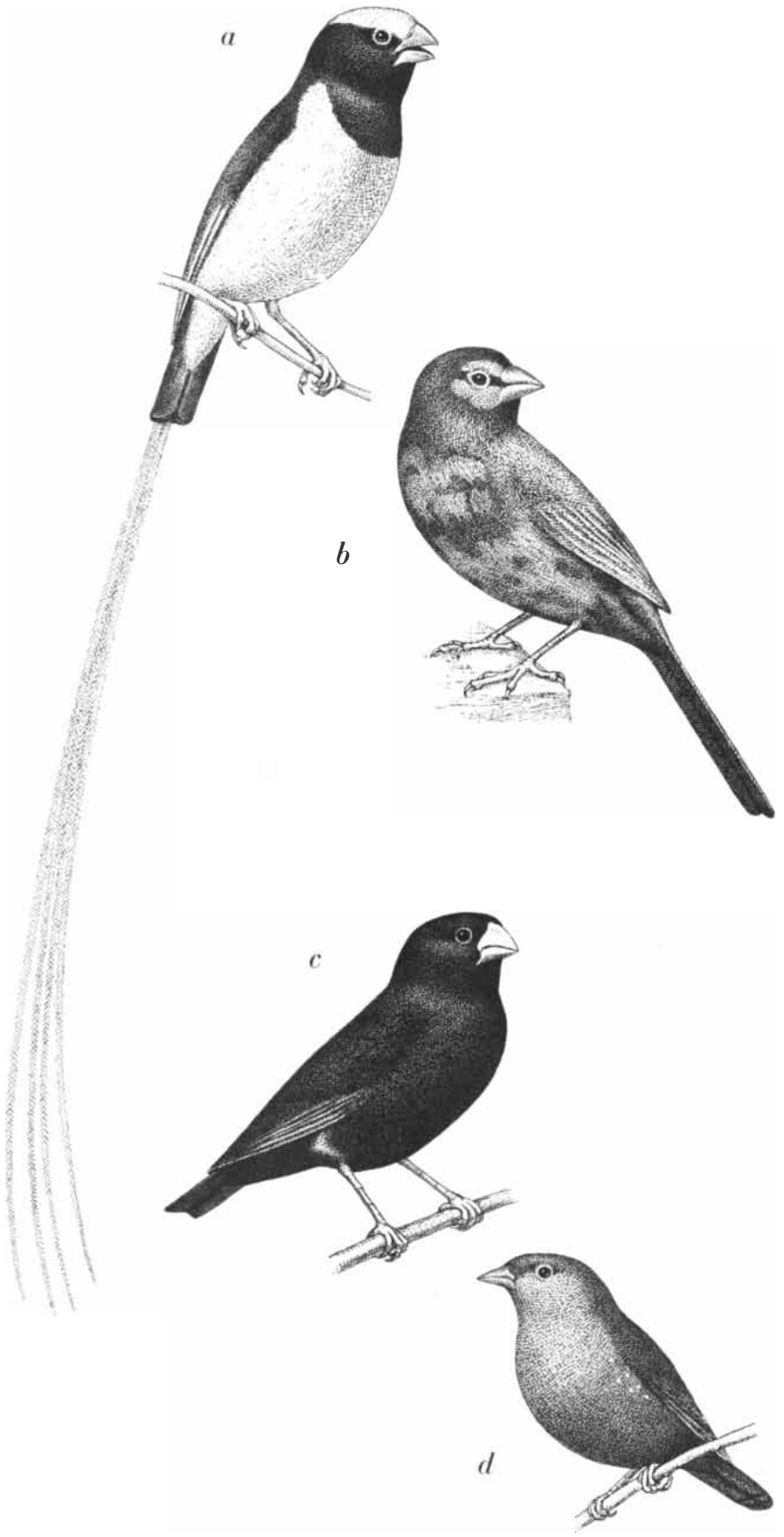



**YOUNG PARASITES, unlike adult parasites, closely resemble their host siblings in appearance. Profiles at left show a paradise widow-bird nestling (*a*) and a melba-finch nestling (*b*) 13 days after hatching; profiles at right show a straw-tailed widow-bird nestling (*c*) and a purple-grenadier nestling (*d*) 15 days old. The young are independent after five weeks.**

Let us examine in detail the song of one widow-bird species. The straw-tailed widow bird of East Africa (*Tetraenura fischeri*) ranges over the scrub savanna of that region from southern Tanzania northward into Ethiopia and Somalia. The male's breeding plumage is black and yellow and its four conspicuously elongated central tail feathers look like yellowish blades of dry grass. During the breeding season the male perches high in small trees or thornbushes and twitters its song unceasingly from early morning until sunset. The first motif in its repertory is usually the contact call of its finch host, the purple grenadier,




**MIMICRY IN THE NEST extends to the size and shape of the egg laid by the parasite. At left is the egg of the melba finch (*a*) and of the paradise widow bird (*b*); the parasite's egg is only slightly larger and more rounded than the host's. At right is the egg of another host, the red-billed fire finch (*c*), and the egg of its parasite, the black-winged combassou (*d*). As centimeter grid shows, the host and parasite eggs are virtually identical in size and shape.**



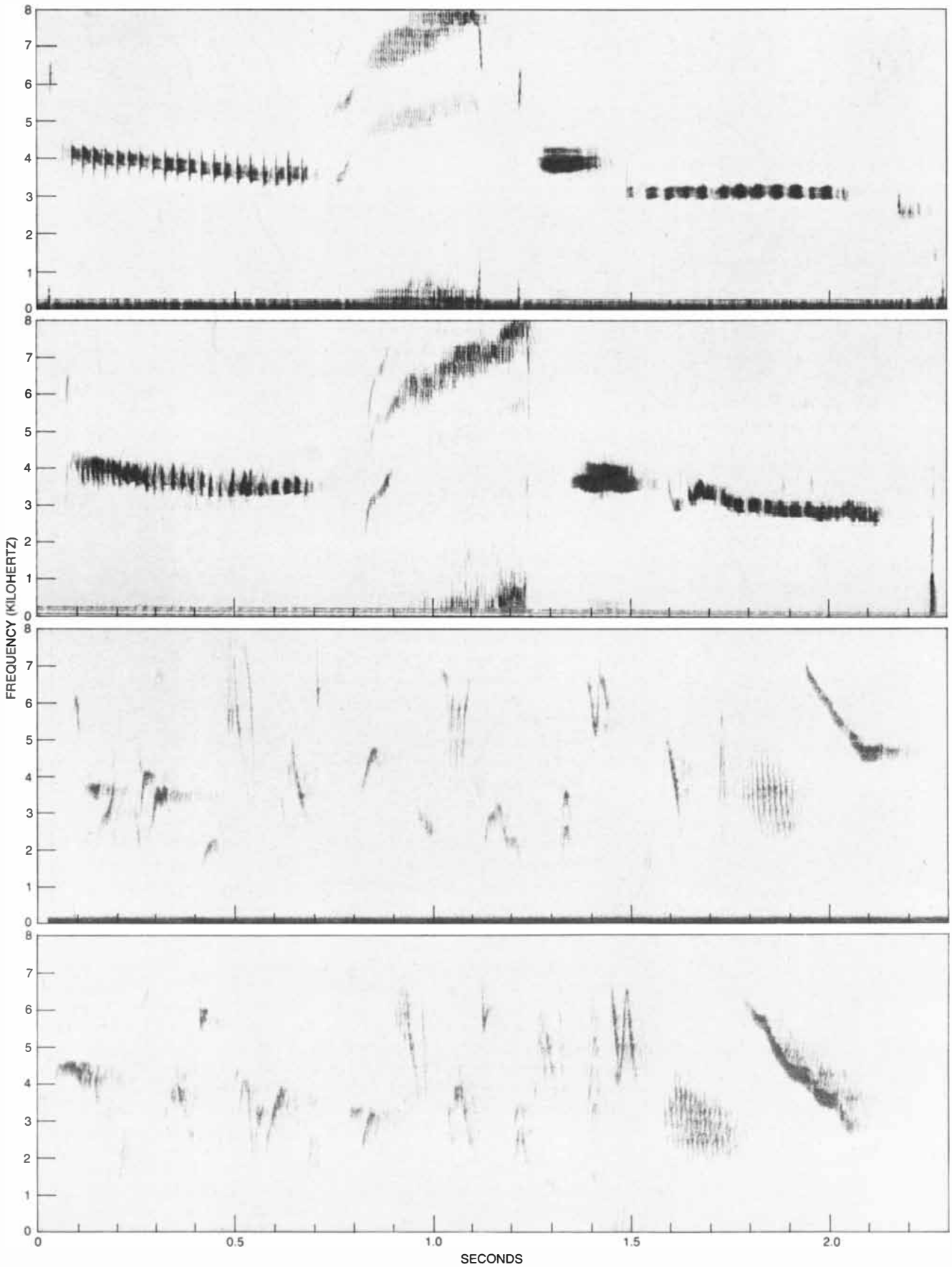

**BREEDING PLUMAGE of parasitic widow birds develops during a four-to-six-week period. Seen here are a male straw-tailed widow bird (*a*) and its host, a purple grenadier (*b*), and below them a male purple combassou (*c*) and its host, a Jameson's fire finch (*d*). The parasite's breeding cycle must synchronize with its host's to ensure synchronous egg-laying; in order to grow breeding plumage, however, a parasite must enter its breeding cycle first.**

*Uraeginthus ianthinogaster;* this is a long trill. The mimicked phrase is repeated several times, increasing and decreasing in loudness. It is followed by other elements of the host's repertory. They include the brief "call" phrase used by the male to attract the female, the male's clacking song, a chase call, whispering nest calls and the shrill begging calls that fledglings use; in all, the mimic's repertory includes a total of eight finch motifs. Each mimicked vocalization is separated from the next by an interval, and each lasts the same length of time as the host's. The finch potpourri is interspersed with three different widow-bird phrases that, like the mimicked host melodies, occur in unpredictable sequences. The eight finch motifs give the male straw-tailed widow bird effective command of the entire vocabulary of the purple grenadier.

The same pattern of host-melody mimicry is apparent in almost all other species of widow birds. The shaft-tailed widow bird of South Africa (*Tetraenura regia*) mimics the repertory of its host, the violet-eared waxbill (*Uraeginthus granatinus*). Each of the five species of the paradise widow-bird genus (*Steganura*) copies the host melody of the particular species of the finch genus (*Pytilia*) it parasitizes. The same is true of the seven species in the genus of short-tailed black combassous (*Hypochera*) that mimic all seven species of the genus of fire finches (*Lagonosticta*). Indeed, in only two instances has it been impossible to demonstrate the presence of host phrases in a widow bird's song; these cases are the two species of the genus *Vidua,* the blue widow bird of East Africa and the pin-tailed widow bird, which is found in the sub-Saharan regions of the continent.

By determining which are the host phrases in a male widow bird's song, it is even possible to predict the identity of the host when it is not known. Such predictions have been made more than once in our laboratory after careful analysis of the songs of captive widow-bird males, and I was later able to prove their accuracy. In the field in East Africa we found the appropriate young parasites among the nestlings of exactly those species of estrildid finches that had been identified as probable hosts.

The use of host melodies in widow-bird singing serves a vital purpose: it is a barrier to hybridization. How this isolating mechanism is acquired is best shown by a review of the widow-bird life cycle. The young nestlings' first weeks of life, as they share the nest of




**SONG MIMICRY, a meticulous imitation of the host male's call, learned by the parasite male while a juvenile, is demonstrated by these paired sonograms. The top sonogram shows the final seconds of a Damara melba finch's song; the sonogram directly below it is mimicry of this part of the finch's song by a male paradise widow bird. The third sonogram shows a two-second segment of the song of the violet-eared waxbill; the fourth is mimicry of this segment of the song by the waxbill's parasite, the shaft-tailed widow bird.**

their host siblings, are ones of constant contact with their foster parents. The parasites become independent at the age of five weeks, but even after that they continue to live for some time among their host siblings. It is only later that they form a flock with other juveniles of their own species.

During these weeks of dependence and association the young widow birds are imprinted to the host species; this imprinting is a decisive factor in their later reproductive behavior. When, at the age of seven to eight weeks, the widow-bird juveniles first engage in premature sexual behavior as all songbirds do, there arises in the parasites a quite narrow and specific interest in the activities of the host species. Whereas other songbirds not only play at building nests during this phase but also feed and court one another, the attention of the juvenile widow bird is concentrated on two aspects of its hosts' adult behavior. The focus of interest depends on the sex of the widow bird.

As the female widow bird approaches sexual maturity her interest is concentrated on the hosts' reproductive cycle. The parasite checks frequently on the progress of the host pair's nest-building. The female also synchronizes her own ovulation with that of the host female; this ensures that her own egg is ready at about the time the host female produces her second or third egg. As a consequence the young parasite will hatch at about the same time that its host siblings do and will be their equal in the competition for food.

The male widow bird's concentration is differently oriented. During the most impressionable phase of its youth its interest is focused on the vocalizations of the male foster parent. The parasite evidently disregards the songs of any other species of bird; as a result, when it reaches maturity, it is able to reproduce its host's sound patterns exactly. Although the female parasite does not sing, she becomes similarly song-imprinted, and a memory of the host's repertory remains with her for life. The imprinting becomes evident when ovulation takes place and the female's sexual drive is aroused. Once the female widow bird is in search of a mate, she "knows" that the only appropriate male partner is one who can recite the same calls and notes she heard from her foster father while she was still a nestling. Only when the courting male corresponds to this prototype will the female widow bird allow mating.

Superficially it might seem that the male's signal is merely a kind of lure, for example "I grew up among the same host species that you did!" In my view the signal contains a message with a much deeper biological meaning; it might be paraphrased as "I have inherited the same adaptive characteristics that you have and so our offspring will have the same chances of survival."

It is clear that the adult widow birds' general adaptation to their hosts' song repertory, breeding rhythms and feeding habits, along with their nestlings' perfect mimicry of foster-sibling markings, movements and calls, can only have been the result of an extended evolutionary process. As we have come to know the identity of more and more host species of estrildid finches some of the mechanisms of the process responsible for these unusual adaptations have become apparent. It seems that in effect the potential for the evolution of a new species of widow bird has arisen in the past only when a new species of host finch has evolved.

Consider the relations between hosts and parasites from a taxonomic viewpoint. Two species of the widow-bird genus *Tetraenura* are known: the straw-tailed widow birds and the shaft-tailed. The estrildid finches parasitized by these two related widow-bird species are also two species of the same genus, *Uraeginthus*. In turn five species of paradise widow birds are the parasites of five species of another finch genus, and the seven species of combassous are linked to seven species of fire finch. Even though the specific hosts of the two widow birds of the genus *Vidua* remain unknown, we are certain that they are one or another of the species in the finch genus *Estrilda*.

Such species-to-species relations can be understood only if one assumes that a long time ago the widow birds began to arise from a single viduine prototype species that had started to parasitize a single species of estrildid finches. When this ancient estrildid species began to evolve into several species, the widow birds were forced to follow suit, because only by doing so could they evolve mouth markings sufficiently like those of the evolving new species of potential hosts. Of course, each widow-bird population that failed to achieve such an adaptive process was destined to become extinct, since its mismarked offspring would starve to death in the nests of hostile hosts. In this way the evolution of the adaptively successful widow-bird species both parallels and reflects the evolution of their chosen hosts among the numerous species of estrildid finches.

# Everybody knows somebody who's been helped.


The United Way is reaching out. It goes into every neighborhood. To every age group. Across every income level.
To help. In a thousand ways.
A neighbor. A relative. Maybe someone down the street who's out of work. Or sick. Or in trouble.
Chances are the United Way helps someone you know.
Maybe someday it'll be helping you.

**Thanks to you it's working. The United Way.**

A Public Service of This Magazine & The Advertising Council

# The Coordination of Eye-Head Movements

*The sequence of events in the nervous system that coordinates the movements of the eyes and the head in fixating a visual target has been clarified by recent experiments with monkeys*

by Emilio Bizzi

How the central nervous system produces coordinated motor output has long been one of the major problems of neurophysiology. What accounts for both the graceful performance of the trained athlete and the ordinary, but hardly less remarkable, movements of everyday life? For movement to be coordinated an appropriate set of muscles must be selected, each of the contracting muscles must be activated in the proper temporal relation to the others and a precise amount of inhibition must be delivered to each of the muscles that will oppose the intended movement. In addition to triggering the contraction of a given set of muscles the nervous system must monitor the effects of its commands; it must have a way of coordinating movements of the various body parts; it must be able to decide when to terminate a given phase in a motor sequence and proceed to a new one, and so on.

For three-quarters of a century, dating back to the early studies of Charles Scott Sherrington, investigators have held different views concerning how temporally patterned sequences of neuromuscular events are programmed by the nervous system. Many workers have stressed the importance of sensory feedback in eliciting and coordinating a motor output. The feedback is viewed as originating with various types of sensors located in the muscles, the tendons and the joints. These receptors inform the central nervous system about various aspects of a movement, such as its velocity, amplitude and force. According to some investigators, sensory feedback from each phase of movement is important in eliciting each subsequent motor output. Therefore a coordinated motor performance could be described as being differentiated into many parts along the time dimension, so that each part is reflexly triggered by the sensory components of its predecessor. Thus, to use Sherrington's words, "coordination is in part the compounding of reflexes."

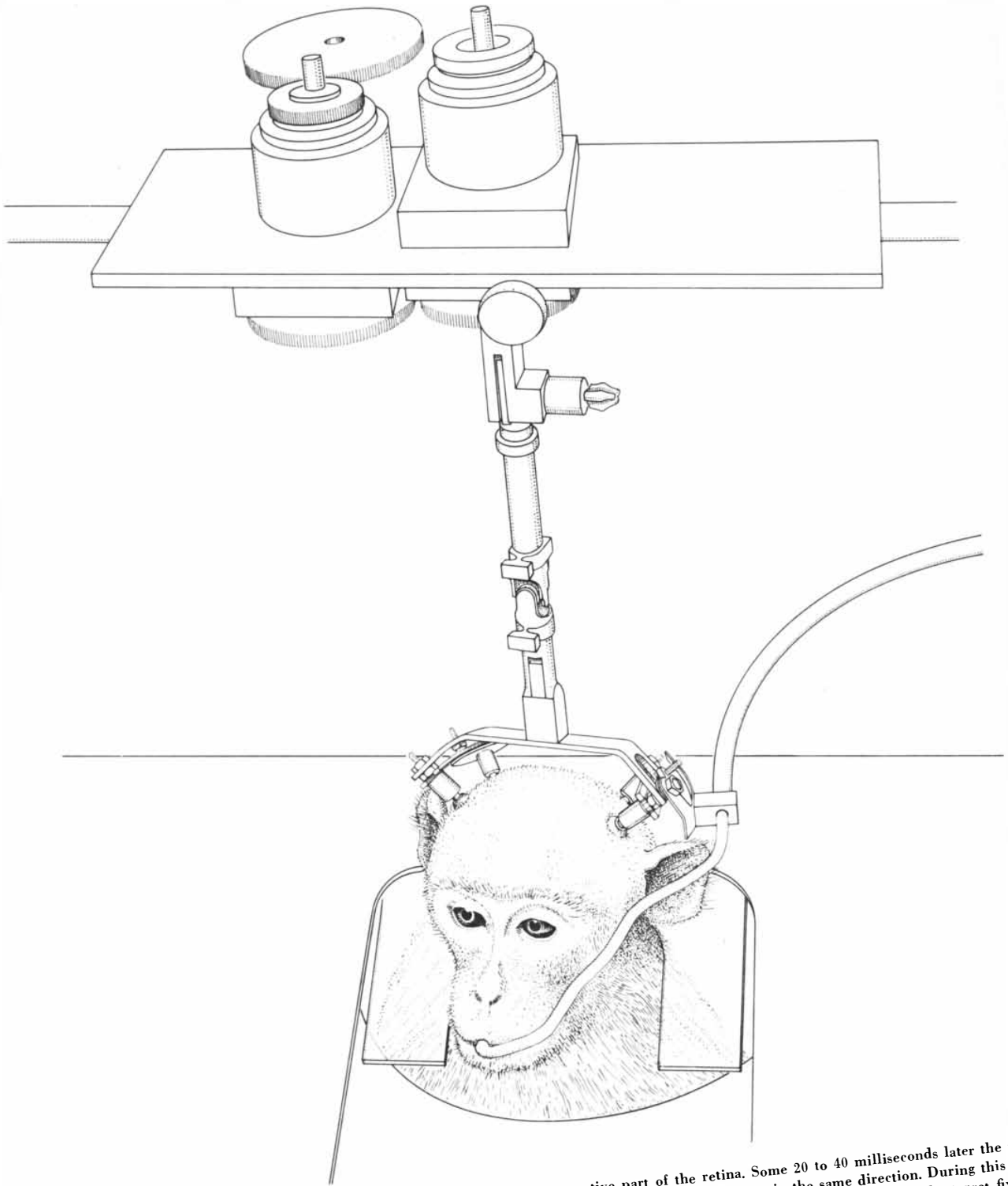
At the other extreme some investigators hold that the nervous system already incorporates all the information necessary for the selection of the muscles involved in a given movement and the sequential activation of those muscles. The motor system is viewed as a network capable of playing stored patterns in a predetermined manner following the presentation of an adequate input. In invertebrates, for example, the work of Cornelis A. G. Wiersma, Donald Kennedy and Donald M. Wilson has shown the importance of central, built-in mechanisms for the generation of motor patterns.

In vertebrates, however, and particularly in mammals, the contribution of centrally programmed patterns of motor command and their possible modification by feedback coming from peripheral sensors is still a matter of speculation. Hence neurophysiologists are still faced with the vexing question: What is the role of peripheral feedback? Does it provide a signal to time the release of the next phase of a movement? Does it facilitate some aspect of the motor pattern? Does it maintain or complete the next phase? Or does it add a new phase to those movements that were centrally initiated?

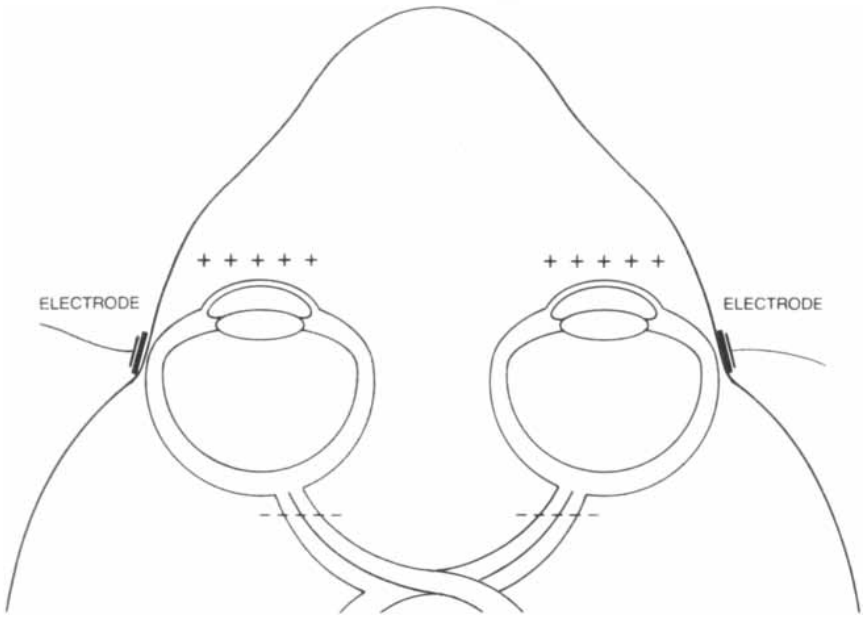
In an attempt to answer such questions we have been studying the coordination of eye-head movements in monkeys in our laboratory at the Massachusetts Institute of Technology. My colleagues in this effort have been Johannes Dichgans, Ronald E. Kalil, Piero Morasso and Vincenzo Tagliasco. We have investigated the spatial and temporal characteristics of the motor programs underlying the orderly sequence of eye and head movements in monkeys, and we have shown how reflex sensory feedback, generated by the turning of the head, interacts with the centrally initiated programs and thereby gives rise to "coordinated" eye-head movements.

Consider the eye and head movements that follow the unexpected presentation of a visual target. In man and monkeys the appearance of a target in the visual field is usually followed by an orderly sequence of such movements. First, a fast eye movement (called a saccade) carries the most sensitive part of the retina, the fovea, to the image of the target. Second, after a delay of between 20 and 40 milliseconds the head turns in the same direction. Since the eyes have moved first, and with a higher velocity than the head, their lines of sight reach and fixate the target while the head is still moving. Then for the duration of the head movement the eyes maintain their target fixation by performing a rotational movement that, by being counter to the movement of the head, allows the fovea to remain constantly on the target it has just acquired. This maneuver is termed compensatory eye movement [*see bottom illustration on page 102*].

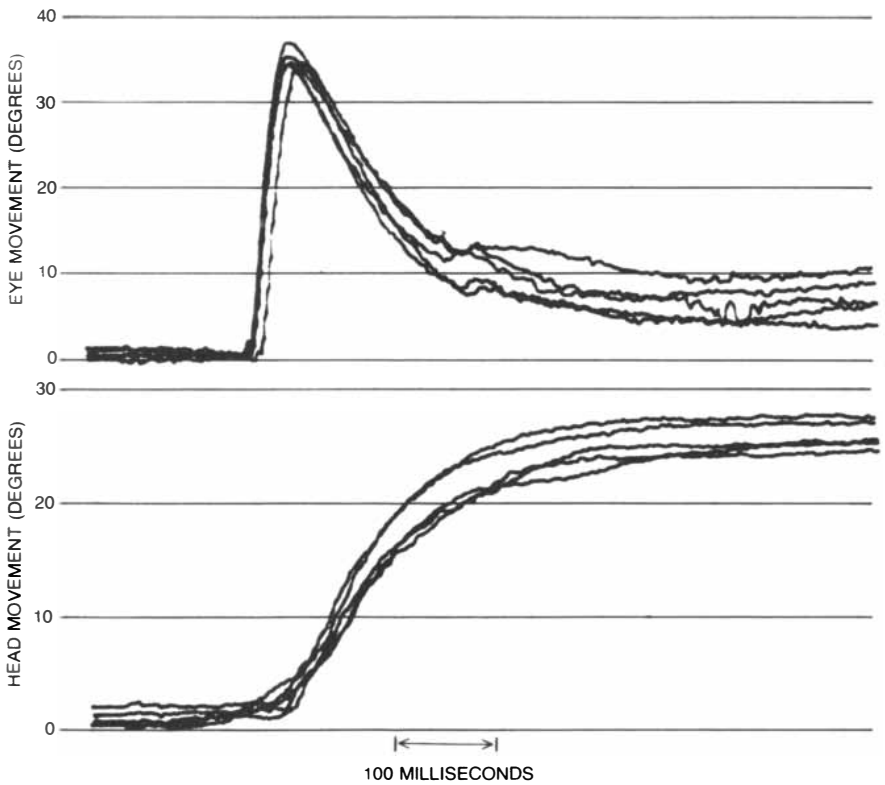
To achieve this orderly sequence of movements, that is, to direct the eyes and the head toward the target and ultimately fixate the target with the fovea, the subject must make a number of computations. To begin with, the subject must compute the angular distance between the initial lines of sight and the position of the target that is to be acquired. Although we do not have a clear understanding of how the angular distance is computed by the cortical and subcortical visual areas of the brain, we do know that a signal corresponding to



**EYE-HEAD COORDINATION STUDIES** are conducted in the author's laboratory at the Massachusetts Institute of Technology with the aid of the experimental apparatus illustrated here. When a visual target (*not shown*) is suddenly presented, it attracts the monkey's attention and sets in train an orderly sequence of eye and head movements. A fast eye movement, called a saccade, turns the eyes so that the target is centered on the fovea, the most sensitive part of the retina. Some 20 to 40 milliseconds later the monkey's head begins turning in the same direction. During this head movement the eyes must counterrotate to keep the target fixated. The amount of head-turning is transmitted to a recorder by the lightweight apparatus clamped to monkey's head. Eye movement is recorded by electro-oculography (*see top illustration on next page*). Tube in monkey's mouth gives drops of water as a reward.




**MEASUREMENT OF EYE MOVEMENT by electro-oculography takes advantage of the fact that the retina is positively charged and the choroid that lies behind it is negatively charged, even under resting conditions. The charges are kept separate by the external limiting membrane (the membrane of Bruch) that lies between the retina and the choroid. As a result the resting eye acts as a dipole in which the cornea is some 10 to 30 microvolts positive with respect to the back of the eye. Eye position can be recorded by placing a pair of electrodes on the inner and outer corner of each eye (for horizontal movements) or above and below each eye (for vertical movements). When the eyes are aimed straight ahead, the potentials between similarly placed pairs of electrodes are balanced. Any rotation of the eye, however, brings the more positive cornea closer to one electrode of each pair than to the other. Resulting difference in potential serves as an index of eye position.**




**COORDINATED EYE-HEAD RESPONSE to the sudden appearance of a target is shown in five superposed tracings. The vertical scales indicate the amplitude of each movement in degrees starting from the straight-ahead position. The eyes make an initial saccade of some 30 degrees, then counterrotate as the head begins turning toward the target. As a result when the head is fully turned, the eyes are once again pointed nearly straight ahead.**

that distance is translated into both the oculomotor system and the head-motor system at about the same time. In fact, the electrical activity recorded from eye-muscle and neck-muscle fibers shows that motor commands are delivered almost synchronously to those muscles. From this it follows that both the oculomotor and the head-motor control systems must be making use of the same angular-distance information at approximately the same time. As a result amplitudes of eye and head movements are produced that are well correlated with the angular distance of the target.

We have shown that this finding is valid in the case where a head movement begins from the straight-ahead position and the eyes are centered in the head at the time the target is presented. Usually, however, the eyes will not be centered when a target appears. How, then, does the head-movement control system know the position of the eye in its orbit so that a proper degree of head-turning can be generated? The head-movement control system must have access to eye-position information and then combine it with the angular-distance signal to achieve a coordinated eye-head movement. Information about the position of the eye can be supplied by sensors sensitive to stretching (muscle spindles) located in the eye muscles or by internal monitoring of the oculomotor positional commands.

The activation of eye and neck muscles leads not only to movements of the eyes and head but also to the activation of a number of sensory receptors, technically referred to as proprioceptors. These include neck-muscle spindles, neck-tendon organs, receptors located in the joints of vertebrae in the neck and receptors located in the vestibule of the inner ear. All will give rise directly or indirectly to nerve impulses whenever the head is turned. The question, then, is whether or not these sensory receptors can modify the ongoing eye and head motor programs. The following observations will make the point clear. If we compare the amplitude of the eye movement in the acquisition of a target when a monkey's head is allowed to turn and when it is held fixed, we observe that the amplitude of the eye movement is greater when the head is held fixed [*see illustration on opposite page*]. Clearly the amplitude of eye movement must be less when the head is free to turn if an overshoot of the target by the combined eye-head movement is to be prevented.

Given this experimental observation,

we must next ask what mechanism is responsible for the modulation of saccadic amplitude. Two hypotheses can be considered: Either the program responsible for fast eye saccades is modified every time the head is ready to move or reflex activities initiated by the head-turning itself apply a corrective action to the ongoing centrally programmed saccade. Our experimental evidence supports the second hypothesis. We find that afferent, or incoming, signals arising from the sensitive receptors in the vestibule of the inner ear are solely responsible for modulating the saccadic amplitude. The vestibular impulses responsible for modulation of eye movements originate when the cupula (a gelatinous cup into which hair cells protrude) in the semicircular canal is deflected by movement of the endolymph (the fluid in the canal). These impulses are first transmitted to nearby clusters of nerve cells, the vestibular nuclei. From there, by way of direct and polysynaptic pathways, they impinge on nerve cells of the oculomotor nuclei. Positive evidence of the crucial role of the vestibular afferent signals was demonstrated in monkeys by surgically interrupting the pathway linking the vestibular receptors to the vestibular nuclei. For several weeks after the operation (before the monkey had learned to compensate) the saccade amplitude during head-turning was identical with the saccade amplitude in the absence of head movement. This resulted in a remarkable overshooting of the target because the unmodulated eye movement was simply added to the head movement [*see middle illustration at top of next page*].

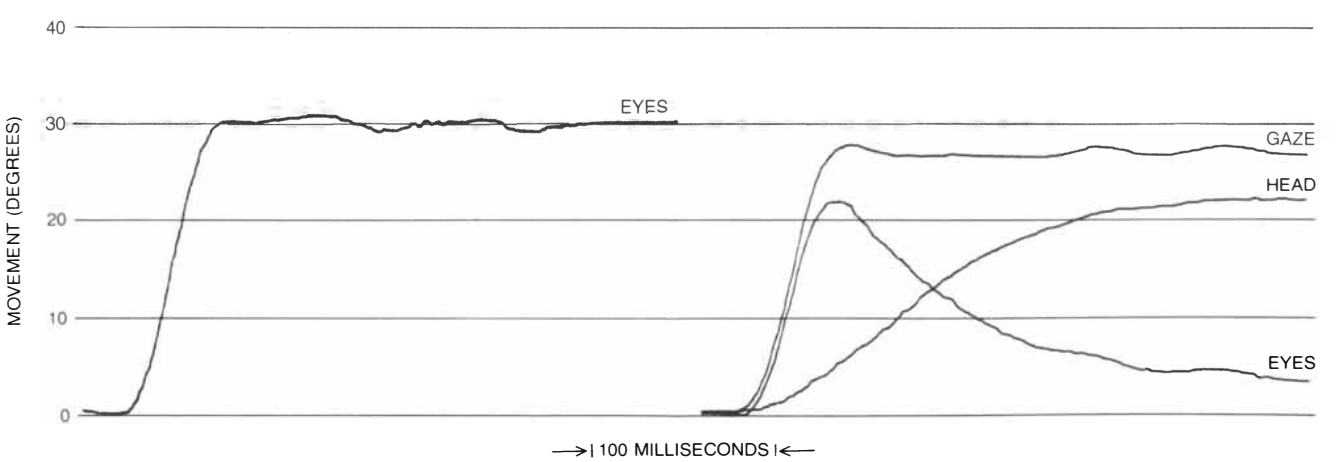
Having established that saccades are adjusted in scale by the vestibular activity initiated by the head movement, let us turn to the question of why this reflex-adaptive arrangement is more advantageous to the animal than one based on a centrally preprogrammed modification of saccadic parameters. Clearly the reflex mode of organization greatly simplifies the task of the motor-programming systems required for eye-head coordination. The eye and head movements can be programmed independently, since the vestibular system "automatically" nullifies any displacement of the fovea from the target as a result of head movement. Furthermore, by relying on vestibular reflexes that monitor the actual movement of the head, the resulting adjustment of eye movements will be able to compensate for all the unpredictable peripheral loads and resistances that might change the course of the centrally initiated (intended) head movement.

The modification of saccade characteristics is one aspect of the interaction of central programming and reflex activities. Although this interaction plays a decisive part in the process of target acquisition by a combined eye-head movement, the role of feedback from peripheral sensory organs (vestibular and neck afferents) extends beyond saccadic modulation to control and generate compensatory eye movements.

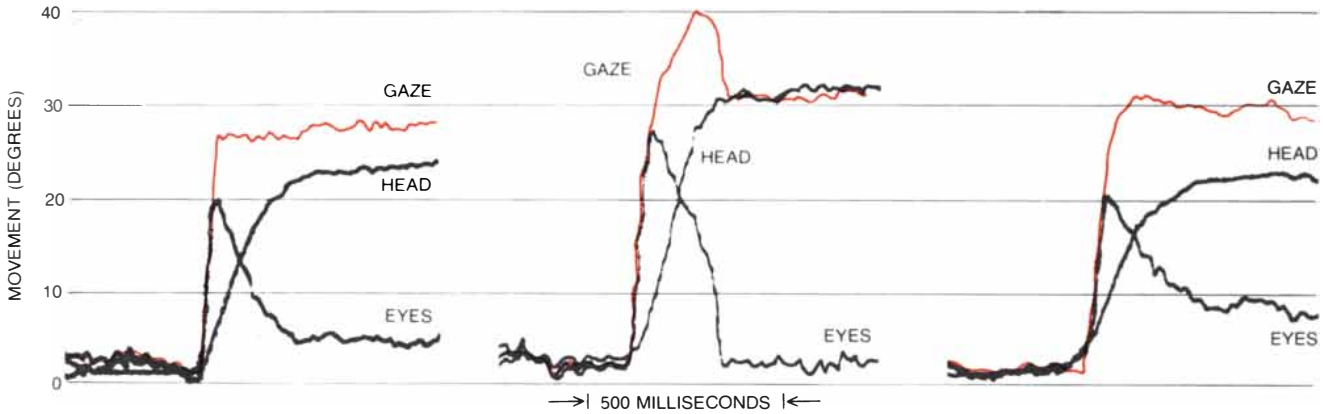
Since the eyes move first and with a higher velocity than the head, their lines of sight reach and fixate the target while the head is still moving. Then for the duration of the head movement the eyes maintain their fixation by executing a compensatory rotational movement that is counter to the movement of the head and compensates for it.

One can easily observe such eye movements induced by head movements by asking a friend to fixate a visual target and turn his head from side to side while maintaining fixation. You will see that, as his head moves, his eyes rotate in an equal and opposite way so that they remain on target. Such compensatory eye movements can be extremely rapid. Try fixating your index finger while rapidly turning your head, and then move your index finger with equal rapidity while keeping your head motionless. In the first case the vestibular system quickly moves the eyes in the direction opposite to that of head movement and the target is seen clearly, but in the second case the slower visual corrective loop to the eye muscles cannot compensate for the movement of the target. The target is consequently reduced to a blur as the eyes try in vain to match its velocity.

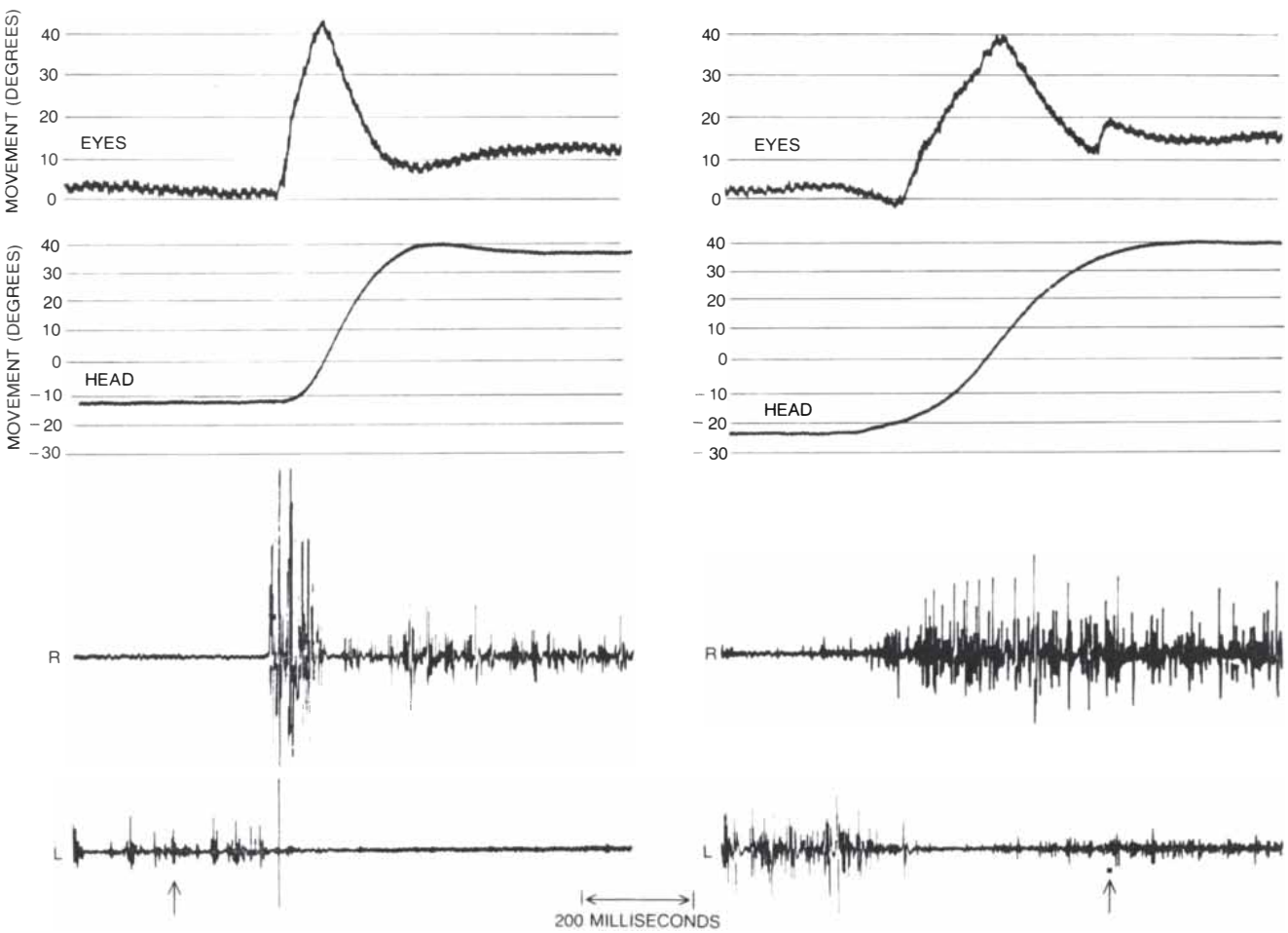
Compensatory eye movements have been studied by several investigators. Although it is generally agreed that such eye movements are strongly influenced by visual and vestibular reflexes, including feedback signals from various proprioceptors, it has been hypothesized that compensatory eye movements are initiated centrally and hence are not primarily dependent on feedback information. In our own recent work we have found no evidence for central initiation. We have been able to demonstrate, however, that compensatory eye movements result from the reflex action of the vestibular system. As a consequence of the head movement vestibular receptors are stimulated, and their activity induces a compensatory eye movement that enables the fovea to remain fixed in relation




**HEAD-FIXED AND HEAD-FREE RESPONSES to a suddenly appearing target are compared in two sets of traces. When the monkey's head is held fixed, the eye saccade must do the entire job of fixating the target (*left*). When the head is free to turn (*right*), fixation is achieved by a combined movement, labeled "gaze," that represents the sum of the eye saccade and the head movement. Notice that the initial saccade is smaller in the second case because the turning of the head accomplishes part of the shift in the gaze.**




**ROLE OF SENSORS IN INNER EAR** in the coordination of eye-head movements was demonstrated in the author's laboratory by studies in which the vestibule of the inner ear of monkeys was surgically inactivated. The set of three traces at the left shows how a normal monkey shifts its gaze to a target by combining eye and head movements. Motion receptors in the inner ear initiate feedback signals that modulate the amplitude of saccadic eye movements and help the eyes to compensate, that is, counterrotate, smoothly as the head turns toward the target. The traces in the middle show what happens when a monkey is presented with a similar visual target 40 days after surgical removal of its vestibular apparatus. Since the initial eye saccade is uncorrected as the monkey begins turning its head, the monkey's gaze strongly overshoots the target. Even in this short period, however, some degree of compensatory eye movement has already developed. Traces at the right were made 120 days after surgery. Now, even without the help of vestibular feedback, saccadic eye movements have been "recalibrated" through experience, so that the gaze no longer overshoots.




**DIFFERENT STRATEGIES** of eye-head coordination appear in different behavioral contexts. Whenever a visual stimulus is suddenly presented to a monkey (*left*), the animal first turns its eyes, then its head, in the direction of the target, which in this case appears 30 degrees to the right at eye level. The arrow indicates the time of target presentation. Trace *R* shows the burst of impulses that signals activation of the neck muscle (the splenius capitis) that turns the head to the right. Impulses in the antagonist muscles (*L*) simultaneously cease. A monkey uses a different strategy in the coordination of eye-head movements, however, when it turns toward a given position in anticipation of the appearance of a similar target (*right*). The author calls this mode "predictive." Now the head begins to move before the target appears and before the start of the eye saccade. Moreover, the head movement is instituted by a gradual increase in the activity of the agonist neck muscles (*R*) and a decrease in the activity of the antagonist muscles (*L*).

to a point in visual space while the head is rotating.

Summing up, we are now in a position to outline a realistic scheme for how movements of the eye and head are coordinated when a visual target is being acquired. The coordination involves a sequence of events in which feedback signals provide a closed loop, thereby making possible the correction of errors [*see illustration on next page*]. The sequence begins with the detection of a target somewhere in the visual field. Motor programs involving the head and the eyes are activated and respond by sending impulses to eye and neck muscles. This results in a saccadic eye movement and a head movement that activates vestibular receptors, which in turn generate a compensatory eye movement. The compensatory eye movement allows the fovea to remain fixed in relation to a point in visual space during head-turning. The fixation allows a second visual sampling, then a third and so on, with opportunities for correcting errors at each sampling.

If our hypothesized closed loop correctly describes the coordination of eye-head movements, it is clear that the role of the motor program stored in the central nervous system is simply to initiate, in an impulsive manner, movements of the eyes and head. Since there is no central programming of saccadic adjustment and of compensatory eye movement, it follows that the functional, or behavioral, coordination of head and eyes is the joint result of a central initiation (following a stored program) modified by the crucial intervention of modulating signals triggered by receptors in the vestibule of the inner ear. This conclusion somewhat simplifies our views of the neural mechanisms underlying motor coordination insofar as, contrary to common assumptions, we find no need to postulate a special central population of "executive" neurons with exclusive responsibility for coordinating the eyes and the head.

So far I have described the timing and the characteristics of the coordination of eye-head movements that are elicited by the appearance of a visual target, and have presented our evidence for the conclusion that the programs for eye-head coordination are not present in the central nervous system in their entirety. In what follows I shall take up two additional and interrelated topics. The first has to do with the fact that there are other modes or programs of the coordination of eye-head movements in addition to the one I have been discussing, and I shall indicate the relevance of distinguishing among such modes to the continuing research on the central control of movement. The second topic is related to the impressive capacity for functional reorganization that is displayed by eye-head motor programs following lesions in key structures such as the vestibular system.

Concerning the first topic, it is well known that in primates there are several other modes of the coordination of eye-head movements, such as those observed while the animal is tracking a moving object or scanning the visual environment. In our laboratory we have recently analyzed a particularly interesting mode of coordination that we call predictive because we observe it only after the animal has memorized a set of reward contingencies and is able to make an appropriate predictive movement that anticipates the presentation of a visual stimulus.

Under these conditions the timing of eye and head movements and the pattern of neck-muscle activation are different from those found in the visually triggered mode [*see bottom illustration on opposite page*]. In fact, during the predictive coordination of eye-head movements the head begins to move well before the saccade of the eye is initiated. In addition the head movement is achieved by a gradual increase in activity of the agonist muscles (the muscles that initiate turning), which is accompanied by a decrease in the activity of the antagonist muscles (the muscles that oppose turning). This pattern of reciprocal, gradual agonist-antagonist activity is in marked contrast to the bursts of muscle impulses invariably recorded from the agonist muscles when a monkey is presented with an unexpected visual stimulus.

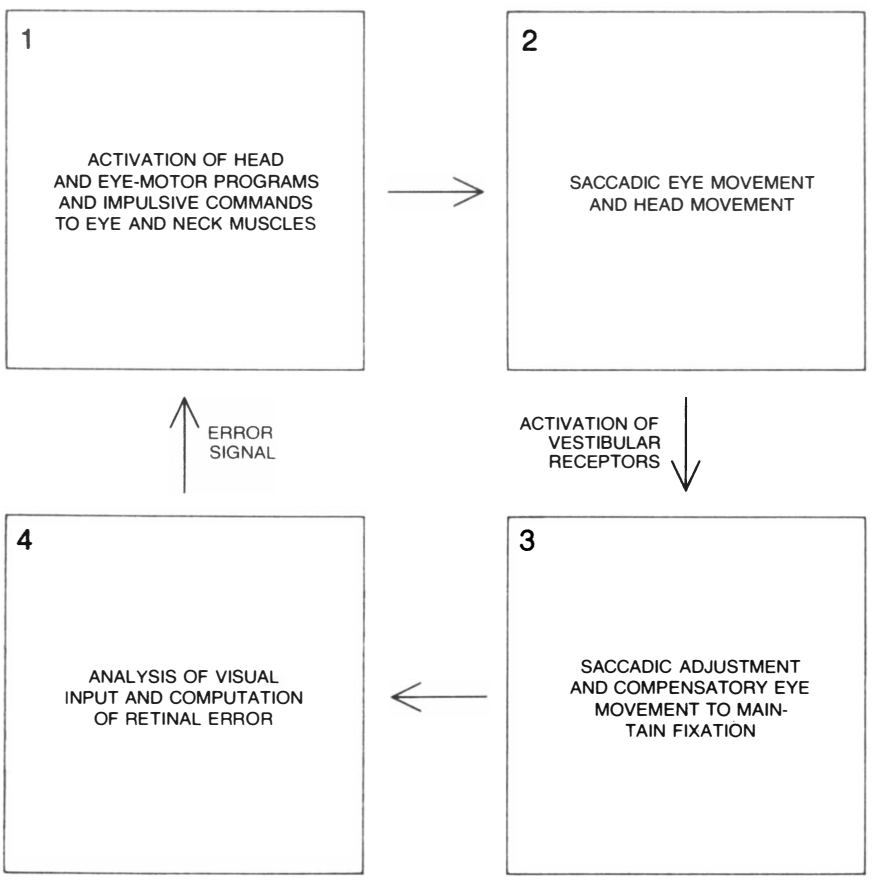
The fact that motor output subserving the coordination of eye-head movements is not fixed but instead exhibits distinctive patterns depending on the specific behavioral situation raises tantalizing questions. For instance, are there separate areas in the brain for the programming of different modes of coordination? If there are, are the various modes of coordination subserved by a totally different and separate neural network? How is the switching from one mode to another accomplished?

These are formidable questions for the neurophysiologist interested in outlining realistic models for the various kinds of motor coordination. Nevertheless, the investigation of such problems is no longer beyond reach. In fact, new

techniques of recording from a single cell in an intact trained animal such as those developed by Edward V. Evarts of the National Institute of Mental Health, combined with traditional neurophysiological and behavioral approaches, might enable us to unravel some aspects of the patterns of motor activity in the cortical and subcortical areas of the brain. I believe the possibility of bringing different strategies of motor coordination under control in the laboratory is relevant to the functional analysis of the motor system.

The notion of strategies of movement is in fact a powerful tool for interpreting single-cell recordings from cortical and subcortical motor areas. Since different strategies of movement are characterized by different spatial-temporal patterns of commands to the muscles, more compelling correlations between single-neuron activity and some aspects of motor behavior can be established. In addition the distinction between strategies of movement is an aid in the interpretation of selective brain lesions of those structures that are deemed important in the regulation of these movements.

Concerning the second topic, we were led to investigate plastic changes in the central organization of the eye-head motor system by our discovery of the decisive importance of the vestibular input in determining the nature of the coordination of eye-head movements. It was natural to wonder how that coordination would be affected if the vestibular input to the brain stem were eliminated. Are there other mechanisms that can take over the vestibular function? Our aim was not only to ascertain the degree of functional recovery of sensorymotor coordination but also to understand the mechanism underlying the recovery of the coordination of eye-head movements.




**CONTROL SCHEME FOR EYE-HEAD COORDINATION is believed to begin (*1*) with a central-nervous-system motor program that initiates the movement of the eyes and head when a visual target appears unexpectedly. The initial saccadic eye movement and head movement (*2*) are evidently based on a rapid computation of the target location carried out by cortical and subcortical regions of the brain in a manner not yet understood. As the head begins to turn, sensitive motion receptors in the vestibule of the inner ear initiate signals that lead to counterrotation of the eyes in order to maintain fixation of the target (*3*). During these coordinated eye-head movements the visual input is analyzed and an error computation is made (*4*). The result is an error signal that closes the control loop.**

Our results have shown that a variety of several basically different mechanisms are developed and jointly brought into play. Among them are the development of new eye programs, including one that provides compensatory eye movements. We have already seen that in the normal monkey compensatory eye movements are achieved by way of vestibular impulses. Two or more months after monkeys were surgically deprived of vestibular sensors, however, centrally programmed compensatory eye movements were found to contribute to ocular stability during active turning of the head. Thus the oculomotor system is capable of taking on, albeit in a crude and inadequate form, functions previously elicited by vestibular activity.

Another mechanism contributing to the remarkable recovery in the coordination of eye-head movements that occurs within the first two to three months following vestibulectomy entails a "recalibration" of saccadic eye movements with respect to visual input. As I have indicated, immediately after vestibulectomy the gaze of a monkey whose head is free to turn will badly overshoot a visual target because the saccade amplitude is no longer modulated by corrective feedback signals from the vestibular apparatus. In other words, the eye movement appropriate for fixating the target when the head is held fixed is simply added to the head movement.

After two to three months, however, vestibulectomized monkeys learn to make saccades that are smaller than normal as they turn their head in the direction of a visual target, thereby reducing the tendency of the gaze to overshoot its mark [*see illustration at right at top of page 104*]. On the other hand, if the head of a vestibulectomized monkey is restrained when it is unexpectedly presented with a visual target perhaps 30 degrees away from its straight-ahead axis of vision, its eyes will continue to make a saccade of 30 degrees, enabling it to fixate the target. Hence the oculomotor system is recalibrated selectively: when the head is restrained, the saccades are normal; when the head is free to turn, the saccades become smaller.

These two oculomotor functions are an important part of the mechanism underlying recovery of the coordination of eye-head movements. They provide a striking example of the remarkable plasticity of the central motor apparatus, a plasticity that comes into play whenever the organism is forced to compensate for a handicap or deficit imposed on it by events over which it has no control.

# DP SCIENCE DIALOG

Notes and observations from IBM which may prove of interest to the scientific community.

*Dr. Alec Broers examines an X-ray image-forming device with the new low-loss scanning electron microscope.*

## Exploring An Invisible Realm

With their incredible densities, computer components are rapidly approaching dimensions no longer visible with even the best of optical microscopes. Now, two IBM scientists have made a major advance in electron optics, a key to further technological progress. Dr. Oliver Wells and Dr. Alec Broers have developed a new method for operating the scanning electron microscope which may allow researchers to see surface details as small as 10 angstroms—or a few atoms wide—on many objects.

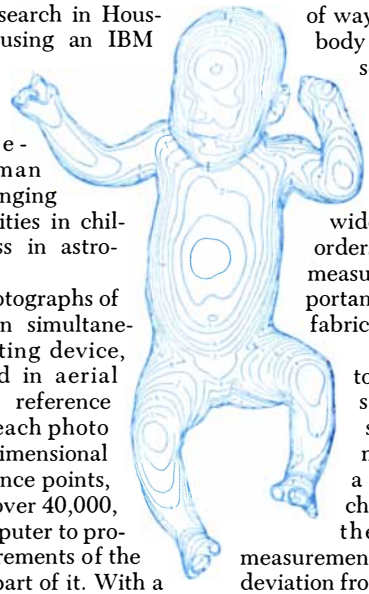
Among the samples being examined are specimens ranging from blood cells to minute electronic components to the experimental X-ray image-forming device pictured above, a development being pursued by Michael Hatzakis and Dr. David Sayre, also members of the Thomas J. Watson Research Center in Yorktown Heights, N.Y.

Whatever the eventual applications of their development, both Drs. Wells and Broers agree that, in areas from medicine to metallurgy, "This new low-loss electron method gives us a chance to examine the world around us, to discover what was perhaps unknown before."

In a conventional surface scanning microscope, now a standard research

*(Continued on Page 2)*

# New Aid to Medicine



At the Texas Institute for Rehabilitation and Research in Houston, scientists are using an IBM computer to produce highly accurate, three-dimensional measurements of the human body for studies ranging from spinal deformities in children to weight loss in astronauts.

Overlapping photographs of the body are taken simultaneously. Then a plotting device, like the type used in aerial mapping, identifies reference points common to each photo to create a three-dimensional image. These reference points, which can number over 40,000, are fed into the computer to produce precise measurements of the entire body or any part of it. With a computer driven plotter, such measurements can be shown in a number of ways—the contour map of the body pictured above, or cross sections of the body, or graphs showing how body volume is distributed from head to foot—to assist doctors in their diagnosis and treatment of a wide variety of medical disorders. Three-dimensional body measurements also provide important data for the automatic fabrication of artificial limbs.

Dr. R. E. Herron, director of the Institute's Biostereometrics Laboratory, says: "The new method makes it relatively easy for a physician to measure a child's growth pattern. Using the computer-generated measurements, he can determine any deviation from the normal range as part of a routine physical examination."

# The Future of Venice

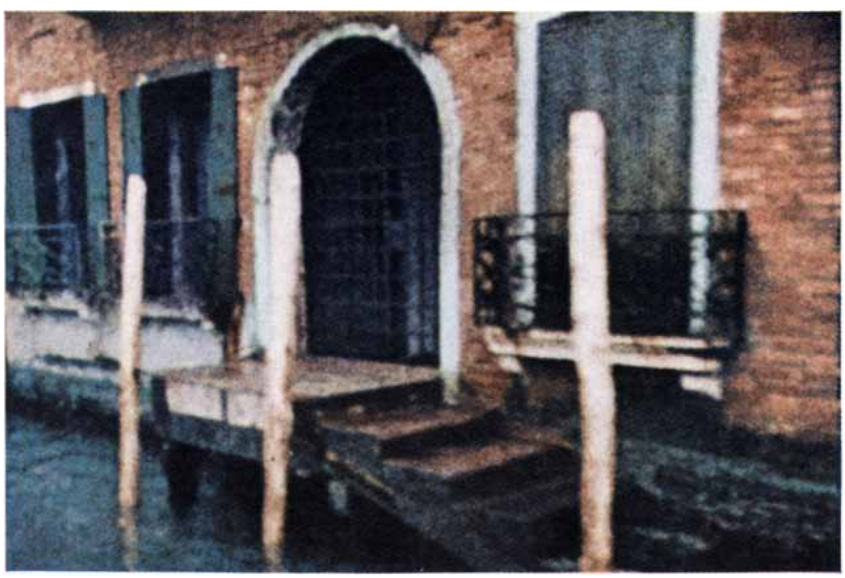
A study team with members drawn from the IBM Scientific Center in Venice and Italy's National Research Council is preoccupied—as are many people—with the future of Venice.

Aqua Alta, or high water, is a phenomenon that plagues the City of Canals. A product of storms in the Adriatic, Aqua Alta floods the Venetian lagoon many times each year. The problem is further complicated because the lagoon bed on which Venice rests is sinking year by year.

To study the complex interaction of wind and water and tides, a seagoing laboratory belonging to the National Research Council is collecting data the study partners can incorporate into a computer model of the problem. With a mathematical substitute for the real situation, team members can test different theories and hypotheses.

One paradox has already been resolved: Venice is sinking because water is disappearing beneath it—drawn off by the fresh-water wells which supply a nearby industrial area.

It is not likely that the sinking will soon stop, or that Aqua Alta will cease to be a threat. But the study is giving scientists a clear analysis of the forces at work on Venice.



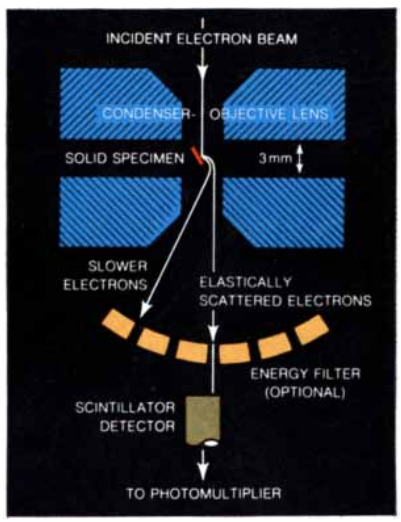
# Invisible Realm...

*(Continued from first page)*

tool, a beam of electrons scans a specimen. Secondary electrons are ejected from each point on the surface as the beam moves back and forth. At the same time, a spot is scanned across a cathode ray tube. The strength of the ejected electron signal controls the brightness of the CRT beam, resulting in a picture of the surface under study.

The microscope's resolution, or ability to clearly distinguish smaller features, is limited because the specimen must be focused as much as a centimeter from the electromagnetic lens to collect electrons efficiently.

In another type of standard microscope—called a transmission microscope




*Resolution is higher in the low-loss scanning microscope because a shorter focal-length lens can be used.*

—objects can be placed as close as a tenth of a centimeter to the lens. But because in this case the electrons pass right through the specimen to form the picture, only thinly-sliced samples can be examined.

Drs. Wells and Broers have combined the two techniques so that the surface of solid samples can be seen with extremely high resolution. While the sample is placed in the same position as in the transmission microscope, the key difference in their development is the use of low-loss electrons—original electrons from the beam that lose very little energy on striking a solid surface. Low-loss electrons which bounce off the sample surface are pulled back by the lower half of the lens and are collected to provide the output signal.

The new electron microscope has been in its present form since the Spring of 1973, although it had gone through several stages of development before that time. It is currently capable of useful magnification up to 300,000 times.

# A Sixty-Year-Old Forest Simulated in a Minute

How can you log a forest without causing soil erosion and dwarfism or destroying the atmosphere for campers, hikers and fishermen? Up until recently clear-cut answers have not always been available—not only because of the complexity of forest ecosystems but simply because trees do not grow fast enough for controlled experiments.

Now with the help of a computer simulator one can "grow" a two-and-a-half-acre portion of a forest at the rate of a year a second and immediately see the effects of a wide variety of simulated conditions. This development, according to one of the originators, allows research studies to be made which would ordinarily require centuries in an actual forest.

The project developed out of a cooperative effort between the Yale University School of Forestry and Environmental Sciences and IBM's Thomas J. Watson Research Center in Yorktown Heights, New York. Dr. Daniel B. Botkin, a Yale ecologist and two IBM researchers—Dr. James R. Wallis, a hydrologist and Dr. James F. Janak, a theoretical physicist—worked together on a mathematical model for forest growth to simulate environmental factors and various properties of each of the tree species in one ecosystem, so that hypotheses about the interactions could be made and tested.

Dr. Botkin and others collected the original data at the Hubbard Brook Ecosystem site in the White Mountains of New Hampshire, which contains 13 different species from sugar maple and white birch to mountain ash and red spruce. They then worked up a number of relatively simple equations to represent many of the interrelated conditions which affect the growth rate of a tree—soil quality, climate, topography of the plot and competition from other trees.

*Ecologists collected millions of measurements, everything from soil structure to water analysis, for original computer model.*

These key equations were included in the subroutine, Grow, along with two other subroutines—Birth and Kill. These took into consideration the annual growth increment for each tree, random planting of new saplings and killing off of other species to reflect the cumulative effect of weather, plant succession and competition.

While the original model deals with a limited number of variables, it has the capacity to handle an infinite number. Pollution, insect infestations—even the rate at which deer nibble young systems."

Dr. Wallis notes: "While the present simulator reflects conditions of a forest in New Hampshire, it is especially adaptable to many other ecosystems.

The computer model developed by the three scientists is the first of its kind. Other computer programs have simulated simpler patterns of growth and development—projecting, for example, productivity in a corn crop—but none before had been able to reproduce the major characteristics of a system as complicated as a multi-specied, mixed-age forest.

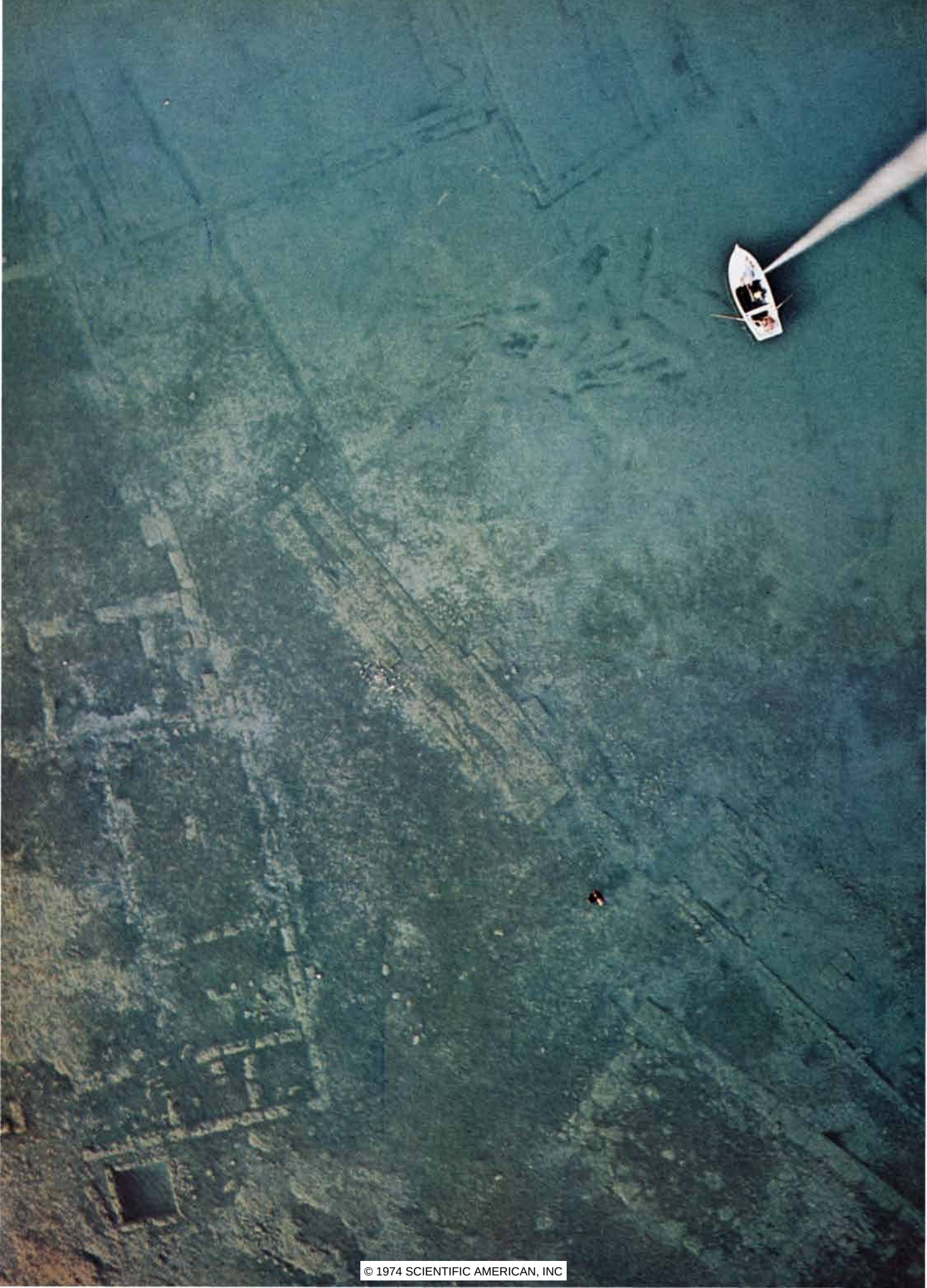
One of the most enthusiastic users of the forest simulator approach has been a multi-university research project on the West Coast funded by the federal government. Scientists there, along with researchers from Washington State's Department of Natural Resources, hope to adapt the forest model to the entire western region from Alaska to Southern California.

It is possible in the not-too-distant future that foresters will be using the computer on an everyday basis to plan reforestation of land swept by fire or clear cut by loggers. In that event the computer will be able to furnish such information as how many trees to plant, which species, and how often they should be thinned—all depending on the particular ecosystem of the region.

America's forests are one of our most valuable natural resources. With computer simulation, the best conditions for maximizing the growth of our forests can be determined more accurately.

*DP Science Dialog is concerned with topics which may prove of interest to the scientific community. Your comments and suggestions are welcome. Just write: Editor, DP Science Dialog, IBM Data Processing Division, 1133 Westchester Avenue, White Plains, N.Y. 10604.*

IBM®



# The Excavation of a Drowned Greek Temple

*Geological changes in the level of the Mediterranean shoreline have submerged a number of classical sites. The investigation of one of these shallow-water ruins suggests that they would all reward study*

by Michael H. Jameson

The introduction of self-contained underwater breathing apparatus some 20 years ago opened up the new domain of underwater archaeology. Now it seems that certain geological events of the past open up an extension of that domain into what can be called shallow-water archaeology. So far the principal beneficiaries of this kind of archaeology have been students of classical Greek and Roman times. The reason is that the sites that have been investigated are all in the Mediterranean area.

The shoreline of the Mediterranean has had its geological ups and downs over the past few thousand years. For example, the ruins of the port of Ephesus in Turkey today lie some miles inland from the coast. An even more dramatic example is the Temple of Serapis at ancient Puteoli, near Naples. The temple now stands, as it did when it was built, on dry land. High up on its stone columns, however, are numerous holes that were cut into the stone by burrowing marine animals; evidently the temple was once submerged for a considerable length of time.

My own acquaintance with shallow-water archaeology began at a site in Greece. At the beginning of the classical Greek period a small city-state arose near the tip of the Argolid peninsula in the Peloponnesus. The town grew up along the shoreline of a sheltered circular harbor that was smaller then than it is today. The settlement was called Halieis, or "the salty places," probably because nearby sea-level flats were used as salt pans. The population of Halieis probably never exceeded 4,000; the inhabitants presumably supplemented their agricultural resources, meager in this arid peninsula, by fishing, exporting salt and dealing in the valuable purple dye of antiquity that came from the marine snail *Murex,* which was plentiful in their waters.

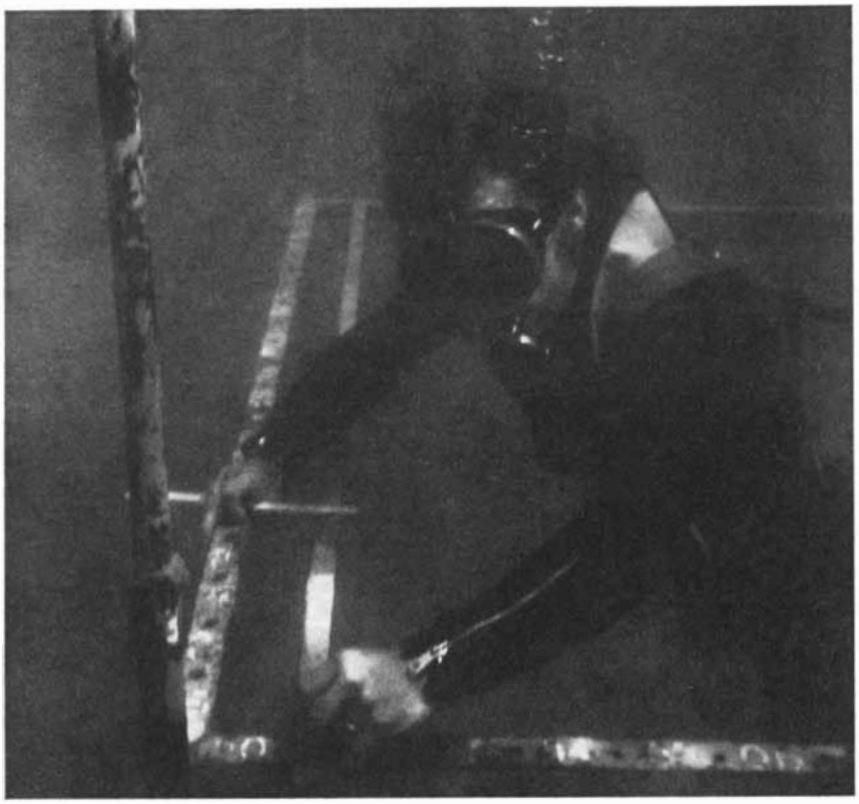
About 470 B.C. the town opened its gates to a company of refugees from Tiryns, a neighboring principality that had been overrun by forces from the larger city-state of Argos. The newcomers seem to have invigorated Halieis. For most of the next two centuries the port, protected by a system of walls and a hilltop citadel, served as a valuable naval base during a period that witnessed continual struggles among the larger Greek powers. Then, sometime after the death of Alexander the Great in 323 B.C., Halieis was mysteriously abandoned.

Beginning in 1962 my department at the University of Pennsylvania, the Department of Classical Studies, undertook with the support of the University Museum a conventional dry-land excavation of the ruins of Halieis, which lie in and around the modern town of Porto Cheli [*see bottom illustration on page 113*]. Tracing the line of the ancient city walls down to the water's edge, we soon found that a considerable part of Halieis, some four and a half acres in extent, was now submerged in the harbor. The depth of water over the ruins ranged from a few inches to more than 10 feet. As we were to learn later, another 40-acre area outside the city walls was similarly covered by the harbor shallows.

To a dry-land archaeologist a drowned site seems at first a total loss. Actually it has certain compensating advantages. For one thing, wood and other organic materials are better preserved under water than on land. Furthermore, Halieis, like most such cities, was built mostly of sun-dried mud bricks that were set on stone foundations. On land when such buildings fall into ruins, the mud brick eventually melts away, and as often as not the foundation stones themselves are later stolen to be used in other construction. At Halieis submersion had protected many of the foundation stones from thieves, and at the same time wave action had swept away such remnants of the mud-brick superstructures as had not already been broken down by rain. To clear the debris from a comparable area on land, using normal archaeological methods, would take years of work, whereas in the harbor, under the right conditions, entire sections of the city lay waiting to be revealed.

The question was how. The octopus fisherman who slowly sculls across the harbor at dawn knows how to find the large blocks of stone in whose crevices his quarry hides. The shallow harbor waters, however, are made murky by rich

**SANCTUARY OF APOLLO, built outside the walls of ancient Halieis, is seen in the aerial photograph on the opposite page. Constructed on what was at the time a harborside site, the sanctuary was later drowned by subsidence. Many of the foundation stones of sanctuary structures escaped later reuse because of their inaccessibility under water, making the original ground plan relatively easy to reconstruct. Out-of-focus object leading to the small boat is the tether for the balloon that suspended the remotely controlled camera above the site.**



**CONTROLLED EXCAVATION under water was achieved by placing a square frame, mounted on adjustable legs, over each intended trench. Here a diver is seen adjusting the horizontal orientation of a frame with a carpenter's level. The position of each two-meter-square frame was noted with respect to the overall site grid, and its depth below sea level was determined. As the excavation under each frame progressed the location of uncovered artifacts was recorded with respect to both horizontal coordinates and vertical position.**

**FOUNDATION PLANS were prepared by noting the size, shape and position of each of the foundation stones on waterproof paper. Here a diver, having measured a stone, sketches its position on a sheet that already shows the location of foundation stones surveyed earlier.**

organic growth and are easily rendered altogether opaque by wave action. Only occasionally does the combination of a flat calm and sunlight at the right angle provide a glimpse of the drowned streets and houses. Indeed, most of Porto Cheli's inhabitants had no suspicion of what lay just below the harbor surface.

In 1967, following an earlier season of snorkel surveying, our group at Porto Cheli was joined by a team from Indiana University. The work load was divided; we would be responsible for the shallow-water research and the Indiana team for the work ashore. Julian and Eunice Whittlesey provided the perspective we needed for our endeavors by suspending a camera from a tethered hydrogen balloon, which a swimmer would guide into position in the early-morning hours when the harbor was at its calmest. The very first roll of film showed us a detail of the drowned city wall, the outline of one of a pair of towers, that we had not been able to make out because we had been too close to it in the shallow water. The Whittleseys' balloon-supported camera has provided similar insights for five seasons now.

Our shallow-water work in both 1967 and 1968 was concentrated on recovering the plan of the drowned part of Halieis by a combination of surveying and utilizing a high-pressure water jet to clear away the silt and sand that obscured many of the foundation stones. The motor, pump and hose for the water jet we kept aboard a chartered local fishing boat. The boat's owner, an octopus fisherman, took us across the harbor one day and showed us a large complex of drowned buildings that he had located in his fishing. The complex lay about half a mile from the main gate in the city wall, and its presence had not been suspected up to that time. The Whittleseys promptly made balloon photographs that gave us some notion of the extent of the new find. We left Porto Cheli determined to survey the complex in the near future and if possible to excavate it.

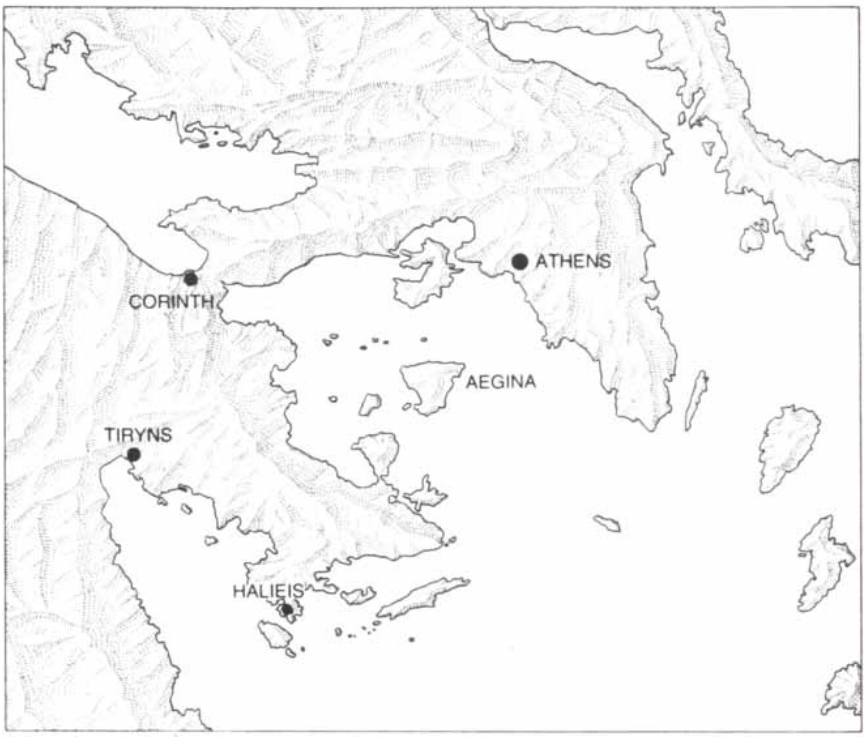
When we returned in 1970, we carried with us new means of clearing away bottom debris: light, portable dredges that had originally been developed for gold miners and that had already proved to be particularly useful in shallow-water work undertaken by archaeologists from the University of Chicago and Indiana University at Kenchreai on the Isthmus of Corinth. The pump for each dredge, which we mounted on a scaffold near the shore, sent a stream of seawater through a hose to the working head of the dredge. The head

was a galvanized iron pipe; the pumped water entered the pipe at a 45-degree angle, producing a powerful suction at the rear of the working head. This part of the apparatus was kept 100 feet or so away from the area under excavation in order to avoid having the discharge wash back and obscure the excavator's vision. Lengths of quick-clamping irrigation pipe led from the dredge head to the work area; the last length of the suction apparatus was made of corrugated rubber, thus allowing the excavator flexibility of movement. The diver could feed mud, small stones and any plant debris that he freed from the bottom into the rubber tube; the suction promptly deposited the waste materials a good distance away. As an added benefit the suction also continuously drew off the obscuring silt that clouded the water as a result of a diver's least motion, so that the area being excavated remained relatively clear.

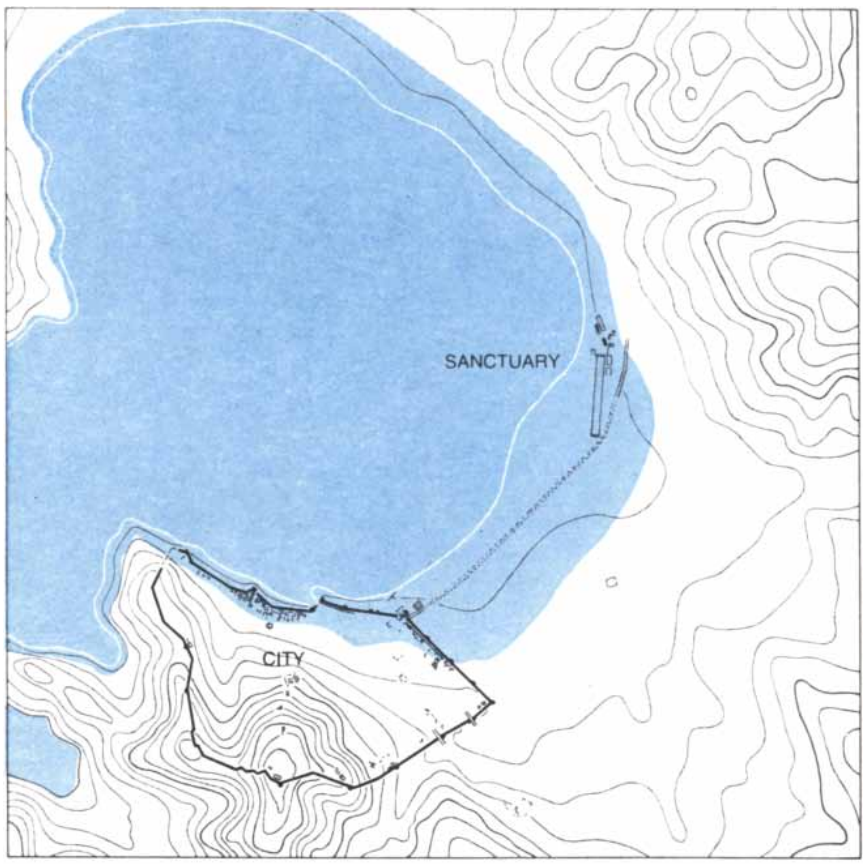
We began work at the complex by establishing over the target area a submerged grid of 10-meter squares that was tied into the Greek ordnance survey grid ashore. The intersections of the grid lines were marked by nails, each identified by colored flagging, driven into the harbor bottom. The grid provided a framework for the stone-by-stone plan of all the foundations in the complex and also fixed the location of the various trenches we intended to excavate in order to sample the stratigraphy of the site.

Stratigraphic excavation is, of course, the *sine qua non* of all archaeology. It requires the removal of successively deeper layers of earth and the assignment of the artifacts discovered in each layer to a general time interval. Often the excavator's downward progress is dictated by minor variations in soil texture and color that, for example, make possible the detection of intrusions from above, such as storage pits or the foundation trenches for later walls, which may cut into older strata. In shallow-water excavation all these considerations can be taken into account, although detection of soil differences is more difficult and allowance must be made for disturbances in levels caused by the incoming water. The uppermost level of mud and sand under water corresponds to the disturbed plow soil of a cultivated field on land and needs to be cleared before the significant levels in a trench are reached.

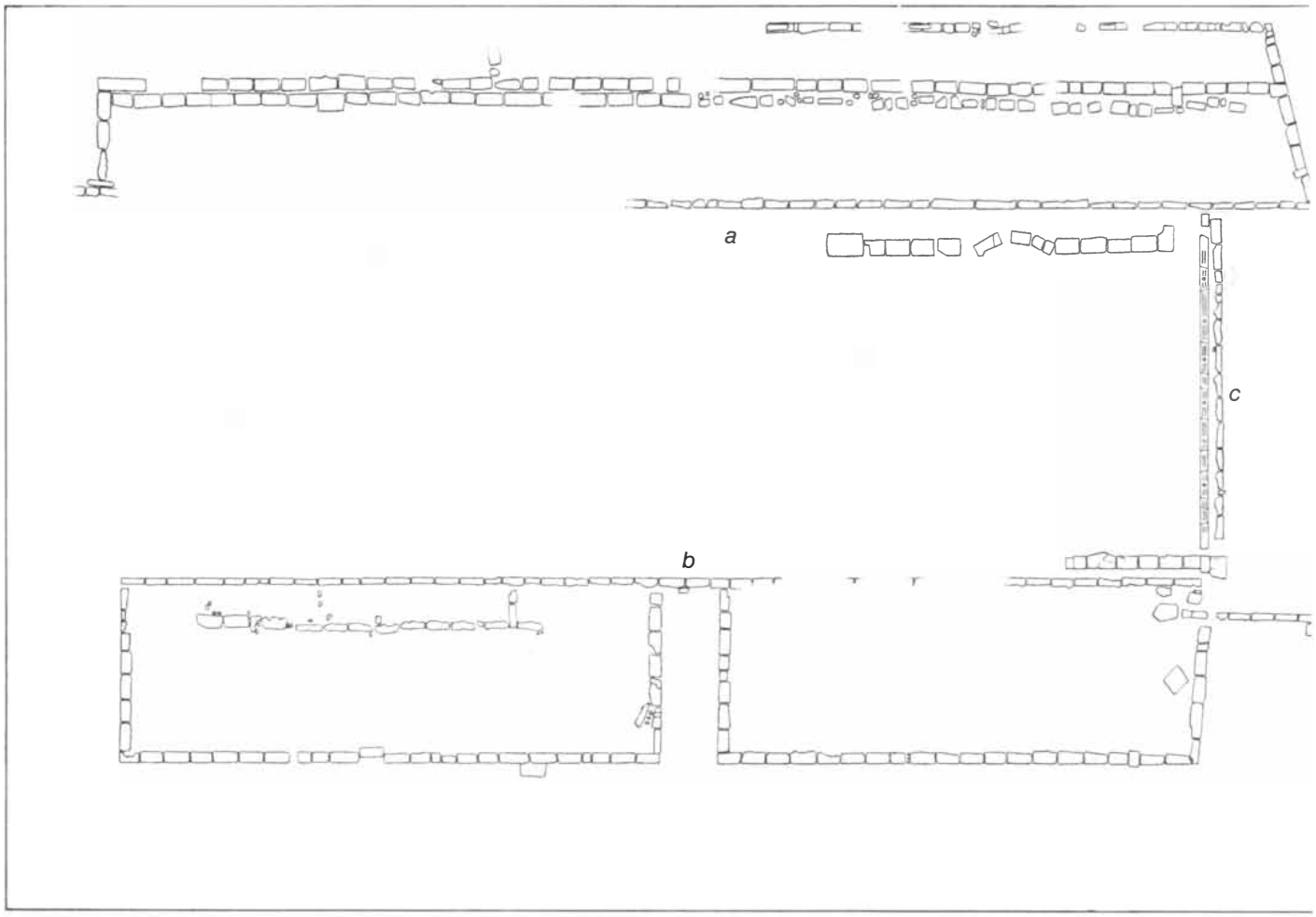
Over each proposed trench we fitted an iron frame two meters square, secured in position by a two-meter leg at each corner. The frame was free to ride up or down on the legs. The legs in turn had




**ANCIENT HALIEIS stood near the southern tip of the Argolid peninsula, a region of the Peloponnesus roughly equidistant between Athens and Corinth. Today its sheltered harbor is the site of the modern town of Porto Cheli, where few knew of the drowned ruins.**




**SUBMERGED PART of the harbor at Porto Cheli today that was dry land during the first millennium B.C. is the area between the innermost contour line (*white*) and the present shoreline (*color*). A considerable part of the city wall and the settlement within it (*bottom*) was thus preserved, as were traces of the road from city to sanctuary and much of the sanctuary complex (*right of center*). A plan of the sanctuary appears on the next two pages.**



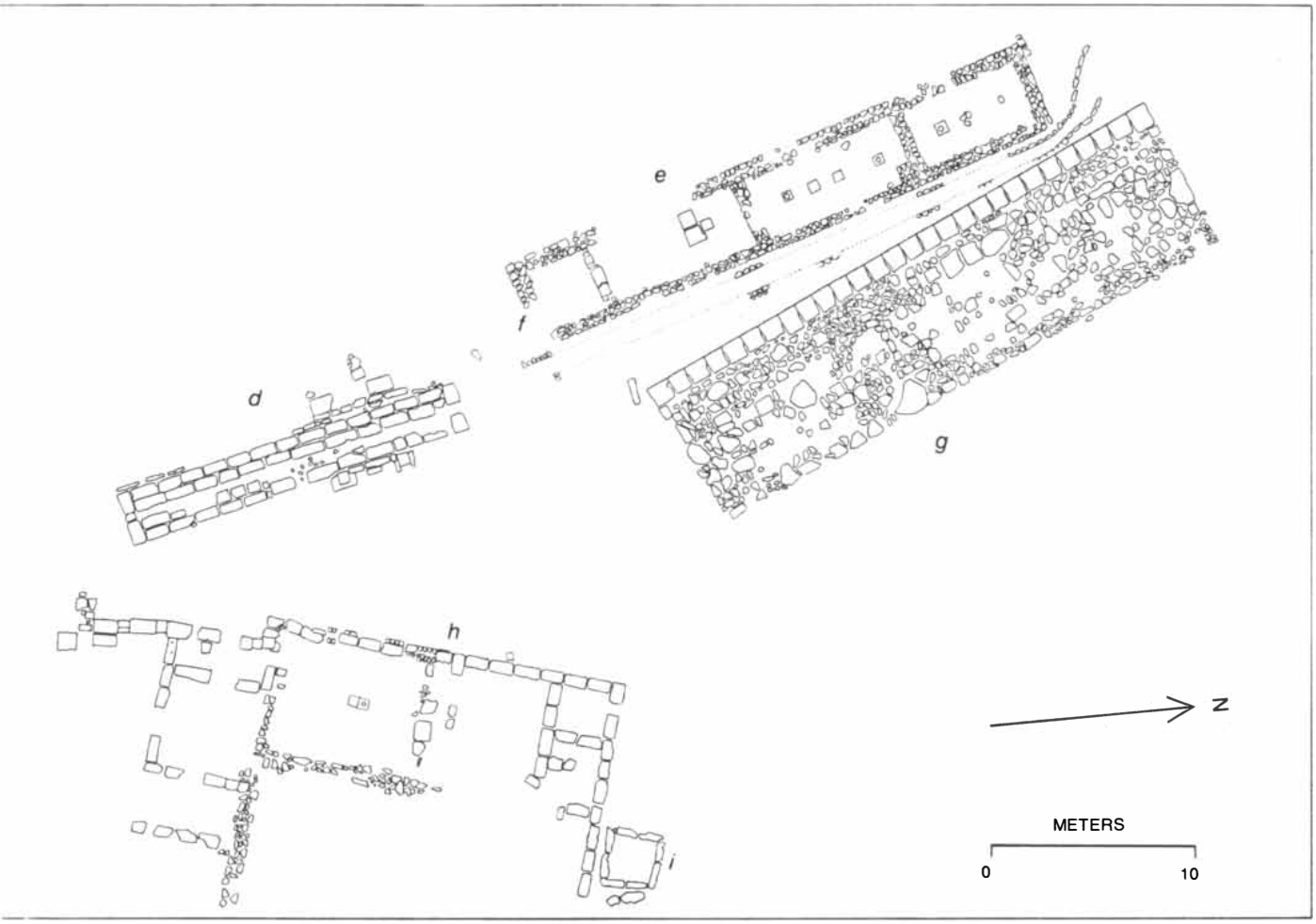

**PRINCIPAL COMPONENTS of the sanctuary of Apollo are (*from left*) "grandstand" foundations (*a*, *b*) to the east and west of the near end of a footrace track; one track starting line (*c*); foundations of an open-air altar (*d*) that stood between the starting line and the front of an early three-chambered temple of Apollo (*e*), with an entrance porch (*f*); massive foundations of a later, more**

been set up in a known position with respect to the overall grid system. The frame was made horizontal with the help of a carpenter's level, and its depth below the surface, which we called its negative elevation, was determined. This required placing the bottom end of a stadium rod on the leveled frame and then taking a reading on the part of the rod that projected above the water, using a transit set up at a known elevation on the shore.

The frame method and the excavation procedures associated with it, mainly devised by James A. Dengate, were ingenious adaptations of techniques that had been developed under other circumstances. Each diver worked from one to two hours at a time, no more than twice a day. Floating at neutral buoyancy under the frame, he would dig up the bottom with a small hand pick, a knife or his bare hands. For the most delicate work, where the excavator ashore might use a small trowel or a brush, the underwater excavator moves his hand back and forth to establish a light current. When an object is located, its position with respect to the frame's edges is measured and recorded (on a waterproof form that bears a scale drawing of the frame) and its vertical position is determined by dangling a plumb bob from a movable bar on the frame and measuring the distance from the bob to the bar with a meterstick. The object is then put in a plastic bag with a numbered tag, and the same number is recorded on the waterproof paper next to its indicated position. Large objects such as roofing tiles and rubble are not bagged but are put in tagged baskets for later examination ashore. Potentially perishable objects such as those made of wood or metal are freed of most of their salt content by immersion in a tank of fresh water ashore and then are subjected to various techniques of preservation. The diver types up his notes immediately on returning to shore, so that each three-diver team's series of notes provides a consecutive record of the excavation of the trench. One man from each team is responsible for writing a final report on each trench, for drawing sections of the four vertical trench faces and for analyzing the finds.

As we surveyed the site, sluicing away mud and sand from the foundation stones and digging test trenches here and there, the character of the structures that had once stood between the edge of the harbor and the ancient coast road leading to the main gate of Halieis gradually became clear. They had formed a sanctuary outside the city: a large open-air altar for animal sacrifices, a temple to shelter the deity in whose honor the sacrifices were made, various auxiliary buildings and, quite unexpectedly, a runners' racecourse that must have been built for the quasi-religious athletic games so beloved of the classical Greeks.

Indirect evidence suggested that the sanctuary's deity was Apollo, the prophet-musician son of Zeus who was widely worshiped throughout Greece. He is known to have been the chief god of Halieis, and his image often appeared on the city's coins. Furthermore, in a treaty between Halieis and Athens, in-




**ambitious building (*g*) that probably replaced the early temple; foundations of a third building (*h*) that may have housed competitors in the games at the racecourse or served as dining quarters for those attending religious festivals, and a well (*i*) that held many drinking cups. The southern starting line of the racecourse was only 167 meters away, making it one of the shortest in Greece.**

scribed on stone in 423 B.C., the text provides that another copy of the treaty be erected in the sanctuary of Apollo at Halieis, so that we know he had a temple there. A marble statue found in fragments among the temple ruins points to the same conclusion. A discovery made during our second season resolved the question beyond further doubt. The divers recovered the remains of three long pieces of angled iron, welded together by corrosion [*see top illustration on page 118*]. We identified them as a set of three temple keys; in Greek art such keys are often shown in the hands or on the shoulder of priests. In spite of the corrosion it was just possible to make out, inscribed at the top of one key, the name Apollo.

Today, after our third season of excavation, the history of the Halieis sanctuary can be reconstructed in considerable detail. The harborside settlement that was to become a city-state started to expand during the eighth century B.C., just as Greece in general was reawakening after its dark ages. It was at this time that the first two elements in the future complex were built. A narrow, south-facing temple that measured 27 meters from end to end but only 4.46 meters from side to side was raised adjacent to a flat stretch of shore suitable for footraces and other sports. The outdoor altar was placed between the front of the temple and the racecourse. No doubt some primitive wood statue of Apollo looked out through the entrance of the temple, overseeing both the altar and the racecourse beyond it.

At later times other buildings were added to the sanctuary, farther to the east and facing the altar. A formal entranceway was also built, leading from the coast road to the open space between the altar and the racecourse. Although added buildings are regularly encountered at most Greek sanctuaries, their architecture is rarely specialized enough to provide a clue to their function. Here the additional buildings near Apollo's temple contained a series of small rooms. This suggests that the buildings provided facilities for dining in connection with religious festivals or were used to house competitors in the games. The remains of many drinking cups, found in a nearby well, would support either interpretation.

All the later structures to the east of the temple testify to the disadvantage of a site's not being sufficiently drowned. They were built on higher ground and were therefore more exposed than the other buildings as the shoreline sank. It must have been all too easy to bring in a boat and load it with the largest and best blocks of stone. In any event we can speak only tentatively about what once stood in that part of the complex. In this connection the original temple, which was the earliest structure at the site and was built of the simplest materials, is also the best preserved; because it was sunk the deepest it was the least pillaged. Certainly a part of the original temple survived into the fourth century B.C., when the city met its end.

We found it possible to re-create many details of the temple's original appearance by studying the debris of the col-

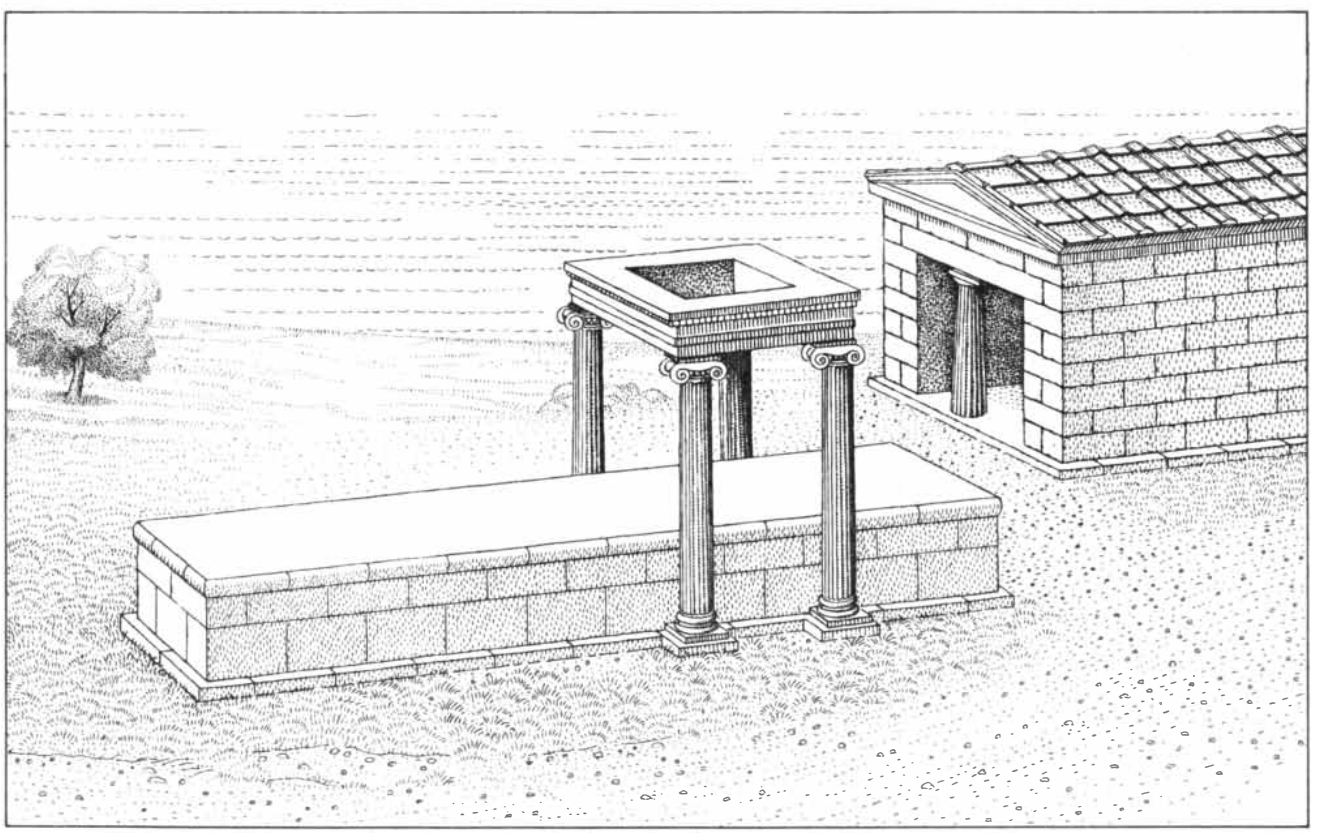
lapsed walls and roof. From the base upward the temple walls were built of two rows of limestone slabs; as many as four courses of the slabs are still standing in some places. The slabs were roughly finished on their outer face only. Judging by the heavy fall of masonry along the outer face of the walls their upper courses may also have been stacked stone slabs rather than the more usual mud brick. In any event solid single blocks of stone ran through the width of the walls at the four corners. The lowest course included a narrow projecting ledge, and on the inside of the building semicircular pieces of limestone were set at 1.5-meter intervals. These bases were evidently the supports for wood columns; the upper part of the columns must have lent additional strength to the upper part of the walls.

The temple floor was paved in places with irregularly shaped slabs of sandstone, but in the main it was composed of chunks of soft, clayey limestone that formed a hard-packed layer. Round slabs of stone stood along the main axis of the temple; these were evidently the bases for a central row of wood columns that helped to support the roof. Considerable support must have been required: the roof was made of heavy tiles. Many of these were found where they had fallen when the roof collapsed. Flat rectangular "pan tiles," 30 inches long, 20 inches wide and two inches thick, were arranged side by side; the joints between adjacent rows were concealed under narrow, angular "cover tiles" some six inches wide. Not all pan tiles and cover tiles are of the same size, shape and fabric, a fact that suggests roof repairs at intervals of many years. As a finishing touch the temple walls were plastered on the inside and perhaps on the outside too. On the inside, at least, the plaster bore painted designs; the remains of the designs, however, are too fragmentary to allow reconstruction.

The narrow temple building had a porch at the south end, facing the altar, and was divided into three chambers. The front chamber would have held a now vanished image of Apollo. The broken marble statue we found there evidently represents a much later replacement. Between the stone base for the statue and the rear wall of the first chamber is a space that seems to have served as a temple treasury. There we found quantities of iron spits, which would have been indispensable for roasting the meat of sacrificed animals. Spits also served as a primitive form of money that continued in circulation in nearby Sparta long after conventional coinage had been introduced elsewhere in Greece. In addition to the spits we found 18 silver coins of various denominations that had been minted between 550 and 525 B.C. on the island of Aegina (modern Aiyina), which lies between Halieis and Athens. The coinage of Aegina was widely circulated throughout Greece up to the time of the island's subjugation by Athens in the middle of the fifth century B.C.

Sunk into the floor of the temple treasury were two large masses of heavily corroded iron. An X-ray plate of one of the masses shows the outline of two axe-adze heads and one double-axe head. Like the iron spits, these tools would have had some monetary value, but another interpretation of their presence in the treasury is possible. One of my colleagues, an Orientalist, tells me that in the ancient Near East the tools that were used in the building of a temple were often thereafter dedicated to the resident deity; such may have been the case here.

Among other finds in the treasury area



**ALTAR AND TEMPLE** are shown in this reconstruction as they may have appeared in the fifth century B.C. The temple porch has been omitted. The temple walls were built of roughly dressed stone and were smoothly plastered on the inside. The roof, covered with heavy terra-cotta tiles, was supported by the walls and by pillars along the walls and along the central axis of the building. Animals were sacrificed to Apollo both on the plastered top of the outdoor altar and within the second of the three temple chambers.

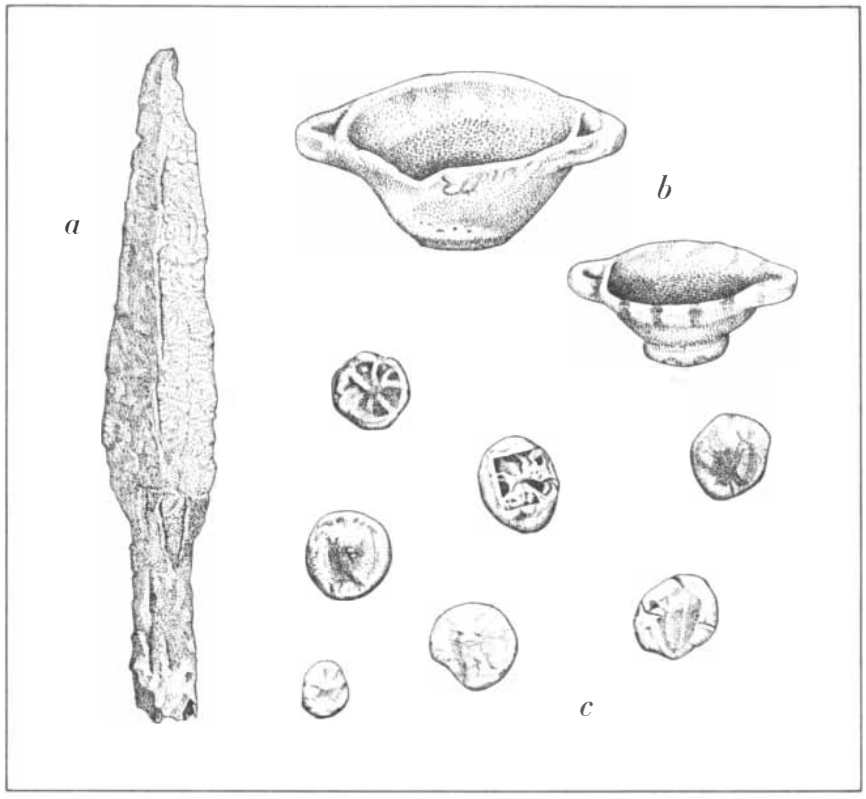
were some pieces of amber, evidently imported from the Baltic, bits of ostrich eggshell imported from Africa and an inscribed bronze plaque made early in the fifth century B.C. Most of the inscription is illegible, but the parts of the text that can be clearly read refer to the imposition of fines. Presumably the plaque contained a warning not to misuse the temple treasures.

The middle chamber of the temple may have served as an inner sanctum where oracular and purification rites were performed. The presence there of numerous horn cores from sheep or goats and the bones of young pigs is suggestive of such a ritual function. A shrine of Apollo on Crete and also one of Hermes there had altars made of piled-up horns; Apollo's altar of horns on Delos was known throughout classical Greece. Similarly, the sacrifice of piglets was a common practice, particularly in a purification context; that was how Orestes was purged of guilt in Apollo's celebrated shrine at Delphi after slaying his mother, Clytemnestra. Also found in the middle chamber were the heads and torsos of three bronze animal figurines: a goat, a ram and a bull. The bronzes were designed to fit as finials on the tips of wood shafts, but their function is open to speculation.

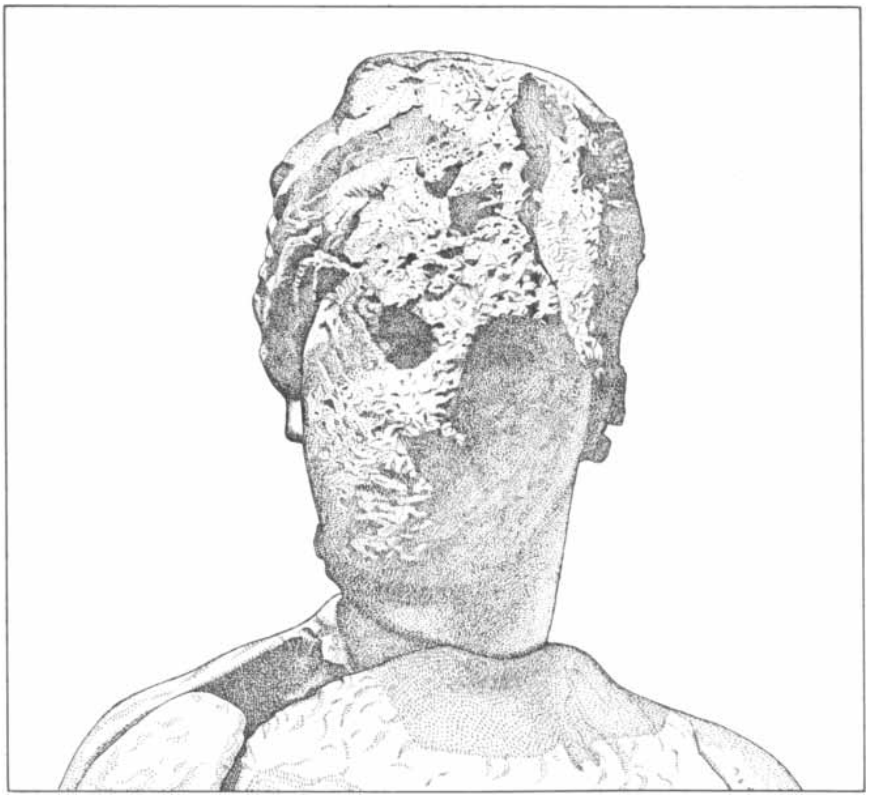
The temple's third and last room was evidently a storage area. There we found the remnants of at least 5,000 miniature pottery wine cups, a common offering of the poor. Throughout the temple we also found knives, swords and spearheads of iron. These evidently represented the offerings of people of greater means.

The arrival of the refugees from Tiryns early in the fifth century B.C. may have prompted certain temple renovations. At about this time the old circular bases for the temple's axial row of pillars were for the most part replaced with rectangular blocks of a finer stone and some of the new bases were topped with circular plinths. The columns they supported, however, continued to be made of wood; no stone column drums have been found in the ruins. The positions of the new pillar bases make it plain that not all was well with the temple roof. They are located irregularly, probably wherever the sagging roof most needed support.

Other changes perhaps inspired by the refugees from Tiryns include a new temple entrance, built from the porch into the forechamber and distinguished by a handsome stone threshold. It is tempting to see in a rectangular loop of iron found nearby the keyhole of a vanished wood entrance door into which the




**ARTIFACTS recovered from the early temple include weapons made of iron (*a*), probably the offerings of well-to-do worshipers, and miniature wine cups (*b*), a common temple offering of the poor. Silver coins (*c*) were found in the part of the first chamber that appears to have been a repository for temple treasure; they had been minted on the island of Aegina.**

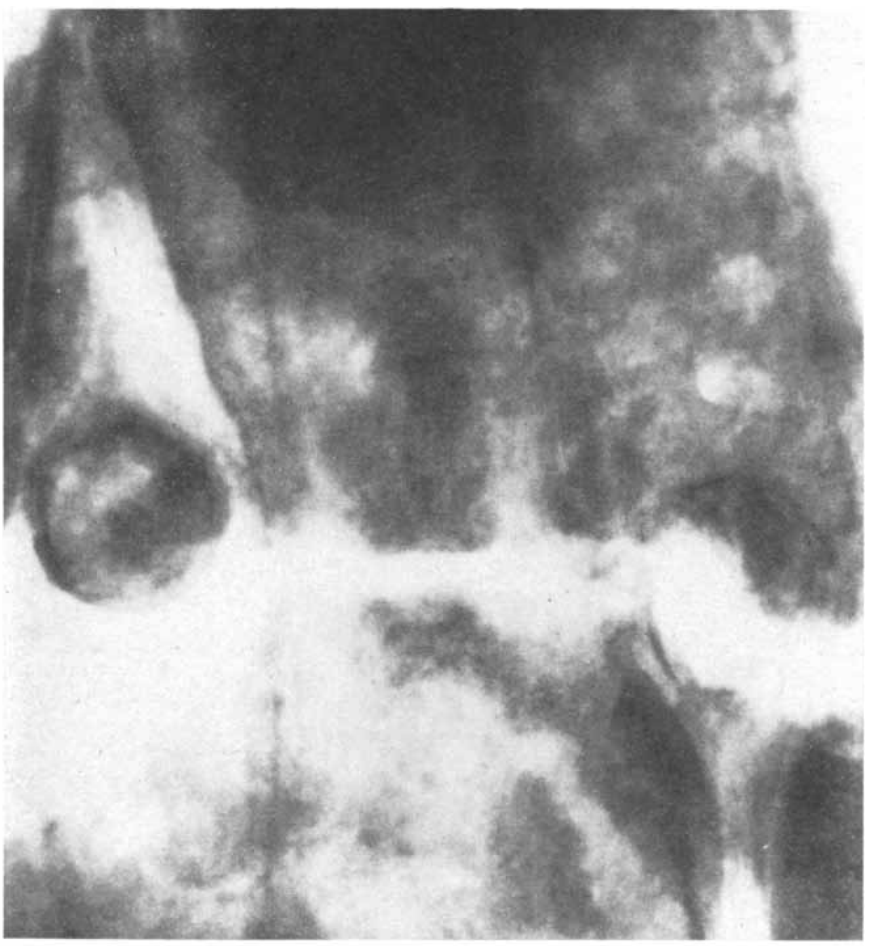


**STATUE OF APOLLO, found in fragments in the first chamber of the temple, imitates the sculptural style of the early fifth century B.C. but was actually made at some later date. It may have replaced an earlier statue, possibly destroyed in an Athenian attack in 460 B.C.**

CORRODED IRON BARS, curiously bent, are the remains of three temple keys. It proved possible in spite of the corrosion to read the name Apollo inscribed on one of the keys.

great iron temple keys fitted and to identify a second iron object as a door pull. That the people from Tiryns were involved in these various innovations is suggested by the fact that the letters on both the temple key and the bronze treasury plaque are written in the alphabet of the Argive plain, the Tirynthians' former home, and not in the alphabet in common use at this end of the Argolid peninsula.

When was the temple of Apollo built and how long did it stand? We were able to salvage some of the pine timbers used in the construction of the temple. After treatment for the effects of long immersion a sample of the wood provided a carbon-14 date (adjusted for correlation with tree-ring dates) of 780 ± 59 B.C. The date is indicative of the time the pine was felled. If we assume that the construction utilized newly cut timbers rather than reused ones, this range of dates suggests that the temple was built in the period from the latter part of the ninth century B.C. to the latter part of the eighth century. The more recent date seems the most plausible: the oldest pottery recovered from the ruins dates back to the late eighth century and early seventh. In the more advanced parts of Greece at that time cult buildings as primitive as the Halieis temple were being replaced by the first of the Doric temples, built of well-cut masonry and surrounded by a colonnade of pillars. In the quiet backwater of Halieis, Apollo's old shrine was to be preserved for centuries longer. Few objects found in the ruins, however, can be dated later than the middle of the fifth century B.C.



X-RAY PLATE of one large mass of corroded iron found with the temple treasure reveals two axe-adzes and one double axe. The tools may have been used in constructing the temple.

We know from recorded history that Athens made an unsuccessful assault on Halieis in 460 B.C., and it may be that the unprotected sanctuary by the shore was damaged at that time. Whether or not the Athenians desecrated Apollo's temple, it appears to have received little attention thereafter. At some later time a small building was constructed in what had been the temple entrance. Either before or after that (the chronology is still not clear) the marble statue we found was placed on a foundation of reused blocks in the outermost chamber and marble lustral basins were set on each side of it. The statue imitates early fifth-century sculpture in its style but it was clearly made at a later time. The impression one has is that the old temple was now a kind of antique, damaged but not destroyed, that was being preserved more as a sacred relic than as a functioning religious building. Eventually, as we know from the condition of the timbers we recovered, the temple was destroyed by fire.

Meanwhile the town of Halieis flourished, and considerable money and energy were invested elsewhere in the sanctuary. The outdoor altar in front of the old temple was rebuilt on a grander scale, and four pillars were erected there, probably to support a canopy near one end. The blocks of fine, hard limestone that composed the upper part of the rebuilt altar were almost all pillaged in later years for use elsewhere.

Immediately to the east of the old temple a larger and more ambitious building was constructed. Like the rebuilt altar, this modern structure was later robbed of much of its stone; today even its orientation is in doubt. I would guess that it was a new temple that faced the rebuilt altar in much the same way the old temple did, and that it was intended to fill the same function. Its foundation was massive: some of the stone blocks are three meters long. We assume that the building walls were made of mud brick because in the vicinity we recovered heavy baked bricks: 30 inches long, 20 inches wide and nearly four inches thick. These bricks were probably

used to make the top course of the wall on which the roof beams rested. The roof itself was tiled, but the tiles differ in design from those used in the old temple. Wide and curved, they are of the type known as Lakonian. The few artifacts we recovered from the ruins—wine cups and iron weapons—are some support for the interpretation that the structure was a temple.

The racecourse was also improved by the construction of a starting line at each end. The starting lines consisted of fitted stone blocks, one for each of the track's six lanes; parallel grooves indicated where the runners should place their feet and sockets allowed the mounting in each lane of a wood *hysplex*, or starting gate. The starting lines are 167 meters apart, which makes the Halieis stadium one of the smallest in classical Greece; a one-lap race would have covered little more than 180 yards, whereas the Olympic *stade* was more than 200 yards. The existence of two starting lines allowed a one-lap race to begin at the city end and finish near the sanctuary altar and a two-lap race to begin and end near the altar. Foundations were built for some kind of spectator accommodation on each side of the starting line at the altar end. We found the starting line at the city end of the stadium by probing along the east wall of the sanctuary until the wall ended; the line at the altar end was discovered by accident when the discharge from one of our dredges exposed one of the starting-line blocks.

The fate that befell Apollo's sanctuary remains as unexplained as the fate of Halieis itself. As far as we know the sea did not encroach on the sanctuary and the city until well after the beginning of the Christian era. Before then, of course, minor changes in the height of the land could have contaminated the local wells with salt water and forced abandonment of the area. It is perhaps more likely that some powerful warlord, perhaps Demetrius Poliorcetes, the "Besieger of Cities," deciding about 300 B.C. to eliminate a naval base that might be used against him, destroyed the fortifications and drove out the inhabitants.

Halieis was a modest town with a modest sanctuary. By an accident of geological change it has been possible to study there aspects of a classical Greek city-state that are seldom accessible at dry-land sites. In the course of the work, methods of reconnaissance and of excavation in shallow water have been advanced. Our experience suggests that many more rich opportunities for archaeology exist in the precarious zone between the sea and the land.

# MATHEMATICAL GAMES

## *On the paradoxical situations that arise from nontransitive relations*

by Martin Gardner

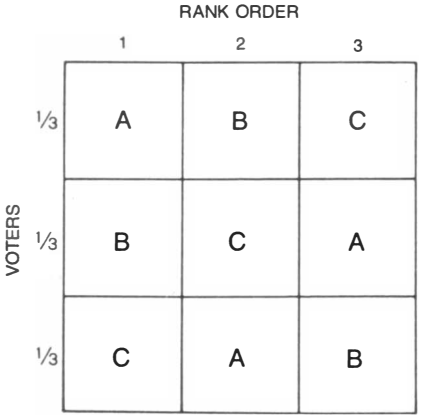
Whenever a relation $R$ that applies to $xRy$ and $yRz$ also applies to $xRz$, the relation is said to be transitive. For example, "less than" is transitive among all real numbers. If 2 is less than $\pi$ and the square root of 3 is less than 2, we can be certain that the square root of 3 is less than $\pi$. Equality also is transitive: if $a = b$ and $b = c$, then $a = c$. In everyday life such relations as "earlier than," "heavier than," "taller than," "inside of" and hundreds of others are transitive.

It is easy to think of relations that are not transitive. If $A$ is the father of $B$ and $B$ is the father of $C$, it is never true that $A$ is the father of $C$. If $A$ loves $B$ and $B$ loves $C$, it does not follow that $A$ loves $C$. Familiar games abound in transitive rules (if poker hand $A$ beats $B$ and $B$ beats $C$, then $A$ beats $C$), but some games have nontransitive (or intransitive) rules. Consider the children's game in which, on the count of three, one either makes a fist to symbolize "rock," extends two fingers for "scissors" or all fingers for "paper." Rock breaks scissors, scissors cut paper and paper covers rock. In this game the winning relation is intransitive.

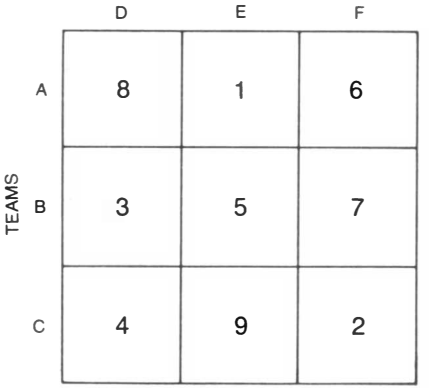
Occasionally in mathematics, particularly in probability theory and decision theory, one comes on a relation that one expects to be transitive but actually is not. If the nontransitivity is so counterintuitive as to boggle the mind, we have what is called a nontransitive paradox.

The oldest and best-known paradox of this type is a voting paradox sometimes called the Arrow paradox after Kenneth J. Arrow because of its crucial role in Arrow's "impossibility theorem," for which he shared a Nobel prize in economics in 1972. In *Social Choice and Individual Values* (1951) Arrow specified five conditions that almost everyone agrees are essential for any democracy in which social decisions are based on individual preferences expressed by voting. Arrow proved that the five conditions are logically inconsistent. It is not possible to devise a voting system that will not, in certain instances, violate at least one of the five essential conditions. In short, a perfect democratic voting system is in principle impossible.

As Paul A. Samuelson has put it: "The search of the great minds of recorded history for the perfect democracy, it turns out, is the search for a chimera, for a logical self-contradiction.... Now scholars all over the world—in mathematics, politics, philosophy and economics—are trying to salvage what can be salvaged from Arrow's devastating discovery that is to mathematical politics what Kurt Gödel's 1931 impossibility-of-proving-consistency theorem is to mathematical logic."




*The voting paradox*




*Tournament paradox based on magic square*

Let us approach the voting paradox by first considering a fundamental defect of our present system for electing officials. It frequently puts in office a man who is cordially disliked by a majority of voters but who has an enthusiastic minority following. Suppose 40 percent of the voters are enthusiastic supporters of candidate $A$. The opposition is split between 30 percent for $B$ and 30 percent for $C$. $A$ is elected even though 60 percent of the voters dislike him.

One popular suggestion for avoiding such consequences of the split vote is to allow voters to rank all candidates in their order of preference. Unfortunately this too can produce irrational decisions. The matrix in the illustration at the left on this page displays the notorious voting paradox in its simplest form. The top row shows that a third of the voters prefer candidates $A$, $B$ and $C$ in the order $ABC$. The middle row shows that another third rank them $BCA$ and the bottom row shows that the remaining third rank them $CAB$. Examine the matrix carefully and you will find that when candidates are ranked in pairs, nontransitivity rears its head. Two-thirds of the voters prefer $A$ to $B$, two-thirds prefer $B$ to $C$ and two-thirds prefer $C$ to $A$. If $A$ ran against $B$, $A$ would win. If $B$ ran against $C$, $B$ would win. If $C$ ran against $A$, $C$ would win. Substitute proposals for men and you see how easily a party in power can rig a decision simply by its choice of which paired proposals to put up first for a vote.

The paradox was first recognized by the Marquis de Condorcet in 1785 and is known in France as the Condorcet effect. Lewis Carroll, who wrote several pamphlets on voting, rediscovered it. Most of the early advocates of proportional representation were totally unaware of this Achilles' heel; indeed, the paradox was not fully recognized by political theorists until the mid-1940's, when Duncan Black, a Welsh economist, rediscovered it in connection with his monumental work on committee decision making. Today the experts are nowhere near agreement on which of Arrow's five conditions should be abandoned in the search for the best voting system. One surprising way out, recommended by many decision theorists, is that when a deadlock arises, a "dictator" is chosen by lot to break it. Something close to this solution actually obtains in certain democracies, England for instance, where a constitutional monarch (selected by chance in the sense that lineage guarantees no special biases) has a carefully limited power to break deadlocks under certain extreme conditions.

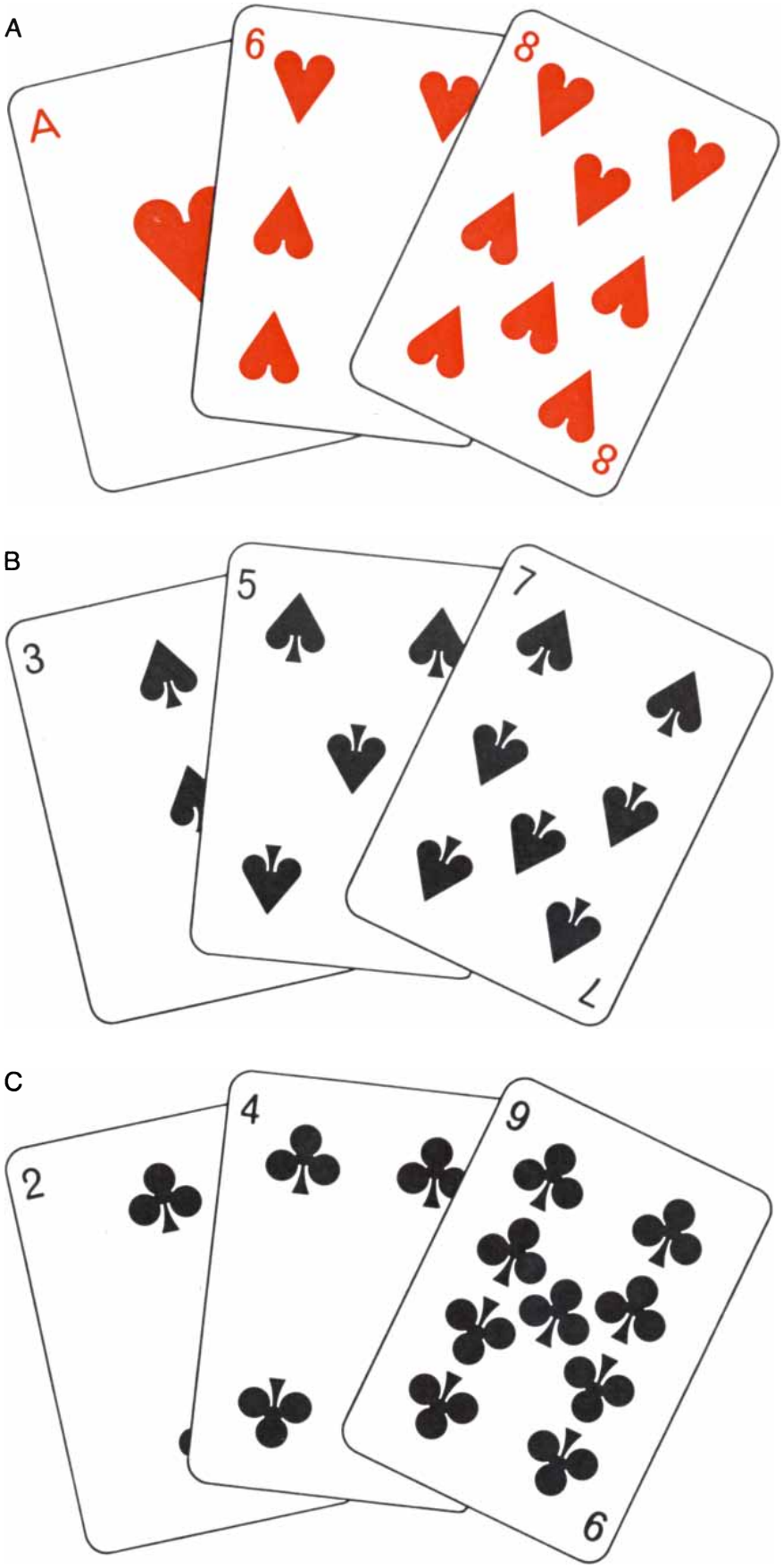
The voting paradox can arise in any situation in which a decision must be made between two alternatives from a set of three or more. Suppose that *A*, *B* and *C* are three men who have simultaneously proposed marriage to a girl. The rows of the matrix for the voting paradox can be used to show how she ranks them with respect to whatever three traits she considers most important, say intelligence, physical attractiveness and income. Taken by pairs, the poor girl finds that she prefers *A* to *B*, *B* to *C* and *C* to *A*. It is easy to see how similar conflicts can arise with respect to one's choice of a job, where to spend a vacation and so on.

Paul R. Halmos once suggested a delightful interpretation of the matrix. Let *A*, *B* and *C* stand for apple pie, blueberry pie and cherry pie. A certain restaurant offers only two of them at any given meal. The rows show how a customer ranks the pies with respect to three properties, say taste, freshness and size of slice. It is perfectly rational, says Halmos, for the customer to prefer apple pie to blueberry, blueberry to cherry and cherry to apple.

Experts differ on how often nontransitive orderings such as this one arise in daily life, but some recent studies in psychology and economics indicate that they are commoner than one might suppose. There are even reports of experiments with rats showing that under certain conditions the pairwise choices of individual rats are nontransitive. (See Warren S. McCulloch, "A Heterarchy of Values Determined by the Topology of Nervous Nets," *Bulletin of Mathematical Biophysics*, Volume 7, 1945, pages 89–93.)

Similar paradoxes arise in round-robin tournaments between teams. Assume that nine tennis players are ranked in ability by the numbers 1 through 9, with the best player given the number 9 and the worst the number 1. The matrix in the illustration at the right on the opposite page is the familiar order-3 magic square. Let rows *A*, *B* and *C* indicate how the nine players are divided into three teams with each row comprising a team. In round-robin tournaments between teams, where each member of one team plays once against each member of the others, assume that the stronger player always wins. It turns out that team *A* defeats *B*, *B* defeats *C* and *C* defeats *A*, in each case by five games to four. It is impossible to say which team is the strongest. The same nontransitivity holds if columns *D*, *E* and *F* of the matrix are the teams.

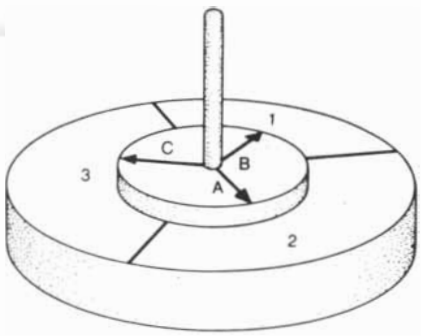
Many paradoxes of this type were



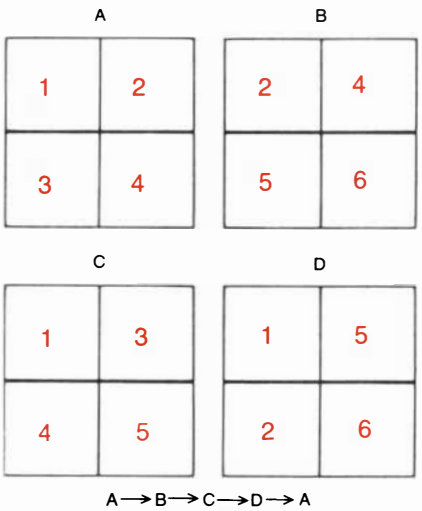

A → B → C → A

***Nontransitive sucker bet based on magic square:*** **A → B → C → A**

jointly investigated by Leo Moser and J. W. Moon. Some of the Moser-Moon paradoxes underlie striking and little-known sucker bets. For example, let each row (or each column) of an order-3 magic-square matrix be a set of playing cards, say the ace, six and eight of hearts for set *A*, the three, five and seven of spades for set *B* and the deuce, four and nine of clubs for *C* [*see illustration on preceding page*]. Each set is randomized and placed face down on a table. The sucker is allowed to draw a card from any set, then you draw a card from a different set. The high card wins. It is easy to prove that no matter what set the sucker draws from, you can pick a set that gives you winning odds of five to four. Set *A* beats *B*, *B* beats *C* and *C* beats *A*. The victim may even be allowed to decide each time whether the high or the low card wins. If you play low card wins, simply pick the winning pile with respect to a nontransitive circle that goes the other way. A good way to play the game is to use sets of cards from three decks with backs of different colors. The packet of nine cards is shuffled each time, then separated by the backs into the three sets. The swindle is, of course, isomorphic with the tennis-tournament paradox.




*Nontransitive top*




*Nontransitive bingo cards*

| B \ A | HH | HT | TH | TT |
|---|---|---|---|---|
| HH | | 1/2 | 1/4 | 1/2 |
| HT | 1/2 | | 1/2 | 3/4 |
| TH | 3/4 | 1/2 | | 1/2 |
| TT | 1/2 | 1/4 | 1/2 | |

***Probabilities of* B *winning***

Nontransitivity prevails in many other simple gambling games. (See this department for December, 1970, for the description of a set of nontransitive dice.) In some cases, such as the top designed by Andrew Lenard [*see top illustration at left*], the nontransitivity is easy to understand. The lower part of the top is fixed but the upper disk rotates. Each of two players chooses a different arrow, the top is spun (in either direction) and the person whose arrow points to the section with the highest number wins. *A* beats *B*, *B* beats *C* and *C* beats *A*, in each case with odds of two to one.

In a set of four bingo cards designed by Donald E. Knuth [*see middle illustration at left*] the nontransitivity is cleverly concealed. Two players each select a bingo card. Numbers from 1 through 6 are randomly drawn without replacement, as they are in standard bingo. If a called number is on a card, it is marked with a bean. The first player to complete a horizontal row wins. Here, of course, the numbers are just symbols; they can be replaced by any set of six different symbols. I leave it to the reader to work out the probabilities that show card *A* beats *B*, *B* beats *C*, *C* beats *D* and *D* beats *A*. The game is transitive with three players, but the winning odds for the four possible triplets are surprising.

One of the most incredible of all nontransitive betting situations, discovered (appropriately) by a mathematician named Walter Penney, was given as a problem in the *Journal of Recreational Mathematics* (October, 1969, page 241). It is not well known, and most mathematicians simply cannot believe it when they first hear of it. It is certainly one of the finest of all sucker bets. It can be played with a penny, or as a side bet on the reds and blacks of a roulette wheel, or in any situation in which two alternatives are randomized with equal odds. We shall assume that a penny is used. If it is flipped three times, there are eight equally probable outcomes: *HHH*, *HHT*, *HTH*, *HTT*, *THH*, *THT*, *TTH* and *TTT*. One player selects one of these triplets and the other player selects a different one. The penny is then flipped repeatedly until one of the chosen triplets appears as a run and wins the game. For example, if the chosen triplets are *HHT* and *THT* and the flips are *THHHT*, the last three flips show that *HHT* has won. In brief, the first triplet to appear as a run wins.

One is inclined to assume that one triplet is as likely to appear first as any other, but it takes only a moment to realize that this is not the case even with doublets. Consider the doublets *HH*, *HT*, *TH* and *TT*. *HH* and *HT* are equally likely to appear first because after the first *H* appears it is just as likely to be followed by an *H* as by a *T*. The same reasoning shows that *TT* and *TH* are equal. Because of symmetry $HH = TT$ and $HT = TH$. *TH* beats *HH* with odds of three to one, however, and *HT* beats *TT* with the same probability. Consider *HT* and *TT*. *TT* is always preceded by *HT* except when *TT* appears on the first two flips. This happens in the long run only once in four times, and so the probability that *HT* beats *TT* is 3/4. The bottom illustration on this page shows the probability that *B*, the second player, will win for all pairs of doublets.

When we turn to triplets, the situation becomes much more surprising. Since it does not matter which side of a coin is designated heads, we know that $HHH = TTT$, $TTH = HHT$, $HTH = THT$ and so on. When we examine the probabilities for unequal pairs, however, we discover that the game is not transitive. No matter what triplet the first player takes, the second player can select a better one. The upper illustration on the opposite page gives the probability that *B*, the second player, defeats *A* for all possible pairings. To find *B*'s best response to a triplet chosen by *A*, find *A*'s triplet at the top, go down the column until you reach a probability shown in color, then move left along the row to *B*'s triplet on the left.

Note that *B*'s probability of winning is, at the worst, 2/3 (or odds of two to one) and can go as high as 7/8 (or odds of seven to one). The seven-to-one odds are easy to comprehend. Consider *THH* and *HHH*. If *HHH* first appears anywhere except at the start, it must be preceded by a *T*, which means that *THH* has appeared earlier. *HHH* wins, therefore, only when it appears on the first

three flips. Clearly this happens only once in eight flips.

Barry Wolk of the University of Manitoba has discovered a curious rule for determining the best triplet. Let $X$ be the first triplet chosen. Convert it to a binary number by changing each $H$ to 0 and each $T$ to 1. Divide the number by 2, round down the quotient to the nearest integer, multiply by 5 and add 4. Express the result in binary, then convert the last three digits back to $H$ and $T$.

Nontransitivity holds for all higher $n$-tuplets. A chart supplied by Wolk gives the winning probabilities for $B$ in all possible pairings of quadruplets [*see illustration on next page*]. Like the preceding two charts and charts for all higher $n$-tuplets, the matrix is symmetric about the center. The upper right quadrant is the lower left quadrant upside down, and the same holds for the upper left and lower right quadrants. The probabilities for $B$'s best responses to $A$ are shown in color.

In studying these figures Wolk discovered another kind of anomaly as surprising as nontransitivity. It has to do with what are called waiting times. The waiting time for an $n$-tuplet is the average number of tosses, in the long run, until the specified $n$-tuplet appears. The longer you wait for a bus, the shorter becomes the expected waiting time. Pennies, however, have no memory, so that the waiting time for an $n$-tuplet is independent of all previous flips. The waiting time for $H$ and $T$ is 2. For doublets the waiting time is 4 for $HT$ and $TH$ and 6 for $HH$ and $TT$. For triplets the waiting times are 8 for $HHT$, $HTT$, $THH$ and $TTH$; 10 for $HTH$ and $THT$, and 14 for $HHH$ and $TTT$. None of this contradicts what we know about which triplet of a pair is likely to show first. With quadruplets, however, contradictions arise with six pairs. For example, $THTH$ has a waiting time of 20 and $HTHH$ a waiting time of 18. Yet $THTH$ is more likely to turn up before $HTHH$ with a probability of 9/14, or well over 1/2. In other words, an event that is less frequent in the long run is likely to happen before a more frequent event. There is no logical contradiction involved here, but it does show that "average waiting time" has peculiar properties.

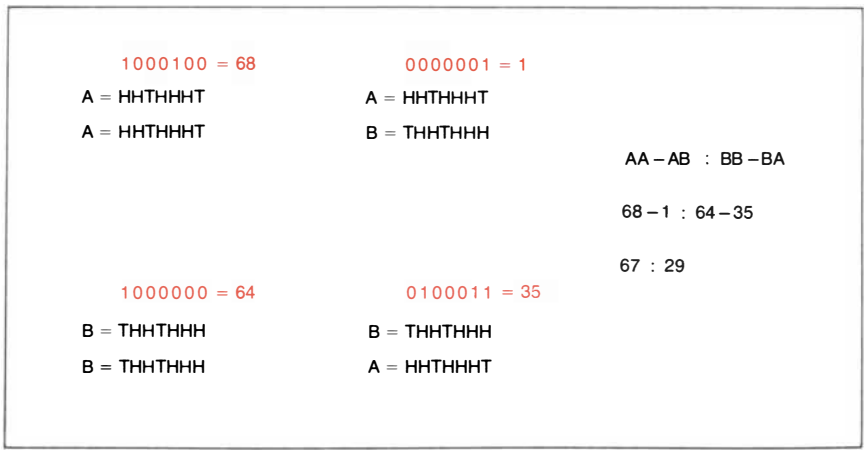
There are many ways to calculate the probability that one $n$-tuplet will precede another. You can do it by summing infinite series, by drawing tree diagrams, by recursive techniques that produce sets of linear equations and so on. One of the strangest and most efficient techniques was recently devised by John Horton Conway of the University of Cambridge.

| A \ B | HHH | HHT | HTH | HTT | THH | THT | TTH | TTT |
|---|---|---|---|---|---|---|---|---|
| HHH | | 1/2 | 2/5 | 2/5 | 1/8 | 5/12 | 3/10 | 1/2 |
| HHT | 1/2 | | 2/3 | 2/3 | 1/4 | 5/8 | 1/2 | 7/10 |
| HTH | 3/5 | 1/3 | | 1/2 | 1/2 | 1/2 | 3/8 | 7/12 |
| HTT | 3/5 | 1/3 | 1/2 | | 1/2 | 1/2 | 3/4 | 7/8 |
| THH | 7/8 | 3/4 | 1/2 | 1/2 | | 1/2 | 1/3 | 3/5 |
| THT | 7/12 | 3/8 | 1/2 | 1/2 | 1/2 | | 1/3 | 3/5 |
| TTH | 7/10 | 1/2 | 5/8 | 1/4 | 2/3 | 2/3 | | 1/2 |
| TTT | 1/2 | 3/10 | 5/12 | 1/8 | 2/5 | 2/5 | 1/2 | |

***Probabilities of* B *winning in a triplet game***

I have no idea why it works. It just cranks out the answer as if by magic, like so many of Conway's other algorithms.

The key to Conway's procedure is the calculation of four binary numbers that Conway calls leading numbers. Let $A$ stand for the 7-tuplet $HHTHHHT$ and $B$ for $THHTHHH$. We want to determine the probability of $B$ beating $A$. To do this write $A$ above $A$, $B$ above $B$, $A$ above $B$ and $B$ above $A$ [*see illustration below*]. Above the top tuplet of each pair a binary number is constructed as follows: Consider the first pair, $AA$. Look at the first letter of the top tuplet and ask yourself if the seven letters, beginning with this first one, correspond exactly to the first seven letters of the tuplet below it. Obviously they do, and so we put a 1 above the first letter. Next look at the second letter of the top tuplet and ask if the six letters, starting with this one, correspond to the *first six* letters of the tuplet below. Clearly they do not, and so we put 0 above the second letter. Do the five letters starting with the third letter of the top tuplet correspond to the first five letters of the lower tuplet? No, and




***John Horton Conway's algorithm for calculating odds of* B's n-*tuplet beating* A's n-*tuplet***

so this letter also gets 0. The fourth letter gets another 0. When we check the fifth letter of the top *A*, we see that *HHT* does correspond to the first three letters of the lower *A*, and so the fifth letter gets a 1. Letters six and seven each get 0. The "*A* leading *A* number," or *AA*, is 1000100, in which each 1 corresponds to a yes answer, each 0 to a no. Translating 1000100 from binary to decimal gives us 68 as the leading number for *AA*.

The bottom illustration on the preceding page shows the results of this procedure in calculating leading numbers *AA*, *BB*, *AB* and *BA*. Whenever an *n*-tuplet is compared with itself, the first digit of the leading number must of course be 1. When compared with a different tuplet, the first digit may or may not be 1.

The odds in favor of *B* beating *A* are given by the ratio $AA - AB : BB - BA$. In this case $68 - 1 : 64 - 35 = 67 : 29$. As an exercise the reader can try calculating the odds in favor of *THH* beating *HHH*. The four leading numbers will be $AA = 7$, $BB = 4$, $AB = 0$ and $BA = 3$. Plugging these into the formula, $AA - AB : BB - BA$ gives odds of $7 - 0 : 4 - 3$, or 7 : 1, as expected. The algorithm works just as well on triplets of unequal lengths. If the formula is applied, for example, to $A = HT$ and $B = H$, it becomes $3 - 0 : 1 - 1 = 3 : 0$, or a probability of 1 that *H* beats *HT*.

I conclude with a problem by David L. Silverman, who was the first to introduce the Penney paradox in the problems department that he edits for the *Journal of Recreational Mathematics*. The reader should have little difficulty solving it by Conway's algorithm before the answer appears next month. *TTHH* has a waiting time of 16 and *HHH* has a waiting time of 14. Which of these tuplets is most likely to appear first and with what probability?

The four tangram problems given last month, involving holes that do not touch one another or the border, are

| A / B | HHHH | HHHT | HHTH | HHTT | HTHH | HTHT | HTTH | HTTT | THHH | THHT | THTH | THTT | TTHH | TTHT | TTTH | TTTT |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| HHHH | | 1/2 | 2/5 | 2/5 | 3/10 | 5/12 | 4/11 | 4/11 | 1/16 | 3/8 | 3/8 | 3/8 | 1/4 | 3/8 | 7/22 | 1/2 |
| HHHT | 1/2 | | 2/3 | 2/3 | 1/2 | 5/8 | 4/7 | 4/7 | 1/8 | 9/16 | 9/16 | 9/16 | 5/12 | 9/16 | 1/2 | 15/22 |
| HHTH | 3/5 | 1/3 | | 1/2 | 3/5 | 5/7 | 1/2 | 1/2 | 5/12 | 5/12 | 9/16 | 9/16 | 5/14 | 1/2 | 7/16 | 5/8 |
| HHTT | 3/5 | 1/3 | 1/2 | | 3/7 | 5/9 | 2/3 | 2/3 | 5/12 | 5/12 | 9/16 | 9/16 | 1/2 | 9/14 | 7/12 | 3/4 |
| HTHH | 7/10 | 1/2 | 2/5 | 4/7 | | 1/2 | 1/2 | 1/2 | 7/12 | 7/12 | 5/14 | 1/2 | 7/16 | 7/16 | 7/16 | 5/8 |
| HTHT | 7/12 | 3/8 | 2/7 | 4/9 | 1/2 | | 1/2 | 1/2 | 7/16 | 7/16 | 1/2 | 9/14 | 7/16 | 7/16 | 7/16 | 5/8 |
| HTTH | 7/11 | 3/7 | 1/2 | 1/3 | 1/2 | 1/2 | | 1/2 | 1/2 | 1/2 | 9/16 | 5/12 | 7/12 | 7/12 | 7/16 | 5/8 |
| HTTT | 7/11 | 3/7 | 1/2 | 1/3 | 1/2 | 1/2 | 1/2 | | 1/2 | 1/2 | 9/16 | 5/12 | 7/12 | 7/12 | 7/8 | 15/16 |
| THHH | 15/16 | 7/8 | 7/12 | 7/12 | 5/12 | 9/16 | 1/2 | 1/2 | | 1/2 | 1/2 | 1/2 | 1/3 | 1/2 | 3/7 | 7/11 |
| THHT | 5/8 | 7/16 | 7/12 | 7/12 | 5/12 | 9/16 | 1/2 | 1/2 | 1/2 | | 1/2 | 1/2 | 1/3 | 1/2 | 3/7 | 7/11 |
| THTH | 5/8 | 7/16 | 7/16 | 7/16 | 9/14 | 1/2 | 7/16 | 7/16 | 1/2 | 1/2 | | 1/2 | 4/9 | 2/7 | 3/8 | 7/12 |
| THTT | 5/8 | 7/16 | 7/16 | 7/16 | 1/2 | 5/14 | 7/12 | 7/12 | 1/2 | 1/2 | 1/2 | | 4/7 | 2/5 | 1/2 | 7/10 |
| TTHH | 3/4 | 7/12 | 9/14 | 1/2 | 9/16 | 9/16 | 5/12 | 5/12 | 2/3 | 2/3 | 5/9 | 3/7 | | 1/2 | 1/3 | 3/5 |
| TTHT | 5/8 | 7/16 | 1/2 | 5/14 | 9/16 | 9/16 | 5/12 | 5/12 | 1/2 | 1/2 | 5/7 | 3/5 | 1/2 | | 1/3 | 3/5 |
| TTTH | 15/22 | 1/2 | 9/16 | 5/12 | 9/16 | 9/16 | 9/16 | 1/8 | 4/7 | 4/7 | 5/8 | 1/2 | 2/3 | 2/3 | | 1/2 |
| TTTT | 1/2 | 7/22 | 3/8 | 1/4 | 3/8 | 3/8 | 3/8 | 1/16 | 4/11 | 4/11 | 5/12 | 3/10 | 2/5 | 2/5 | 1/2 | |

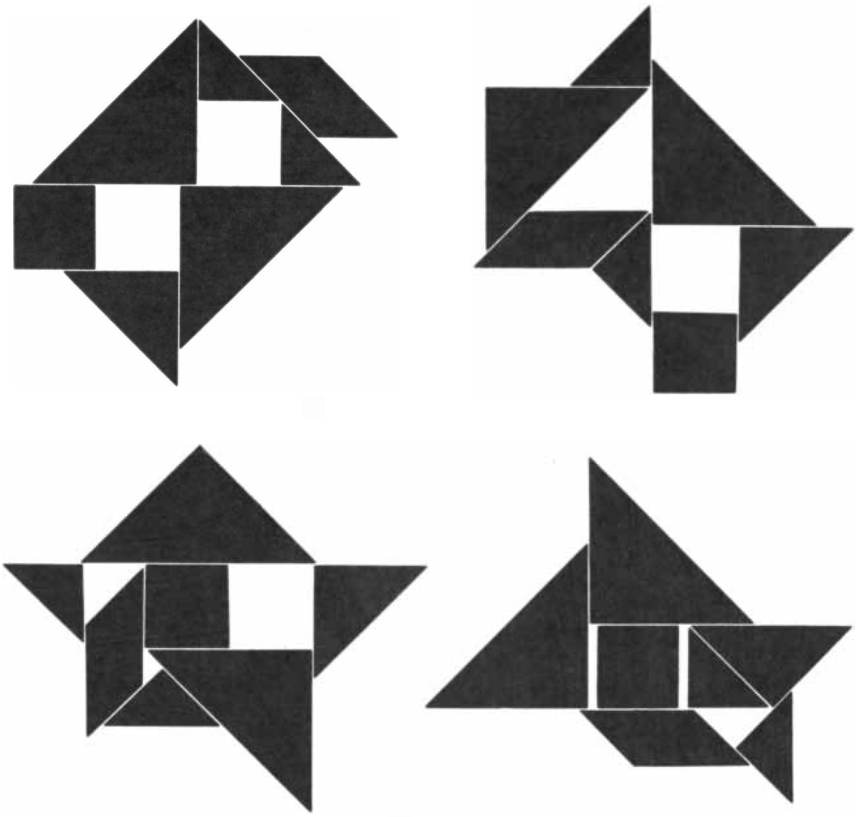
***Probabilities of* B *winning in a quadruplet game***

solved as shown in the upper illustration at the right. At top left is a way of making two unit square holes. At top right are two holes, one a unit square and the other a triangle of area 1. At bottom left is a way of making a square hole and two triangular holes. (The fit is extremely close. The top triangle's hypotenuse is longer than the length it must span by only .121+.) At bottom right are two rectangular holes and one triangular hole.

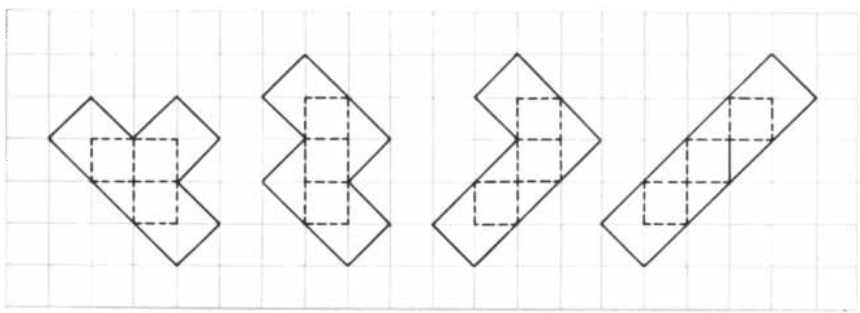
Ronald C. Read has proved that a tangram that is snug except for one or more holes cannot have more than one hole if the holes do not touch one another or the border. The smallest possible hole is equal to a small triangular tan. No matter how two such holes are placed so that they do not touch, at least 17 triangles are required to isolate them from a tangram border. Since the seven tans are made of 16 such triangles, completing the required tangram is impossible. For nonsnug tangrams it appears that no more than three holes not touching one another or the border are possible, although I know of no proof.

The square is the only snug tangram with all its sides irrational. This is how Read proves it. As I explained last month, an irrational tangram drawn on graph paper with the unit square tan oriented orthogonally would have all its edges making 45-degree angles with the matrix lines. All corners clearly must be either 90 degrees or 270 degrees. Since each side is a multiple of $\sqrt{2}$ and the total area is 8, it follows that any irrational snug tangram will be a tetromino composed of four squares, each $\sqrt{2}$ on the side.

There are five tetrominoes. One of them, the square, we know can be formed. Each of the other four is easily proved impossible by placing the square tan in each of three possible positions [*see lower illustration at right*] and then exploring ways of completing the tetromino. The first tetromino is ruled out at once because there is no way to place the two large triangles. In the other three cases, for each position of the square tan there are at most only four ways to place the two large triangles. In every case, after the square and two large triangles are placed there is no spot for the rhomboid. Thus the square is the only possible irrational snug tangram.



*Solutions to last month's hole tangrams*



*Impossibility proof for the nonsquare tetrominoes*

Dr. Matrix' magic-pyramid problem in June intrigued many readers. Hiram Fuller Gutgasz pointed out that the solution is unique in a stronger sense than I supposed. It is unnecessary to restrict the labeling to positive numbers no greater than 10. The solution is unique for *any* set of eight distinct positive integers. Two readers, J. A. McCallum and Sheldon B. Akers, each asked themselves if a magic pyramid could be constructed with eight consecutive positive integers. Each found the only solution and proved it unique. The numbers are 4, 5, 6, 7, 8, 9, 10, 11, with a magic constant of 24.

July's article on hex and star numbers asked if a hex could be a cube. David Chess, Sin Hitotumatu and Wesley Johnston were the first of many readers to point out how easily this can be answered in the negative. As I showed in the article, every hex is the difference between two consecutive cubes, therefore the question is the same as asking if the formula $(x + 1)^3 - x^3 = y^3$ has an integral solution. When this is written $x^3 + y^3 = (x + 1)^3$, we see at once that it is a case of Fermat's last theorem, with exponents of 3, which was shown long ago to have no integral solution.

In mentioning that the sum of the first $n$ consecutive stars is square only when $n$ is 1 or 169, I got my facts a bit garbled. This was first established in 1973 by John Harris on the basis of 1942 results reported by Louis J. Mordell on page 271 of *Diophantine Equations*.

Harris and G. J. Westerink each corrected another error. I said that all square stars begin with 1, but this is not true. The 12th square star begins with 9, and thereafter the numbers can start with any positive digit.

# THE AMATEUR SCIENTIST

## *Electrostatic motors are powered by the electric field of the earth*

Conducted by C. L. Stong

Although no one can make a perpetual motion machine, anyone can tap the earth's electric field to run a homemade motor perpetually. The field exists in the atmosphere between the earth's surface and the ionosphere as an electric potential of about 360,000 volts. Estimates of the stored energy range from a million kilowatts to a billion kilowatts.

Energy in this form cannot be drawn on directly for driving ordinary electric motors. Such motors develop mechanical force through the interaction of magnetic fields that are generated with high electric current at low voltage, as Michael Faraday demonstrated in 1821. The earth's field provides relatively low direct current at high voltage, which is ideal for operating electrostatic motors similar in principle to the machine invented by Benjamin Franklin in 1748.

Motors of this type are based on the force of mutual attraction between unlike electric charges and the mutual repulsion of like charges. The energy of the field can be tapped with a simple antenna in the form of a vertical wire that carries one sharp point or more at its upper end. During fair weather the antenna will pick up potential at the rate of about 100 volts for each meter of height between the points and the earth's surface up to a few hundred feet. At higher altitudes the rate decreases. During local thunderstorms the pickup can amount to thousands of volts per foot. A meteorological hypothesis is that the field is maintained largely by thunderstorms, which pump electrons out of the air and inject them into the earth through bolts of lightning that continuously strike the surface at an average rate of 200 strokes per second.

Why not tap the field to supplement conventional energy resources? Several limitations must first be overcome. For example, a single sharp point can draw electric current from the surrounding air at a rate of only about a millionth of an ampere. An antenna consisting of a single point at the top of a 60-foot wire could be expected to deliver about a microampere at 2,000 volts; the rate is equivalent to .002 watt. A point-studded balloon tethered by a wire at an altitude of 75 meters might be expected to deliver .075 watt. A serious limitation appears as the altitude of the antenna exceeds about 200 meters. The correspondingly higher voltages become difficult to confine.

At an altitude of 200 meters the antenna should pick up some 20,000 volts. Air conducts reasonably well at that potential. Although nature provides effective magnetic materials in substances such as iron, nickel and cobalt, which explains why the electric-power industry developed around Faraday's magnetic dynamo, no comparably effective insulating substances exist for isolating the high voltages that would be required for electrostatic machines of comparable power. Even so, electrostatic motors, which are far simpler to build than electromagnetic ones, may find applications in special environments such as those from which magnetism must be excluded or in providing low power to apparatus at remote, unmanned stations by tapping the earth's field.

Apart from possible applications electrostatic motors make fascinating playthings. They have been studied extensively in recent years by Oleg D. Jefimenko and his graduate students at West Virginia University. The group has reconstructed models of Franklin's motors and developed advanced electrostatic machines of other types.

Although Franklin left no drawing of his motor, his description of it in a letter to Peter Collinson, a Fellow of the Royal Society, enabled Jefimenko to reconstruct a working model [*see illustration on opposite page*]. Essentially the machine consists of a rimless wheel that turns in the horizontal plane on low-friction bearings. Each spoke of the "electric wheel," as Franklin called the machine, consists of a glass rod with a brass thimble at its tip. An electrostatic charge for driving the motor was stored in Leyden jars. A Leyden jar is a primitive form of the modern high-voltage capacitor. Franklin charged his jars with an electrostatic generator.

The high-voltage terminals of two or more Leyden jars that carried charges of opposite polarity were positioned to graze the thimbles on opposite sides of the rotating wheel. The motor was started by hand. Thereafter a spark would jump from the high-voltage terminal to each passing thimble and impart to it a charge of the same polarity as that of the terminal. The force of repulsion between the like charges imparted momentum to the wheel.

Conversely, the thimbles were attracted by the oppositely charged electrode of the Leyden jar Franklin placed on the opposite side of the wheel. As the thimbles grazed that jar, a spark would again transfer charge, which was of opposite polarity. Thus the thimbles were simultaneously pushed and pulled by the high-voltage terminals exactly as was needed to accelerate the wheel.

Franklin was not altogether happy with his motor. The reason was that running it required, in his words, "a foreign force, to wit, that of the bottles." He made a second version of the machine without Leyden jars.

In this design the rotor consisted principally of a 17-inch disk of glass mounted to rotate in the horizontal plane on low-friction bearings. Both surfaces of the disk were coated with a film of gold, except for a boundary around the edge. The rotor was thus constructed much like a modern flat-plate capacitor.

Twelve evenly spaced metal spheres, cemented to the edge of the disk, were connected alternately to the top and bottom gold films. Twelve stationary thimbles supported by insulating columns were spaced around the disk to graze the rotating metal spheres. When

Franklin placed opposite charges on the top and bottom films and gave the rotor a push, the machine ran just as well as his first design, and for the same reason. According to Franklin, this machine would make up to 50 turns a minute and would run for 30 minutes on a single charge.

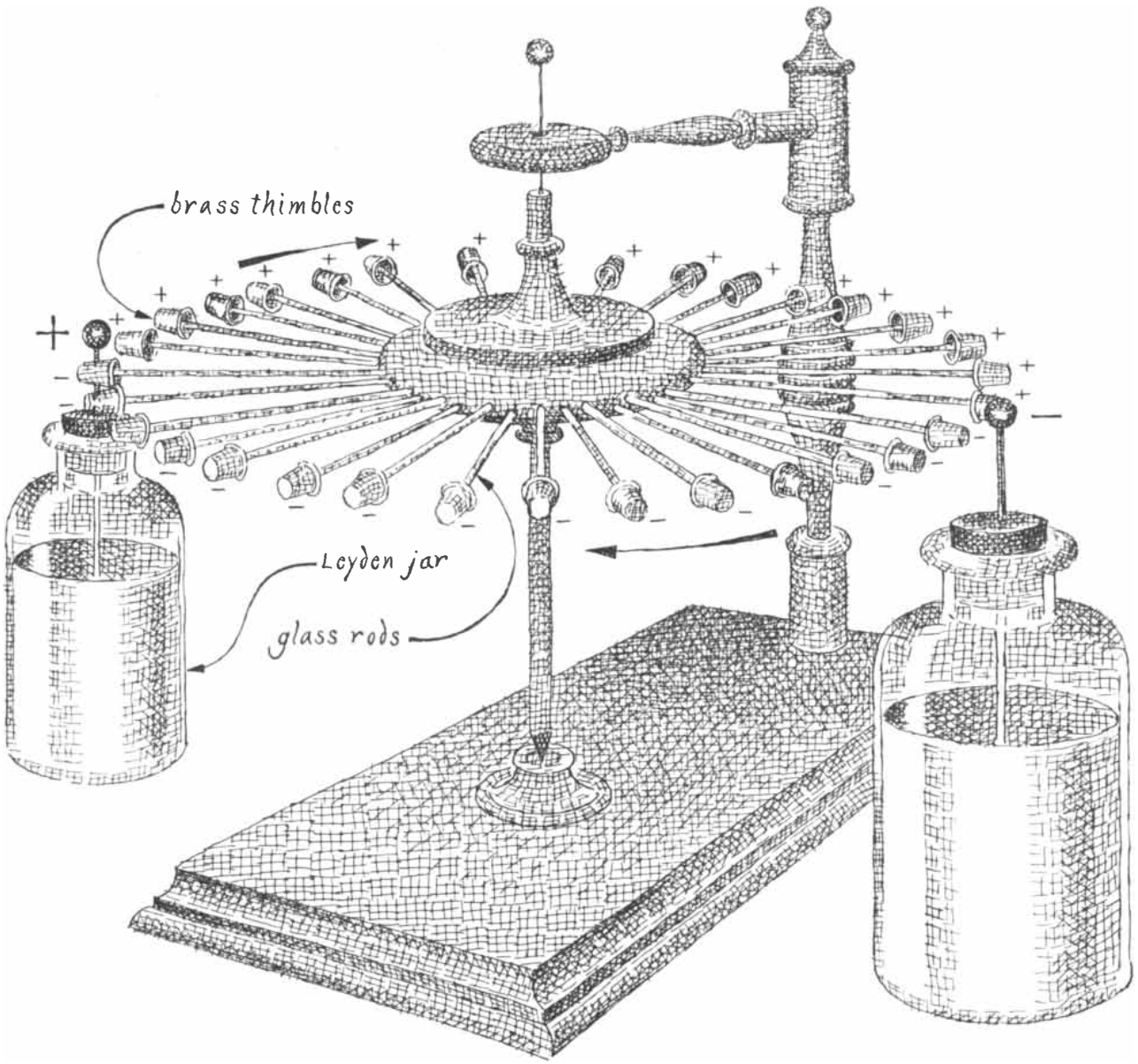
Jefimenko gives both motors an initial charge from a 20,000-volt generator. They consume current at the rate of about a millionth of an ampere when they are running at full speed. The rate is equivalent to .02 watt, which is the power required to lift a 20-gram weight 10 centimeters (or an ounce 2.9 inches) in one second.

Jefimenko wondered if Franklin's motor could be made more powerful. As Jefimenko explains, the force can be increased by adding both moving and fixed electrodes. This stratagem is limited by the available space. If the electrodes are spaced too close, sparks tend to jump from electrode to electrode around the rotor, thereby in effect short-circuiting the machine. Alternatively the rotor could be made cylindrical to carry electrodes in the form of long strips or plates. This scheme could perhaps increase the output power by a factor of 1,000.

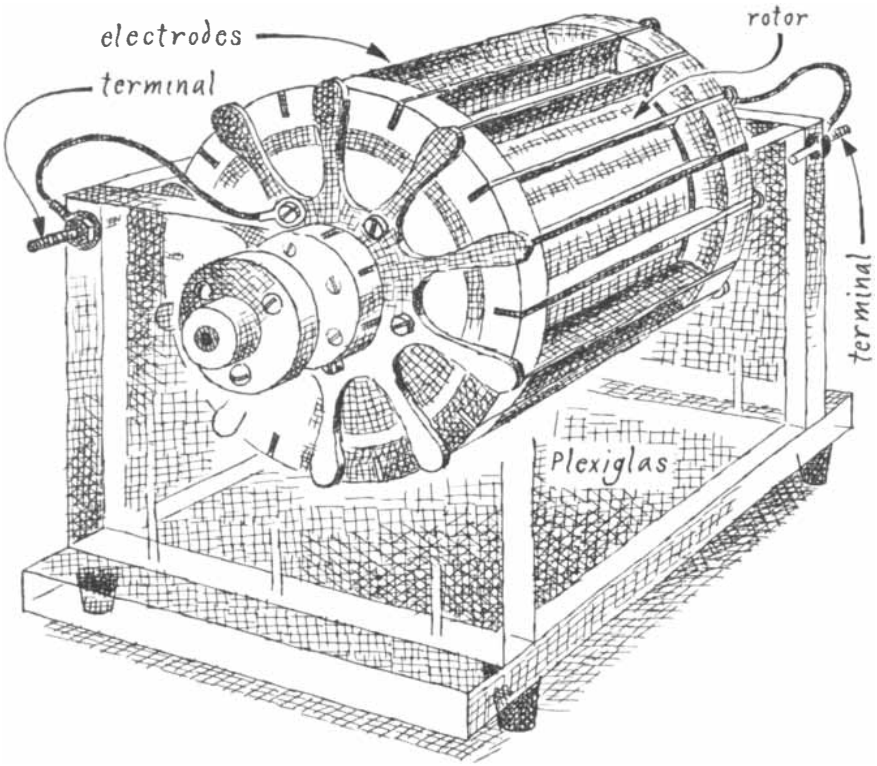
Reviewing the history of electrostatic machines, Jefimenko came across a paper published in 1870 by Johann Christoff Poggendorff, a German physicist. It described an electrostatic motor fitted with a rotor that carried no electrodes. The machine consisted of an uncoated disk of glass that rotated in the vertical plane on low-friction bearings between opposing crosses of ebonite. Each insulating arm of the crosses supported a comblike row of sharp needle points that grazed the glass.

When opposing combs on opposite sides of the glass were charged in opposite polarity to potentials in excess of 2,000 volts, air in the vicinity of the points on both sides of the glass was ionized. A bluish glow surrounded the points, which emitted a faint hissing sound. The effect, which is variously known as St. Elmo's fire and corona discharge, deposited static charges on both sides of the rotor.

Almost the entire surface of the glass acquired a coating of either positive or negative fixed charges, depending on the polarity of the combs. The forces of repulsion and attraction between glass so




***Benjamin Franklin's electrostatic motor***




*Oleg D. Jefimenko's corona motor*

charged and the combs were substantially larger than they were in Franklin's charged thimbles. The forces were also steadier, because in effect the distances between the combs and the charged areas remained constant. It should be noted that adjacent combs on the same side of the glass carried charges of opposite polarity, so that the resulting forces of attraction and repulsion acted in unison to impart momentum to the disk, as they did in Franklin's motor.

By continued experimentation Poggendorff learned that he should slant the teeth of the combs to spray charge on the glass at an angle. The resulting asymmetrical force made the motor self-starting and unidirectional. When the teeth were perpendicular to the glass surface, the forces were symmetrical, as they were in Franklin's motor. When the machine was started by hand, it ran equally well in either direction.

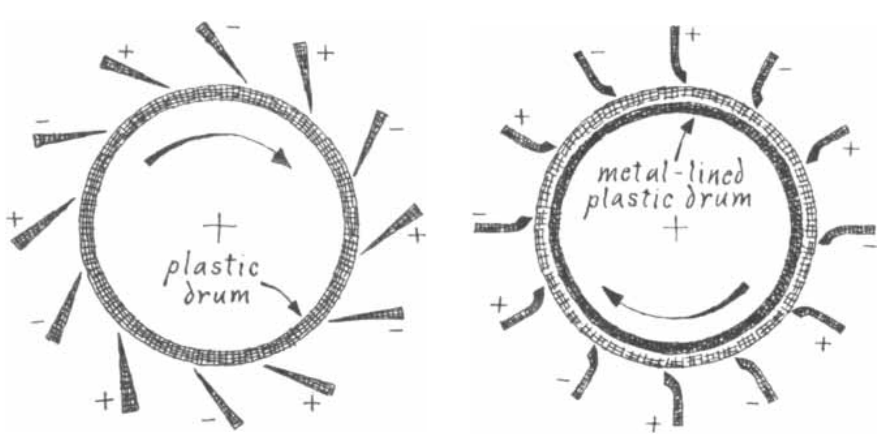
Poggendorff was immensely pleased by the rate at which his machine converted charge into mechanical motion. He concluded his paper with a faintly odious reference to Franklin's device. "That such a quantity of electricity must produce a far greater force than that in the [Franklin] electric roasting spit," he wrote, "is perfectly obvious and nowadays would not be denied by Franklin himself. With one grain of gunpowder one cannot achieve so much as with one hundred pounds."

Electrostatic motors are now classified in general by the method by which charge is either stored in the machine or transferred to the rotor. Poggendorff's machine belongs to the corona type, which has attracted the most attention in recent years. Although its measured efficiency is better than 50 percent, Poggendorff regarded it merely as an apparatus for investigating electrical phenomena. He wrote that "it would be a sanguine hope if one wanted to believe that any useful mechanical effect could be achieved with it."

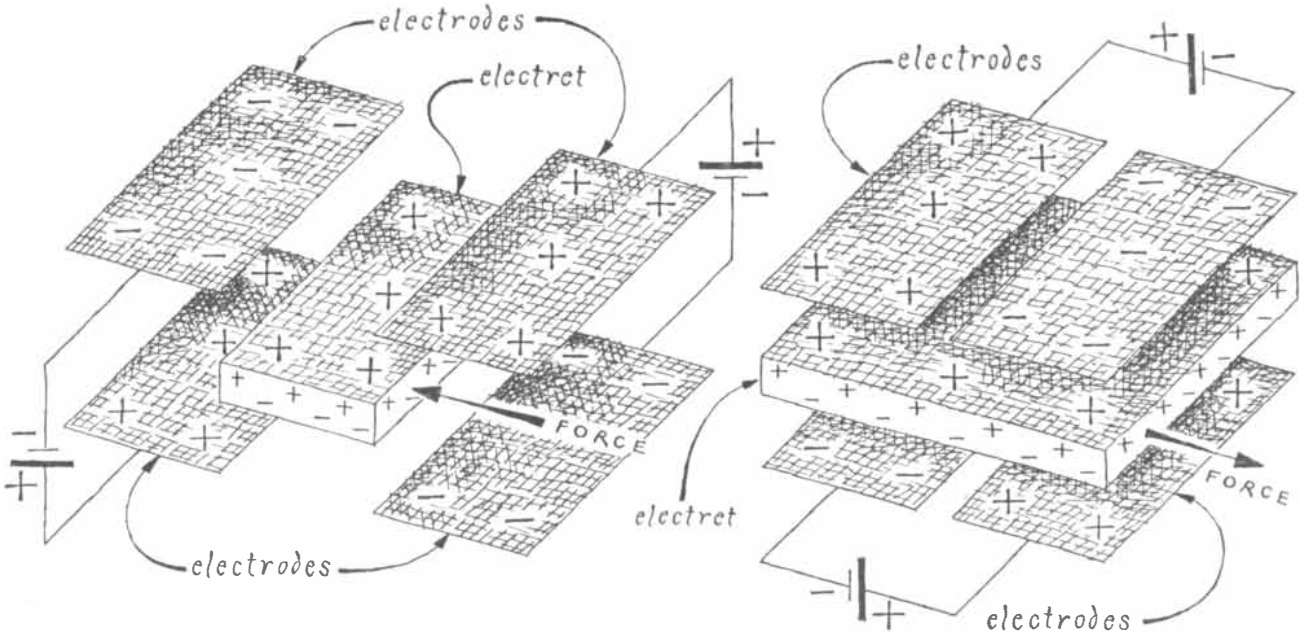
Poggendorff's negative attitude toward the usefulness of his design may well have retarded its subsequent development. A modern version of the machine constructed in Jefimenko's laboratory has an output of approximately .1 horsepower. It operates at speeds of up to 12,000 revolutions per minute at an efficiency of substantially more than 50 percent. In one form the modern corona motor consists of a plastic cylinder that turns on an axial shaft inside a concentric set of knife-edge electrodes that spray charge on the surface of the cylinder [*see top illustration on this page*]. Forces that act between the sprayed charges and the knife-edge electrodes impart momentum to the cylinder.

Machines of this kind can be made of almost any inexpensive dielectric materials, including plastics, wood and even cardboard. The only essential metal parts are the electrodes and their interconnecting leads. Even they can be contrived of metallic foil backed by any stiff dielectric. The shaft can be made of plastic that turns in air bearings. By resorting to such stratagems experimenters can devise motors that are extremely light in proportion to their power output. Corona motors require no brushes or commutators. A potential of at least 2,000 volts, however, is essential for initiating corona discharge at the knife-edges.

A smaller and simpler version of the machine was demonstrated in 1961 by J. D. N. Van Wyck and G. J. Kühn in South Africa. This motor consisted of a plastic disk about three millimeters thick and 40 millimeters in diameter supported in the horizontal plane by a slender shaft that turned in jeweled bearings. Six radially directed needle points grazed the rim of the disk at equal intervals. When the machine operated from a source of from 8,000 to 13,000 volts, rotational speeds of up to 12,000 revolutions per minute were measured.




***Conventional electrodes*** **(left)** ***and improved electrodes*** **(right)**




*Scheme of A. N. Gubkin's electret motor* (left) *and the slot-effect motor* (right)

I made a corona motor with Plexiglas tubing two inches in diameter and one and a half inches long. It employed stiff-backed single-edge razor blades as electrodes. The bore of the tube was lined with a strip of aluminum foil, a stratagem devised in Jefimenko's laboratory to increase the voltage gradient in the vicinity of the electrodes and thus to increase the amount of charge that can be deposited on the surface of the cylinder. I coated all surfaces of the razor blades except the cutting edges and all interconnecting wiring with "anticorona dope," a cementlike liquid that dries to form a dielectric substance that reduces the loss of energy through corona discharges in nonproductive portions of the circuit.

The axial shaft that supports the cylinder on pivot bearings was cut out of a steel knitting needle. The ends of the shaft were ground and polished to 30-degree points. To form the points I chucked the shaft in an electric hand drill, ground the metal against an oilstone and polished the resulting pivots against a wood lap coated with tripoli.

The bearings that supported the pivots were salvaged from the escapement mechanism of a discarded alarm clock. A pair of indented setscrews could be substituted for the clock bearings. The supporting frame was made of quarter-inch Lucite. The motor can be made self-starting and unidirectional by slanting the knife-edges. Those who build the machine may discover, as I did, that the most difficult part of the project, balancing the rotor, is encountered after assembly. The rotor must be balanced both statically and dynamically.

Static balance was achieved by experimentally adding small bits of adhesive tape to the inner surface of the aluminum foil that lines the cylinder until the rotor remained stationary at all positions to which it was set by hand. When the rotor was balanced and power was applied, the motor immediately came up to speed, but it shook violently. I had corrected the imbalance caused by a lump of cement at one end of the rotor by adding a counterweight on the opposite side at the opposite end of the cylinder. Centrifugal forces at the ends were 180 degrees out of phase, thus constituting a couple.

The dynamic balancing, which is achieved largely by cut-and-try methods, took about as much time as the remainder of the construction. To check for dynamic balance suspend the motor freely with a string, run it at low speed and judge by the wiggle where a counterweight must be added. Adhesive tape makes a convenient counterweight material because it can be both applied and shifted easily.

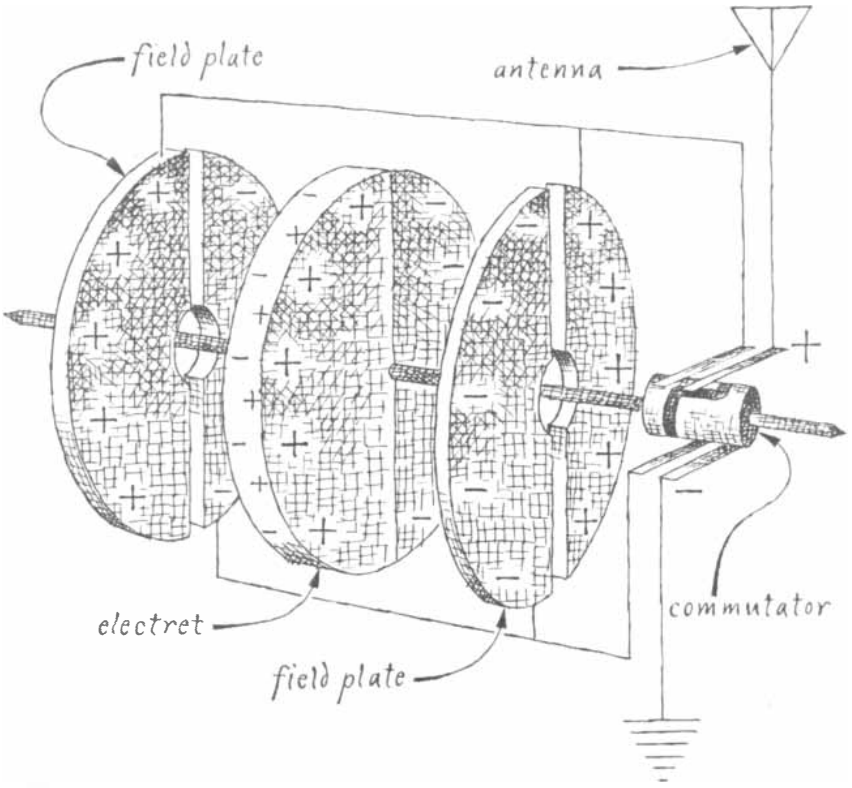
I made the motor as light and frictionless as possible with the objective of operating it with energy from the earth's field. The field was tapped with an antenna consisting of 300 feet of No. 28 gauge stranded wire insulated with plastic. It is the kind of wire normally employed for interconnecting electronic components and is available from dealers in radio supplies.

The upper end of the wire was connected to a 20-foot length of metallic tinsel of the kind that serves for decorating a Christmas tree. The tinsel functioned as multiple needle points. Strips cut from window screening would doubtless work equally well.

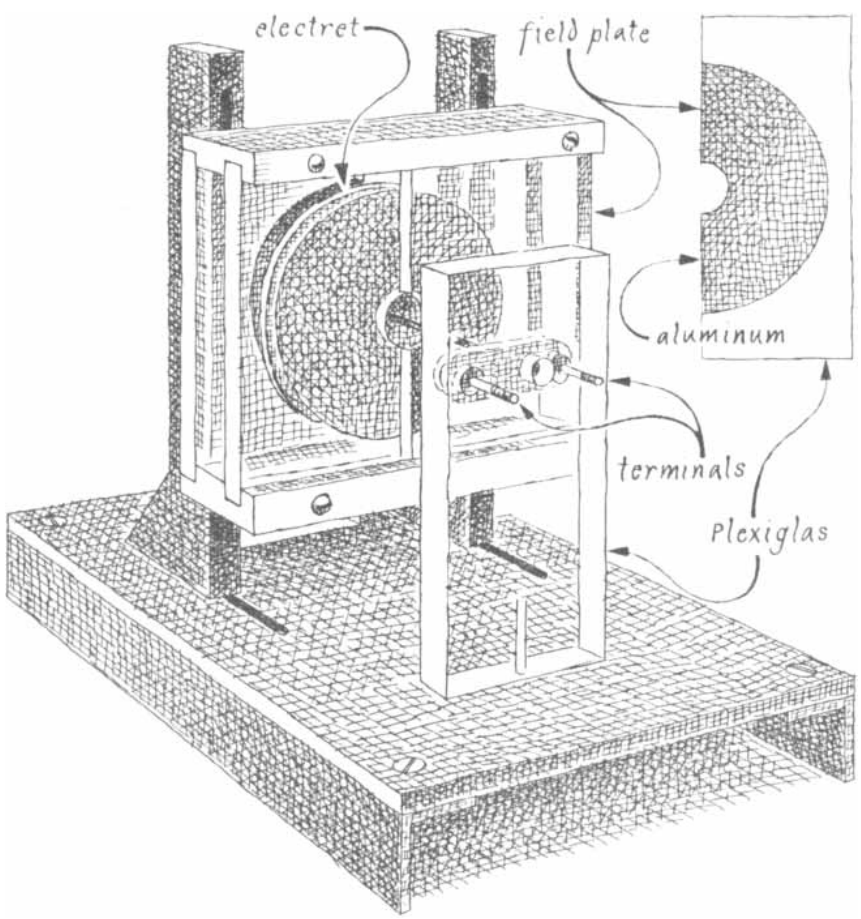
The upper end of the tinsel was hoisted aloft by a cluster of three weather balloons. Such balloons, each three feet in diameter, and the helium to inflate them are available from the Edmund Scientific Co. (300 Edscorp Building, Barrington, N.J. 08007). The weight in pounds that a helium-filled balloon of spherical shape can lift is roughly equal to a quarter of the cube of its radius in feet. To my delight the motor began to run slowly when the tinsel reached an altitude of about 100 feet. At 300 feet the rotor made between 500 and 700 revolutions per minute.

A note of warning is appropriate at this point. Although a 300-foot vertical antenna can be handled safely in fair weather, it can pick up a lethal charge during thunderstorms. Franklin was incredibly lucky to have survived his celebrated kite experiment. A European investigator who tried to duplicate Franklin's observations was killed by a bolt of lightning. The 300-foot antenna wire can hold enough charge to give a substantial jolt, even during fair weather. Always ground the lower end of the wire when it is not supplying a load, such as the motor.

To run the motor connect the antenna to one set of electrodes and ground the other set. Do not connect the antenna to an insulated object of substantial size,




*Circuit arrangement for the slot-effect electret motor*




*Jefimenko's slot-effect electret motor*

such as an automobile. A hazardous charge can accumulate. Never fly the balloon in a city or in any other location where the antenna can drift into contact with a high-voltage power line. Never fly it below clouds or leave it aloft unattended.

A variety of corona motors have been constructed in Jefimenko's laboratory. He has learned that their performance can be vastly improved by properly shaping the corona-producing electrodes [*see bottom illustration on page 128*]. The working surface of the rotors should be made of a fairly thin plastic, such as Plexiglas or Mylar. Moreover, as I have mentioned, the inner surface of the cylinder should be backed by conducting foil to enhance the corona. Effective cylinders can be formed inexpensively out of plastic sewer pipe. Corona rotors can of course also be made in the form of disks.

One model consists of a series of disks mounted on a common shaft. Double-edged electrodes placed radially between adjacent disks function much like Poggendorff's combs. This design needs no foil lining or backing because a potential gradient exists between electrodes on opposite sides of the disks. It is even possible to build a linear corona motor, a design that serves to achieve translational motion. A strip of plastic is placed between sets of knife-edge electrodes slanted to initiate motion in the desired direction.

Notwithstanding the problem of handling potentials on the order of a million volts without effective insulation materials, Jefimenko foresees the possibility of at least limited application of corona power machines. In *The Physics Teacher* (March, 1971) he and David K. Walker wrote: "These motors could be very useful for direct operation from high-voltage d.c. transmission lines as, for example, the 800 kV Pacific Northwest-Southwest Intertie, which is now being constructed between the Columbia River basin and California. It is conceivable that such motors could replace the complex installations now needed for converting the high-voltage d.c. to low-voltage a.c. All that would be required if corona motors were used for this purpose would be to operate local low-voltage a.c. generators from corona motors powered directly from the high-voltage d.c. line."

As Jefimenko points out, a limiting factor of the corona motor is its required minimum potential of 2,000 volts. This limitation is circumvented by a novel electrostatic motor invented in 1961 by a Russian physicist, A. N. Gubkin. The

# SCIENCE/SCOPE

Indonesia will unite its 5,000 islands with a synchronous communications satellite system -- first in the Eastern Hemisphere -- which will provide nationwide television for education and entertainment and island-to-island telephone service for most of its 120 million people. Mountains, jungles, and vast stretches of ocean have made it difficult to link existing telecommunications with conventional microwave relays and submarine cables. Lack of communications in remote locations has hindered development of many mineral-rich areas of the 3,000-mile archipelago.

Work is now under way at Hughes to define the Indonesia system, which includes two satellites and 50 earth stations. The satellites are similar to those now in service for Telesat Canada and Western Union, which were also built by Hughes. Operations are scheduled to begin in mid-1976.

A mobile dockside trainer for U.S. Navy ship combat teams is now being built by Hughes under a recently awarded contract. Called MISTER (for Mobile Integrated System Trainer, Evaluator, and Recorder), it will allow the Navy to prepare and execute simulated fleet exercises by simultaneously presenting realistic tactical problems to the combat teams of two ships and to evaluate their performance. Up to five MISTERS can be tied into one exercise to provide coordinated task-force-level training. Because the exercises are controlled by minicomputers, the combat teams train at actual battle stations without need for ship deployment or electromagnetic radiation, which results in a significant reduction in fuel and other operating costs.

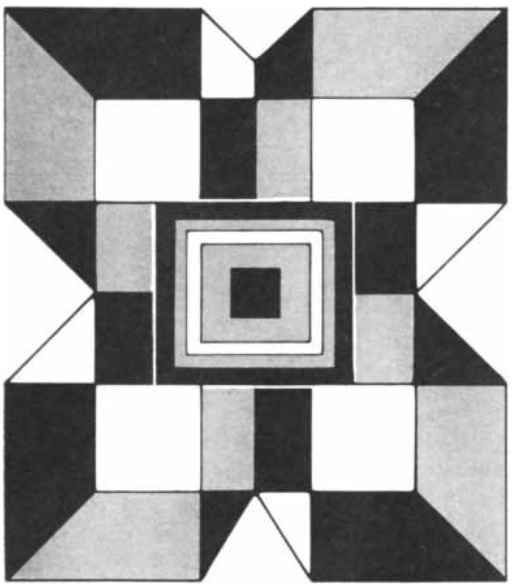
The Hellfire (Helicopter-Launched Fire and Forget) missile, which Hughes is now developing under contract to the U.S. Army Missile Command, is an advanced modular missile for use against tanks, vehicles, and ground emplacements. The next-generation missile beyond the Army's present TOW, Hellfire has a greater range and can use four different seeker-heads -- laser, radar frequency-infrared, optical contrast, or imaging infrared -- which give it an all-weather, day-night capability. Hughes has one year to build a prototype missile, launcher, functional cockpit mockup, and other equipment for a competitive evaluation.

Systems Engineers and Analysts: Our Field Service and Support Division has personnel requirements for design and developmental work on the MISTER program for: logic design engineers (BS or MS in EE or Physics, digital design experience); system engineers (BSEE or Physics, training simulator experience); real-time programmer/analysts (technical degree, real-time simulation experience). U.S. citizenship required. Please send your resume to: D.I. Wingate, Hughes Aircraft Company, P.O. Box 90515, Los Angeles, CA 90009. An equal opportunity M/F employer.

X-ray lithography for batch production of microelectronic devices has been brought closer to practical application by Hughes research scientists. Their technique involves exposing X-ray sensitive polymeric resists through a high-resolution mask generated by scanning electron-beam systems. Advantages of X-ray lithography include large-area parallel exposure, 0.2-micron resolution, off-contact exposure, insensitivity to dust and contamination, use of positive or negative resists, uniform exposure with depth, and no requirement to place mask or substrate in a vacuum.

# Announcing
# The Portable University

**THE PORTABLE UNIVERSITY** is a new series of books offering concise, nontechnical insights into topics of current and continuing interest. All of the authors are experts in their fields and they share with their readers a wealth of fascinating, hard-to-get information, the results of years of research and study. This is information that even the most knowledgeable layman rarely sees; you'll learn more from one satisfying reading of any of these books than from weeks of your own research. And if you want to learn more about a topic after reading the book, you'll have a head start—each contains a helpful, annotated reader's guide.



**CONCEPTUAL BLOCKBUSTING** by Professor James L. Adams. Conceptual blocks are mental walls that get between you and the solutions to your problems. This book is a fresh, practical, and often humorous guide to breaking down those walls and improving your creative abilities.

**CHINA The People's Republic of China and Richard Nixon** by Professor Claude A. Buss examines the many factors that contributed to the decision by the People's Republic and the United States to begin establishing normal diplomatic relations.

**SOME MUST WATCH WHILE SOME MUST SLEEP** by Dr. William C. Dement reports on the scientific study of sleep, dreams, and sleeping disorders. Sleep appears to be a highly dynamic state with a complexity rivaling waking consciousness.

**JAPAN The New Superstate** by Professor Nobutaka Ike discusses Japan's astonishingly rapid growth from the ashes and ruin of World War II to a modern industrial giant. What features of Japanese society contributed to this growth and what are that nation's options for the future?

**HUMAN SEXUALITY Sense and Nonsense** by Dr. Herant Katchadourian is a straightforward, nonjudgmental survey of what scientists know and don't know about the biological, psychological, and social aspects of the many forms of human sexuality.

**IS MAN INCOMPREHENSIBLE TO MAN?** by Professor Philip H. Rhinelander. Can we agree on a definition of "human nature?" Here is a thoughtful, contemporary survey of philosophy from Aristotle and Thucydides to Marx, McLuhan, Marcuse, and Skinner that never loses its focus on the question: who is man?

—Each title is available for $5.95 cloth and $2.95 paper and comes with a full money-back guarantee.
—New titles will be regularly added to the series.
—These books were published originally as a part of *The Portable Stanford*, a series of books published by the Stanford Alumni Association. They have never before been offered to the general reading public.
Illustration based on a drawing by Jim M'Guinness

**W. H. FREEMAN AND COMPANY**
660 Market Street, San Francisco, California 94104
58 Kings Road, Reading, England RG1 3AA

motor is based on an electret made in 1922 by Mototaro Eguchi, professor of physics at the Higher Naval College in Tokyo.

An electret is a sheet or slab of waxy dielectric material that supports an electric field, much as a permanent magnet carries a magnetic field. Strongly charged carnauba-wax electrets are available commercially, along with other electrostatic devices, from the Electret Scientific Company (P.O. Box 4132, Star City, W.Va. 26505). A recipe for an effective electret material is 45 percent carnauba wax, 45 percent water-white rosin and 10 percent white beeswax. Some experimenters substitute Halowax for the rosin.

The ingredients are melted and left to cool to the solid phase in a direct-current electric field of several thousand volts. The wax continues to support the field even though the external source of potential is turned off [see "The Amateur Scientist," SCIENTIFIC AMERICAN, November, 1960, and July, 1968]. The electret reacts to neighboring charges exactly as though it were a charged electrode, that is, it is physically attracted or repelled depending on the polarity of the neighboring electrode.

Gubkin harnessed this effect to make a motor. The rotor consisted of a pair of electrets in the shape of sectors supported at opposite ends of a shaft. The center of the shaft was supported transversely by an axle. When the rotor turned, the electrets were swept between adjacent pairs of charged metallic plates, which were also in the form of sectors.

The plates were electrified by an external source of power through the polarity-reversing switch known as a commutator. The commutator applied to the electrodes a charge of polarity opposite to the charge of the attracted electret. As the electret moved between the attracting plates, however, the commutator switched the plates to matching polarity. The alternate push and pull imparted momentum to the rotor in exact analogy to Franklin's motor.

Gubkin's motor was deficient in two major respects. The distances between the electrodes and the electrets were needlessly large, so that the forces of attraction and repulsion were needlessly weak. Moreover, during the electret's transit between electrodes its surfaces were unshielded. Unshielded electrets attract neutralizing ions from the air and lose their charge within hours or days.

Both inherent deficiencies of Gubkin's motor have been corrected in Jefimenko's laboratory by taking advantage of what is termed the slot effect. Instead of sandwiching the electret alternately between pairs of metal plates, Jefimenko employs opposing pairs of adjacent plates [*see illustration on page 129*]. The adjacent plates are separated by a narrow slot. When adjacent plates carry charges of opposite polarity, the electret experiences a force at right angles to the slot and in the plane of the electret. The strength of the force is at a maximum because the electret is close to the electrodes. Simultaneously the electrodes function as shields to prevent the neutralization of the electret by free ions.

Motors based on the slot effect can be designed in a number of forms. One design consists of an electret in the shape of a wafer-thin sheet of Mylar supported by a flat disk of balsa wood 100 millimeters in diameter and three millimeters thick. (A long-lasting charge is imparted to the Mylar by immersing it in a field of a few thousand volts from an electrostatic generator after the motor is assembled.) This rotor is sandwiched between four semicircular sectors that are cross-connected [*see top illustration on page 130*].

The electret is mounted on a four-millimeter shaft of plastic that turns in jeweled bearings. The conducting surfaces of the commutator consist of dried India ink. The brushes are one-millimeter strips of kitchen aluminum foil. The motor operates on a few microwatts of power.

Jefimenko has demonstrated a similar motor that was designed to turn at a rate of about 60 revolutions per minute and develop a millionth of a horsepower on a 24-foot antenna having a small polonium probe at its upper end. (By emitting positive charges probes of this type tap the earth's field somewhat more efficiently than needle points do.) The performance of the motor easily met the design specifications. The charm of these motors lies in the fact that, although they do not accomplish very much, they can run forever.

Paul R. Burnett (2401 32nd Street, S.E., Washington, D.C. 20020) suggests a simple procedure to extract nitrogen from air for use in the laser described in this department in June. Simply burn natural gas, propane or even alcohol and let the aspirator of the laser pull the combustion products successively through a bed of lime and a bed of anhydrous calcium sulfate respectively. Nitrogen of adequate purity will be drawn into the laser. A cold trap chilled with dry ice and acetone or any other desiccant can be substituted for the calcium sulfate.

# Early man... math ...the revolution in earth sciences ...life itself... history of science ...computers...

to keep up with everything that's happening in every scientific area that interests you — get set now for a great reading spree!

# Take any 3 books all for only $3.95

(values to $55.45) as a member of The

# Library of Science

(if you will join now and agree to accept only 3 more books — at handsome discounts — within the next 12 months)

**HOW THE CLUB OPERATES** • The Book Club **News,** describing the coming Main Selection and Alternate Selections, will be sent to you 15 times a year at three to four week intervals. • If you wish to purchase the Main Selection, do nothing and it will be shipped to you automatically • If you prefer one of the Alternates, or no book at all, simply indicate your decision on the reply form always enclosed with the **News** and mail it so we receive it by the date specified • The **News** is mailed in time to allow you at least 10 days to decide if you want the coming Main Selection. If, because of late mail delivery of the **News,** you should ever receive a Main Selection without having had the 10-day consideration period, that Selection may be returned at Club expense • After completing your trial membership, you will be entitled to take advantage of our fabulous bonus plan

**The Library of Science** 2-47A
Riverside, New Jersey 08075

Please accept my application for membership and send me the three volumes indicated, billing me only $3.95 for all three. I agree to purchase at least three additional Selections or Alternates during the first year I am a member, under the club plan described in this ad. Savings range up to 30% and occasionally even more. My membership is cancelable any time after I buy these three books. A shipping and handling charge is added to all shipments. **Send no money. Members are billed when books arrive.**

**3 books for $3.95**

Some expensive books count as 2 selections.

Name _____

Address _____

City _____ State _____ Zip _____

Book selections purchased for professional purposes may be a tax-deductible expense. (Offer good in Continental U.S. and Canada only. Prices slightly higher in Canada.)

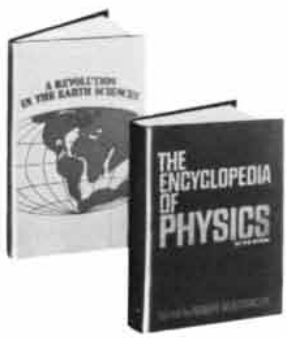
**55200. THE INCREDIBLE EGG.** *Dal Stivens.* Australia's pre-eminent nature writer explores the range of mating, nesting, and protective behavior that has brought tens of thousands of backboned, egg-producing creatures alive and well through millions of years of unending competition for food and space **$9.95**

**62400. MESSAGES OF THE BODY.** *John P. Spiegel and Pavel Machotka.* Fascinating look at how we communicate without speech — through gestures, facial expressions, posture. The first truly scientific study. Crammed with examples and experiments. **$17.95**

**87350. WORD PLAY.** *Peter Farb* A fascinating examination of language as behavior. 'It is so absorbing, so curious, amusing, and enlightening, that we almost inadvertently learn a great deal about linguistics.' — Christopher Lehmann-Haupt in the *New York Times* **$8.95**

**78820. THE SOCIAL BEHAVIOR OF THE BEES.** *Charles D Michener* One of the world's foremost authorities on bees offers the first definitive work on group phenomena among these most amazing insects. A publishing classic. *Counts as 2 of your 3 books.* **$25.00**

**74145. RIEGEL'S HANDBOOK OF INDUSTRIAL CHEMISTRY.** *James A. Kent* The one and only classic that draws on the expertise of the foremost authorities in every field of chemical processing. A complete broad-scale reference work covering chemistry applied to fertilizers, pharmaceuticals, dyes, plastics and a host of other materials. *Counts as 2 of your 3 books* **$37.50**




(retail prices shown)

**74080. A REVOLUTION IN THE EARTH SCIENCES.** *Anthony Hallam* A noted English geologist presents a serious yet non-technical history of the fascinating theory of continental drift. .concise and well done — *Choice* **$10.00**

**44941. THE ENCYCLOPEDIA OF PHYSICS. Second Edition.** *Edited by Robert H. Besancon.* In this updated and expanded edition of a modern classic, over 300 distinguished authorities in 344 articles offer incisive, authoritative information on every major topic in physics today. Outsized volume contains more than 1000 pages. *Counts as 2 of your 3 books* **$37.50**

**44000. EINSTEIN AND THE GENERATIONS OF SCIENCE.** *Lewis S. Feuer* What really accounts for the shattering insights of Einstein, Bohr, de Broglie, Heisenberg and other great scientists? Lewis Feuer points to a generation conflict as the true cause of genius. A ground-breaking book with a bold and controversial explanation **$12.95**

**53920. HUMAN EVOLUTION. Second Edition.** *Dr. Bernard Campbell* Captivating view of human development which synthesizes the major findings of modern research and theory on evolution — anatomical, physiological, and psychosocial **$12.50**

**58620. THE LIFE GAME.** *Nigel Calder* Reporting on the dazzling research of the New Biology, the author of *The Violent Universe, The Mind of Man,* and *The Restless Earth* explores the ultimate biological question: Why do we exist? Provocative reading. **$12.50**

**67630. PEDIGREE: WORDS FROM NATURE.** *Stephen Potter and Laurens Sargent.* The spellbinding history of the origins of thousands of words making up the vocabulary of natural science. Written with gusto and rare narrative skill and full of unusual, interesting information. **$9.95**


**41300. CULTS OF UNREASON.** *Christopher Evans.* A treasure of information on unreason today — the mystics, the flying saucer cults, the gurus, the humorous and often unbelievable attempts to ignore man's knowledge of science. **$7.95**

**63061. MOONS AND PLANETS.** *William K. Hartmann.* An outstanding introduction to current knowledge of the solar system. A masterpiece of factual information linking astronomy, geology, meteorology, biology and philosophy **$12.95**

**39635. COMETS, METEORITES, AND MEN.** *Peter Lancaster Brown* Delightfully informative presentation of comets and meteorites: their nature, what they tell about the solar system, their role in astrology and legend, and how they are discovered, named and calculated. Wealth of hard-to-find facts, practical advice for star gazers, and ribtickling anecdotes **$12.50**

**36520. THE BONE HUNTERS.** *Url Lanham* Absorbing history of the nineteenth-century adventurer-scientists who startled the scientific world, shattered many near-sacred illusions, and laid the foundations of modern paleontology. Filled with rare old photographs. **$12.95**

**56322. INTRODUCTION TO CALCULUS AND ANALYSIS. Volume I and Volume II.** *Richard Courant and Fritz John.* Two distinguished mathematicians present a lucid and up-to-date introduction to this fascinating subject, stressing the interaction between mathematical analysis and its applications, as well as the reciprocity between differential and integral calculus. *Counts as 2 of your 3 books* **$29.90**

**39830. COMPUTABILITY AND LOGIC.** *George Boolos and Richard Jeffrey* A lucid introduction to the more advanced areas of logic. The Skolem-Lowenheim theorem, representability in Q, first- and second-order logic, monadic vs. dyadic logic — these and several dozen other stimulating topics are all succinctly and clearly discussed **$10.95**

**60670. MATHEMATICS APPLIED TO THE DETERMINISTIC PROBLEMS IN THE NATURAL SCIENCES.** *C. C. Lin and Lee H. Segel* Two distinguished mathematicians provide a well-written account of the physical side of applied math. A handy reference for anyone using math **$15.95**

# Keys that solve complex mathematical calculations as easily as simple arithmetic.

**On/Off** – Slide to the right applies power, slide to the left removes power. Power-on indicated by a number in the display.

**DEG/RAD Switch.** Interprets displayed angle in degrees or in radians.

**ARC.** When pressed as prefix to sin, cos, or tan key, tells calculator to determine the inverse trig function.

**Sine.** Determines the sine of the displayed angle.

**Cosine.** Determines the cosine of the displayed angle.

**Tangent.** Determines the tangent of the displayed angle.

**Clear.** Clears the calculator. The memory is not affected.

**Hyperbolic Function Key.** Instructs the calculator to determine the hyperbolic function of the displayed value when pressed as a prefix to the sin, cos, or tan key.

**Angle Change Key.** If the DEG/RAD switch is set for degrees, pressing the D/R key instructs the calculator to convert the displayed angle from radians to degrees. If the switch is set for radians, pressing this key instructs the calculator to convert the displayed angle from degrees to radians.

**Natural Logarithm.** Determines logarithm to the base e of the displayed number.

**e to the x Power.** Raises e to the power of the displayed number.

**Common Logarithm.** Finds the logarithm to base 10 of the displayed number.

**Square.** Squares the number displayed.

**Square Root.** Finds the square root of the number displayed.

**Reciprocal.** Finds the reciprocal of the number displayed.

**Factorial.** Finds the factorial of the number displayed.

**The xth Root of y.** Finds the xth root of a y number.

**Store.** Stores the displayed quantity in the memory.

**Recall.** Retrieves stored data from the memory.

**Sum and Store.** Adds the displayed number algebraically to the number in the memory.

**Exchange.** Tells calculator to exchange the x and y quantities in $y^x$ or $\sqrt[x]{y}$ before the function is processed. Operands in times (x) and divide (÷) can also be changed with this key.

**y to the x Power.** Raises a number to a power.

**Clear Entry.** Clears last keyboard entry.

**Enter Exponent.** Enters the subsequent number as exponent of 10.

**PI.** Enters pi to 13 significant digits. Display indicates value rounded off to 10 significant digits.

**Digit Keys.** Enter numbers 0 through 9 to a limit of a 10-digit mantissa and a 2-digit exponent.

**Decimal Point.** Decimal is assumed to be to the right of any number entered unless it's positioned in another sequence with this key.

**Change Sign.** Changes the sign of either the mantissa or exponent.


**Display.** Power-on and numerical information. Provides indication of a negative number, decimal point, overflow, underflow and error.

**Overflow & Underflow Indications.** Display flashes when number entered or calculation result is entered or calculation is larger than $\pm 9.999999999 \times 10^{99}$ and when number entered or calculation is closer to zero than $\pm 1. \times 10^{-99}$.

**Fast Rechargeable Battery Pack.** Provides 4 to 6 hours' operation without recharging. Recharging for about 3 hours restores full charge.

**Fast and easy. No calculator in its class is so easy to master.**

The SR-50 lets you key the problem just as you would state it. A unique register system provides a sum of products capability directly at the keyboard. This ability to store the first product while the second is being calculated is in addition to the memory accessed by the memory keys. The efficiencies are suggested by this simple problem:

$(2 \times 3) + (4 \times 5) = 26$

TI's Algebraic Entry Method:

2 [×] 3 [+] 4 [×] 5 [=] 26

Reverse Polish Entry Method:

2 [↑] 3 [×] 4 [↑] 5 [×] [+] 26

# The reason the SR-50 has so much math power at the price lies beneath the keyboard.

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

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

TI is steeped in calculator technologies from start to finish, making all critical parts and controlling quality every step of the way. And that's the key to the exceptional quality and value of the SR-50.

**15-day no-risk evaluation**
What's more, we'll let you experiment with your SR-50 for 15 days at no risk. Then if you change your mind, we'll refund your money without question.

Don't put off getting your SR-50 any longer. The SR-50 has the math power today's work world requires. In the office or in the classroom. In the lab or on the drawing board.

# Evaluate an SR-50 for 15 days at no risk.

Mail this coupon today. Try the SR-50 for 15 days at no risk. Send this order form and enclose a check, money order, or company purchase order of $149.95 for each SR-50. Use the SR-50 for 15 days. Then, if you decide it's not the best value you've seen, just return it in the original carton with all accessories for a full refund.

**TO: Texas Instruments Incorporated**
**P.O. Box 3640, M/S 84M, Dallas, Texas 75221**

Dear Sir:

I enclose a ( ) check ( ) money order ( ) company purchase order for $_____ for the purchase of _____ SR-50(s). I understand that I'll get my money back if anytime during the 15-day trial I'm not completely satisfied.

Name _____

Address _____

City/State/Zip _____

Company _____

Please add state and local taxes where applicable.*

SEMI

*States requiring submission of taxes: AZ, CA, CO, FL, IL, IN, KY, MA, MI, MN, NJ, NM, NY, PA, TN, TX, UT, VA, WA.


The SR-50 includes AC adapter/charger which operates on either 115V/60 Hz or 230V/50 Hz; cushioned vinyl carrying case; and user's manual/application guide.

**Full year warranty.**
Texas Instruments warrants each scientific calculator for a period of one year against defective parts and workmanship. If you ever do need service, you can get it through TI's Customer Service Center.

TEXAS INSTRUMENTS
INCORPORATED

# BOOKS

## *The background light of the night sky and the architecture of the pyramids*

by Philip Morrison

THE LIGHT OF THE NIGHT SKY, by F. E. Roach and Janet L. Gordon. D. Reidel Publishing Company ($16.50). "As Earth rotates within the sunlight sheath" blue day gives way to the sunset drama until at last the stars come out. The darkening sky is still blue to the camera, although not to the eye, even in mid-twilight; the sun's rays are still scattered by the particles in the high air, but those particles are so sparse that the skylight intensity has fallen far below the color threshold of the eye. At the official end of twilight the nearest rays of the sun illuminate the zenith 200 miles above the ground; sometimes a satellite up there is caught for a while in the beam. In the absence of the moon and clouds the night sky is 400,000 times fainter than the sky at sunset. But even the midnight sky is not truly dark; that can easily be proved by looking at a part of the sky where no stars appear to the unaided vision and then holding a black card up to it.

Photons rain down on us detectably from the entire universe. This little textbook—varied, personal and comprehensive—tells where they come from and how we know. Photometry by eye, camera and digitized photomultiplier, stations in orbit, in the oasis of Tamenrasset and two miles above the sea on a Hawaiian crater (where the authors work) and star counts by William Herschel and by Bart Bok and his Harvard "star-counting brigade" are discussed, with sciences from free-radical chemistry to cosmology, at a level that requires some previous knowledge but little mathematics. Results are graphically displayed (there are 70 figures) and the scientific literature is widely cited.

Five different components of the night-sky light are assessed, a paradigm of successive approximation. Brightest of all shines the fine dust of our solar system, spread out in the orbital plane of the planets and illuminated by the sun. The frontispiece shows the oval patch of the zodiacal light seen from the vantage of the Hawaiian crater edge. The stuff of the zodiacal cloud is as ever changing as a river; it is steadily lost to space and the sun and is resupplied in a cycle of 100,000 years, probably from the loss of mass by comets and the grinding collisions of asteroidal material. The cloud is much denser than its gossamer kin, the interstellar dust; if the interstellar material were as dense as the solar, the stars of Orion would be invisible.

The stars are the next component. They have been sampled ever since William Herschel counted off to his devoted sister Caroline the stars drifting into his eyepiece. (His data compare well with modern numbers to his limit of about the 14th magnitude.) By extrapolation and sampling we can reckon that the sphere of stars (mostly in the Milky Way) equals in visible light about 100 stars as bright as Sirius.

Our location in space is fortunate in many ways. If we lived in a globular cluster, the sky would dazzle us with bright stars. With any luck the Sirius of that sky would blaze out nearly 10 times as bright as our Venus and would be easily visible by day, but the galaxy would be overwhelmed, and perhaps we would not yet have learned of other galaxies. (Of course, radio intercommunication on a few months' round-trip time basis might be possible; "the local SPACE organizations" would be busy planning interstellar travel.)

The bright, colorful polar aurora, excited in the upper air by fast particles energized by streams from the sun, distinctly masks faint stars whenever it occurs. Fortunately for astronomers it is sporadic and zonal; the treatment here is brief. There is, however, the ubiquitous night glow. It is dynamic and shifting, but it is worldwide and not sporadic. It is not as bright as the aurora; it is striking only to instruments. That is because it shines mainly in the deep red, where our eyes are only a tenth as sensitive as they are in the auroral green. The astronaut sees it as a luminous arc on the curved limb of the earth; he looks through the region of its origin edgewise, so that it is effectively 30 times thicker than it is when it is viewed at the zenith from the ground. The main source of energy for this glow (there are several components) is the night-long chemical reaction among the sunlight-excited atoms of the thin air some 90 kilometers up. Most of its photons are infrared: vibratory emissions of the hydroxyl radical, probably excited by the reaction of atomic hydrogen with ozone. If this reaction were capable of emitting green photons with fair efficiency, it would cause a fairly bright permanent aurora and deep-sky optical astronomy could hardly exist.

The Milky Way contributes another diffuse light, a pedestal on the graph above which the individual star spikes stand like a grove of bamboo. That is a sign of galactic dust. An entire chapter of the book discusses the distribution of matter by size in our galaxy, from gas atoms, dust, pebbles, boulders, rocks, comets, moons and asteroids, rocky and gassy planets to stars, dwarf and giant. Between the dust and the stars all we have is a tiny local sample and a penchant for smooth interpolation. This simple and unreliable model predicts the existence of five faint stars closer to us than Alpha Centauri, and a large number of major and minor planets!

It is striking that all the main sources of sky light (apart from the light of cities and the moon) are of the same order of magnitude in brightness. There is one grand exception. The still unclaimed prize of those who gauge the glow of the night sky is the real cosmic light: the superimposed emission of the distant galaxies. They spread out in all directions, until we are looking so far out that the passage of light takes us back to the time before the galaxies and their stars were born out of cosmic hydrogen. An upper limit can be set at some 5 percent of the total sky light. Plausible cosmological estimates put the expected value rather lower: about 1 percent. Already we have moved our stations out to cislunar space,

well beyond the aurora and the night glow. It is not much use to think of moving among the stars, unless we can hope to leave the galaxy, reducing all starlight to a mere patch. Our most plausible expedition, then, is one to the outer planets; there the zodiacal light is much fainter. A large telescope looking toward the galactic pole from such precincts is our best hope for a clear view of the cosmic light. The radio observers may do better here on the earth, but let the visible observers dream of an instrument in orbit around Jupiter or its satellite Ganymede: "Put down this book and take... a cursory glance at the night sky—on a clear night, you can see forever!"

The book is an outstanding intermediate text, as refreshing and as wide as the night sky and certainly worth its cost of 13 cents a page. Its two-page biographies and informal air are winning additions, although one misses two or three pages about the instrumentation these artist-experts use.

THE PYRAMIDS, by Ahmed Fakhry. The University of Chicago Press ($3.25). THE RIDDLE OF THE PYRAMIDS, by Kurt Mendelssohn. Praeger Publishers ($12.95). Professor Fakhry's book is a small paperback, yet it is careful and comprehensive and imparts the pride and enthusiasm of a man who lived and worked for decades at the pyramids; his desert dwelling was almost literally within their shadows. It is an admirable summary, replete with dates, diagrams, maps and many photographs. One remarkable chapter treats the Sphinx, "the great man-headed lion" the Egyptians still regard as "a source of endless wisdom and hope for the future." The builders of the Second Pyramid at Giza excavated pyramid stone from a quarry where the Sphinx now stands. Then the Sphinx was carved out of a great core of softer rock that had been left behind as unusable. Rather than simply haul away the unsightly mass whose natural form may have suggested a crouching beast, the architects and their workmen brilliantly transmuted the "eyesore into a sublime monument." Ever since that time, over the entire ebb and flow of Egyptian history, it has been sand-buried and recovered, embellished, vandalized or celebrated.

Fakhry, who died last year, was a professional of the most devoted kind, full of learning and yet resilient enough to welcome the bold proposal to radiograph the Second Pyramid of Giza with cosmic ray muons. (No internal tomb is hidden there, the spark chambers say, too recently for this book.) The second of these books, which has color plates and a couple of valuable air views to augment more conventional photographs and drawings, is the novel work of an amateur Egyptologist. He is no amateur at science but a well-known low-temperature physicist of Berlin and Oxford. He too has watchfully climbed his pyramids; he presents here a persuasive and novel set of physicist's inferences from study of those man-made hills.

The essential background of the new arguments is as follows. The rulers of Egypt built pyramids off and on for two millenniums. Just five of them dwarf the rest in volume. (The others "were well within the ordinary budget," Mendelssohn drily remarks.) The first of all was the Step Pyramid, built under Zoser. Then came the pyramid at Meidum and the two at Dahshur, all finished under Snefru. Snefru's son Khufu built the Great Pyramid at Giza, 15 miles south of Dahshur. The architects of Giza had been apprentices at Dahshur and the grand style was continued under Khufu's son and heir, but there it stopped. Few later works have as much as a tenth the bulk of these five, built within two centuries, and mainly in the 100 years centered on 2600 B.C., the pyramid age.

The central clue is at Meidum, where the pyramid "looks like a colossal tower standing at the edge of a hill." It has no smooth sloping walls but three great steps surmounting a conical mound of rubble. The rubble has always been thought to be the work of stone robbers, but it is far too bulky. No city is near to have used up so much stone. Although robbers have nibbled at the pile, that hill of broken debris is the ruin of a pyramid. Like the mine tip at Aberfan in Wales, the pyramid of Meidum suddenly collapsed while it was still unfinished and poured the bulk of its mantle laterally outward into a roughly circular heap of broken limestone.

A more detailed analysis of the construction shows how it happened. A pyramid of ideal blocks bears its load stably, with each vertical column held firmly on its base. The compressive load is some 25 atmospheres: high, but not high enough to cause the failure of good stone. Irregular blocks, however, such as those we see through a hole into the Meidum pyramid, touch at only a few points of contact. The loading is heavy, and the material flees it by thrusting outward. One day, perhaps after a lubricating rainstorm like Aberfan's, the growing outer mantle flowed catastrophically out and down.

So much is structural inference. The next pyramid, however, is unique; it is

the Bent Pyramid, so named because at one level its slope sharply decreases from 54 degrees (about like Meidum) to 43 degrees. The story unfolds: This change was made not to save stone (the next pyramid was built entirely at the shallower angle) but to forestall another collapse. A detailed study of the inner structure and workmanship of the two pyramids supports the case strongly. The pyramids at Dahshur fit a logical account at last, and the boldness and skill of the builders of the steep Great Pyramid emerge even more strongly. (Its masonry is remarkably well dressed, although at great cost in labor.)

The argument continues. For the builders at Dahshur to have bent their pyramid for safety, they must have been partway along, working on a course some 50 meters above the ground before the mantle of Meidum was complete. The pyramid-building schedules overlapped. Snefru must have built pyramids all his life, and so his son must have done. No king, however great, needs three tombs; the pyramids were not built solely as tombs, although they served as tombs. Their great centrally directed work force could not be disbanded and recalled each generation. It remained mobilized for a century. "The pyramids do not represent an aim in itself but the means to achieve an aim: the creation of a new form of society. These huge heaps of stone mark the place where man invented the state." Once the state had been formed, once the peasant villages had been joined under a civil service and a rule of common purpose, pyramids entered the usual budget. Post-Khufu, like post-Apollo, the work force dwindled and the project was normalized. Indeed, Professor Mendelssohn proposes that we too need to build a "space pyramid"; we need a great common task, and with fine Tory sentiment he feels no useful task can fill the bill. Any great useful project, he fears, can be ruined by unexpected side effects. Only outer space is large enough and useless enough, he says, to become the pyramid of our powerful era!

Not as novel a piece of social criticism as perhaps it seems (compare Lewis Mumford, *passim*), the archaeological origins of this argument are fascinating. There is much more, even a plausible, if not compelling, theory of the origin of pyramid slopes, so close to $4/\pi$ and to $3/\pi$, among surveyors apparently ignorant of the transcendental. Travelers to Mexico will particularly profit by a chapter on those pyramids, viewed here as striking confirmation. There too a large-scale effort justified and tested a strong central administration and a new

concept of the state, as enduring as the pyramids.

PSYCHOLINGUISTICS AND READING, by Frank Smith. Holt, Rinehart and Winston, Inc. ($4.95). The "monstrous national effort" of reading pedagogy in the 1970's is one more load borne by U.S. schools in their effort to mend the damage done by the structural defects of the American economy, now and in the past. Apart from all this, a bit anxious but genuinely engaged in helping children learn to read, this Toronto author has assembled an anthology, lively, rather polemic and full of implications. Teachers and parents directly concerned need these 16 chapters (a dozen of them recent reprints by other authors with complementing and background material), and anyone can learn in this intimate theater a good deal about the state of knowledge in the great arena of current views of the human mind.

The point is sharp: Reading is not decoding from sound, nor is it primarily a visual skill. It centers on a deeper level: meaning. The printed page does not even supply most of the information; we read actively, building internal models that we guide, check and elaborate as the string of bits flows by. In order to learn to read realistically you have to know the language, and you do not know language unless you are aware of meaning. (It goes without saying here that it is unlikely you can learn much until you want to learn.) Bilingual readers absorb texts that are mixed out of two languages, without slowing down. They read them aloud with errors that are translations and syntactic improvements. Indeed, a skilled and successful reader practices a sophisticated art that gives him a statistical hierarchy of options. He can always drop down in sophistication, and generally he slows down as the task varies, say for reading mathematics, poetry or proof. (No proofreader in English would miss "Son cheval, suivi by two hounds.") Meaning tells you quickly that "The none tolled hymn..." has its troubles. The model of simple sequential decoding can explain reading only at its lowest and slowest levels, just as a naïve stimulus-response theory can explain pigeon-pecking better than it can language acquisition.

Two small surprises follow this line of analysis. First of all the text is pretty convincing that the fluent reading of alphabetic English is not so very different from reading ideographic Chinese. One mostly recognizes the words as wholes and does not spell them out laboriously letter by letter. It is the digitizing not of reading but of writing that is the real gain of an alphabetic way of recording. Next, that crazy English orthography ("ghoti" for "fish" and the like) is also in part rehabilitated. Carol Chomsky, in a paper reprinted here, observed tellingly that the conventional written symbols mean to convey meaning clearly; they need not also bring unique sounds. The troubles implied in the words "cilia," "kill" and "sill" are not the same as those in "medicate" and "medicine." ("Words that *are* the same *look* the same.") Spelling is more conservative than pronunciation, and meaning often sticks with it. Medikate? Medisine? Who would be helped?

The author ends with a wry little chapter in which he gives 12 golden rules for reading teachers, all highly plausible, many even genuinely popular and in his reasoned view all dead wrong. "Provide immediate feedback." It is a bad idea, because teachers who pounce at once on each misreading will never find out who is reading for meaning; such a learner tends to go back and self-correct his errors as he comes to understand the text. (There is experimental proof, in case you doubt it.) One rule Smith believes in: Respond to what the child is trying to do. One can do that only with understanding, in reading as in anything else worthwhile. Psycholinguistics is no panacea; heaven forfend a new methodological fad. What we do need is understanding, and it seems closer along this route than along any other on the still uncertain map.

DARWIN ON MAN: A PSYCHOLOGICAL STUDY OF SCIENTIFIC CREATIVITY, by Howard E. Gruber, together with Darwin's early and unpublished notebooks, transcribed and annotated by Paul H. Barrett. E. P. Dutton & Co., Inc. ($20). Three men wrote this book in intimate collaboration. One (Barrett) is a biologist at Michigan State, another (Gruber) is a Rutgers psychologist. They "met, as it were, across the library table." On that table in the University of Cambridge Library sat the manuscript notebooks compiled within a couple of years long ago by a promising young scientist, a man of independent means then pondering marriage and building in his mind the framework of what is perhaps the most important single theoretical structure in all science.

Darwin's making of notebooks was no personal idiosyncrasy. The six leather-bound pocket books he used were "well-made manufactured articles," and his method of work had roots in industrial practice, family tradition and British em-

piricism. Although pages were later excised (mostly much later by Darwin for refiling), many of them still exist, and the good paper and ink preserve this intimate record of a mind. The handwriting is the main technical barrier, surmounted by "our having lived with it for a number of years." Four of the notebooks, centering on the transmutation of species, have been published within the past decade or so, but two of them, marked by Darwin later as being "full of Metaphysics on Morals," are newly published here.

*Darwin on Man* has two sections. The first, a little more than half the book, offers an absorbing account of the thought of that London biologist in two crucial years, with ample supporting discussion of the intellectual setting and subtle reflection on the structure of theory Darwin put together. That work was no mere solving of problems, no single great insight, but rather a productive reconstruction of a few important "invariant" elements, joined fruitfully at last by a man who knew pigeon breeding as well as the Galápagos finches. In the second section the two notebooks are given to us in full, with careful annotation by Barrett, the transcriber, and a running commentary by Gruber. Other unpublished material of Darwin's ("Old and USELESS Notes," Darwin called some of it) and a number of key extracts from already published material complete the book.

A diagram of Gruber's sets out "Darwin's changing world view." Before sailing on the *Beagle* Darwin saw the world as two creations, the organic world perfectly adapted to the physical. The *Beagle*—and Charles Lyell's geology—made it plain that the physical world was in continuous lawful change and that the world of life had to follow. With the theory of coral reefs, drafted toward the end of Darwin's voyage, he saw that organisms influenced the environment as well; the coupling was strong and the steady-state universe of Lyell therefore incomplete. Before Darwin had begun to write in the notebooks he had held, as Lamarck had, that the fitness of the living world was secured by the physical induction of biological change. Finally, in the two years before his marriage the theorist, worried but undaunted by his gradually arriving at a more materialist view of man and hence of nature, came to see both worlds, evolving and interacting. The physical world evolved by the laws of geology, the living world by the laws of variation and natural selection. The shifting equilibrium meant change. Only the origin of it remained for the Creator, if he existed, somewhere outside the system of his works. For a

time, it seems, Darwin toyed with the continuous generation of life, before his rereading of Malthus gave him the last push to his deeper structure. In November, 1838, his notebook says: "Three principles will account for all (1) Grandchildren like grandfathers (2) Tendency to small change especially with physical change (3) Great fertility in proportion to support of parents." It was clear enough, even though the puzzling black box of variation waited for genetics to open it.

Most of the text is given over to Darwin on man, important because it early and plainly goes to his perception of the unity of life, in which anatomy and behavior, and with behavior mind, were woven into one fabric. "The problem of the mind cannot be solved by attacking the citadel itself.—the mind is function of body.—we must bring some *stable* foundation to argue from." Explicit self-awareness, trips to the primates at the zoo, a marvelous paper on the behavior of his first-born infant William and finally the late books *The Descent of Man* and *The Expression of the Emotions in Man and Animals* all bear witness to the depth and intensity all his life of the ideas first examined in the intimate notebooks presented here. They are by no means "Victorian" but rather are medically frank. (Darwin's father was, after all, a rationalist practicing country doctor, no prude and no diplomat.) "Sexual desire makes saliva to flow / yes, *certainly* / curious association... ones tendency to kiss, & almost bite, that which one sexually loves." There is room for study of the unconscious personality of this young man, but what we have before us is quite the reverse, a documented and clear effort to follow the branching and flowering of his explicit ideas. The authors plainly love him, this man transparent and open to his friends, yet assuming a quiet, nonpolemic, guarded stance in a world where his views meant intellectual warfare. Here is no neurotic victim of a tyrant father, no timid recluse, no hypochondriac, whether in these years or later. Thought, like the kingdom of life, grows and evolves slowly; the book focuses on "the inner work of synthesis" and not on the claimed *satori* of one chance page. It is interesting to recall that Darwin read Malthus in quiet London reflection years after his field contact with the Galápagos; his counterpart Alfred Russel Wallace thought of Malthus over many years but produced his synthesis of natural evolution only while he was living in the very islands where he made his observations on the diversity of tropical species. Wallace late in life excluded man from evolution; the mind, he held, was spiritual in its nature and origin.

Darwin's life, the scientist's Ramayana, to be reviewed and celebrated in every form and style, is well served by this book. Darwin admirers and all students of the 19th century will need it. Other readers in search of how ideas arise will find it a prize case study, as free of simplistic naïveté as of strong preconception. And what biologist can resist the man who wrote in his notes: "The grand Question which every naturalist ought to have before him when dissecting a whale or classifying a mite, a fungus or an infusorian is 'What are the Laws of Life.' "

BEHAVIOUR OF MICRO-ORGANISMS, coordinated by A. Pérez-Miravete. Plenum Press ($19.50). No scientific editor reviewed these 25 papers, so that the name appearing in the editor's place is that of the scientist who coordinated all the publications reporting the Mexico City Congress of Microbiology a few years ago, from which these contributions stem. The authors—microbiologists from all over the U.S. and from Europe and Japan—are solely responsible, although the papers have been given a pleasingly consistent appearance.

Here is psychology at its farthest limit. The known behavior of single-celled forms—bacteria, algae, pollen, protozoa and spermatozoa are all included—is not complex. Motion is almost the only response we can detect; it can be more or less speedy, a mere creep or as slow as growth. A pollen tube just microns across grows down a length of the cornstalk to hit its target embryo nearly two feet away within one day. There is chemical guidance, all right, but the means remain obscure. Early workers saw that animal sperm cells similarly found their egg targets, but they steadily confused the effects of such simple factors as egg secretions, which tended to immobilize even rather distant sperm, with a crowd of successful spermatozoon navigators. Even in 1966 it was still learnedly held that cells as small and simple as the sperm—from which each reader gets half of his or her genetic legacy—could not seek out their target save by random means. Drawings made from frame-by-frame films here clearly exhibit sperm (of certain marine hydroids) swimming straight as a die to a tiny mass of agar impregnated with an extract of the egg.

Bacteria detect chemical gradients and move toward them. They are discriminating; the colon bacillus *Escherichia coli,* for example, can detect at

least eight different substances. How? We do not know. Whether protozoa can learn, and even what that might mean, is still not well established. When the protozoon no longer responds to repeated mechanical shock (a controlled pinging at the edge of the microscope slide), it is learning very little; on a more familiar scale we might suspect it is just getting tired. Can it be conditioned like Pavlov's dog? Can it discriminate several stimuli? Not yet known. There are internal clock rhythms on this scale of being, and they can be reset and entrained. Light too can induce motion in cells too small to boast an eye; some blue-green algae steer toward the light, although some simply bask in it once they get there by a random walk. *Euglena*, a true nucleated alga, possesses an eye spot, and it can seek or shun light as it chooses (if it can choose). It is no crude detector; only four photons absorbed are enough to induce a perceptible reaction!

The collection includes five papers on the interesting nature of the flagella or the cilia. Except for the jellylike amoebas, which appear to contract internally and thereby to ooze themselves along the substrate toward the sticky ends of their changing pseudopods, flagella or cilia are the source of motion in all single cells. They have a common structure: external spiral microtubular filaments of a peculiar cablelike cross section and a common protein chemistry. This structure moves not by contraction—all lengths are fixed—but by the rigid sliding of one microtubule against another. The structures can bend, and even support a wave of bending, by the sliding of one or another filament in or out along an outer wall. The complexity of this ultramicromechanism, and the degree of completeness with which it is being analyzed against experimental test of detailed predictions, are clear testimony of the rising power of molecular biology. Its success now ranges from the governing DNA double helix, rather static in its archival role, to the intricate $9 + 2$ helical cabling of the restless flagella. So equipped, cells row their way to food, moving without the slightest turbulence or persistent flow, at very low Reynolds numbers, through the watery medium that on their scale is more viscous than the thickest syrup is to us.

The preface tells us that no book has appeared on single-cell behavior since 1906, when the classic volume of H. S. Jennings came out. Progress on a wide front is not yet marked, but the latent power of these new ideas and techniques is boldly present on these pages.

# BIBLIOGRAPHY

*Readers interested in further reading on the subjects covered by articles in this issue may find the lists below helpful.*

## THE INTERNATIONAL CONTROL OF DISARMAMENT


SECURITY IN DISARMAMENT. Edited by Richard J. Barnet and Richard J. Falk. Princeton University Press, 1965.

STUDIES IN WAR AND PEACE. Michael Howard. Viking Press, 1971.

THE GAME OF DISARMAMENT. Alva Myrdal in *Impact of Science on Society,* Vol. 22, No. 3, pages 230–231; July–September, 1972.

CHEMICAL DISARMAMENT: SOME PROBLEMS OF VERIFICATION. Stockholm International Peace Research Institute, 1973.


## THE STRUCTURE OF EMISSION NEBULAS


ATOMS, STARS, AND NEBULAE. Lawrence H. Aller. Harvard University Press, 1971.

PLANETARY NEBULAE. L. H. Aller. *Sky and Telescope,* 1971.

ASTROPHYSICS OF GASEOUS NEBULAE. D. E. Osterbrock. W. H. Freeman and Company, 1974.

PLANETARY NEBULAE. J. S. Miller in *Annual Review of Astronomy and Astrophysics,* Vol. 12; in press.


## HOW CILIA MOVE


STUDIES ON CILIA—III: FURTHER STUDIES ON THE CILIUM TIP AND A "SLIDING FILAMENT" MODEL OF CILIARY MOTILITY. Peter Satir in *The Journal of Cell Biology,* Vol. 39, pages 77–94; 1968.

ADENOSINE TRIPHOSPHATE-INDUCED SLIDING OF TUBULES IN TRYPSIN-TREATED FLAGELLA OF SEA-URCHIN SPERM. Keith E. Summers and I. R. Gibbons in *Proceedings of the National Academy of Sciences of the United States of America,* Vol. 68, No. 12, pages 3092–3096; December, 1971.

CILIA AND FLAGELLA. M. A. Sleigh. Academic Press, 1974.


## NITROGEN FIXATION


THE SYNTHESIS OF AMMONIA FROM ITS ELEMENTS. Fritz Haber. Nobel Lecture, June 2, 1920.

FIXATION OF ATMOSPHERIC NITROGEN IN A GAS HEATED FURNACE. Nathan Gilbert and Farrington Daniels in *Industrial and Engineering Chemistry,* Vol. 40, No. 9, pages 1719–1723; September, 1948.

RADIATION CHEMISTRY OF THE FIXATION OF NITROGEN. P. Harteck and S. Dondes in *Science,* Vol. 146, No. 3640, pages 30–35; October 2, 1964.

BIOLOGICAL NITROGEN FIXATION. R. W. F. Hardy and R. C. Burns in *Annual Review of Biochemistry,* Vol. 37, pages 331–358; 1968.


## THE DIMENSIONS OF STAIRS


ACCIDENTAL INJURIES DUE TO FALLS. Dan D. Gowings in *National Safety Congress Transactions,* Vol. 6, pages 59–62; 1960.

TELEMETERING ELECTROMYOGRAPHY OF MUSCLES USED IN WALKING UP AND DOWN STAIRS. J. Joseph and R. Watson in *Journal of Bone and Joint Surgery,* Vol. 49 B, No. 4, pages 774–780; 1967.

OPTIMUM DIMENSIONS FOR DOMESTIC STAIRCASES. Joan S. Ward and William Beadling in *Architects' Journal,* Vol. 151, No. 8, pages 513–520; February 25, 1970.

PEDESTRIAN PLANNING AND DESIGN. John J. Fruin. Metropolitan Association of Urban Designers and Environmental Planners, 1971.


## MIMICRY IN PARASITIC BIRDS


CHARACTER DISPLACEMENT. W. L. Brown, Jr., and E. O. Wilson in *Systematic Zoology,* Vol. 5, pages 49–64; 1956.

HOMAGE TO SANTA ROSALIA, OR WHY ARE THERE SO MANY KINDS OF ANIMALS? G. E. Hutchinson in *The American Naturalist,* Vol. 93, No. 870, pages 145–159; May–June, 1959.

EVOLUTION IN CHANGING ENVIRONMENTS. Richard Levins. Princeton University Press, 1968.

ON THE METHODS OF RESOURCE DIVISION IN GRASSLAND BIRD COMMUNITIES. Martin L. Cody in *The American Naturalist,* Vol. 102, No. 924, pages 107–147; March–April, 1968.


## THE COORDINATION OF EYE-HEAD MOVEMENTS


CENTRAL CONTROL OF MOVEMENT—NEUROSCIENCES RESEARCH PROGRESS BULLETIN: VOL. 9. Edited by E. V. Evarts, E. Bizzi, R. Burke, M. DeLong and W. T. Thach, 1971.

EYE-HEAD COORDINATION IN MONKEYS: EVIDENCE FOR CENTRALLY PATTERNED ORGANIZATION. Emilio Bizzi, Ronald E. Kalil and Vincenzo Tagliasco in *Science,* Vol. 173, No. 3995, pages 452–454; July 30, 1971.

ADJUSTMENT OF SACCADE CHARACTERISTICS DURING HEAD MOVEMENTS. P. Morasso, E. Bizzi and J. Dichgans in *Experimental Brain Research,* Vol. 16, pages 492–500; 1973.

MECHANISMS UNDERLYING RECOVERY OF EYE-HEAD COORDINATION FOLLOWING BILATERAL LABYRINTHECTOMY IN MONKEYS. J. Dichgans, E. Bizzi, P. Morasso and V. Tagliasco in *Experimental Brain Research,* Vol. 18, pages 548–562; 1973.


## THE EXCAVATION OF A DROWNED GREEK TEMPLE


BALLOON OVER SARDIS. *Archaeology,* Vol. 20, No. 1, pages 67–68; January, 1967.

SHALLOW-WATER EXCAVATION AT KENCHREAI. Joseph W. Shaw in *American Journal of Archaeology,* Vol. 71, No. 3, pages 223–231; July, 1967.

BALLOON OVER HALIEIS. Julian Whittlesey in *Archaeology,* Vol. 21, No. 1, pages 66–67; January, 1968.

EXCAVATIONS AT PORTO CHELI AND VICINITY, PRELIMINARY REPORT, I: HALIEIS, 1962–1968. Michael H. Jameson in *Hesperia,* Vol. 38, No. 3, pages 311–342; July–September, 1969.

ARCHAEOLOGY UNDER WATER. George F. Bass. Penguin Books, 1972.


## MATHEMATICAL GAMES


THE THEORY OF COMMITTEES AND ELECTIONS. Duncan Black. Cambridge University Press, 1958.

VOTING AND THE SUMMATION OF PREFERENCES: AN INTERPRETATIVE BIBLIOGRAPHICAL REVIEW OF SELECTED DEVELOPMENTS DURING THE LAST DECADE. William H. Riker in *The American Political Science Review,* Vol. 55, No. 4, pages 900–911; December, 1961.

TOPICS ON TOURNAMENTS. John W. Moon. Holt, Rinehart and Winston, 1968.


## THE AMATEUR SCIENTIST


ATMOSPHERIC ELECTRICITY. J. Alan Chalmers. Pergamon Press, 1968.

ELECTROSTATIC MOTORS: THEIR HISTORY, TYPES AND PRINCIPLES OF OPERATION. Oleg D. Jefimenko. Electret Scientific Company, 1973.

ELECTROSTATICS AND ITS APPLICATIONS. Edited by A. D. Moore. John Wiley & Sons, 1973.

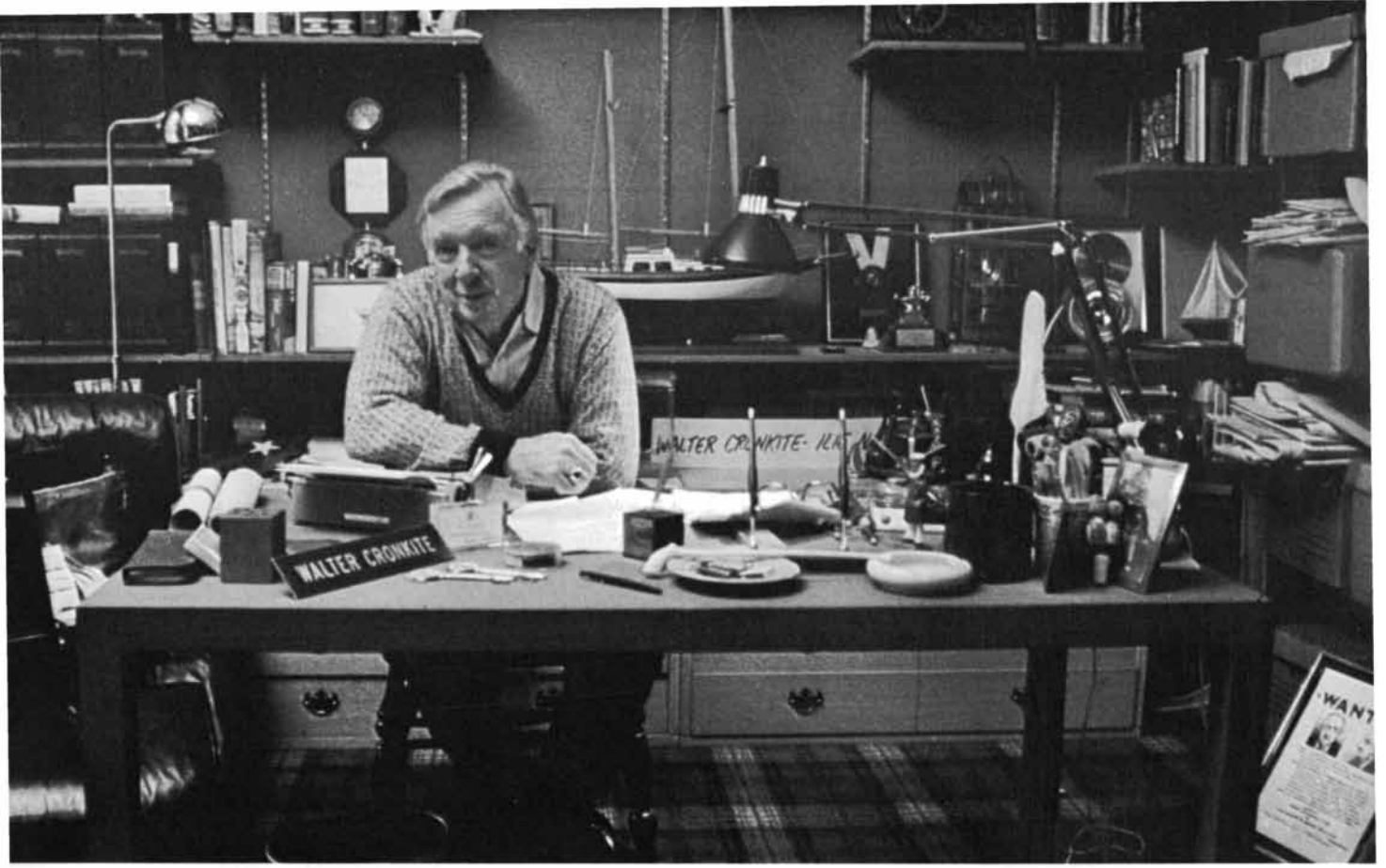



PHOTO: DAN WYNN

# Q. Mr. Cronkite, why is it good business for a U.S. Corporation to help in the education process of people in less developed countries?

**A.** We are not alone on this planet. Our future is closely related to the continuing development of all people. Through massive relief and technical assistance projects we are providing food and clothing to deserving people... but that is not enough. They also are hungry for knowledge and for long range self-development tools.

Education is the key to progress in less developed countries, as it was in the development of our own country. School and reference books, audio-visual material, self-help information, even fun books for children are a scarce commodity in most developing countries... and especially in their own languages of Bengali, Urdu, Arabic, Farsi and Spanish.

The cooperation of U.S. Corporations is needed in this educational process to create a better informed world. Through education the standard of living of people in developing countries can be raised, with far-reaching results for them and indirectly for all of us.

There is an organization that can assist your corporation in this vital educational process.

Franklin Book Programs, Inc. is a nonprofit organization for international book publishing and educational materials development that has been serving successfully since 1952. It is an organization which I have been associated with for the past 8 years, as a Member of the Board of Directors and presently as a Member of the Corporation.

You and your Corporation Contribution Committee should know more about the worldwide educational scope of Franklin. I hope that you will join with me in supporting their efforts to help others help themselves.

Walter Cronkite

*To receive an annual report and for further information, please call 212-683-2500 or write to:*

**Franklin Book Programs, 801 Second Avenue, New York, N.Y. 10017**

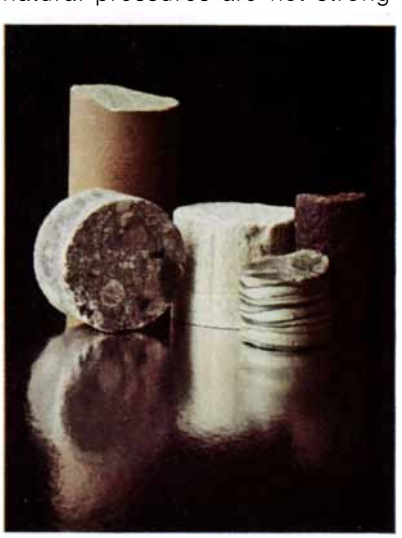
# More oil from

## How Exxon is squeezing additional oil out of a 55-year-old field.

A popular myth has it that oil is found in large underground lakes or pools. Actually, when oil and natural gas are discovered they are inside sedimentary rock. They are trapped in the tiny pores of limestone or sandstone as far as five miles underground.

Usually, when a well is drilled into oil-saturated rock, natural underground pressures force the oil to *flow through the rock* to the well and then to the surface.

However, in many oil fields these natural pressures are not strong enough to move the oil to the well. Or, over time, they may have diminished to the point where they must be supplemented by additional pressure. This additional pressure is provided by "secondary recovery" methods.



*Sample drilling cores taken from oil-saturated rock formations in Alaska, Texas and Saudi Arabia. Oil and gas must come from and flow through rock like this before they reach the well bore.*

**Forcing water deep into the oil-bearing rock.**

The most widespread secondary recovery method in use throughout the oil industry is *waterflooding,* which was introduced over 30 years ago. Water is pumped into the oil-bearing rock to flush out more of the oil than would be produced by natural, or primary forces.

On the average, natural pressures move only about 15-20% of a field's oil from the underground rock up to the surface. By using secondary recovery technology, the final yield can be increased to an average of about 35%.

**The Hewitt Field—a case in point.**

The Hewitt oil field, discovered in 1919, near Ardmore, Oklahoma, is operated by Exxon and several other companies. Originally, this field produced nearly 30 thousand barrels of oil a day. However, the depletion of natural underground pressures gradually reduced Hewitt's production to the point where the field needed waterflooding.

In 1969, Exxon and other owners began a cooperative secondary recovery waterflood project, which has since cost about $20 million.

Engineers now estimate that the Hewitt field ultimately will produce 35% of the original oil in place. Without the use of waterflooding, total recovery had been estimated at only 20%.

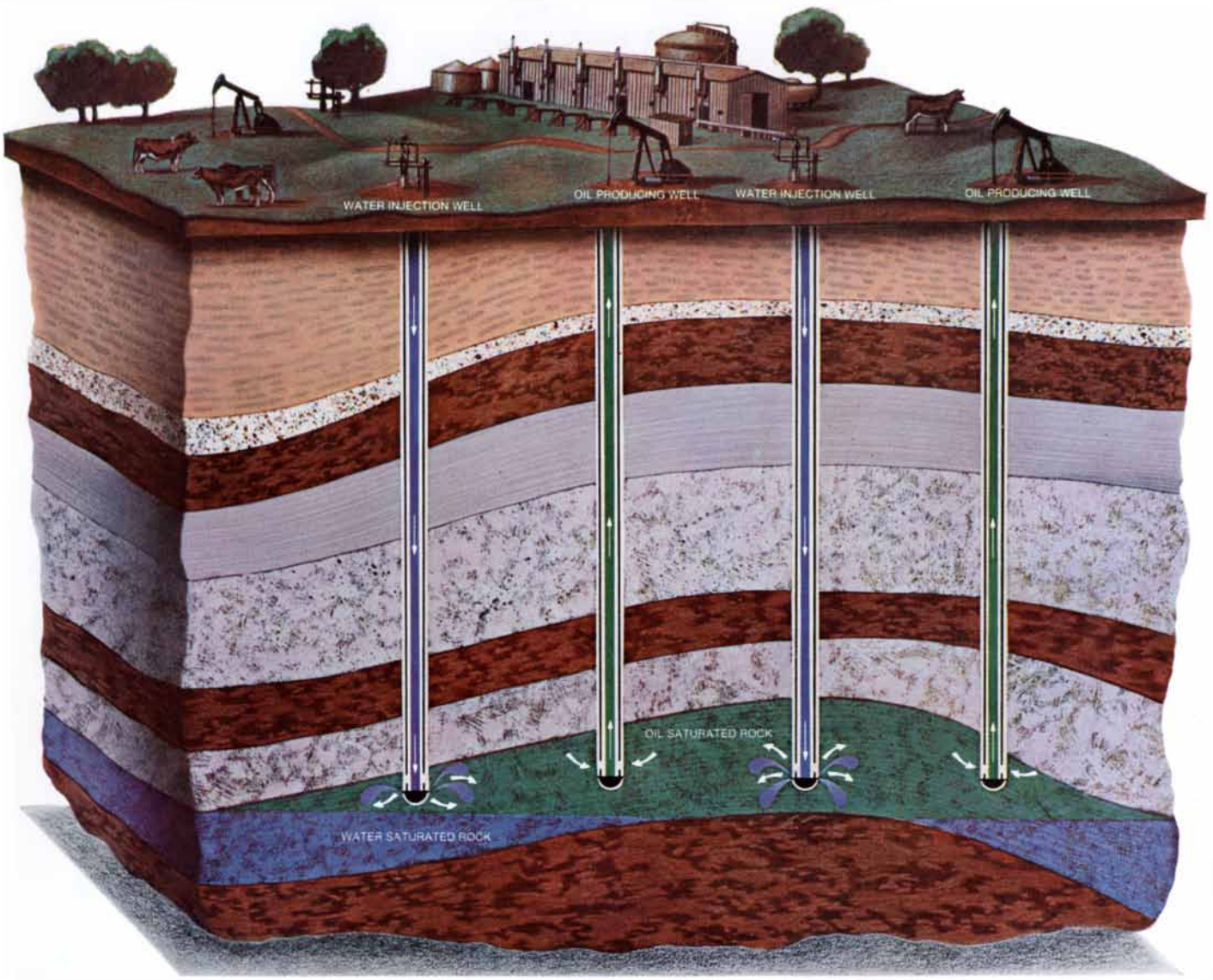
Sometimes, but not at Hewitt, natural gas is injected instead of water. However, the efficiency of waterflooding usually makes it more attractive.

**An experiment using heat and subterranean "detergents."**

Even after an oil field is subjected to secondary recovery techniques, such as water and gas injection, a large percentage of the oil still remains locked in the pores of the rock. So for many years, Exxon has been experimenting with tertiary recovery methods designed to recover more of this oil.

In one experimental process, a "detergent" or surfactant is pumped into the rock. This agent works like a laundry detergent that removes oil from a shirt—it flushes out the droplets of oil which still remain in the pores of the rock. In some cases, particularly where the oil is thicker, sometimes nearly the consistency of molasses, steam injection also is being tested. The hot steam heats the oil, making it flow more freely.

# old wells.




*Artist's conception of the Hewitt oil field located near Ardmore, Oklahoma. Large quantities of water are being injected into the oil-saturated rock to significantly increase the oil production.*

In perhaps the most dramatic oil-recovery technique, air is injected into the rock formation holding the oil. This injected air supports controlled underground combustion, which heats the oil, causes it to flow more readily and drives it to the well.

In addition, Exxon is testing and evaluating several other tertiary techniques.

**A realistic look at the future.**

To date, about 450 billion barrels of oil have been found and documented in the United States. Even after additional recovery efforts, like the waterflooding at Hewitt, are applied, some 300 billion barrels of this oil will still remain locked in the pores of rock. Today, there is no way to recover all of this oil.

But scientists believe that experimental "tertiary" techniques like "detergents" and combustion, could someday produce a significant amount of additional oil which America will need in theyears ahead.

EXXON

# Dodge is right on target with the '75 Dart Special Edition. It's a lot of luxury in a little car.

If you've found that a lot of small cars today are pretty small on comfort and style, then the '75 Dart Special Edition is the small car for you. It has an array of standard features that are usually anything but standard. Seats covered with crushed velour upholstery. Plush carpeting on the floor, halfway up the door, and even in the trunk. Map pockets are standard. So are color-keyed wheel covers. And front disc brakes. And power steering. Even the radio is standard.

In other words, on the '75 Dart Special Edition, luxury comes as standard equipment.

And, as on all the '75 Darts, there are a lot of money-saving engineering features such as Electronic Ignition and electronic voltage regulator and, now for 1975, an optional fuel pacer that can help you save gas.

Today, over one million satisfied owners are already sold on Dart. The '75 Dart Special Edition is one beautiful way to make sure there'll be one million more.


CHRYSLER MOTORS CORPORATION

Extra care in engineering makes a difference in Dodge. Depend on it.

Dodge
Dodge Trucks