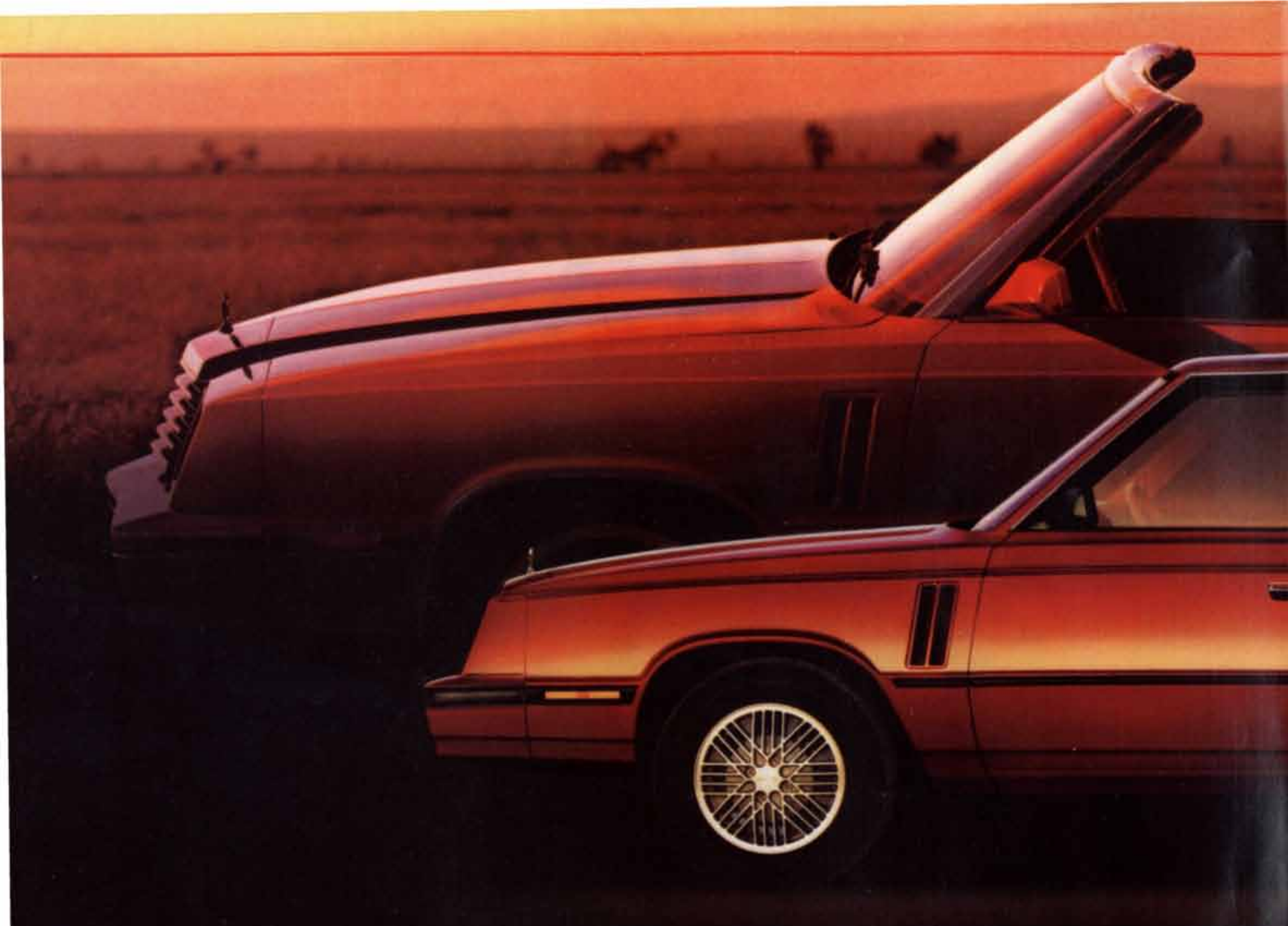
SCIENTIFIC AMERICAN

HUMAN-POWER LAND VEHICLES

$2.50

December 1983




600 ES INTERIOR


600 CONVERTIBLE
600 ES
UNCOMMON CARS WITH SOME REVOLUTIONARY THINGS IN COMMON.
WE ARE DODGE.
Dodge
AN AMERICAN REVOLUTION.
compare. Your mileage may vary depending on speed, distance, weather. Actual hwy mpg lower.
BUCKLE UP FOR SAFETY.


Product of France. Made with fine cognac brandy 80 proof. © 1983 Carillon Importers, Ltd., N.Y.
Grand Marnier
Liqueur
Paris
France
Produce of France
Triple Orange
May your winter be a wonderland.
Grand Marnier.
For delivery anywhere call 800-528-6148 (except where prohibited by law).

Established 1845 SCIENTIFIC AMERICAN December 1983 Volume 249 Number 6

ARTICLES

DEPARTMENTS

SCIENTIFIC AMERICAN (ISSN 0036-8733), PUBLISHED MONTHLY BY SCIENTIFIC AMERICAN, INC., 415 MADISON AVENUE, NEW YORK, N.Y. 10017. COPYRIGHT © 1983 BY SCIENTIFIC AMERICAN, INC. ALL RIGHTS RESERVED. PRINTED IN THE U.S.A. NO PART OF THIS ISSUE MAY BE REPRODUCED BY ANY MECHANICAL, PHOTOGRAPHIC OR ELECTRONIC PROCESS, OR IN THE FORM OF A PHONOGRAPHIC RECORDING, NOR MAY IT BE STORED IN A RETRIEVAL SYSTEM, TRANSMITTED OR OTHERWISE COPIED FOR PUBLIC OR PRIVATE USE WITHOUT WRITTEN PERMISSION OF THE PUBLISHER. SECOND-CLASS POSTAGE PAID AT NEW YORK, N.Y., AND AT ADDITIONAL MAILING OFFICES. AUTHORIZED AS SECOND-CLASS MAIL BY THE POST OFFICE DEPARTMENT, OTTAWA, CANADA, AND FOR PAYMENT OF POSTAGE IN CASH. SUBSCRIPTION RATE: $24 PER YEAR IN THE U.S. AND ITS POSSESSIONS; $33 PER YEAR IN ALL OTHER COUNTRIES. POSTMASTER: SEND ADDRESS CHANGES TO SCIENTIFIC AMERICAN, 415 MADISON AVENUE, NEW YORK, N.Y. 10017.

SCIENTIFIC AMERICAN

CORRESPONDENCE

Offprints of more than 1,000 selected articles from earlier issues of this magazine, listed in an annual catalogue, are available at $1.25 each. Correspondence, orders and requests for the catalogue should be addressed to W. H. Freeman and Company, 4419 West 1980 South, Salt Lake City, UT 84121. Offprints adopted for classroom use may be ordered direct or through a college bookstore. Sets of 10 or more Offprints are collated by the publisher and are delivered as sets to bookstores.

Photocopying rights are hereby granted by Scientific American, Inc., to libraries and others registered with the Copyright Clearance Center (CCC) to photocopy articles in this issue of SCIENTIFIC AMERICAN for the flat fee of $2 per copy of each article or any part thereof. Such clearance does not extend to the photocopying of articles for promotion or other commercial purposes. Correspondence and payment should be addressed to Copyright Clearance Center, Inc., 21 Congress Street, Salem, MA 01970. Specify CCC Reference Number ISSN 0036–8733/83. $2.00 + 0.00.

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

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

Subscription correspondence should be addressed to Subscription Manager, SCIENTIFIC AMERICAN, P.O. Box 5969, New York, NY 10017. The date of the last issue on your subscription is shown in the upper right-hand corner of each month's mailing label. For change of address notify us at least four weeks in advance. Please send your old address (if convenient, on a mailing label of a recent issue) as well as the new one.

Name ______________________

New Address ______________________

Street ______________________

City ______________________

State and ZIP ______________________

Old Address ______________________

Street ______________________

City ______________________

State and ZIP ______________________

THE COVER

The painting on the cover shows the Vector Single, a fully streamlined tricycle that the rider pedals in a supine position, that is, while essentially lying on his back. The vehicle embodies a number of advances that have been made in recent years in the aerodynamics of bicycles and other pedaled machines (see "The Aerodynamics of Human-powered Land Vehicles," by Albert C. Gross, Chester R. Kyle and Douglas J. Malewicki, page 142). The Vector Single was designed and built by a team headed by Allan A. Voigt, president of Versatron Research, Inc. According to Voigt, the tricycle is theoretically capable of attaining a speed of almost 62 miles per hour on level ground with an input of one horsepower from the rider. A companion vehicle, the Vector Tandem, pedaled by two riders positioned back to back, holds the speed record of 62.92 m.p.h., which was achieved with an input of slightly more than 1 h.p. from each rider. It averaged 50.5 m.p.h. for 40 miles on an interstate highway.

THE ILLUSTRATIONS

Cover painting by Ted Lodigensky

Page	Source
20–28	Ilil Arbel
50	Wärtsilä
51–55	Walken Graphics
57–63	National Aeronautics and Space Administration, U.S. Geological Survey
64–66	Ian Worpole
67	John T. Trauger, California Institute of Technology
69	Caroline Quillian Stubbs
70–74	Alan D. Iselin
95	Nelson Max, Lawrence Livermore National Laboratory
96–100	Irving Geis
101	Richard E. Dickerson, Molecular Biology Institute, University of California at Los Angeles
102	Irving Geis
103	Richard E. Dickerson, Molecular Biology Institute, University of California at Los Angeles
104	Irving Geis
105	Richard E. Dickerson, Molecular Biology Institute, University of California at Los Angeles
109–111	Irving Geis
113–129	George V. Kelvin
131–140	Tom Prentiss
143	Bruce A. Sargeant
144–148	Walken Graphics
149–151	Albert E. Miller
152	Walken Graphics
154–162	Gabor Kiss
165–166	Frode Wissing
167–170	Michael Goodman

LETTERS

Sirs:

Derek Bickerton, in his article "Creole Languages" [SCIENTIFIC AMERICAN, July], argues very convincingly for what is undoubtedly the most important idea in modern linguistic theory: the innate knowledge in human beings of the structure of language. Creole languages are without question evidence of the strongest kind for this position, since, as Bickerton argues, how else are we to explain the strong structural similarities between all creoles no matter what their linguistic and cultural derivation? Bickerton's discussion of child language is also most suggestive: it seems that children are in some sense expecting to encounter a creole grammar when they begin to learn language, and they work with this idea until they see evidence in their mother tongue that negates the original assumptions.

I should like, however, to take Bickerton to task on his criticism of Noam Chomsky's view of acquisition. First of all, Bickerton maintains that the acquisition mechanism of transformational grammar is conceived of as "selecting" the appropriate grammar from a range of possible grammars. Actually this notion of selection is a rather outdated one: it was the view of acquisition put forth in Chomsky's *Aspects of the Theory of Syntax* (1965). In current, more sophisticated views of transformational grammar, for example that of Chomsky's *Lectures on Government and Binding* (1981), the notion of selection has given way to one of parameter setting whereby a child, given a priori a set of parameters, or "switches," learns the grammar of his language by setting each parameter appropriately as linguistic data are encountered. This is an important difference: selection is vague from a computational point of view, and there can be many interpretations (some of which would suggest algorithms that are undecidable in finite time) on how this selection is to be made. Parameter setting represents a concrete claim, one that Robert C. Berwick has successfully embedded in an implemented model of acquisition. Thus what Bickerton is criticizing is a view of language learning that is no longer even held.

Second, and more important, Bickerton objects to the Chomskyan approach because in it, he asserts, there is no preferred grammatical model, one the child will aim for and utilize until evidence in the language he is learning forces him to change his model. This assertion is false. The idea that the child has no preferences has been taken as an *idealized* model in transformational grammar.

Similarly, there is the idealization of instantaneity: the obviously factually false model whereby the formal language learner acquires knowledge of his mother tongue by an instantaneous exposure to a large body of data. But these are *not* part of a theory of acquisition. They are (as Chomsky himself puts it) merely idealizations, like that of an ideal gas in physics. In fact, for many years linguists have had a notion of markedness by which specifications on features or parameters could be characterized as being either marked (less likely, or less highly valued) or unmarked (more likely, or more highly valued). One of the central ideas of markedness theory is that the unmarked specification is the preferred one from the point of view of acquisition and the one the child will select unless he has some evidence to the contrary.

Creole languages in this view would represent the maximally unmarked case: having no counterevidence, the child assumes the unmarked settings, which give him a creole grammar. Hence within the transformational-grammar view of acquisition we can explain the effects Bickerton's theory seeks also to explain. Nevertheless, it should be pointed out that the parameter-setting model is quite different from Bickerton's theory, which maintains that the creole grammar is the *only* encoding of linguistic knowledge given to the child. Without a formal statement of exactly what to do given an input of contrary evidence, we are left to guess, in Bickerton's approach, as to exactly how a child does proceed if he is exposed not to Hawaiian Creole but rather to English.

In contrast, Chomsky's theory, given the markedness conventions, makes a specific claim on how this should occur. It does not state that children will not make mistakes, or even start off with creole-like grammars; it does give a formal model (by providing a set of parameters any of which can, if necessary, be set to the marked position) for what exactly a child will do in the common case that his mother tongue is not a creole. Bickerton's approach does not, and thus cannot, account for the obvious ability of children to learn noncreole languages.

In summary, current theories of acquisition by way of parameter setting and markedness are marvelously adapted to accounting for the facts of natural language, creole and noncreole alike.

RICHARD SPROAT

Department of Linguistics
and Philosophy
Massachusetts Institute of Technology
Cambridge

Sirs:

Of the points Richard Sproat raises the first has little substance. I fail to

see how I misrepresented Chomsky by stating that the child must select an appropriate grammar from the range made available by universal grammar, since the "parameter setting" model of Chomsky's *Lectures on Government and Binding* is, as Sproat admits, simply one of the many ways in which such a selection process might be instantiated. I fully agree with Sproat that this latest model is an advance on the earlier one, which was quite vague, but it is still a selection model. Thus statements to the effect that I am criticizing a view of language learning no longer held are calculated to mislead a nonspecialist audience into supposing there has been some radical change in Chomsky's approach to language acquisition, whereas in fact there has been merely a filling out of an original sketch.

The second point is both more substantive and more interesting, but again Sproat leads off with a misrepresentation. A careful rereading of my article fails to reveal anything that could be construed as a claim that within Chomsky's approach there is no preferred grammatical model. Indeed, I show in my most recent article ("The Language Bioprogram Hypothesis," to appear in *The Behavioral and Brain Sciences,* June, 1984) that on one interpretation creole grammar *could* consist of the unmarked parameter settings in Chomsky's latest model, just as Sproat suggests. The situation is much more problematic, however, than Sproat seems to believe.

In the first place Chomsky himself considers (*Lectures on Government and Binding*, page 9) that his model may not work without "indirect negative evidence," that is, if unmarked rules and structures are not exemplified in the input the child receives, he may select marked alternatives to these. Neither Chomsky nor Sproat, to the best of my knowledge, has carried out the thought experiment of seeing how "indirect negative evidence" would work in the situations I have described. Pidgin input to the first creole generation exemplified few if any rules or structures, marked or unmarked. In such circumstances, according to Chomsky's own statement, the child would be driven to abandon unmarked options and replace them with marked ones. These too, however, would remain unconfirmed. According to this model, then, the child of pidgin-speaking parents could not have learned any language—a conclusion that is patently incorrect.

A more radical argument against the "parameter setting" model is that the hypothesis of a single innate grammar might eventually give us a better theory of syntax and how it is acquired. At present markedness and the various parameter settings of Chomsky's model have no principled foundation; they are not deduced from a central theory but simply induced from empirical observations of the variety of natural-language data.

If universal grammar consisted of a single syntactic rule system narrowly constrained by a small set of principles, however, Chomsky's parameter settings might fall out quite naturally from relaxing one or another (or varying combinations of) these constraining principles to varying degrees. If this proved to be the case, not only would we have an explanatory theory of parameter settings but also there would be a rational explanation of the acquisition of noncreole languages, something Sproat fails to find in my article.

In fact, even now the position is not quite as Sproat describes it. Both Chomsky's model and mine contain things that are innate and things that have to be learned: in Chomsky's terminology "core grammar" and "the marked periphery." Since all languages have at least some creole characteristics, the child equipped with creole grammar but faced with a noncreole language would by no means be in the helpless position Sproat seems to suppose; he would just be less well endowed and have to do more old-fashioned learning than the Chomsky-equipped child. How much of language is innate and how much is learned is still a very open (and empirical) question, in spite of grandiose talk of "theories... marvelously capable of accounting for the facts."

I do not think, however, that either model is explicit enough at this stage. If I am right in suspecting that the range of natural-language grammars can be derived by relaxing the constraints on creole grammar in various ways, we might get a bit closer to a viable theory of the acquisition of language. The child equipped with creole grammar but doomed by fate to learn language *X* would simply have to observe what constraints were relaxed in *X*, and to what extent, in order to arrive at the core grammar of *X*. If in addition that child was equipped with something like the preemption principles proposed by Steven Pinker ("How Do Children Unlearn Their Mistakes? A Hypothesis from Learnability Theory," presented at the meeting of the Society for Research in Child Development in Detroit in April), which would oblige him to reject the output of the creole grammar wherever it differed from that of the target grammar (instead of simply allowing the two outputs to coexist in his speech), then he would have an apparatus powerful enough to learn any natural language in finite, indeed brief, time.

DEREK BICKERTON

Department of Linguistics
University of Hawaii at Manoa
Honolulu

WHILE SOME CARS MIGHT MAKE IT TO THE NEXT DECADE, STATISTICS SHOW VOLVOS COULD MAKE IT TO THE NEXT CENTURY.

Today people are trying to hold on to their cars longer than ever.

And while statistics show the average life expectancy of today's cars is eleven years, that figure pales in comparison to the life expectancy of today's Volvo.

Because statistics show the average life expectancy of a Volvo is over sixteen years.*

Which could mean over sixteen years of comfortable driving in seats equipped with adjustable lumbar supports that relieve tension and road fatigue.

Years of effortlessly maneuvering through parking lots due to power assisted rack and pinion steering that cuts one of the smallest turning circles of any car.

Years of clean air made possible by a Fresh Air Ventilation System that exhausts stale air from the passenger compartment, and helps prevent odors and fumes from ever getting inside.

If all this sounds like a sensible way to transport yourself to the next century, buy a Volvo.

Who knows?

By the year 2000 we may have a car that'll get you through that century, as well.

VOLVO

A car you can believe in.

*Based upon an actuarial analysis of 1981-1982 U.S. Registration Data conducted by Ken Warwick & Associates, Inc. Due to many factors including maintenance, driving conditions and habits, your Volvo may not last as long. Then again, it may last longer. Summary available at your Volvo dealer. © 1983 Volvo of America Corporation.

50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

DECEMBER, 1933: "The blame for the dearth of satisfactory television transmitters can be laid directly at the door of the Federal Radio Commission. This body has ruled that advertising shall not be a part of any television program and that the latter shall be 'for experimental purposes only.' This kind of caution is commendable in that it may prevent any repetition of the blasts of advertising that have done so much to degrade sound broadcasting. On the other hand, the restriction prevents the operation of self-supporting television transmitters, and without them television receivers become only curiosities, depending for their programs on a few scattered stations operating on perfunctory schedules."

"The doctrine of Nordic superiority is an offshoot of Aryanism, the chief exponent of which was Count Joseph Arthur de Gobineau, a French aristocrat who died in 1882. Gobineau maintained that one race alone, the Aryans, has been the creator and sustainer of all that is great and good in civilization. The idea of an Aryan race was based on the discoveries of similarities in the languages of the Indo-European group, which led to the theory that all these languages were derived from a common stem, the Aryan language. Gobineau and his disciples assumed that the existence of an Aryan language implies existence also of an Aryan race. Having created this mythical race, they attributed to it all virtue and excellence, and they saw in it the source of every great civilization of antiquity and of modern times. The Nordics were represented as descendants of the original Aryans who settled in northern Europe and from whom in turn came the Teutonic and Anglo-Saxon peoples. In spite of all efforts no one has ever been able to produce the slightest bit of real evidence that any such race ever existed. There is no necessary relation between language and race, and the very use of the term 'Aryan' in a racial sense—as the Germans are using it today—has no justification whatever."

"It is very probable that there is a great deal of free hydrogen in Jupiter and Saturn. Hydrogen is enormously abundant in the sun, and these planets have sufficient gravitative power to prevent even so light a gas from diffusing away from their surface into space. The clouds that form the visible surface must arise from the condensation of some substance that comes down at about 100 degrees below zero centigrade. Why these clouds, particularly on Jupiter, show such conspicuous and various colors—white, brown and red—is a problem for the future."

"Drilling an oil well to a depth of more than 11,000 feet, as has recently been done on the edge of California's famous Kettleman Hills, gives rise to three pertinent questions: Will it prove possible to go still deeper? What are the problems of such deep drilling? Will those who risk their fortune find oil in those superheated subterranean rocks? Oil men have been heard to discuss the possibilities of extending their operations down to three miles. They think that they can reach this distant horizon, but that few operators will attempt to go so deep until the present visible supply of oil dwindles to the point where rising prices will justify the high cost."

SCIENTIFIC AMERICAN

DECEMBER, 1883: "According to the recent annual report of the Edison Electric Light Company, the First District Works in New York, Pearl Street Station, is now running up to its full capacity. It has 9,811 incandescent lamps in actual use, and it is connected and has wired for 12,379 lamps. Altogether the Edison company now has 246 plants at work and 61,366 lamps."

"The storage and transportation of petroleum is in the hands of two companies, whose pipes cover the entire field of Pennsylvania and convey the oil to reservoirs hundreds of miles distant. The larger of these companies, the United Pipe Lines, is controlled by the Standard Oil Company. Its lines are six in number; two run from Olean to Communipaw, on New York Bay, another runs to Buffalo, one to Cleveland, one to Pittsburg and the sixth to Milton Station on the Reading Road. Its pipes in the aggregate are more than 3,000 miles long, and it owns more than 600 tanks with an aggregate storage capacity of 20,000,000 barrels. The oil that is carried by the pipe lines is crude petroleum. The refining necessary to fit the oil for its commercial uses is done principally at Cleveland, Buffalo, Oil City, Pittsburg and in the vicinity of New York City. The bulk of the petroleum exported is refined oil."

"Two years ago M. Gaston Tissandier, struck by the advantages that are presented by the new dynamo-electric motors from an aerostatic point of view, operated at the Exhibition of Electricity a small directable electric balloon. Since that time he has constructed a new bichromate of potash pile of light weight and capable of yielding considerable electricity for a short time. Soon afterward M. Albert Tissandier began a special study of the mode of constructing a large elongated balloon that should offer the indispensable conditions of lightness and stability in the atmosphere. The resulting balloon is 28 meters in length by 9.2 meters in diameter. The car contains the motor, which is formed of 24 bichromate of potash elements that actuate a Siemens dynamo connected with the propeller through the intermedium of gearing. The motor has a power of 100 kilogram-meters, equivalent to that of 10 men, and drives the propeller, which is about three meters in diameter, at the rate of 180 revolutions per minute. The balloon was inflated on the 8th of October by means of a large hydrogen apparatus. The Messrs. Tissandier ascended slowly at 20 minutes past three o'clock in the afternoon amid the applause of numerous spectators. The balloon remained for a long time over the Bois de Boulogne and was easily swerved to the right and left under the action of the propeller and rudder."

"Sir Charles William Siemens, the well-known scientist, engineer and electrician, died in London on the 20th ult. of rupture of the heart. Dr. Siemens was born at Lenthe, Hanover, in 1823. In 1843 and 1844 he visited England to introduce a gilding and silvering method of his brother Werner's. He ever afterward made England his home, becoming a naturalized citizen in 1859. Between 1856 and 1861 he worked out, with his brother Frederick, the regenerative gas furnace. While working upon this furnace Sir William also sought to make steel and iron direct from the ore. With this object in view he constructed his sample steel works at Birmingham in 1866. Thereafter he continued to produce steel upon the open hearth of his regenerative gas furnace. In 1868 he originated the Landore Siemens steel works, which produce upward of 1,000 tons of cast steel per week. Ever since 1848 Dr. Siemens had been interested in telegraphy and had occupied a prominent position in the development of electrical devices. In 1858 he established, with his brothers Werner and Carl and Dr. Halske of Berlin, the works now known as those of the Siemens Brothers, in London, Berlin and St. Petersburg. He planned and had built the steamer *Faraday* for laying ocean cables, and at their Woolwich factory the brothers manufactured several of the Atlantic cables. In the department of electric lighting he was esteemed second to no one in England. His name is connected with many other inventions."

10 QUESTIONS TO ASK YOURSELF BEFORE YOU BUY A BUSINESS COMPUTER.

The toughest thing about investing in the best computer is finding which computer is best for you. Because computers can be just as different as the people they work for.

That's why we've devised a list of 10 basic, unbiased questions to help simplify the choice of which computer will get along with you.

1 HOW MUCH SHOULD I SPEND?

As little as possible, and still get a fully functioning computer.

That would not include machines you see advertised for under $1,000. They're mainly designed for home entertainment and balancing your checkbook in full color.

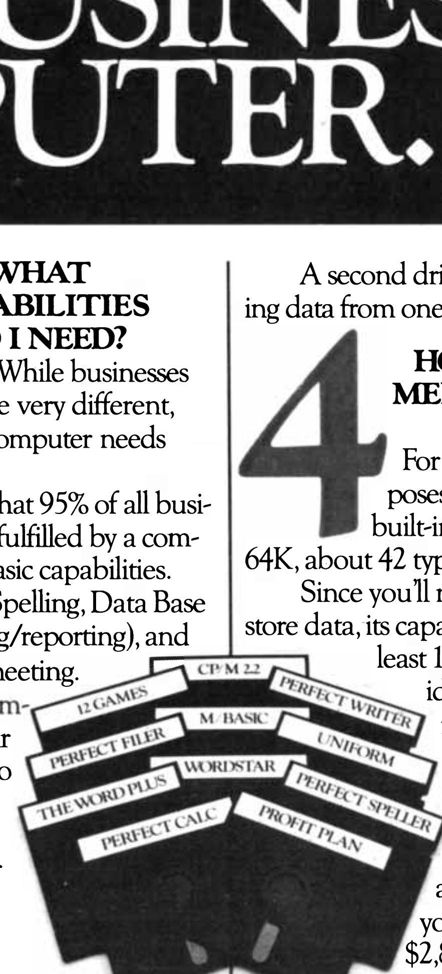
For a serious business system, with all the capabilities you need, you can spend as little as $1,600, including software. Or as much as $5,000. And up.

Of course this gives you a choice of more than 500 different makes and models. So rather than devote the next 5 years of your life trying 500 different computers, a good short list would include computers from IBM, Kaypro, Apple and Tandy—the leading brands.

2 WHAT CAPABILITIES DO I NEED?

While businesses can be very different, their computer needs usually aren't.

We've found that 95% of all business needs can be fulfilled by a computer with three basic capabilities. Word Processing/Spelling, Data Base Management (filing/reporting), and Financial Spreadsheeting.



If the same computer has a popular 'operating system,' so much the better.

It will have thousands of other programs available to fill more specialized needs.

3 HOW MUCH HARDWARE DO I NEED TO GET STARTED?

Beyond the basic computer, you need a monitor and at least one disk drive, the device that stores information. For word processing or data base management, you will need a computer with 2 drives. Such powerful programs require the storage capacity of their own disk drive.

A second drive also eases copying data from one disk to the other.

4 HOW MUCH MEMORY WILL I NEED?

For most business purposes, a machine with a built-in memory (RAM) of 64K, about 42 typed pages.

Since you'll need a disk drive to store data, its capacity should be at least 140K. Of course, the idea is to get the most memory for the money. For example, when you compare the hardware of equivalent systems, an IBM PC will give you 320K for about $2,800. A Kaypro II will give you 400K for about $1,595. And an Apple IIe will give you 286K for about $2,400. So once again, the choice depends upon your needs.

5 DO I NEED AN 8-BIT OR 16-BIT SYSTEM?

A 16-bit system costs more, gives you a little more speed, and can run longer programs. Unfortunately, only a handful of programs take advantage of 16-bit capacity.

75% of all microcomputers sold today are of the 8-bit variety, indicative of their ability to satisfy the needs of most businesses.

6 WILL I NEED A PRINTER?

Most people do, sooner or later. It's a fast, efficient way of getting information from your computer screen to other people. For letters, for reports, for balance sheets and for records. That's why it's important to make sure the computer you want has a built-in printer connection. Otherwise when you go to buy a printer, you'll also find yourself paying for an interface.

7 HOW MUCH TRAINING WILL I NEED TO USE MY COMPUTER?

You won't have to learn to write programs to use your computer. Somebody's already done that for you. You will have to learn how your machine operates, which may take a few hours.

And you'll learn how to run programs. The time that it takes depends on their complexity. The ideal way to save time is having a family of programs that share the same commands. Each program in the series will take less time to learn.

8 WHERE CAN I GO FOR GOOD SOLID ADVICE?

Do the same thing you would do if you were shopping for a stereo or an expensive camera. Ask an enthusiast. A computer buff. You probably know one. If not, call a consultant. In the long run, he or she could save you a lot of money by directing you to the right combination of hardware and software. And by advising you whether to buy a computer now or wait for prices to come down further.

It is true that the price of serious business computers is drifting down an average of 10-12% a year. The average system costing $3,000 today may cost $2,700 next year. Yet, you could lose thousands of dollars of increased productivity, waiting to save $300.

SPECIFICATIONS

Microprocessor Z-80	Perfect Filer
Operating System CP/M 2.2	Perfect Calc spreadsheet
User Memory 64K	Wordstar word processing
Disk Drives: 2 drives, 400K, unformatted	The Word Plus
Interfaces 1 Serial, 1 Parallel	Profit Plan spreadsheet
Keyboard Detached, 63-key with numeric keypad	M-Basic
Software included: Perfect Writer word processing	12 Games
Perfect Speller	Uniform—allows computer to 'read' and 'write' TRS-80, Osborne, Xerox disks
	Dimensions Height: 8 inches; Width: 18 inches; Depth: 15½ inches; Weight: 26 lbs. (portable)

9 WHY IS THERE SUCH A GREAT DIFFERENCE IN THE PRICE OF COMPUTERS?

Some differences are based on capabilities. Business computers capable of playing joystick games cost more than ones that stick to business. And, as we mentioned before, 16-bit computers cost more than 8-bit systems.

A more important difference is the way computers are manufactured and sold. Most start with the basic keyboard and microprocessor, for a basic price.

You then pay more for a monitor. More for disk drives. More for communication interfaces. More for software. All optional extras that can run up the price two or three times.

However, at Kaypro we don't consider monitors, disk drives and software to be optional extras. So we make and sell completely integrated systems with all the hardware you need. All the software you need. All for $1,595.

Of course, a lot of other computer companies feel you should pay extra for everything you get.

10 WHICH IS THE SMARTEST COMPUTER FOR ME TO BUY?

All things considered, we'd consider it a Kaypro. And since this is a Kaypro ad, you wouldn't think we'd suggest anything else. But fortunately for us, our computer gives us a lot to be biased about. Off the shelf, it gives you everything you want for 95% of your business needs. Plus the ability to run 3,000 more programs. It's the complete business computer.

What do you get complete? 64K RAM, Z-80 microprocessor. CP/M, the most widely used business software operating systems. A 9" easy-to-read, green monitor. Two disk drives (400K capacity). Interfaces for both a printer and communications. A detachable keyboard, with a numeric keypad. And one of the leading computers in the $1,000 to $5,000 price range.

A Kaypro may not be the best computer for everyone. But for $1,595 complete, it gives you the best solution to serious business problems.

CALL 800-447-4700 FOR THE DEALER NEAREST YOU.

Registered Trademarks: Apple — Apple Computer, Inc., IBM — IBM Corp., CP/M — Digital Research, Inc., Z-80 — Zilog, M-Basic — Microsoft, Inc., Tandy, TRS-80 — Tandy Corporation, Osborne — Osborne Computer Corporation, Xerox — Xerox Corporation. Prices based on published information as of July 15, 1983. © 1983 Kaypro Corporation.

THE AUTHORS

JOHN D. HARBRON ("Modern Icebreakers") is a foreign affairs analyst for the Thomson Newspaper Group in Canada. His career as a professional writer has been closely connected with his long-standing interest in ships and the sea. After being graduated from the University of Toronto and doing postgraduate work at the University of Havana he entered the Royal Canadian Navy. His first navy job was as naval officer and professor of naval history at the Canadian Naval College at Royal Roads, B.C. After retiring from the service with the rank of commander he worked for *Business Week* and the *Toronto Telegram* and served as correspondent for several British and Canadian shipping publications; he left the *Telegram* to join the Thomson group. Harbron is a recipient of the Maria Moors Cabot Medal of the Columbia University Graduate School of Journalism and a commander of the Order of Isabel la Católica of Spain.

TORRENCE V. JOHNSON and LAURENCE A. SODERBLOM ("Io") are respectively an astronomer and a geophysicist; they were members of the National Aeronautics and Space Administration teams that obtained the first closeup views of many bodies in the solar system. Johnson got a B.S. at Washington University and a Ph.D. in planetary science from the California Institute of Technology. He was on the staff of NASA's Planetary Astronomy Laboratory and the National Research Council before taking up his present job as senior research scientist at the Jet Propulsion Laboratory of Cal Tech. He was a member of the outer-planets-probe working group and the Voyager imaging-science team. Soderblom received his Ph.D. in planetary science and geophysics at Cal Tech; he received his undergraduate education at the New Mexico Institute of Mining and Technology. Since getting his doctorate he has worked for the U.S. Geological Survey, in which he holds the rank of supervisory physical scientist.

PETER S. WELLS ("An Early Iron Age Farm Community in Central Europe") is associate professor of anthropology at Harvard University and associate curator of European archaeology at the Peabody Museum. He has been associated with Harvard almost continuously since his undergraduate days. He got a B.A. at Harvard College in 1970 and went on to earn his Ph.D. in anthropology from Harvard University in 1976; his graduate studies included work at the University of Tübingen. His main research interest is economic and social change in the Bronze Age and the Iron Age. The excavations in Bavaria that are described in his article are currently being expanded to include other prehistoric settlements in the region to yield a broader picture of the development of society in central Europe in the final millennium B.C. Wells's research is supported by the National Geographic Society, the National Science Foundation, Earthwatch and the Center for Field Research.

RICHARD E. DICKERSON ("The DNA Helix and How It Is Read") is professor of biochemistry and geophysics and director of the Molecular Biology Institute at the University of California at Los Angeles. He attended the Carnegie Institute of Technology as an undergraduate before receiving his Ph.D. in physical chemistry from the University of Minnesota. After a year of postgraduate study at the Cavendish Laboratory of the University of Cambridge he returned to the U.S. to take up an appointment at the University of Illinois. After four years there he moved to Cal Tech. In 18 years at Cal Tech his main research interests were analyzing the crystal structure of the electron-transport protein cytochrome *c* and the study of the evolution of molecules. Since going to U.C.L.A. in 1981 his interests have shifted to the analysis of the structure of short stretches of DNA and the complexes such molecular structures form with antitumor drugs and the proteins that act to repress the reading of the double helix.

BRYCE S. DEWITT ("Quantum Gravity") is professor of physics at the University of Texas at Austin and director of the Center for Relativity there. He holds three degrees from Harvard: a B.S., an M.A. and a Ph.D. in physics. After working briefly at several institutions he became senior research physicist at the Radiation Laboratory of the University of California at Berkeley. He left Berkeley to join the department of physics at the University of North Carolina at Chapel Hill. He was at North Carolina for 15 years, leaving in 1972 to move to Texas. In addition to these positions he has held visiting appointments at many institutions, including the École d'Été de Physique Théorique at Haute Savoie in France, the University of Osaka, Stanford University, the University of Oxford and the Institute for Theoretical Physics in Santa Barbara.

ARTHUR T. BERGERUD ("Prey Switching in a Simple Ecosystem") is research associate at the University of Victoria in British Columbia. He got a B.S. at Oregon State College in 1956 and an M.S. in biology at the University of Wisconsin five years later. After receiving his master's degree he went to Newfoundland to join the provincial Department of Wildlife. In his 11 years with the department he rose from district biologist to director of wildlife management. In 1966 he moved to the University of British Columbia to continue his formal education. He earned his Ph.D. in biology from British Columbia in 1968; his thesis was on the caribou population of Newfoundland.

ALBERT C. GROSS, CHESTER R. KYLE and DOUGLAS J. MALEWICKI ("The Aerodynamics of Human-powered Land Vehicles") share an interest in the topic of their article. It is one they approach from quite different professional perspectives. Gross, a management consultant and journalist, is president of Action Research Associates, Inc. He has two bachelor's degrees: one in American history from Columbia College (1969) and one in psychology from San Diego State University (1975). After getting his second degree he continued the study of experimental psychology at the University of California at San Diego, which granted him an M.A. in the subject in 1976. Kyle is professor of mechanical engineering at California State University at Long Beach. His bachelor's degree is from the University of Arizona; his M.S. and Ph.D. in engineering are from the University of California at Los Angeles. Outside the laboratory his avocation is the aerodynamics of human-powered vehicles, a subject on which he has often published. He is cofounder and past president of the International Human Powered Vehicle Association (IHPVA). Streamlined bicycles of his design have been employed in setting five world speed records for human-powered vehicles. Malewicki is director of Aerovisions, Inc. He has an M.S. in aeronautical and astronautical engineering from Stanford University. He is a consultant who works mainly for Aerochem, Inc., on aerospace projects, including ones related to the space shuttle. His work has included stress analysis of two-man submarines and performance analysis of hot-air balloons designed to set altitude records. He has also designed motorcycles with which distance and altitude records for jumping have been set.

DONALD D. HOFFMAN ("The Interpretation of Visual Illusions") is assistant professor of cognitive science and computer science at the University of California at Irvine. He got a B.A. in 1978 from the University of California at Los Angeles before going on to earn his Ph.D. in computational psychology in 1983 from the Massachusetts Institute of Technology.


"...the new Mark is a worthy road machine that can hold its own against the higher priced German imports. And if it can't beat them at their own game right out of the box, it can give them a real good run..."
Tony Assenza, Motor Trend.
"Inside the Mark, I found a restyled but refreshingly simple interior with bright LED displays for speed, fuel consumption and miles traveled."
Jim Dunne, Popular Science.
"We're not saying the LSC...is a luxury racer, at least not as it's now being presented. We do say, however, we can't think of another car in this class, domestic or imported, that has its manners and comfort."
Burge Hulett, AutoWeek.
"The box is gone, and in its place is a sleek, 5-passenger luxury touring coupe. Inside and out, the new Mark is as different and innovative as it looks."
John McElroy, Road & Track.
CONTINENTAL MARK VII LSC.


Is it better to give Wild Turkey
or to receive Wild Turkey?
WILD TURKEY
101 Proof 8 Years Old
WILD TURKEY® / 101 PROOF / 8 YEARS OLD
AUSTIN NICHOLS DISTILLING CO. LAWRENCEBURG, KENTUCKY © 1983

COMPUTER RECREATIONS

On the finite-state machine, a minimal model of mousetraps, ribosomes and the human soul

by Brian Hayes

The most powerful computers have neither hardware nor software; they are built out of pure thought stuff. Among these abstract machines the most celebrated is the one invented in 1936 by the British mathematician Alan Mathison Turing. It can do more than any computer made of mere silicon ever could; indeed, it can compute anything that can be computed. A related class of conceptual computers lack the omnipotence of the Turing machine, but they are no less interesting. They are called finite-state machines or finite-state automata, and they establish the minimum specifications of a working digital computer.

Properly defining a finite-state machine calls for a degree of mathematical rigor that is not appropriate here. The nature of the concept can be made clear, however, by means of a few examples. When I went out looking for finite-state machines, I found an excellent specimen in a station of the Lexington Avenue subway in New York. It is a turnstile, an old one made not with the compact steel tripod of current practice but with four oak crossarms, worn smooth by a river of hands and hips.

The turnstile has two states: locked and unlocked. Suppose it is in the locked state, so that the arms cannot be turned. Putting a token into the slot alters the internal mechanism in some way that allows the arms to move; in other words, the token induces a transition to the unlocked state. Rotating the arms by 90 degrees causes another transition that restores the turnstile to the locked state. The transitions are shown schematically in the upper illustration on the next page. The states of the system are represented by nodes (boxes) and the transitions by arcs (arrows) between them.

In the finite-state analysis of the turnstile, inserting a token and pushing on the arms are the possible inputs to the system. The response of the machine depends both on the input and on the state at the time of the input. Pushing on the crossarm when the turnstile has not yet received a token will not get you a ride on the subway. Inserting a token when the arms are already unlocked is also futile, although in a slightly different way. The second token is accepted, but it has no effect on the state of the machine; one person is admitted and then the turnstile locks again. Three or four tokens in sequence are likewise accepted but buy only one ride. Skeptics may want further evidence before accepting the generalization that all tokens after the first have no effect, but they will have to supply their own tokens.

The reason the turnstile cannot give multiple rides for multiple tokens is that it has no means of counting the tokens it has received. Its only form of memory is a rudimentary one: by changing from one state to the other it "remembers" whether the most recent input was a token or a push on the crossarms. All earlier inputs are lost. It is worth noting that this forgetfulness can never work to the disadvantage of the city. It could be worse: a turnstile could be designed to change state after every token, regardless of the present state, in which case two tokens in a row would admit no one.

The turnstile illustrates most of the essential properties of a finite-state machine. Obviously the machine must have some states, and there can be only a finite number of them. There can be inputs and outputs associated with any state. The states must be discrete, or clearly distinguishable, and the transitions between them must be effectively instantaneous. In these matters much depends on the point of view: day and night are discrete states if one is willing to define sunrise and sunset as instantaneous processes. The set of states, the inputs and the outputs constitute the entire machine; there can be no auxiliary devices, and in particular no facilities for the storage of information.

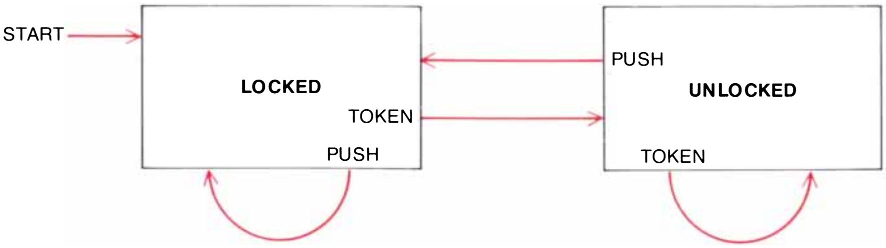
The rules for building a finite-state machine allow some scope for variation. There are deterministic and nondeterministic machines, Moore machines and Mealy machines. In a deterministic machine a given input in a given state invariably has the same result; in a nondeterministic machine there can be several possible transitions. In the Moore machine (named for Edward F. Moore) each state has a unique output. In the Mealy machine (named for G. H. Mealy) the outputs are associated with the transitions rather than the states. It turns out, however, that the variety of architectures is something of an illusion. Any task that can be done by one kind of finite-state machine can be done by the other kinds as well, although the number of states needed may vary. Here I shall discuss mainly deterministic Moore machines, which have the simplest structure.

When you start looking for finite-state machines, you find them everywhere. Coin-operated devices are favorite textbook examples. Some vending machines are less rapacious than the subway turnstile: once they have received the proper amount of money they enter a state in which all additional coins are rejected. The coin-operated device with the largest number of possible states is surely the Las Vegas slot machine. In principle it is deterministic, but finding an input (a coin and a pull on the handle) that will cause a transition to a particular final state is nonetheless challenging.

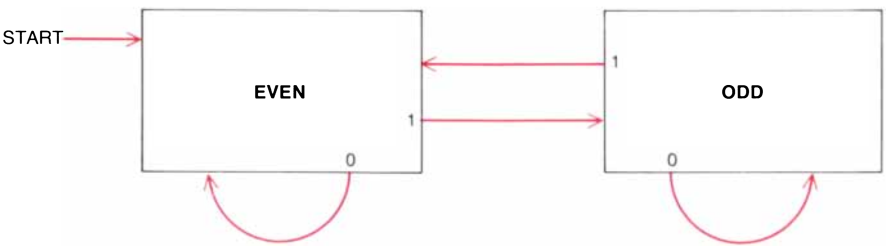
Many household appliances can be regarded as finite-state machines, although they tend to be rather dull ones. A clothes washer goes through an inflexible sequence of states—filling, agitating, rinsing, spinning—and the few meaningful inputs, such as pulling the plug out of the electric outlet, generally have the same effect in all the states. Similarly, a traffic light has a small repertory of states, which repeat indefinitely. To me the most boring of all finite-state machines is a digital clock. If it displays the month, the date and the passage of hours, minutes and seconds, it has some 31 million states; in the course of a year it visits each state exactly once.

A mousetrap is a finite-state machine; the mouse, usually to its misfortune, triggers a transition from the cocked state to the sprung state. A combination lock is a finite-state machine with many possible inputs, only one of which causes a state transition. A telephone has states that might be labeled on hook, off hook, waiting, dial tone, dialing, ringing, connected and out of order. An automobile can demonstrate vividly that the effect of an input varies according to the present state of the system. What happens when you press the accelerator pedal to the floor? It depends. Is the engine running? Is the clutch engaged? Is the parking brake off? Is the transmission in gear? Is it in forward or reverse? Is the garage door open?

In the living cell the molecular system made up of the ribosome and the vari-



A state-transition diagram for a subway turnstile



The parity-testing machine

ous species of transfer RNA operates as a finite-state machine. The inputs are the four nucleotide bases of messenger RNA, designated by the abbreviations *U, A, G* and *C*. The outputs are the 20 amino acid components of proteins. A chain of nucleotides is recognized as a valid input to the machine only if it begins with the "start" signal *AUG*. Thereafter the machine reads the input stream continuously, changing state as each codon, or triplet of nucleotides, is recognized. The three special codons *UAA, UAG* and *UGA* are "stop" signals: when one of them is encountered, the machine halts. Many other biological systems can usefully be represented as finite-state machines; examples that come to mind are the hemoglobin molecule and the promoter and repressor proteins of bacteria.

In the theology of Thomas Aquinas the soul is a finite-state machine, a wonderfully elaborate and fully deterministic one. It is created in a state of jeopardy, as a consequence of original sin. On baptism it enters a state of grace, but certain acts (idolatry, blasphemy, adultery and so forth) induce a transition to a state of sin. Confession, repentance and absolution are then needed to restore the soul to grace. The effect of a final input, death, depends critically on the state of the soul at the moment of death: in a state of grace death leads to salvation but in a state of sin it leads to damnation. The soul machine is actually more complicated than this description suggests. A full account would have to distinguish among the various grades of sin (venial and mortal, actual and habitual) and would have to include other possible states of the soul (such as those associated with limbo and purgatory) and other possible inputs (such as the Last Judgment).

In quantum mechanics even the atom becomes a finite-state machine, and hence so does everything that is made up of atoms. The states of the atom are the allowed energy levels; the inputs and outputs are photons, or quanta of electromagnetic radiation. In a precise description I think the atom would be classified as a nondeterministic Mealy machine with epsilon transitions. It is nondeterministic because the effect of an input cannot be predicted with certainty. It is a Mealy machine because the nature of the output (namely the energy of the photon) is determined by the transition, not by the state entered. Epsilon transitions are those that can take place in the absence of any input; they must be included in the model because an atom can emit a photon and change its state spontaneously.

Is the brain a finite-state machine? As it happens, the modern study of finite-state systems began with a model of neural networks introduced in 1943 by Warren S. McCulloch and Walter Pitts. The neurons of McCulloch and Pitts were simple cells with excitatory and inhibitory inputs; each cell had a single output and two internal states: firing and not firing. The cells could be arranged in networks to carry out various logic functions, including the "and," "or" and "not" functions that are now commonplace elements of electronic logic systems. The equivalence of the idealized neural networks to state-transition diagrams of the kind shown here was established in 1956 by Stephen C. Kleene of the University of Wisconsin at Madison.

Forty years after the work of McCulloch and Pitts it is still subject to dispute whether the brain can reasonably be classified as a finite-state system. Of course the number of neurons is necessarily finite, but that is not the only issue. A real neuron is far more complicated than a two-state cell, and some of its properties may vary over a continuous range rather than being constrained to occupy discrete states. Furthermore, the prohibition of auxiliary information storage in a finite-state model of the brain is awkward at best. If mental life is no more than a succession of instantaneous states, without knowledge of its own history, then what is memory?

The states of mind discussed in psychology, such as boredom, fear, thirst, ecstasy and grief, seem to fit more readily into the apparatus of a finite-state theory. On the other hand, the states are so numerous and the transitions are so poorly understood that the model is useless. Only for lower animals is it possible to draw more than isolated fragments of the state-transition diagram, and in those species the experimenter can have no direct access to the presumed mental states. Indeed, much work of this kind has been done by behaviorists who deny the very existence of mental states.

The case of the digital computer—and here I mean the tangible machine, the hardware—is also problematic. The common mental model of a computer, formulated by John von Neumann, divides the machine into a central processing unit and an array of memory cells. There is no doubt that the finite-state concept can be applied to the various components of the central processor, such as registers, adders and the control mechanism that directs the internal operations of the processor.

The trouble begins when the memory is taken into account. Under the rules for building a finite-state machine no external memory is allowed, and so each cell must be viewed not as a storage facility separate from the processor but as a part of the overall machine state. If all the cells are blank, the computer is in one state; if a single cell is filled, another state is entered, and so on. This conception of the computer is singularly unilluminating, in part because it makes no connection between the state of the machine and what it is doing. Moreover, the number of states is immense. Even a computer of trivial size (100 binary elements), running continuously throughout the age of the universe, could not possibly have worked through all its states.

The primary role of the finite-state machine in computer science is at a higher level of abstraction than the clockwork mechanisms of the hardware. A computer running under the direction of a program is no longer an assemblage of logic gates, registers, memory cells and other electronic paraphernalia; it is a "virtual" machine whose working parts are defined by the program and can be redefined as necessary. Whereas the hardware knows only binary integers and simple commands for moving and manipulating them, the virtual computer deals with far more expressive symbol systems: words, equations, arrays, functions, vectors, codons, lists, images, perhaps even ideas. Finite-state techniques can be valuable in creating the virtual computer, and sometimes the virtual computer *is* a finite-state machine.

Consider a program whose object is to read a series of binary digits (1's and 0's) and report whether the number of 1's received is even or odd. (The task has practical significance; for example, such parity-checking programs are employed to detect errors when digital data are transmitted by telephone.) The program can be constructed as a finite-state machine with two states, as is shown in the lower illustration on the opposite page. Operation begins in the even state, because initially no 1's have been received and 0 is considered an even number. Each 1 in the input stream causes a change of state, whereas a 0 received in either state leaves the state unchanged. Even though the machine cannot "remember" any inputs before the most recent one and certainly cannot count the 1's or 0's, its output always reflects the parity of the input stream.

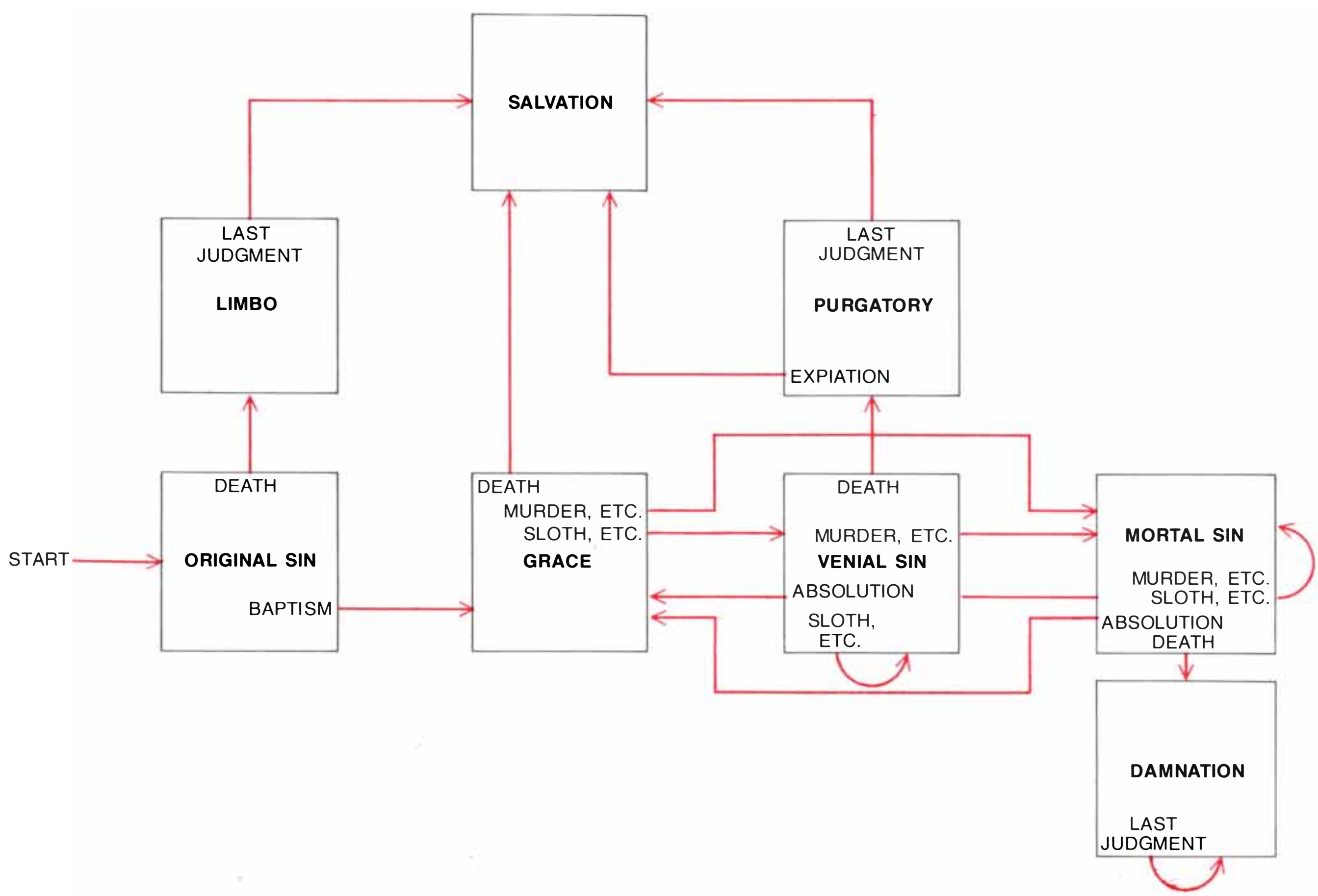
The finite-state model of computation is commonest in programs that deal in some way with text or other information that takes a linguistic form. The preeminent example is found in compilers: programs that translate programming statements in a source language into equivalent statements in a target language, most often the "machine language" of a particular computer. Compilers and other translating programs are essential to the notion of the virtual machine; they mediate between symbols with human meaning and those recognized by the computer.

The part of a compiler that can be designed as a finite-state machine is called the lexical scanner. Like the subway turnstile, it is a token-gobbling device. In this case, however, the tokens are the words, or fundamental lexical units, of the language. The scanner examines each group of characters and determines whether it is a genuine token, such as a command or a number; if it is not, the scanner rejects it as nonsense, just as the turnstile would reject a slug.

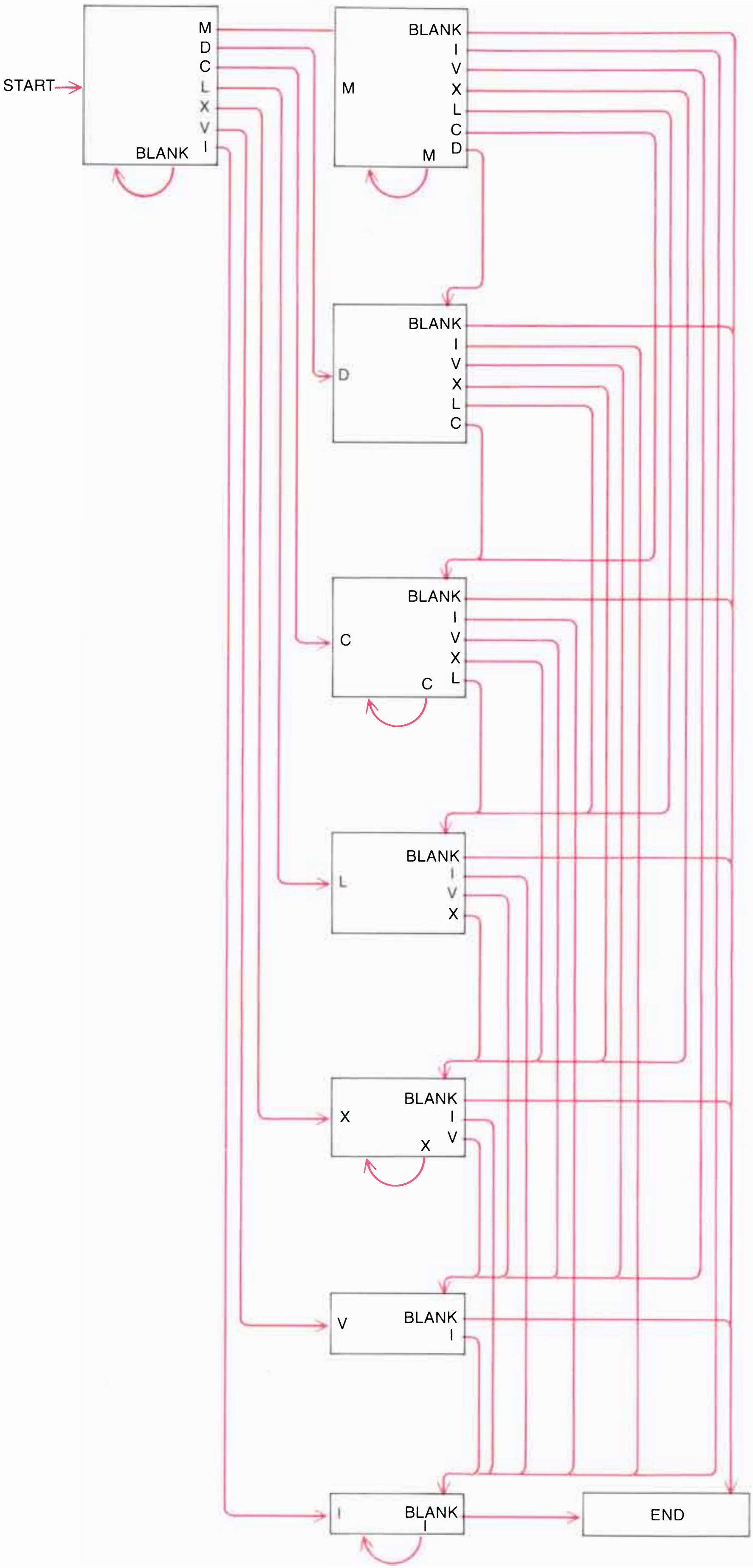
The operation of a lexical scanner can be illustrated by a finite-state machine designed to recognize the tokens of a simple language, albeit one of limited expressive range: the tokens consist exclusively of Roman numerals. Indeed, only Roman numerals of a special form are accepted; they must be given in strict additive notation, so that 9 is represented by VIIII rather than by IX. (There is evidence that the Romans themselves employed the additive notation; the subtractive form is thought to have been a German innovation.)

A state-transition diagram for the Roman-numeral machine is shown in the illustration on the next page. Its alphabet of input symbols includes the letters M, D, C, L, X, V and I as well as the space symbol, or blank. Any initial blanks are simply ignored, but once the first letter is received the program makes an immediate transition to a state identified (for convenience) by the name of the letter. If the first letter is an M, it can be followed by any character from the allowed set, including another M. If the next character is a D, however, the situation is different. From the D state no transition back to the M state is defined, because any series of symbols that includes DM cannot be a well-formed token in the language of additive Roman numerals. Furthermore, there is no transition from the D state to the D state itself, so that DD is also an excluded sequence. (The reason is that the "half value" symbols D, L and V cannot be repeated in proper Roman numerals.)

In the D state the only recognized letters are the lower-valued ones C, L, X, V and I. The same set is accepted in the C



States of the soul in the theology of Thomas Aquinas



A lexical scanner for a language of Roman numerals

state (because C can be repeated), but in the L state only the letters X, V and I are recognized. The rule governing the transitions should be clear. The states are arranged in a hierarchy, and once a given level has been reached the machine can never return to a higher level; in the half-value levels it cannot even remain at the same level. By the time the I state is reached only an additional I or a blank is allowed. The blank, entered at this point or at any other time after the first letter, indicates the end of the token and sends the machine back to its starting state, ready to receive the next Roman numeral.

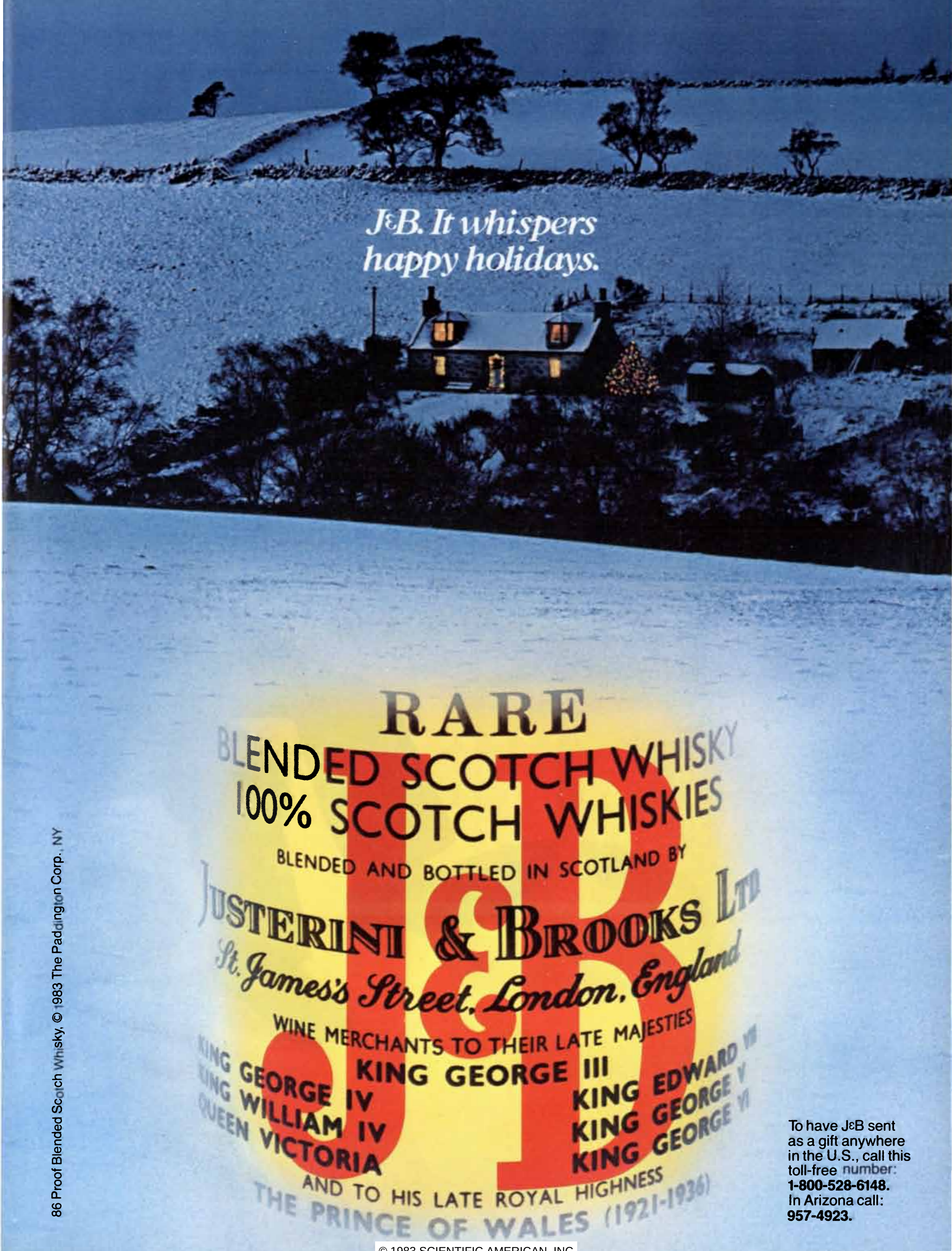
No programming language known to me allows numbers to be entered in Roman form, but virtually all such languages have facilities for handling Arabic numbers. The techniques for recognition are similar, although there is a greater variety of formats. Simple integers such as 137 can be handled in principle by a one-state machine, but the several parts of a number such as $+6.625 \times 10^{-27}$ require a more elaborate lexical analysis.

The ribosome–transfer-RNA system can be regarded as a lexical scanner that recognizes biologically meaningful nucleotide sequences in a molecule of messenger RNA. To be accepted a sequence must begin with a start codon and end with one of the three stop codons; between these boundaries any combination of the input symbols *U, A, G* and *C,* taken three at a time, is allowed.

Lexical analysis is only the first step in the process of compilation. The components of the compiler that are called into action after the lexical scanner are the parser and the code generator. The parser takes as its input the tokens identified by the scanner and analyzes their syntactic relations; this is the closest the compiler comes to understanding the meaning of the program statements it translates. The code generator writes a program in the target language that carries out the functions specified by the parsed statements.

For the toy languages considered here the tasks of the parser and the code generator are trivial. The compiled form of a statement in the Roman-numeral language might be simply the Arabic equivalent of the number. It could be generated by the following strategy. Before a token is scanned a storage cell is specified and is set equal to zero. Then each time the scanner enters the M state 1,000 is added to the value in the cell; for the D state 500 is added, and so on. When the scanning is complete, the memory cell holds the value of the Roman numeral. Note that the toy compiler is no longer a pure finite-state machine, because it has auxiliary storage.

A compiler for the genetic code is even simpler and can be realized entire-


J&B. It whispers happy holidays.
RARE
BLENDED SCOTCH WHISKY
100% SCOTCH WHISKIES
BLENDED AND BOTTLED IN SCOTLAND BY
JUSTERINI & BROOKS LTD
St. James's Street, London, England
WINE MERCHANTS TO THEIR LATE MAJESTIES
KING GEORGE III
KING GEORGE IV
KING WILLIAM IV
QUEEN VICTORIA
KING EDWARD VII
KING GEORGE V
KING GEORGE VI
AND TO HIS LATE ROYAL HIGHNESS
THE PRINCE OF WALES (1921-1936)
J&B
86 Proof Blended Scotch Whisky. © 1983 The Paddington Corp. NY
To have J&B sent as a gift anywhere in the U.S., call this toll-free number: 1-800-528-6148. In Arizona call: 957-4923.

There's a world of information just waiting for you. But to use it, study it, enjoy it and profit from it, you first have to get at it.

Yet the facts can literally be right at your fingertips — with your own telephone, a modem and the IBM Personal Computer.

Plug into the network.

The modem that plugs into your phone takes the codes your IBM Personal Computer understands and turns them into signals that can be transmitted over the phone lines.

On the other end of the phone lines are independent information services* ready for you to access from your home, school, lab or office. So, in minutes, you can tap the resources of over a thousand data bases — and plug into a world of possibilities.

Something for everyone.

At home, for example, you can shop through an electronic catalog. Access mortgage amortization and tax depreciation schedules. Study the wines of France, the foods of Italy. Find property to sell, rent, swap or share. And play hundreds of games.

At school, you can bring news into the classroom — as it happens. You can teach and learn with facts about everything from phrase origins to plane geometry. You can also tap instructional courses and learn how to program the very IBM Personal Computer you're using.

In the lab, you can call up a wealth of mathematical and statistical programs or retrieve scientific and technological data.

In the office, you can plug into the commodities market. Read abstracts of leading publications. Even get census figures to see who's in your major market. And when you're planning a business trip, your IBM Personal Computer can provide airline schedules and car rental information, as well as tell you what the weather is and where the best restaurants are.

Calling all computers.

And these outside data bases aren't your only information link, either.

With assigned passwords, you can also have access to data that's stored in your company's IBM minicomputer or mainframe via your IBM Personal Computer.

You can even exchange a sales forecast with a manager two floors down or two thousand miles away — one IBM Personal Computer to another.

So the IBM Personal Computer really *can* let you call the world. But there's one call to make first. For more information on where to buy the IBM Personal Computer, call 800-447-4700. In Alaska or Hawaii, 800-447-0890. IBM®

The IBM Personal Computer
A tool for modern times

*Information services are available by subscription at additional cost, and are not supplied by IBM.

When a man as discerning as St. Nick gives his personal seal of approval

to a car, you know it must be something to say ho ho ho about. Santa's a man with tough requirements. That's why he chose the Tercel 4WD Wagon. The sleigh is fine for Christmas Eve, but for the rest of the year he needs something more practical. He needs a car roomy enough to carry feed for all his reindeer, yet economical and dependable to pull through treacherous roadways.

If the Tercel 4WD Wagon can fit needs as demanding as Santa's, just think how easily it can accommodate you. If room is what you want, stretch out. No small 4WD wagon can match its roominess. The roof line is higher than any other small wagon, for more headroom. And there's a panoramic rear window for more visibility. Five people can sit comfortably with lots of space for gear in the carpeted cargo area. And if you want more luggage space, just fold down the rear seat backs and get 59.8 cubic feet of space.

If economy is what you want, Tercel's fuel efficient 1.5 liter SOHC engine delivers the best fuel economy of any 4-wheel drive wagon: 41 Estimated Highway, (31) EPA Estimated MPG*.

If you want a tough, dependable car, look no further. The Tercel

OH WHAT A FEELING! TOYOTA

4WD Wagon can pull through almost any road condition from the North Pole to Main St.** And it switches from front-wheel drive to four-wheel drive on command.

New for 1984 is a conventional 2WD Wagon that has all the features Santa wants in his 4WD, but it's for people who don't have to go to the North Pole.

Is the Tercel 4WD Wagon for you? Just ask Santa.

*Remember: Compare this estimate to the "EPA Estimated MPG" of other cars with manual transmission. You may get different mileage, depending on how fast you drive, weather conditions and trip length. Actual highway mileage will probably be less than the "Highway Estimate."
**Optional chassis gravel shields recommended for rough road use.

BUCKLE UP—IT'S A GOOD FEELING!

THE 1984 TERCEL 4WD WAGON. THE OFFICIAL CAR OF SANTA CLAUS!

HO HO HO!

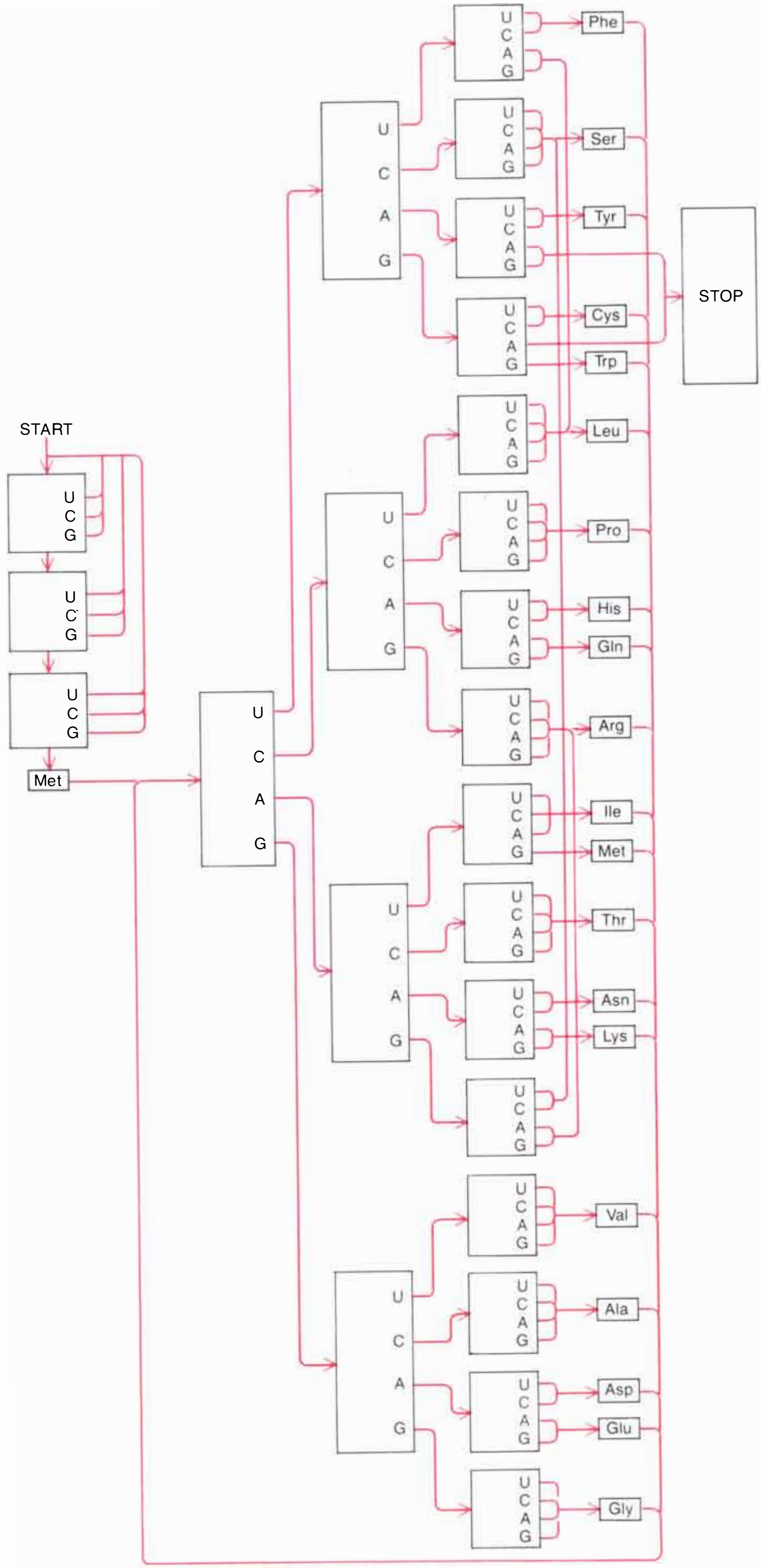
ly within the context of a finite-state system. The compiled "program" is a sequence of the standard three-letter symbols for amino acids; the symbols can be generated as the output of the states of the scanner that recognize codons. The three states corresponding to stop codons have no output.

Creating a compiler for a language large enough to be of general utility is not a casual undertaking, but the underlying architecture of the finite-state machine can at least provide an organizing principle. If the syntax of the language is specified with sufficient precision, part of the work can even be mechanized: it can be done by a compiler compiler, a program whose input is a formal description of a language and whose output is another program that translates statements in the language. As far as I know no one has yet written a compiler compiler compiler.

The identification of tokens by a lexical scanner is in itself a kind of parsing, and the set of all possible sequences of symbols in a token is a kind of language. Indeed, it is an infinite language: unless some artificial limit is put on the length of individual sequences, an infinite variety of recognizable tokens can be formed. How can a machine with only a finite number of parts recognize an infinity of well-formed statements and exclude an infinity of ill-formed ones? The key is in the structure of the language itself. If the statements of an infinite language are to be recognized by a finite-state machine, they must be formed according to strict rules.

The rules were set forth by Kleene in 1956; they define a class of languages called regular languages or regular sets. Kleene proved that a finite-state machine can recognize a language only if it is regular, and further that every regular language can be recognized by some finite-state machine. What is meant by regular can be indicated briefly (although not rigorously) by two rules. First, any finite language is regular and therefore can be recognized by a finite-state machine; after all, one could build a machine with a state for each possible expression of the language. Second, if a language is infinite, it must be possible to parse all its statements by reading one symbol at a time from left to right, or beginning to end, without backtracking or looking ahead. If the acceptability of any symbol is contingent on the presence of another symbol, the governing symbol must be the one immediately to the left.

The second rule is a direct consequence of the limitations of a finite-state machine, which can neither foresee its future states nor keep a record of its past ones; it must choose a state transition based only on the current state and the current input symbol. It is for this



A finite-state machine translates the genetic code into protein reason that the subtractive notation for Roman numerals cannot be handled by a finite-state machine. If the expression XI is read and the machine interprets it as 11, it cannot go back to revise the value when the next character turns out to be V. Many other functions are ruled out by the same limitation. For example, it is not possible to build a finite-state machine that reads a sequence of binary digits and determines whether the number of 1's is equal to the number of 0's. Similarly, although a finite-state machine can add binary numbers, it cannot multiply them; I leave it to the reader to deduce why.

Beyond finite-state machines and regular languages there extends a hierarchy of more powerful machines and more general languages. It is called the Chomsky hierarchy, after the linguist Noam Chomsky, who investigated the various formal languages as possible models of natural language. The more general languages are created by relaxing constraints on the grammatical rules of regular sets; the machines are built by adding memory elements to the basic finite-state model.

The next machine in the series is called the pushdown automaton. It consists of a finite-state machine with the addition of a memory array that has an infinite capacity but a peculiar organization. The memory takes the form of a stack, like a counterweighted stack of cafeteria trays. An item of information can be stored only by putting it on top of the stack; when the information is retrieved, any overlying items must first be removed. Thus the last item in is the first one out.

The language recognized by a pushdown automaton is called a context-free language. In parsing its statements the acceptability of a symbol can depend both on the symbol immediately to the left and on the one immediately to the right. This bidirectional dependency is permissible because any symbols whose interpretation cannot be decided immediately can be stored on the stack until the ambiguity is resolved. Hence a pushdown automaton can work with subtractive Roman numerals, and it can identify expressions with equal numbers of 1's and 0's (or other symbols, such as left and right parentheses). On the other hand, it cannot detect sentences with equal numbers of three symbols, such as 0's, 1's and 2's. Most programming languages are context-free, and the parser of a compiler is generally a pushdown automaton. Many computers include hardware facilities for organizing a part of the memory capacity as a pushdown stack. One programming language, Forth, makes a stack the primary memory structure. Of course, the stack in any real machine cannot have infinite depth.

The context-free languages merit their name because the parsing of any symbol can be influenced directly only by the symbol's two immediate neighbors, not by the wider context in which it is found. Removing this constraint gives rise to a context-sensitive language and once again increases the difficulty of interpretation. Now widely separated symbols can interact; in the worst case it is not possible to interpret the first symbol in an expression until the last one has been read. In exchange for the added complexity somewhat greater capability is gained. A machine based on a context-sensitive language can determine whether an expression includes equal numbers of three symbols.

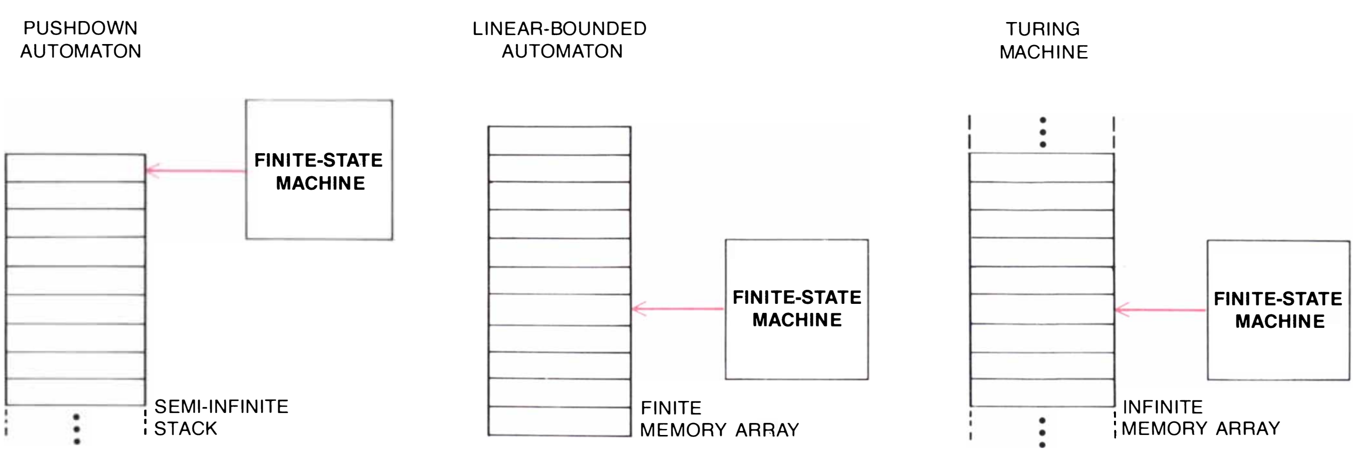
The machine that can recognize a context-sensitive language is a linear-bounded automaton. In addition to the usual finite-state apparatus it has a memory organized in such a way that any storage location can be reached at any time; it is a random-access machine. The memory is only finite in capacity, but it is assumed to be large enough to hold any input the machine receives. The linear-bounded automaton seems a good approximation of the von Neumann model of a digital computer. Oddly, though, the corresponding context-sensitive programming languages seem to be rare; evidently the simpler context-free structure almost always has sufficient expressive power.

All the languages described above have a property in common: they are said to be recursive. What this designation amounts to is that one can imagine a procedure for generating all possible "utterances" in the language in order of increasing length. It follows that there is a guaranteed method of deciding whether any given statement of finite length is a member of the language: simply generate all the statements up to that length and compare them.

There are languages that cannot meet even this minimal standard of tractability. For them there is only one possible recognizing machine: it is the computer of last resort, the Turing machine, a finite-state automaton allowed to roam freely through an unbounded memory. In the description given by Turing the memory is a tape, infinite in both directions and marked off into cells, which the finite-state apparatus can write on, read or erase.

Looking down from the elevated perspective of the Turing machine, the relations of the lesser computing devices become clearer. The linear-bounded automaton is simply a Turing machine with a finite tape. The pushdown automaton has a tape that is infinite in one direction, but the "head" for reading and writing on the tape always remains fixed over the last nonblank cell. The finite-state machine is a Turing machine with no tape at all.

Brand-name-conscious readers, eager to parse nonrecursive languages, may already be out shopping for a Turing machine. They should be warned that the ultimate computer also has its weaknesses. There are languages with grammars so preposterous that even a Turing machine cannot be counted on to recognize their statements in a finite amount of time. So far such languages have found little use in the world of computing machines, but people somehow manage to speak them.



The Chomsky hierarchy of finite and infinite machines

BOOKS

A Christmas survey of books on science and technology for the younger reader

by Philip and Phylis Morrison

This year we found before us a large number of computer books. Most of them turned out to be so specific as to be mere instruction manuals, useful and even necessary but lacking general interest. We hope that deeper treatments will come. Meanwhile we present a collection of other topics, this time richer for older young readers than for the youngest ones.

Laboratory and Field

FINDING YOUR WAY ON LAND AND SEA: READING NATURE'S MAPS, by Harold Gatty. The Stephen Greene Press, Brattleboro, Vt., and Lexington, Mass. ($8.95). A man or woman "with a good sense of direction is, to me, quite simply an able pathfinder ... who can find his way by the use of the five senses ... he was born with ... developed by the blessing of experience and the use of intelligence." Harold Gatty wrote this little classic of pathfinding without map or compass in his Fiji home not long before his death; it was first published in 1958. A ship's officer and a pioneer in aerial navigation, he became a famous air navigator on the occasion of his 1931 round-the-world flight with Wiley Post. He maintained a lifelong interest in finding one's way without benefit of costly gyroscopes. The book conveys the quality of an expert not by book alone but at first hand. It includes a learned look at surprising ethnographic material.

The smell of burning aspen in the Rockies usually means a homestead; the smell of pine smoke might mark a forest fire. Burning peat signals the Falklands many miles at sea; the Cape Verde Islands waft the scent of orange groves afar; the odor of drying coconut often carries out of sight of the atoll whence it comes; seabird colonies broadcast the ammoniac guano smell, welcome enough to a lost sailor. Brigadier R. A. Bagnold, a gifted student of the desert, once found a water hole in broken country "entirely by the smell of one camel, which was picked up eight miles away." In the lifeless desert the background of organic smells is low: the signal-to-noise ratio is good.

Gatty's point is less the utility of our underused sense of smell than it is the value of meticulous attention to local circumstances. He develops the theme brilliantly. Moss on the north side of trees is only the most banal of a dozen different clues from the habit of trees and other plants and from the farmhouse too. These clues are regional rather than general: local knowledge counts. The moss clue is a "dangerous generalization" in the old books. The preferred direction depends on wind and rain; moisture is what the mosses seek. In the Jura Mountains mosses and lichens mark the southwest. Termite nests, ridges on polar snowfields and the form of desert dunes offer more exotic orienting data and exemplify the method.

Only the Micronesians and Polynesians have understood how to read the ocean swells for signs of distant land. One of their wave-pattern diagrams woven in coconut palm is shown. There is a fresh small chapter on ethnocartography; quite a few records exist of the use of maps, both mental and graphic, in a good number of preliterate cultures, often in training contexts. There is even a chapter on finding your way in towns: television aerials, shutters, rust, worn paint, traffic gradients—also excellent training. Distance? Fifty yards is the distance to a man when you first see the white of his eyes, as it was at Bunker Hill; at 800 yards a man looks like a post. An average church steeple can be made out at nine miles.

There are now other books of this ingenious art, but these inexpensive words, with sun tables, star maps and bird pictures, all by a master, are welcome. Swift, lively reading, Gatty should be prized by many a young adventurer out in the woods or in an armchair, from the upper grades onward.

VOLCANO WEATHER: THE STORY OF 1816, THE YEAR WITHOUT A SUMMER, by Henry Stommel and Elizabeth Stommel. Seven Seas Press, Inc., Newport, R.I. ($15). You won't find it in *Old Farmer's Almanac*. Contrary to folk legend, this venerable source of forecasts said not a word about the unseasonable frosts that struck the New England highlands in all 12 months of the year 1816. The facsimile pages are here to prove it. That year deserved its folk name of "eighteen hundred and froze to death."

In this delightful example of science and scholarship done persuasively as one the Stommels (he is a veteran oceanographer from Woods Hole) marshal the evidence to make clear what happened in the strange summer of 1816. From primary evidence, notably the meticulous temperature records kept by a network of college presidents, including those of Harvard and Yale, it is plain that the summer went awry. New Haven had the mean June temperature we normally expect at an inland site 200 miles north of the city of Quebec. The frost-free period at Yale lasted for only 71 days that summer, although the mean value in a cool decade was 126 days. Those hardscrabble northern Yankee farmers who depended on the maize they grew as a staple had very poor crops, and in interior Quebec even the hardier wheat failed. The farmers of Maine and Vermont set out for Ohio and Indiana in droves; although the emigration was not, of course, solely in that year or solely for the reason of climate, graph after graph supports the idea that 1816 was a year of flight.

The probable cause is known: dust in the high air from a remote eruption. The volcano Tambora exploded in April, 1815, depositing ash a foot thick over the island of Bali a couple of hundred miles to the west. It is thought to have been the greatest volcanic explosion in historic times, the volume of ejecta 100 times those of Mount St. Helens and 10 times those of Krakatau. In a fine closing chapter the reader is offered the summary data, in aligned curves over the decades for sunspots, temperature and sources of volcanic dust, and is challenged to form an opinion. Formal statistics help little; the authors, and at least these readers, see the data as suggesting that volcanoes cause cold spells. Whether dust cooling will favor us by compensating for fuel-fed carbon dioxide warming, or whether the soots of nuclear-ignited cities and forests will spell worldwide catastrophe, are questions the future will decide. Any able young readers—or their parents—can learn to appreciate and to share the concerns from this brief, engrossing and scrupulously honest work.

KITES: THE SCIENCE AND THE WONDER, by Toshio Ito and Hirotsugu Komura. Japan Publications, Inc. Distributed by Kodansha International/USA, Ltd., through Harper & Row, Publishers, Inc. ($11.95). Old Japan: the strong, subtle handmade paper of mulberry and the light, tough, elastic lengths of bamboo are worked into kites of high performance and intriguing shape in the hands of masters working in a tradition of centuries. New Japan: over four years


THE GIFT OF LOVE
SARONNO
AMARETTO
di
SARONNO®
ORIGINALE
A LIQUEUR PRODUCED BY
ILLVA SARONNO ITALY
750 ml
56 PROOF
PRINTED IN ITALY
AMARETTO DI SARONNO
ORIGINALE·1525
TO SEND ANYWHERE IN THE U.S. CALL (800) 528-6148. VOID WHERE PROHIBITED.
LIQUEUR 56 PROOF. SOLE U.S. DISTRIBUTOR, FOREIGN VINTAGES, INC., NEW YORK, N.Y. © 1983. A UNIQUE PRODUCT FROM ILLVA SARONNO, ITALY

M E R C E D E S - B E N Z

The new Mercedes-Benz 380 SE. With its 3.8-liter light-alloy gasoline V-8, sophisticated fully independent suspension

The performance renaissance accelerates: introducing the new Mercedes-Benz 380 SE Sedan.

THE NEW 380 SE joins the Mercedes-Benz line at a most propitious moment–just as America rekindles its appetite for mightily responsive automobiles.

An automobile more likely to satisfy that appetite is difficult to conceive.

The 380 SE is, in a word, potent. It combines a 3.8-liter gasoline V-8 engine with a 3,685-lb. curb weight, creating a ratio of power to weight conducive to formidable over-the-road performance.

As invigorating as the 380 SE's turbine-smooth acceleration is its willing throttle response at almost any speed. Power reserves seldom feel less than massive. Yet in stop-and-go urban traffic, the Mercedes-Benz 380 SE feels as docile as the proverbial lamb.

This robust V-8 engine is testimony to how far Mercedes-Benz has advanced performance-engine technology since the age of the cast-iron behemoths.

A LIGHTWEIGHT V-8

For example, the block itself is one of numerous engine components fabricated in *aluminum alloy*–making this V-8 as notable for the weight it saves as for the power it generates.

It is meanwhile as resistant to heat, vibration and wear as a conventional iron-block V-8. Note that the pistons ride on cylinder walls having a silicon surface–a substance so hard that it must be machined by diamond-tipped tools.

Mercedes-Benz engineers introduced fuel injection to passenger cars. In the 380 SE, fuel is metered to the cylinders by a sophisticated third-generation C.I.S. continuous injection system.

Power flows to the rear wheels through a crisp-shifting four-speed torque converter automatic transmission.

The 380 SE's performance is maximized by its extremely low wind resistance. (Why battle the wind, when you can outwit it?) Its 0.37 aerodynamic Cd comes within a whisker of the most ballyhooed aerodynamic sensations of the day.

RESPONSIBLE ROADHOLDING

The 380 SE's forged light-alloy wheels are the most visible elements of a suspension system set up for determined driving.

Mercedes-Benz believes that a performance automobile that was all speed and no handling would be an act of engineering irresponsibility. Thus, you may be as taken with the 380 SE's roadholding as with its power.

You may be equally taken by its steering precision. A power-assisted recirculating-ball steering system lets you guide this substantial machine with pinpoint accuracy. No luxury-car sponginess.

*See your authorized dealer for complete warranty information. ©1983 Mercedes-Benz N.A., Inc., Montvale, N.J.

and four-wheel disc brakes, this is one performance sedan meant to go, to handle, and to stop.

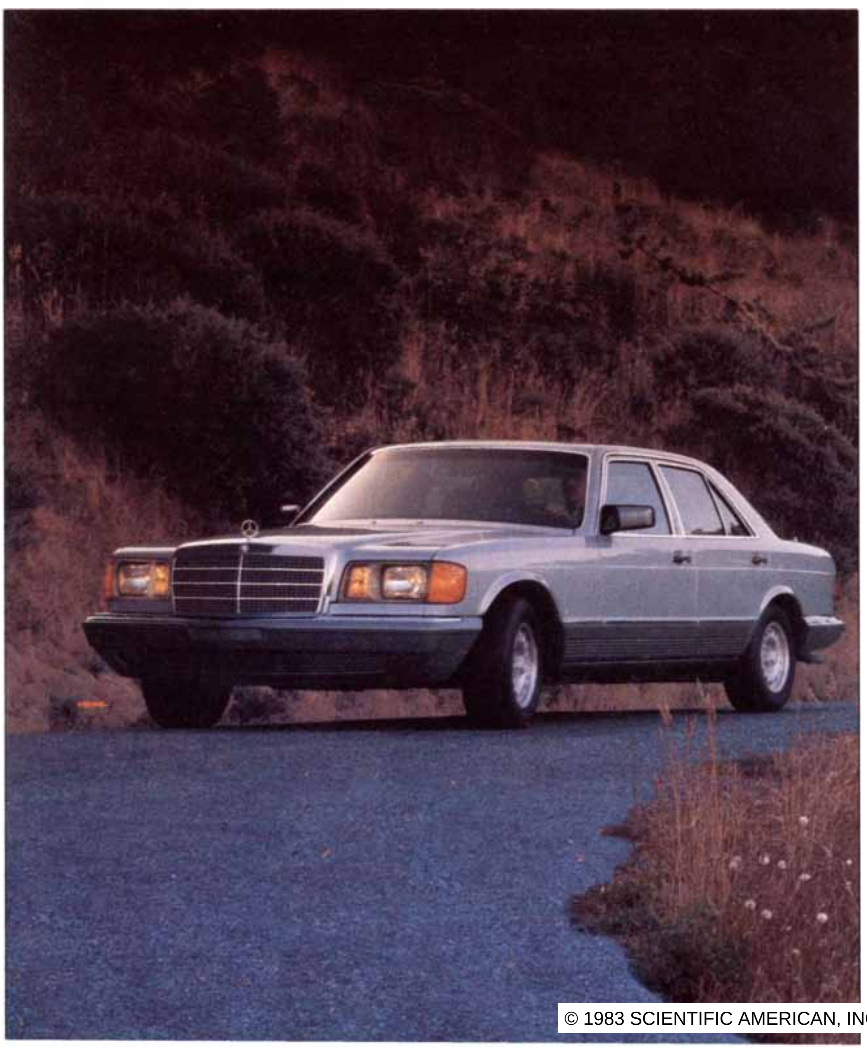
Automobiles built to move this well should also be built to stop with equal responsiveness. The 380 SE is engineered to stop with the power of four disc brakes. Total swept brake area: 456 square inches.

CRAFTSMANSHIP LIVES

The cabin accommodates five persons within 93.6 cubic feet of volume. Their comfort is further enhanced by an array of electric, electronic and other amenities. Especially ingenious are the 10-way electric adjustment controls for each front seat–activated by a tiny facsimile of a seat.

The 380 SE confirms, among other truths, that patiently crafted automobiles are still being produced in the 1980s. It is an example of consummate workmanship–from the precise fit of every panel to its hand-finished Zebrano wood interior trim.

Mercedes-Benz wants your 380 SE to be as pleasurable to own as it is to drive. One proof of this is a new *48-month-or-50,000-mile* limited warranty.*

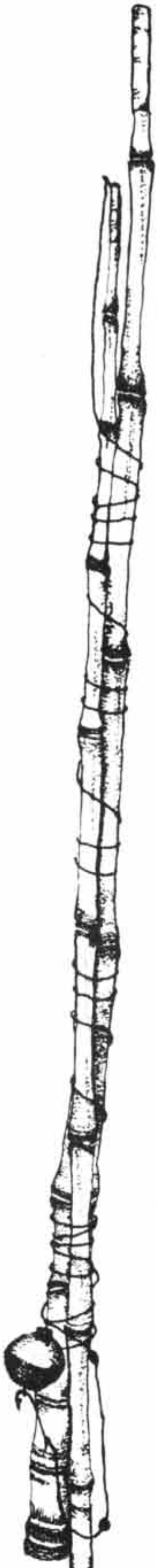


RESALE PERFORMANCE

Mercedes-Benz automobiles have been shown, year after year after year, to retain a high percentage of their original value. This performance can be as reassuring, in its way, as the 380 SE's over-the-road performance.

The ultimate 380 SE accolade may have been voiced by the Editor of *Car and Driver*, in a recent survey of America's most desirable automobiles: "There is nothing like a big Mercedes V-8. Nothing. It is a car that is simultaneously fun to drive, strong as a D-8 Cat, and luxurious enough for all but the irresponsibly sybaritic."

Engineered like no other car in the world



two expert engineers, making and flying together more than 1,000 trial kites, develop a full-blown aerodynamic theory of kites and elaborate a system of trial design, accessible to earnest novices, achievable in materials both old and new, high in performance and shaped after birds, insects and fishes. This is their own account, published in 1979 in Japan and newly translated into serviceable English.

The opening chapter, with many force diagrams and performance curves, puts before a reader who has some grasp of high school mechanics the most complete theory of kites in stable flight yet published. That is followed by a well-argued account, although one much less complete and quantitative, of kites seeking stable flight: the spins and yaws and flapping that plague every builder. The latter half of the brief but full book is less analytic: it looks to construction. Materials are discussed and their properties tabulated; modern wonders such as Tyvek and graphite composites are included, although not nylon sailcloth, a staple of U.S. kite builders. The authors' system of design and construction is outlined, and detailed kite forms are drawn and discussed. A terse final chapter treats wind and the kite builder's adjustments in response to it.

In an empirical art this old and refined the rise of a full theory, to which the largest contribution has been made by these two authors, is not perhaps an obvious event. Their kites perform very well in a free choice of forms, but so do many others. No serious kite builder can afford to overlook this book; it opens an entire road of investigation for clubs and informal teams of kitepersons, whose patience and intuitive skills can now be turned to rewarding engineering at the level of high school physics and well beyond.

If the reader is not misled, the implicit scaling of these formulas suggests that kites built on plausible rules of proportion will fly in wind speeds that decrease with the diameter of the kite only as the square root, and that the altitude the kite can reach under the load of drag and string weight increases in the same slow way. The sky is forgiving to both small kites and large, structure permitting.

THE SCIENCE IN SCIENCE FICTION. General editor, Peter Nicholls; contributors, David Langford and Brian Stableford. Alfred A. Knopf, Inc. ($25). Nearly 100 thoughtful pieces of only two pages each are assembled into related topics by chapters and are supported by a wide choice of illustrations, most of them, although not all, reproduced or redrawn from published sources. The themes are part and parcel of the received doctrines of science fiction: starships, aliens, hyperspace and hyperspeed, time travel, nuclear holocaust,

robots, clones, telekinesis, flying saucers and force shields. The three British authors bring impressive credentials. The general editor is a master of the canon of science fiction; the two contributors were trained respectively as a physicist and a biologist and are both science-fiction writers as well. The illustrations include reality, such as the jet of the radio galaxy M87 and a wide defoliated strip along the DMZ in Vietnam; travesties, such as Erich von Däniken's slightly retouched version of a Mayan relief seen as a drawing of an astronaut; art, such as the bronze she-wolf suckling Romulus and his brother, and plenty of fictional representations, from the great mother ship of *Close Encounters of the Third Kind* to the less ethereal covers of old issues of *Amazing Stories.*

In just about every piece there is a literary mention of the science fiction of the past and present that includes that particular theme, followed by a popular but knowing account of the actual situation as far as scientific knowledge or likely consistent conjecture goes. The entire volume stands as clear evidence of how pervasive the imaginative ideas of science fiction in fact are, of how many of the genuine dilemmas of a world immersed in the implications of a newly powerful science are reflected in the popular genre in all media, sometimes intelligently, sometimes merely for calculated effect.

The next most striking fact is that all these discussions are sensible, even helpful, on their straightforward expository level. Perhaps not everyone would agree with all the conclusions, since they range over human expectations broadly. Not one of the brief essays, however, seems in scientific error or illogical; quite the contrary. A deal of best-selling junk is disposed of by the evidence, and plausible conjecture is compared with what we know and what we only surmise. For example, light speed and time travel, even black holes and tachyons, gain entirely reasonable exposition here.

The book is a bridge between fiction and reality that does its fiction-oriented architects proud. Young readers from the middle grades on would find it absorbing; school libraries need it at any level, a sure winner for topical reference and bibliography, fact as well as fiction.

KITCHEN CHEMISTRY, by Robert Gardner. Illustrated with photographs and with line drawings by Jeff Brown. Julian Messner/Simon & Schuster ($9.29). Stove, sink, counter and refrigerator, the household kitchen is a well-equipped laboratory, better than many a college student can boast for a good deal of chemistry and physics. Using it requires tact, since the house depends on it. Given a degree of thoughtfulness, any young person can find satisfying results in the kitchen.

This small book outlines an entire set of experiments for each of the major pieces of kitchen equipment. At the sink it is capillarity that draws first attention. Bubbles, gas collection and jets are not forgotten, even the electrostatic attraction of a water jet (perhaps not as easy as it looks). Melting ice and boiling water offer a chance (it takes patience) to follow in the famous footsteps of Joseph Black, who first understood and measured those latent heats by simple timing. Electrolysis is proposed safely, relying on battery power. There is a good deal more, all of it solidly workable. This is no mere stew of plausible old proposals but a sound guide to what actually works. (Candles burned in sealed jars go out before they use up all the oxygen in the air, and the way to prove it is here.) The enthusiastic book is overrich; it would take months for a devoted kid to carry out all that is in it.

RAIN SHADOW, by James R. Newton. Illustrated in black and white by Susan Bonners. Thomas Y. Crowell ($10.95). The book begins with a view at the shore, the air heavy with the ocean moisture. Clouds form and float above the evergreen forest; we are in the Cascades. In the high cold clouds the droplets of water bump, grow larger and fall as rain on the mountain slopes nearly every day of the year. Here are ferns and mosses, tall trees draped with green and bushes growing from every decaying log; even the air seems green. We climb to flowery snow-fed alpine meadows, and then to the snowfields themselves, too cold for much life. Now the air has lost its water; the valley beyond lies in a rain shadow. Cactus and sage abound; the farmers can grow crops only because irrigation canals capture some of the river water that flows into the valley from the heights. The case is not unique; a few others are mentioned, where ocean-born rainfall is shadowed by mountains. The clear text, a few sentences to a page, not without useful detail and even a few numbers, is strengthened by the textured and convincing charcoal drawings. The book is a cool pleasure for readers in early grades who like to travel in reality or in imagination.

Plants and Animals

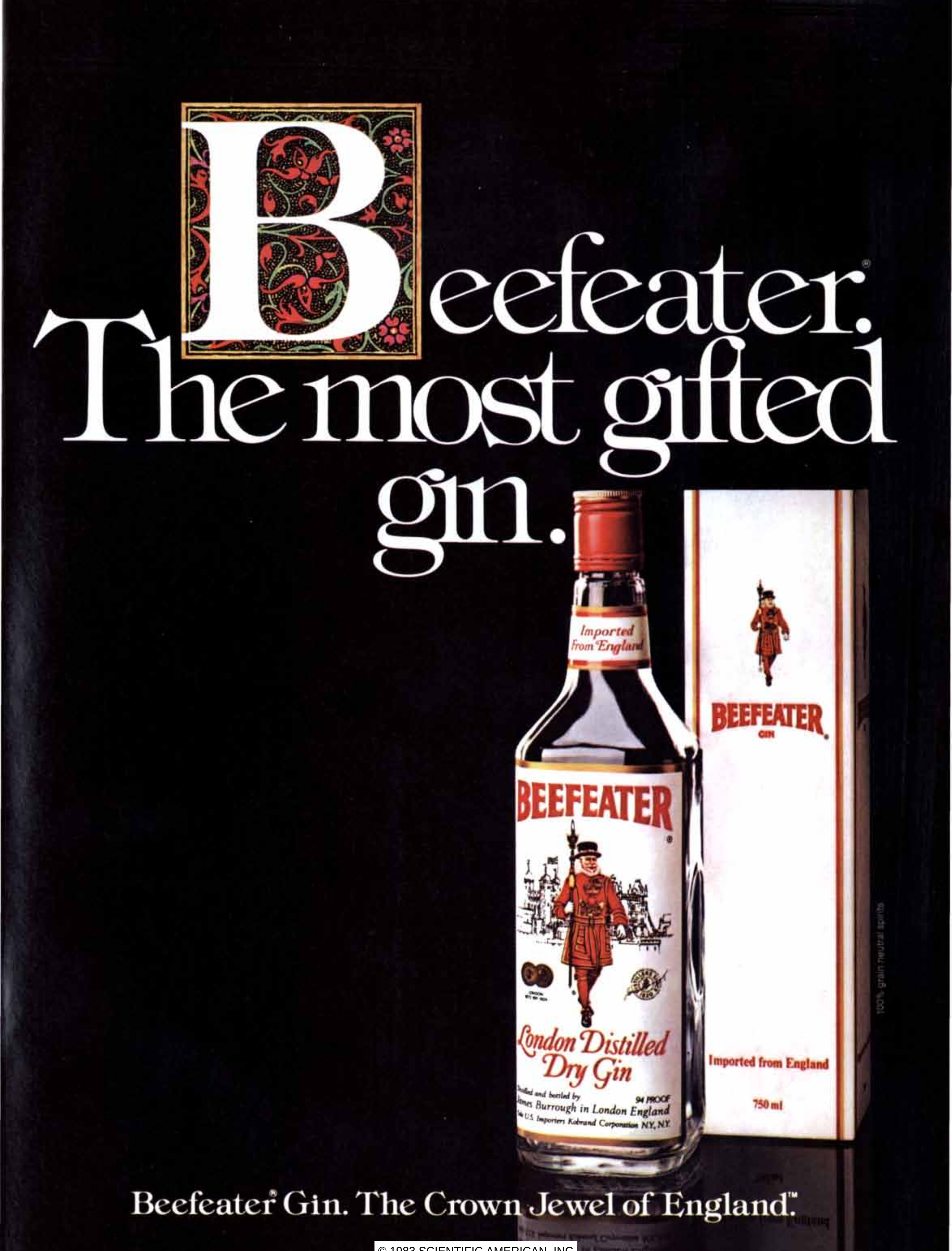
DINOSAURS: AN ILLUSTRATED HISTORY, by Edwin H. Colbert. Hammond Incorporated ($30). THE SUCCESSFUL DRAGONS: A NATURAL HISTORY OF EXTINCT REPTILES, by Christopher McGowan. With original illustrations by Marg Sansom-Markezinis. Samuel Stevens and Company, Toronto and Sarasota ($29.95). Finicky taxonomists fluent in a Hellenic jargon of their own, the vertebrate paleontologists in the dinosaur trade remain admirably broad-minded, lively and articulate scientists. They can write attractively for the general reader, and they do. Here are two top specimens of the art.

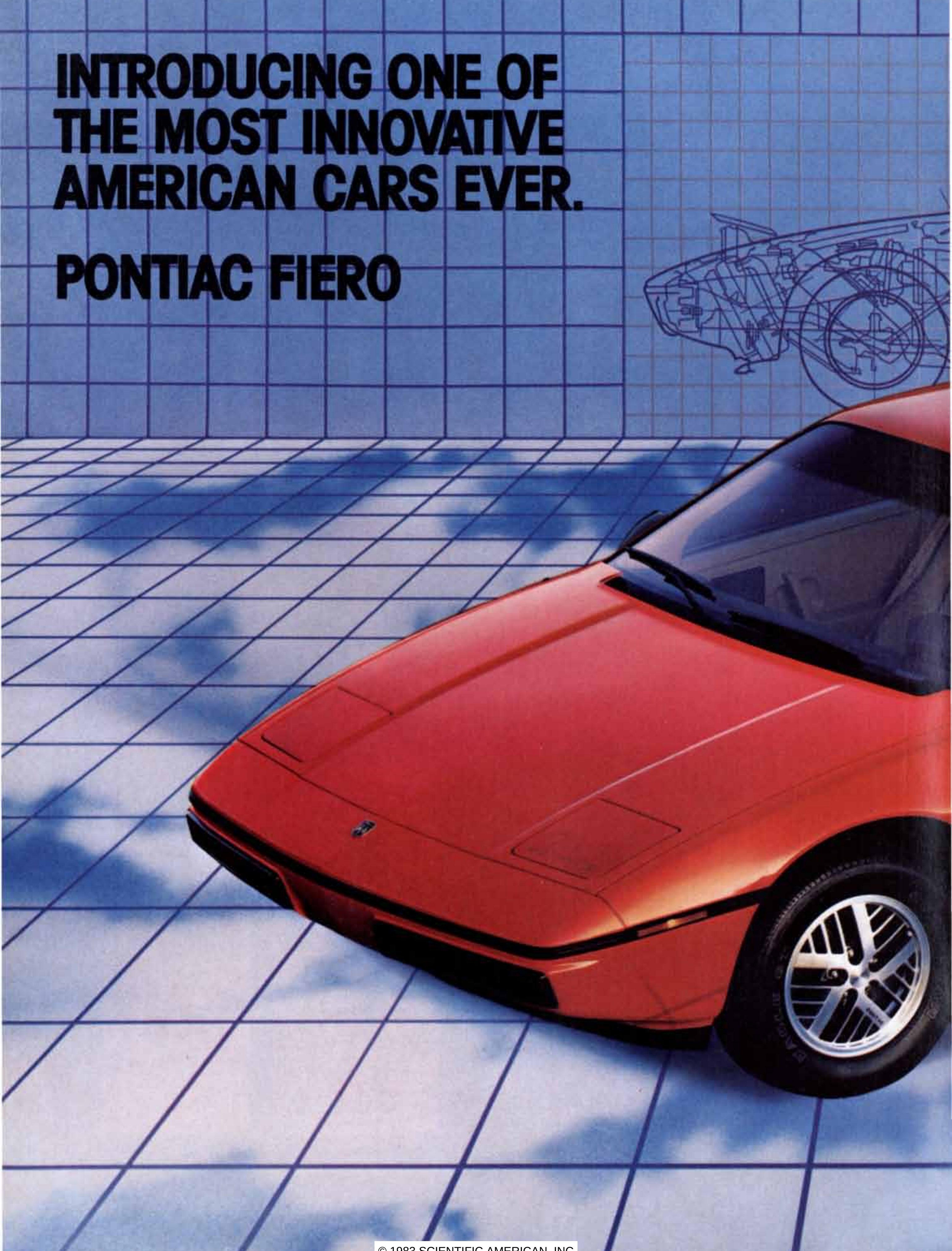
Dr. Colbert, for 40 years responsible for the exhibit halls that display these large fossils at the American Museum of Natural History in New York, is dean of American dinosaur experts. Dr. McGowan is for his part curator of the splendid renewed displays at the Royal Ontario Museum in Toronto; the books are manifestly both authoritative. The Colbert book offers more of what people call production value: its pages are full of colorful reconstructions, some classical, some new. Its opening spread presents a striking color close-up of the Victorian reconstruction of the first dinosaur recognized, the iguanodon, its image made of tile and painted concrete for the Crystal Palace near London in 1854.

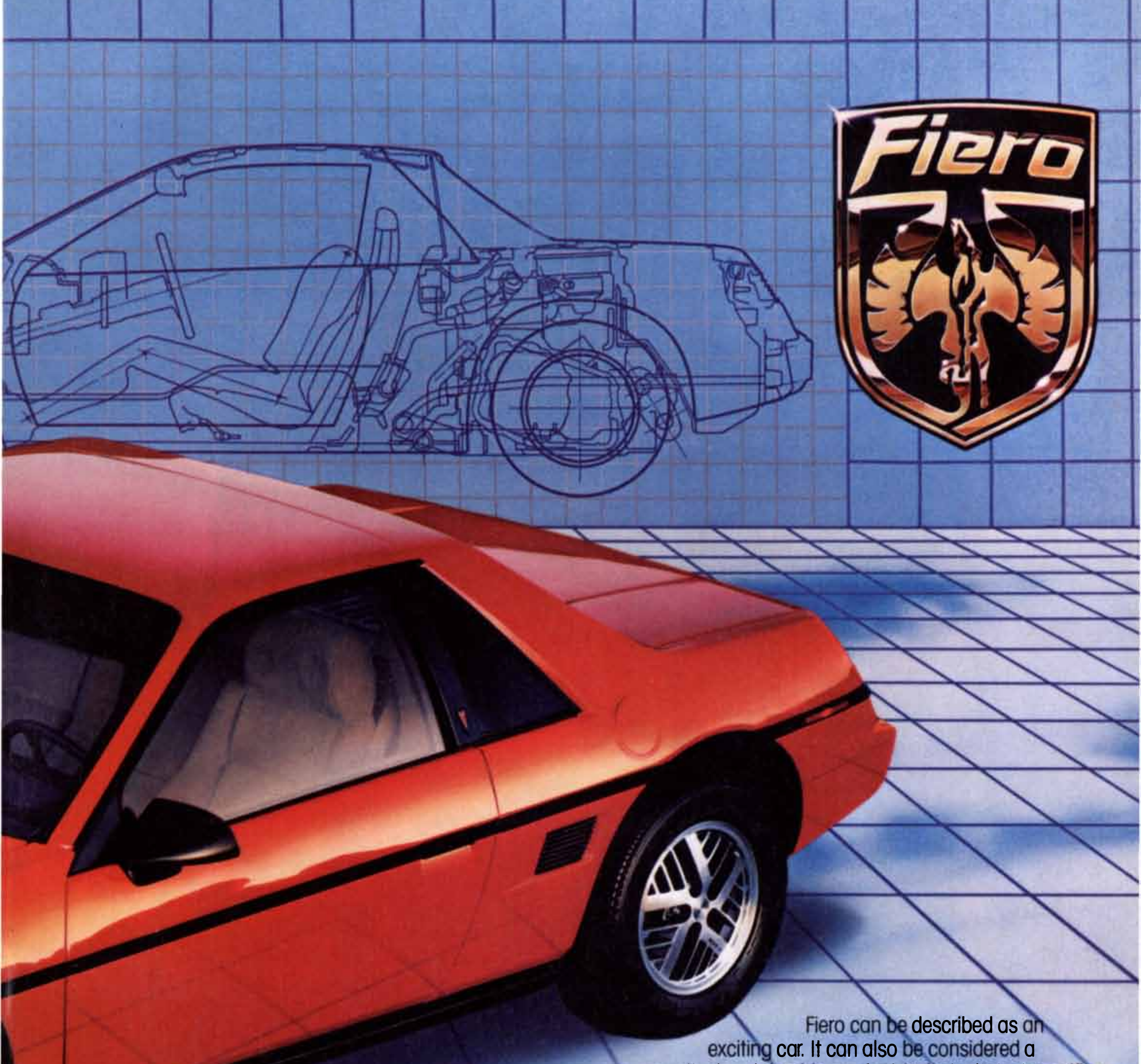
McGowan's book, handsomely illustrated without quite such panache, begins instead with a brief survey of all the orders of reptiles, and it treats of Darwin as historical prelude. McGowan is a specialist in the marine and the airborne reptiles of the past, not strictly speaking dinosaurs at all, so that his book takes a wider look at variety, with its center of gravity in the physical and biological problems of flight, of swimming fast at sea and of the scaling constraints on being a giant. Colbert instead centers his text and pictures on a winning account of the variety of the creatures—all those saurians, large and small, track, egg, skin and bone. McGowan tells, say, of the Reynolds number, a remark of a kind not to be found in Colbert; Colbert details many forms that do not occur in McGowan. The vivid books share an air of sustained high quality; one might see the Colbert, plainly intended for a wide family audience, as a comprehensive introduction to the entire topic, whereas the McGowan, more text than image, is a second book on ancient reptiles for those who seek a little more connection with biology as a whole.

Like evolution itself, the books converge a good deal. Both discuss very well the issues of the warm-blooded and vigorous behavior of some, not all, of the fossil reptiles. Both agree that the easy answer of asteroid collision, or indeed any other known cause, cannot really explain the rapid extinction of the beasts. (Indeed, these writers both think the true cause may be unknowable, but we shall see.)

One image might unite the books. The Red Deer River flows southeastward through the little Alberta town of Drumheller, now the "focal point of the dinosaur cult." The sandstone badlands of this part of Alberta have already yielded about 500 specimens, more perhaps than any other place in the world. New finds weather out of the soft rock formations year after year. The idyll


Beefeater.®
The most gifted gin.
Imported From England
BEEFEATER®
London Distilled Dry Gin
Distilled and bottled by James Burrough in London England
94 PROOF
Sole U.S. Importers Kobrand Corporation NY, NY.
BEEFEATER GIN
Imported from England
750 ml
100% grain neutral spirits
Beefeater® Gin. The Crown Jewel of England.™


INTRODUCING ONE OF THE MOST INNOVATIVE AMERICAN CARS EVER.
PONTIAC FIERO


Fiero

Today, you're doing more than ever. So should your phone.

AT&T introduces GENESIS™ Telesystem.

Genesis telesystem is designed to help organize and streamline your life in so many ways.

The Genesis telesystem remembers all your important numbers, including emergency numbers. And it dials with just a touch. It times your calls and even has a built-in speaker.

But the really terrific thing about the Genesis telesystem is that you can customize it to fit your own needs. Add an optional cartridge, and you've got one-touch access to call forwarding, call waiting, and three-way calling. Or add another cartridge, and you have the convenience of automatic re-dial of busy or unanswered numbers. And that's only half of it.

Soon, you'll be able to add a module with an electronic directory that memorizes names and numbers in alphabetical order. Or the reminder cartridge, to remind you of special occasions and appointments.

Genesis telesystem. The more you do, the more it will help you do it.

We set the standards.

AT&T

See it at your AT&T Phone Center, now at thousands of leading retail stores.

(mosquitoes apart) of this wonderful vocation occurred in the early years of the century, when the collectors drifted down that cool river on long lazy summer days, scanning the high banks from big tented rafts, to find fossil after remarkable fossil. Colbert shows one such floating field camp in a period photograph, and McGowan puts an end to the period, if not to the narrative, in a photograph that shows a thin band of coal marking a great change: below it a world ruled by reptiles, above it one under mammalian sway. At Hell Creek, Mont., an observer can stand with his feet in one era and his head in the other, a world revolution in at most a couple of million years.

WALKER'S MAMMALS OF THE WORLD, by Ronald M. Nowak and John L. Paradiso. Fourth edition, Volume 1 and Volume 2. The Johns Hopkins University Press ($65). The first edition of this grand work of reference appeared in 1964, the outcome of 30 years' work by Ernest P. Walker, then assistant director of the National Zoological Park. His expert continuators have brought the work up to date in a series of new editions; here it is, better than ever, the only reference that offers information on just about every known genus of the mammals, our own class. More than 4,000 of these marvelously diverse species are included, about 1,000 genera, in 1,400 pages.

Plainly there is not much space available for any single species, even for human beings, who are given two and a half pages but no pictures. That so many forms are pictured is the strength of this comprehensive reference for young readers and their libraries. The text is moderately technical, with terms of art for structure and color. The brief articles include as a matter of course, however, the distribution, relationships and interesting behavior for all the genera and most individual species. There is a black-tailed prairie dog (no dog, but a stout, short-legged squirrel) alert in its conical crater, and the reminder that many years ago in western Texas there was a large colony, a town, of the animals said to have sheltered 400 thousand individuals. Here is a shelly three-banded armadillo from Brazil, curled up in Walker's hand and resembling a coconut. There is the whitefin dolphin, which lives along the Yangtze River for 1,000 miles up from the mouth; it has long enjoyed protection by custom, and it is now under legal protection as well, although silting, motorboat collision and water use are reducing its numbers.

It is up to the reader, however, to seek out more interesting stories among the thousands here. A table of distribution among world regions covers every recent genus of mammal, and an indispensable index is found in each volume, with both common and Latin names. More than 3,000 papers and books are cited in detail. Less showy than its authoritative treasures merit, Walker's is a reference bargain for any library where young readers and old enjoy animals.

OAK & COMPANY, by Richard Mabey. Illustrated by Clare Roberts. Greenwillow Books, William Morrow & Co., Inc. ($9.50). "One winter night in its 283rd year, the Oak was hit by a blizzard." The old tree had only a few leaves as more and more branches were shaded out; its weakened roots could not hold the heavy load of snow and the great tree toppled. The foresters came soon and cut it up; the timber was of little use with its crooks and hollows. The tree rotted slowly away, still host to toadstools and insects, as in life it had held a great company, woodpeckers and butterflies, mosses and beetles, now moved to the straight trunks of the Oak's offspring nearby, timber oaks perhaps in the far future.

In this beautifully illustrated book young readers will share the long, slow story of an oak, a living organism that begins as an acorn, matures to bear its own young only in its 49th year, then for a long time endures chance and the elements. Human beings visit and use this English forest thoughtfully over centuries, and an entire community of life centers on and under the great tree, including a million earthworms slowly eating through the litter of the forest floor. The pace is slow and majestic, although we move also to the swifter rhythms of woodpeckers and pigs, bluebells and cepes, coppiced stumps and spangled galls. The quiet account (with a few phrases unfamiliar in American speech) would appeal to thoughtful middle-grade readers.

THE CRANBERRY BOOK, by Elizabeth Gemming. Coward-McCann, Inc. ($9.95). If the Royal Navy banished scurvy by issuing West Indies limes to its men, the limeys had a counterpart in American ships. The whalers and the smart China clippers carried tart red berries, rich in vitamin C, able to stay fresh for two years packed in barrels filled with spring water. This is a well-written and interestingly illustrated little monograph on the cranberry, suited for the right readers in the middle grades, even though it tells more than many might care to know.

Today Massachusetts and Wisconsin grow more than four-fifths of the North American crop. The term cranberry bog is a word of art: cranberries grow well only in drained places, not in true wet bogs. The roots require the acid peat subsoil of a former bog; the runners thrive in a layer of sand; the crop is typically watered by sprinklers. Nowadays the berries are harvested by flooding the ripened vines, churning the shallow water mechanically into turbulence so that the berries come off their stems to float "like a gorgeous ruby carpet."

The Wampanoag used the wild berries well, but there is no record of cranberry sauce at that first Thanksgiving of 1621. Cranberries were a wild resource of no value until Captain Henry Hall of East Dennis first transplanted them in about 1816. The national crop is today some two million barrels per year (a small barrel, peculiar to cranberries, of less than 25 gallons), nearly all consumed as frozen, tinned or bottled products. Visitors to the Cape Cod National Seashore can seek out both wild berries and a small cultivated bog.

ED EMBERLEY'S SCIENCE FLIP BOOKS: THE CHAMELEON/THE CHICKEN; THE FROG/THE HARE; THE BUTTERFLY/THE DANDELION. Little, Brown and Company ($5.95). Six "nature adventures," little dramas of natural history, are presented here in three small, colorful flip books. All the sequenced paintings are carefully and brightly done; for each story there are a few lines of initial text to set the scene, and the rest of some 60 pages are shaped and bound to allow the quick flip that generates a moving picture. The chick pecks out of her egg and grows past a pullet to a hen. The chameleon walks a few slow steps as its color changes. Frog and monarch butterfly metamorphose grandly. The hare's coat turns from winter white to summer brown as it escapes the fox. The dandelion blossoms and sets seed into the winds. Somehow gentle in mood, these hand-sized books are amusing and instructive, whether flipped fast or savored slowly page by page.

FERAL: TAME ANIMALS GONE WILD, by Laurence Pringle. Macmillan Publishing Co., Inc. ($9.95). The rock dove was a wild creature of the cliffs. Buildings of wood and stone offered the birds a new set of cliffs, conveniently near grain and pools of water. By Greek times they were half-tamed. The dovecote housed them at human expense, although they were free to come and go. In compensation the people collected occasional squabs for food. Domestication of the species was clear. Nowadays urban dovecotes are few, and the squabs of the cities are safe. The large pigeon populations are now feral; the mute swan and the monk parakeet, hardy in cold New York City, are other bird species that remain near human settlement, although they have escaped from our direct control.

Bands of feral pigs, omnivorous, fierce and clever, roam the forests of Hawaii, California and the Southeast; there may be two million of them. In the 1870's there were as many as 3.5 million feral horses in the unfenced West. Now their number is down by a factor of some 200, but what to do about the remnant remains a dilemma. The burros of the deserts and the ownerless cats and dogs that wander city streets are further witnesses to defects in the complex unwritten contracts between people and the few animal species that have come to mean much to human life over the past few thousand years.

Laurence Pringle's simple, informed text will engage more than a few readers in the fifth and sixth grades, whose empathy with these introduced living beings, neither chattel nor yet alien, is not a bad basis for a first encounter with knotty ethical and economic problems. The book is dedicated very affectionately to a Burmese cat we see in one photograph, "so wild, so tame, so fine."

Technology Past and Present

FROM HAND AX TO LASER: MAN'S GROWING MASTERY OF ENERGY, by John Purcell. Illustrated by Judy Skorpil. The Vanguard Press ($16.55). In a deft and personal prose, full of unexpected insights arising from widely shared experiences, the author has made a believable and never credulous narrative of the entire history of human technology. It begins with the beginning, the Rift protohumans of two million years ago, then dwells on the fire users of the cave at Chou-kou-tien. The theme of energy provides a minor framework, if an interesting one, for the account. The simple machines of grade school physics are fully as relevant, and they are very usefully discussed. We find stone knife blades and infer the digging stick, a clear lever. Fire making is pleasantly handled: the conservative archaeologist is shown in doubt about the fragment of charred beechwood rod from a Mousterian site in Yugoslavia. Maybe it was a fire-hardened spear. His radical colleague has no such doubts: it is a friction fire stick. Why not fire making in that cold epoch 100,000 years back? "For myself... I am content merely to hazard that some Neanderthal men knew how to *make* a fire when necessary," Purcell writes.

So the text moves, always lighthearted but sensible, parting certainty from plausibility, all the way from early hominids as mere scavengers and not mighty hunters to spoked wheels, possibly invented more than once. Up-to-date always, the material used includes the evidence on flint mining in the Neolithic, with the simple reckoning that establishes the large scale of the mining efforts. In the same way we read of the origins of writing, lately shown by Denise Schmandt-Besserat to be linked with the clay tokens commonly found in the preliterate trading villages of southwestern Asia, three millenniums older than any Uruk seals. Pottery, cast metal, wheels for many uses and much more are surveyed in the same tone: the best evidence, its uncertainties and a sensible trial interpretation, without pretensions. The last tenth of the book treats of the heat engine, of James Watt and his context, and on to radio via Reginald Aubrey Fessenden's pioneer voice broadcasts from Cape Cod in 1906. Lasers end the book, a little trendily perhaps, with a few pages on zapping.

The blend of good sense and wide learning, served up with good humor, makes a remarkable book for young people and their elders as well. It is a contemporary introduction to human history and prehistory, not meant to be comprehensive, yet more central in its choice of topics than any outline of kings and their battles, and as much fun to read. A good deal of the fundamentals of how important things work, from fire flints to infectious diseases to steam engines, comes along with the chronicle.

WINDMILLS, BRIDGES, & OLD MACHINES: DISCOVERING OUR INDUSTRIAL PAST, by David Weitzman. Charles Scribner's Sons ($13.95). Keyed sectional drawings of an Adirondack iron furnace of 1849 (it stands a ruin near Tahawus, N.Y.) and a fine old Atlantic locomotive, a speedy 4–4–2, set the tone of this appealing book for junior-high archaeologists and up. The concise account of the evolutionary growth of industry, say for locomotives from *Tom Thumb* to *Big Boy,* offers not only perspective but also a start toward understanding the technology. The author takes some pains to give a very simple account of how steam engines and canal locks, boilers and smelters, and the structural trusses in bridges and roofs actually work. The simpler versions of the past are a more manageable starting place for the particular go of these devices than the larger and more complex examples of our time. It may seem commonplace to cross an unimpressive little bridge, but "you're a historian, an industrial archaeologist. You walk back a few yards, climb down a grassy embankment a ways, and...." Maybe you've found, like the author, a Whipple truss, modified for cast and wrought iron, the rods going every which way, the oldest iron bridge still in service in California.

The library too is a hunting ground for the search. Hidden in the books and magazines are plenty of old devices, and what people thought of them a century and more ago. "Old copies of *Scientific American* are a good place to start." Even better than the author's similar guide for adults, this small book is full of interest for mere readers and is better still for little teams of kids eager to get out into the field to some purpose.

FIREWORKS! PYROTECHNICS ON DISPLAY, by Norman D. Anderson and Walter R. Brown. Dodd, Mead & Com-

pany ($9.95). It is the family Grucci of Bellport, N.Y., that holds high the banner of American display fireworks. Here they are, with their long racks of wired mortars and the factory too, in photograph and text. The brief history of fireworks opens with some account of the deep past, back to the invention of gunpowder, and proceeds to a very simple account of the kinds of fireworks and how they work. The photographs are more detailed than the text, which settles for a few sentences on fuel and oxidant; the assemblies are better treated than the chemistry.

Three small Iowa towns that burned down from Fourth of July displays gave the impetus to the fireworks control laws; their tale is here, along with other disasters. Handel's *Music for the Royal Fireworks* (in 1749, the display set by the famous Ruggieris) was played to a paroxysm of unscheduled explosions by the light of a burning theater. Fireworks even in these safe and sane States still inflict some 11,000 injuries a year, mostly from oversize illegal fireworks in imprudent hands. Readers in the middle grades will learn a good deal both about the ironies of technology and about the rudiments of how the striking displays work. This book is not at all for budding pyrochemists, however careful; the how-to-do-it chapter is a neat one about taking photographs of that spectacle high in the sky, with a comment on distance measurement by the delay of the sound.

THE MAKING OF RETURN OF THE JEDI, edited by John Phillip Peecher. Ballantine Books ($3.50). Since the spring of this year almost 100 million people have seen the remarkable images made by imaginative engineering that convey the contemporary fairy tale of George Lucas. Out of the Monster Shop in San Rafael in northern California flow the creature costumes of sculptured and fitted plastics, the Pig Guards and the Squidheads, off to the EMI Elstree Studios. Industrial Light and Magic prepares its ingenious models and its paintings on glass. On the largest sound stage in Europe principal photography began in the first weeks of 1982, after elaborate set construction.

This well-informed chronological account of the 30 months from plan to preview centers on the sense of the people involved, with interviews with Lucas himself, with all the stars now so well known and a variety of specialists: the wizards of so many effects, stunts and sounds strange though real. Most of the story tells of telexes, urgent or bantering, of caterers, carpenters and little people recruited by the score to don the furry costumes of the Ewoks of the forest. One learns almost in passing about the "bluescreen" superposition scheme that created the speeder bikes that whiz through the forest, or about the radio-controlled pupils and eyelids of the grotesque tyrant Jabba the Hutt. Jabba was a jellylike slavering puppet large enough to contain three operators, one in each arm and a little one in the tail, each equipped with a video monitor to check on the effect outside. Two more operators worked those eyes from a distance.

Fantasy on such a large scale is exacting work; it is extraordinary to read of the care with which a videotaped study was made, complete with a miniature forest and cutout actors on hand-moved bikes, only as a guide to how the redwoods and the speeders would one day look in full scale but superposed. The actors had to mount stationary speeders in front of the bland bluescreen, of course seeing no trees at all, while they tried to imagine their feelings on a dizzying ride among the great tree trunks.

Pictured fantasy has power. Mark Hamill, who plays the last knight, holds the view that young people regard the film as real: "They don't want to hear about bluescreen techniques. To them, those are actual rocket bikes in an actual forest." This book is an easy-to-read account of the creation of a public myth, absorbing to any screen-struck young reader who wants nonetheless to part image from reality in a galaxy not at all far away.

Symbols and Stories

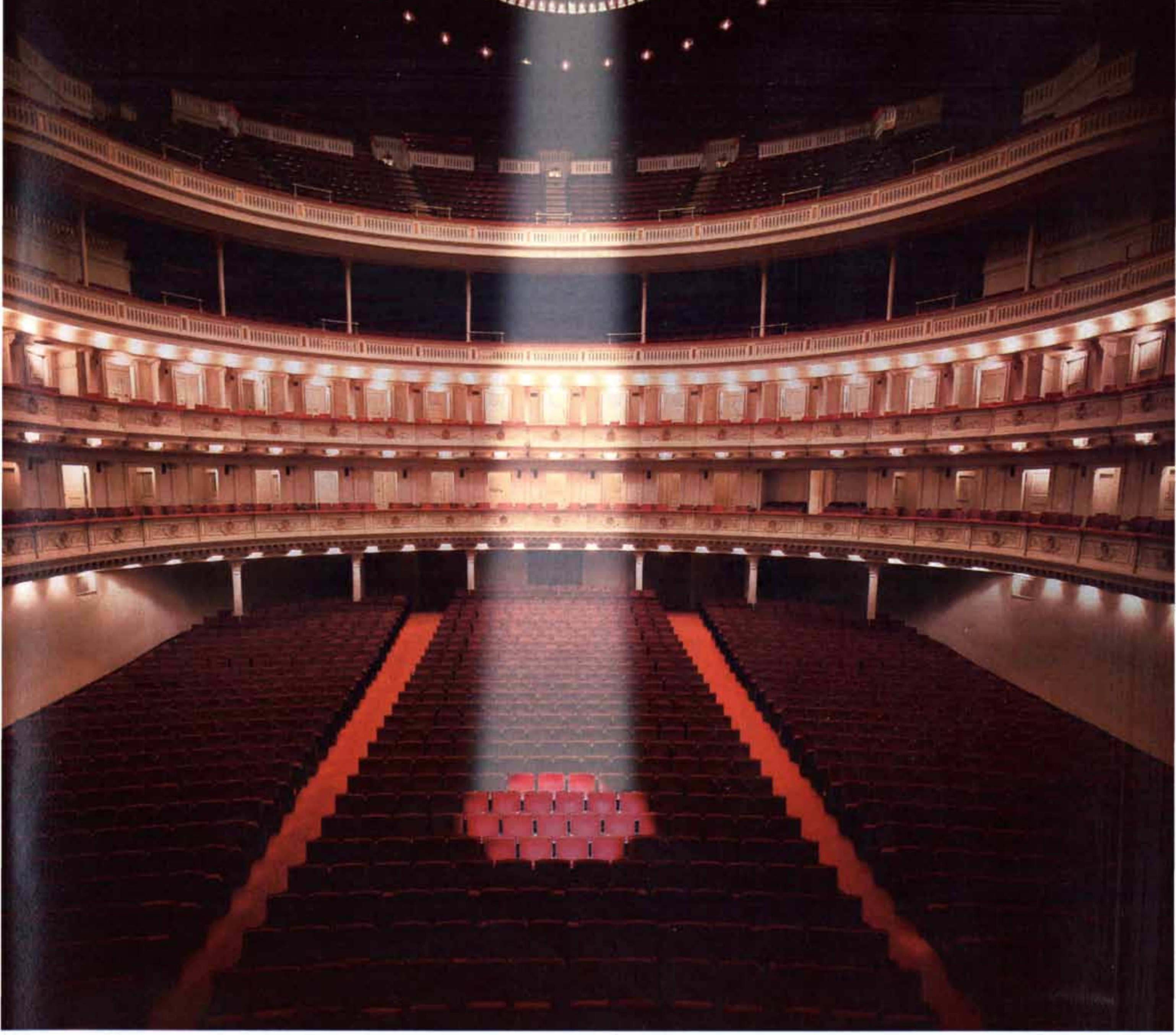
ANNO'S COUNTING HOUSE, by Mitsumasa Anno. Philomel Books, New York ($12.95). ANNO'S MYSTERIOUS MULTIPLYING JAR, by Masaichiro and Mitsumasa Anno. Philomel Books, New York ($10.95). These two books, one without words, one with a few lines of text, glow alike with the fine-grained paintings of the gifted Tokyo artist and traveler Mitsumasa Anno. In the second book his magic found assistance from his son.

There are really two counting houses, one timbered, one stone. They are drawn on the facing pages of the slender book, the facades removed, so that we see the rooms and their contents, including five boys and five girls, each with a number of attributes in common. The facades of the two houses are painted on the two sides of the leaf between every facing pair of cutaway houses. As you open the book anywhere, you see the colorful front of one house and the open interior of the other. The open section has no color, unless in one or another room there stands one of the colorfully dressed boys and girls. The 10 wander from house to house as the pages turn, carrying with them a picture of the Queen, a guitar or something else they love. Die-cut window openings through the facades expose more or less at random a view into one or another room on the page below. On this theme of 10 conserved children and their whimsical possessions there is much to be played out, up to the whole of sums and differences among 10. Anno makes it plain that his ingenious book is meant to lead children to find for themselves the beauty and pleasure in the world of numbers. (That royal portrait ends up in a closet.)

The multiplying jar is an actual setting of a task that is almost set by the nursery-rhyme traveler to St. Ives. Here a magical jar—very blue-flowered—holds a sea with 1 island. On the island are 2 countries, each with 3 mountains, every mountain with 4 walled kingdoms, and so on, until at last we find 9 boxes of 10 jars each. How many jars were there within? The answer comes by way of visual multiplication, dots drawn in arrays, up to two entire pages of dots for the 8! cupboards. There is not enough room for all the dots of the 9! boxes and the 10! jars, but the numbers are worked out and the idea of the factorial exclamation point is explained. This book repays early-grade skill at reading and at figures. Both volumes are arithmetic gems; no better introduction to the deep art of counting may ever have been published than those two houses shared among 10 cheerfully roaming children.

ORDERLY TANGLES: CLOVERLEAFS, GORDIAN KNOTS, AND REGULAR POLYLINKS, by Alan Holden. With photographs by Doug Kendall. Columbia University Press ($19.95). Three-space supports a marvelous variety of interlinked structures. The construction and properties of some of them are here exemplified and analyzed with elegance by a well-known writer on the pleasures of polyhedrons. The first half of this book of fresh recreational geometry, entirely without formulas, the eye always guided by crisply pictured models, treats of useful tangles. The first class of tangles is the cloverleaf highway intersection, realized so often and on such a large scale. Simple rules of traffic flow induce those looped connections; a practical degree of redundancy makes the motorist less likely to get dizzy as he shifts direction through many loops. The 12 one-way connections to link four lanes to each of the other three are in practice cut to eight.

Even more useful is the grand tangle we call woven cloth. Twill and tabby are here, constructed, like most of these structures, as well-made models formed of wood dowels instead of rope or yarn. We see also neat models of knots such as the square knot and the granny, not to mention the Turk's head and other baroque Gordian knots, those that cannot be untied without cutting, and we read something of their theory. A striking chapter treats cat's cradle, that ancient play of string-looped hands, as an example of unknotted rings; models, again in



Sony creates seventh row, center. Forever.

INTRODUCING THE SONY COMPACT DISC PLAYER.

You are looking at an invention so extraordinary that High Fidelity magazine has hailed it as "the most fundamental change in audio technology in more than eighty years."

A piece of audio equipment that permits you to hear something you've never heard before: perfection.

Sony is pleased to raise the curtain on the world's first digital audio compact disc player. The CDP-101.

THE END OF HI-INFIDELITY.

The CDP-101 uses a laser beam to play compact discs that hold far more musical information than conventional records ever could. The result: fidelity and depth of sound that no conventional turntable could ever deliver.

This highest of fidelity remains faithful too. Because the digital discs are read by laser beam, there is absolutely no physical wear. Nor is there any distortion from such annoyances as dust, fingerprints or even scratches. The laser simply sees through them.

Equally ingenious, an infrared remote control lets you select tracks without budging from your armchair.

While an ever-expanding library of compact discs lets you listen to your favorite artists as though you, and your armchair, were centered in the spotlight above.

Maybe most important, the CDP-101 is fully compatible with the conventional equipment you may own.

We suggest you hear the Sony Compact Disc Player soon. For a sound you can't believe, from the audio innovator you assuredly can.

SONY®
THE ONE AND ONLY

wood, reveal the symmetry of the seven stations of the game with unexpected clarity. Finally, polylinks, the author's form of elegant chain, are defined, explored and illustrated in the rest of this highly original book.

MATHEMATICAL SNAPSHOTS, by H. Steinhaus. With a new preface by Morris Kline. Oxford University Press ($7.95). This member of the small set of mathematics brought to the amateur is offered in paperback for the first time, in the version first published in 1969. The same book, then a little smaller in its first appearance in English, was reviewed with enthusiasm by *Scientific American* in 1950.

A dozen swift, concise chapters, with a wealth of simple, clear diagrams or photographs, span a wide class of mathematical results, from Pythagoras to probability. The results are often surprising, occasionally profound and never dull. The key word is snapshot: in a few sentences of straightforward if tight argument and a clever sketch or two Professor Steinhaus makes his points, time and time again leading the reader to direct conviction and delight. Formal proof seldom occurs, no lines of algebra string across the pages, no long series of constructions appears. Open to readers equipped with high school algebra and geometry, the book is at once rich and disarmingly unpretentious. A fair share of the material will have meaning for mathematically inexperienced but interested readers.

To sample the hundreds of problems: how to share a cake with total fairness among any number of people; how to measure the great diagonal of a shoebox with a ruler too long to fit inside; how to trisect an angle to one part in a few hundred; how to list the Platonic solids without relying on the geometry of a plane (by dividing a sphere topologically); how to circumscribe Lake Michigan, or any other closed curve, within a square. This small classic of recreational mathematics is a find for anyone pleased with that witty domain who has not yet encountered these snapshots of what mathematicians really do and genuinely enjoy.

YEH-SHEN: A CINDERELLA STORY FROM CHINA, retold by Ai-Ling Louie. Illustrated by Ed Young. Philomel Books, New York ($10.95). Yeh-Shen, poor girl, did the heaviest chores for her jealous stepmother, there in the home of her late father, a chief of the cave people of southern China. The plot is well known: there is a magic benefactor, here no fairy godmother but a wonderful golden-eyed carp. Indeed, the poor carp was stabbed to death by the wicked stepmother, but its bones remain able to counsel the bright and lovely Yeh-Shen. Sure enough, there is a festival banquet. Yeh-Shen appears at it unexpectedly radiant in a gown of azure with a cloak of kingfisher feathers and a tiny pair of magic golden slippers. The tale unfolds without surprises or pumpkin coaches, the slipper lost as the unknown girl runs off in rags, the slipper reclaimed to fit the right tiny foot as the king finds his true love.

The luminous paintings glow across pages arranged to evoke an old Chinese book, although Ed Young's style in pastels and watercolor is richer than it is spare. The text is based on a Ch'ing edition of an encyclopedic work that retains the T'ang story by Tuan Ch'eng-Shih, more than 1,000 years old; two printed pages of the Chinese text are reproduced. Cinderella is found first written down in Europe in an Italian tale of the 17th century. The sense of the storyteller's unity across the world is strong here, in a very attractive book delightful on its own terms, even if you had never heard of Cinderella.

ZOUNDS! THE KIDS' GUIDE TO SOUND MAKING, by Frederick R. Newman. Illustrated by Elwood H. Smith. With a six-inch floppy stereo disc. Ran-

dom House, Inc. ($4.95). Once adept at the palate grind, the familiar pucker whistle, inhaled speech and the cheek thump—all of which are carefully described here as easy elements of spoken sound beyond everyday speech—any kid is on the way to evoking the night noises of the woods and the home, the fearsome roars and hisses of the jungle, the animal and mechanical noises of the farmyard, an entire band of musical instruments and the strange sounds imputed to space in the best-loved media. For each of these families of sound there is offered a small story, weaving the sounds into a more or less coherent plot. Just as in ordinary speech, the combinatoric power of a limited repertory (not, to be sure, just the basic four mentioned above) is the basis of elaboration. For the sound of a gently landing flying saucer try humming as you produce a pucker whistle, and let both glide down in pitch.

The book is written in a chatty, direct way; the illustrations are bold cartoons of a somewhat outlandish kind, as befits the buhs, ta-tas, ooo-wees and chiks that punctuate the text. The language is simple, although the oral tasks set are not easy for everyone to follow. They are perhaps easier for the aspiring young sound maker than for those nearby! This is an original and amusing guide to a diverting imitative art. The author is a lifelong artist in this domain; the accompanying recording bears cacophonous witness.

WHAT'S WHAT: A VISUAL GLOSSARY OF THE PHYSICAL WORLD, edited by Reginald Bragonier, Jr., and David Fisher. Ballantine Books ($9.95). THE OXFORD-DUDEN PICTORIAL GERMAN-ENGLISH DICTIONARY, edited by the Dudenredaktion and the German Section of the Oxford University Press Dictionary Department. Oxford University Press ($29.50). The scholars of Laputa, to preserve breath and minimize ambiguity, never spoke but carried with them their nouns, that is, an example of everything they might choose to converse about. Next come the big unabridged dictionaries of our youth, with thumbnail drawing after drawing to illustrate at least on paper the strange objects they define. Who has not looked up toughies, such as astragals and topgallants, in this kind of drawing?

These two reference books, one American and one Anglo-German, now aim to fill exactly such a role. The Ballantine book is a version of the Laputan backpacks, of course much more sensibly done through images on paper. Its hundreds of photographs show ice skates and the space shuttle, the tepee and playing cards, and the parts are named. Most of the pages are photographic images, so that the old dictionary thumbnail sketches are not visually invoked. One has the feeling that odd words are sought out and that the blander terms used in context to designate all the pieces of a toaster, say, do not add much knowledge.

The Oxford-Duden book consists almost entirely of crowded ink drawings, quite like those of the big old dictionaries. The pages are rich; in general each page shows a complex of items rather than a collection of single ones. One sees a hotel lobby or a cement factory, and keyed to numbers on each drawing are 50 names, both in English and in German, for key rack and bar man, for hammer mill and clinker cooler. The book was motivated by the belief that the visual invariants would help the task of technical translation; indeed, there is a version in English and Japanese.

The Oxford-Duden indexes about 28,000 terms, some 50 percent more than the Ballantine. The Ballantine, however, looks altogether more familiar to the Yankee eye, and it has a lighter touch; the Oxford-Duden never smiles. Either form of this remarkable mapping of the world is something kids would pore over and get much out of, although not always what was intended.


80 PROOF CALVERT DISTILLERS CO., N.Y.C.
JOHN JAMESON & SON
ESTD 1780
SINE METU
IRISH WHISKEY
JOHN JAMESON
ESTD 1780
SINE METU
IRISH WHISKEY
A BLEND DISTILLED IN IRELAND BY
John Jameson
LIMITED
BOW STREET DISTILLERY
DUBLIN
IMPORTED

Established 1845 SCIENTIFIC AMERICAN December 1983 Volume 249 Number 6

Modern Icebreakers

They have evolved rapidly because of economic pressures to maintain commerce in icebound waters. They feature unusual hull designs and propulsion systems, and most of them are currently built in Finland

by John D. Harbron

A substantial fraction of the world's seaborne commerce, particularly in the waters of North America and northern Europe, moves in the high latitudes where ice is an impediment for several months of the year. The nations engaged in such commerce deal with this impediment with a fleet of about 100 icebreakers, which require for their difficult task not only highly specialized construction but also unusual techniques of seamanship. It seems likely that the exploitation of resources such as oil in several regions in the high latitudes will intensify the demands on the icebreakers and their crews in the coming decades.

The distinctive features of an icebreaker are its shape, its strong bow, its machinery and its propellers. In shape the icebreaker is characteristically short in length and wide in beam, displaying therefore a stubby profile. The shape is accentuated in modern icebreakers built in Finland, whose shipyards have turned out some 60 percent of the world's fleets, by the piling up of the bridge and the crew's quarters above the deck. This design creates space in the hull for the sturdy main engines, the auxiliary engines and other machinery and the propeller shafts.

A look at an icebreaker in profile when the ship is out of the water would reveal another feature of the shape, namely that the draft is quite shallow at the bow, sloping to a greater depth amidships. This shape is made necessary by the fact that an icebreaker does much of its work not by driving head on into the ice but by riding up on it, so that the increasing weight soon causes the ice under the bow to break. The ship's unusual width contributes to the success of the stratagem by making the weight compact. The design also reduces the strains put on the hull by the fact that the ship is in effect repeatedly running aground.

Understandably a vessel doing this kind of work needs a bow that is thick, made typically of steel plate that is about five centimeters (two inches) thick and strongly braced with steel structural members. The hull is also reinforced with thick plate all along the waterline because that area will be often in contact with ice or struck by floes that have been broken loose by the action of the bow. At one time many icebreakers had a weakness in that the welds between the steel plates tended to corrode and fail. Workers at the foundry in the Rautaruukki primary steel mill in Finland developed a microalloyed steel in which the content of manganese, sulfur and silicon was limited compared with conventional steels. The welds in this plate have proved to be resistant to the low-temperature seawater in which icebreakers usually operate.

The propulsion machinery of an icebreaker faces severe demands arising from the extreme variations of load on the engines between the times when the ship is breaking ice and the times when it is backing off for another run or is operating at a forward speed in open water. To meet the heavier demands the engines are designed to reach full speed rapidly. In contrast, the engines of a conventional vessel work up slowly to full speed and could be damaged by having to develop maximum power rapidly. The icebreaker's engines must be robust, reliable and able to deliver full power over a range of revolutions. They also must be economical in fuel consumption, since the ship is often operating for extended periods in areas that are remote from fuel supplies.

In the days of steam the icebreaker had reciprocating steam engines. Now the diesel-electric system predominates. In these vessels diesel engines power electric generators, and the electricity runs the motors that turn the propellers. A few icebreakers with nuclear power have been built by the U.S.S.R., beginning in 1957 with the *Lenin*. The nuclear reactor in such a vessel is a source of heat for generating steam that runs turbines. Nuclear power is advantageous when an icebreaker is likely to do much of its work in areas remote from fueling stations, such as the high Arctic regions of the U.S., Canada and the U.S.S.R., but otherwise it cannot compete economically with diesel-electric systems.

An icebreaker may have as many as four propellers. If there are four, it is likely they will be "wing propellers," meaning that they are mounted on the sides of the ship, two near the stern and two near the bow. Bow propellers work well in first-year ice, that is, ice that has built up in a single winter and is not more than about 30 centimeters (12 inches) thick. (In some places, such as the Canadian Arctic, first-year ice is much more formidable, attaining a thickness of as much as two meters.) The bow propellers draw water from under the ice, weakening the support of the ice sheet and causing pieces of it to break off. Then the propellers wash water and ice away from the bow of the ship, facilitating forward progress. Reversed, they help the ship to move astern, thereby increasing its maneuverability. Bow propellers have proved to be useless in Arctic and Antarctic waters. The Arctic ice is too hard and the Antarctic ice is too thick to be affected by their action.

There is no single answer to the number of stern propellers an icebreaker

FINNISH ICEBREAKER ***Urho*** **opens a new channel through ice in the Gulf of Bothnia, followed by two cargo ships. The** ***Urho*** **was the first ship of a class of icebreakers, all of which have been built by the Wärtsilä shipyard in Helsinki. The ships are fairly large as icebreakers go: 104.6 meters (343 feet) long and 23.8 meters wide. Each of the ships has two propellers at the bow and two at the stern, with an output of 22,000 shaft horsepower. A diesel-electric system furnishes the propulsive power. Finland keeps the gulf open all winter.**

should have. The amount of propulsion they must deliver depends on the severity of the ice conditions where the ship is likely to operate. Several types of icebreaker with twin shafts, including the U.S.S.R.'s *Moskva* class in the 1960's and the Canadian Coast Guard's *R* class in the 1970's, have had a good deal of trouble with bent and broken propeller blades and fractured propeller shafts.

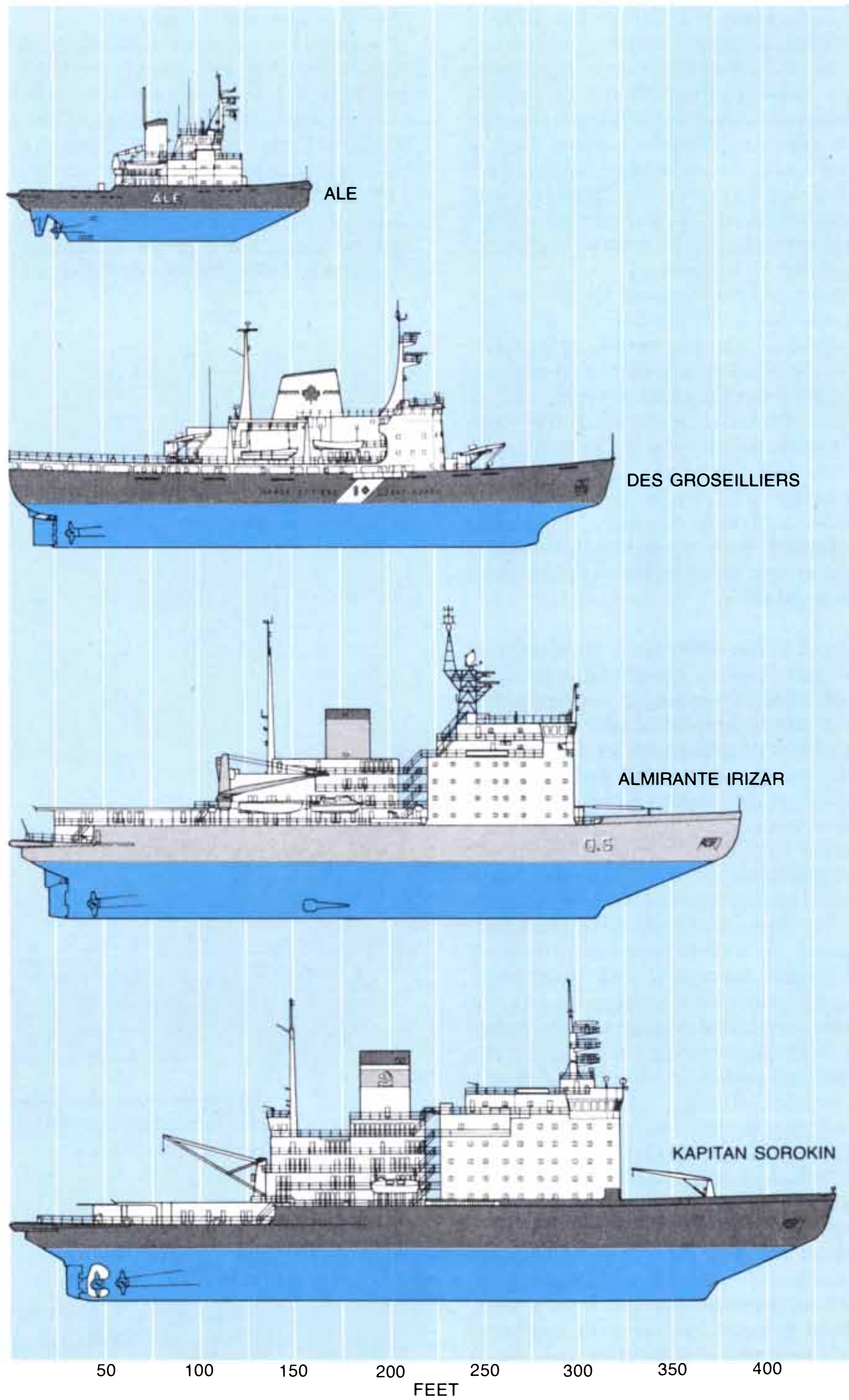
One answer to this problem has been to build icebreakers with three propellers, one in the conventional position at the middle of the stern and two on the wings. This configuration too has had trouble in the two icebreakers of the U.S. Coast Guard's *Polar* class. One reason apparently is that because of the dual role of these ships (to break ice in the Arctic and to serve as supply and research vessels in the Antarctic) they have too much shaft horsepower for the size of their hull. The result is a chronic vibration of the hull. Each vessel has 60,000 shaft horsepower (more than any non-nuclear icebreaker), 42,000 from the gas-turbine engine driving the main shaft for breaking ice and 18,000 from diesel-electric engines driving the two wing propellers for running in open water, as on the long trip to the Antarctic.

Since an icebreaker is highly specialized in construction and faces demanding sea and ice conditions, operating one calls for special types of seamanship. Examples are the two modes of operation in ice, known formally as the continuous mode and the ramming mode. In the continuous mode the ship moves steadily ahead at a speed of from three to five knots through ice that is as much as 1.5 meters (five feet) thick. The technique works if the ice is fairly level and offers only average resistance to the ship. Such conditions are normally encountered in several areas where icebreakers operate regularly, including the Gulf of Bothnia, the Gulf of Finland, parts of the Russian Arctic and the Canadian Arctic, the Great Lakes and the St. Lawrence Seaway.

In the ramming mode the ship rides up on the ice. When the ice breaks and the vessel is again level in the water, it is backed off several times its length and then driven forward to ride up on the ice sheet again. The technique has to be applied when the ice is notably thick, hard or irregular in surface form.

Yet another challenge to seamanship comes when the ship gets stuck in the ice, a not unusual occurrence even though the vessel is designed to break ice. (The *Polar Sea* of the U.S. Coast Guard was embarrassingly beset in the Chukchi Sea north of Alaska from February 18 to May 13, 1981. During much of that time the Coast Guard made the best of the situation by redesignating the ship as a floating research station under the label "Polar Sea Drift Project '81.") Having propellers at both ends of the ship can be helpful by increasing the vessel's maneuverability so that it is less likely to become icebound or has a better chance of breaking free. Another technique still in use on some older icebreakers is heeling, that is, causing the ship to rock from side to side by pumping water or fuel oil back and forth between heeling tanks on opposite sides of the craft. With trimming tanks fore and aft the ship can similarly be made to rock from bow to stern.

On more modern vessels the heeling



FOUR TYPES OF ICEBREAKER are portrayed at a common scale. The *Ale* class is designed for icebreaking assistance and towing duty in rivers, harbors and lakes in Temperate Zone waters such as the Great Lakes and the Baltic Sea. The *Des Groseilliers* is representative of the Canadian Coast Guard's *R* class, operating in the St. Lawrence Seaway and during the summer in the Arctic. The *Almirante Irizar,* built by Wärtsilä for Argentina, serves in the Antarctic both as an icebreaker and as a research ship. The *Kapitan Sorokin* is a member of a class of large icebreakers operated by the U.S.S.R. The sloping bow, heavily plated for repeated encounters with ice, enables an icebreaker to ride up on the ice, which breaks under the weight.

technique is being replaced by an air-bubbling system developed and patented by Wärtsilä, a Finnish shipbuilding firm that is the world's largest builder of icebreakers. The aim of the system is to reduce friction between the hull and the ice or snow. Compressed air is forced through a series of nozzles at intervals near the bottom of the ship. When the air rises along the hull in the form of bubbles, it creates a strong current of water and air that forms a lubricating layer between the hull and the ice. At low speeds (less than two knots) the system reduces friction to about half of what it would be otherwise. As the speed of the ship increases, the effectiveness of the air-bubbling system diminishes, but even at five knots the friction is reduced by about 15 percent.

Most people think of the ice covering the sea as being a fairly homogeneous sheet varying mainly in its thickness. Actually it is found in formations ranging from the impenetrable shelf ice of the Antarctic (some of it 100 years old) to the rough polar packs of the high Arctic and the winter-only ice of the Baltic Sea passages, the Great Lakes and the St. Lawrence Seaway. The people concerned with operating icebreakers have to take these many varieties of ice into account.

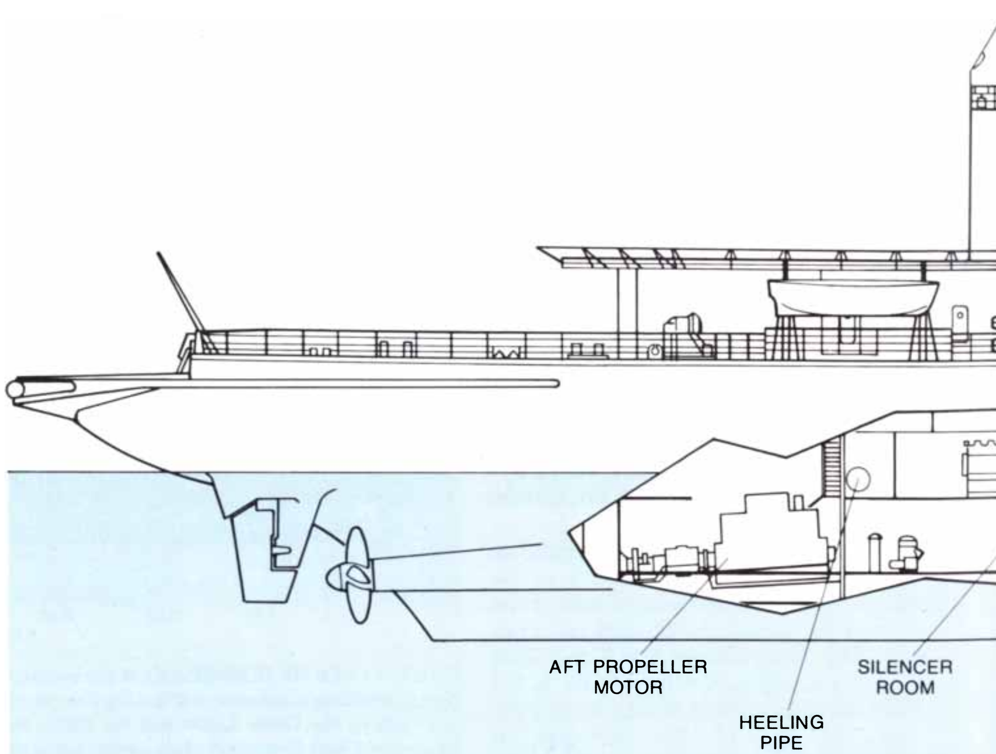
Until about 1966 the U.S., Canadian and Russian government agencies involved with shipping in the Arctic and the Antarctic employed only four rather broad classifications of ice formations: new, young, winter and polar. As the number of icebreakers increased and the means of determining ice conditions expanded to include surveys by aircraft and photographs from satellites, such generalizations proved to be inadequate for the crews that had to function successfully in icebreaking work. Now all the nations concerned with major shipping in areas where ice may be a factor maintain detailed and constantly changing definitions of ice formations. The definitions that follow are from Canadian sources, although the other nations recognize parallel categories.

The definitions begin with "open water," which is described in the *Sea Atlas of Canada* as "water in which the concentration of ice does not exceed one tenth of the total area." Another major formation is first-year ice, meaning ice that has persisted for less than a year. Second-year ice has survived one summer's melting and is building up again in the following winter. Multiyear ice has remained in place through two summers or more.

In addition to these kinds of ice there are other formations that can make an icebreaker's passage either difficult or impossible, depending on the extent of the ice structure. One is rafted ice, in which one sheet of ice has been overridden by another sheet or more than one. Another is hummocked ice, in which ice that has broken is forced upward by pressure and forms hillocks. A ridge is a wall of broken ice that has been formed by the collision of ice masses; its base can be as much as 15 meters below the surface. Finally, snow that has settled on an ice formation can considerably decrease an icebreaker's efficiency.

In general such regions as the Gulf of Bothnia, the Gulf of Finland, the Gulf of St. Lawrence, the Great Lakes and the northwest and northeast coasts of the U.S. have from 30 to 90 centimeters (about one foot to three feet) of ice between the middle of December and the middle of May. Severer conditions are faced by icebreakers in the Canadian Arctic and in the high-latitude waters of the U.S.S.R. and the Scandinavian countries, where efforts are made to maintain shipping for at least part of the ice season. Finland keeps the Gulf of Bothnia open throughout the year because of the need to move goods out of the country's many ports there.

A point that can be made about the Canadian Arctic also applies to other major ice areas. The Canadian Arctic extends some 3,000 miles from Baffin Island to the North Slope of Alaska and about 2,000 miles northward from Canada's Arctic coastline to the North Pole, with major passages and islands in between. The region has an archipelago of diverse ice-covered areas. Each area has its own historic record of ice, wind, sea-condition and other climatologic factors, and any one of them can change



BELOWDECKS MACHINERY of the *Urho* and other icebreakers of its class takes up almost all the space in the hull. It is largely to make room for this equipment that the bridge and the facilities for the crew are all above the deck on many icebreakers, giving them a top-heavy

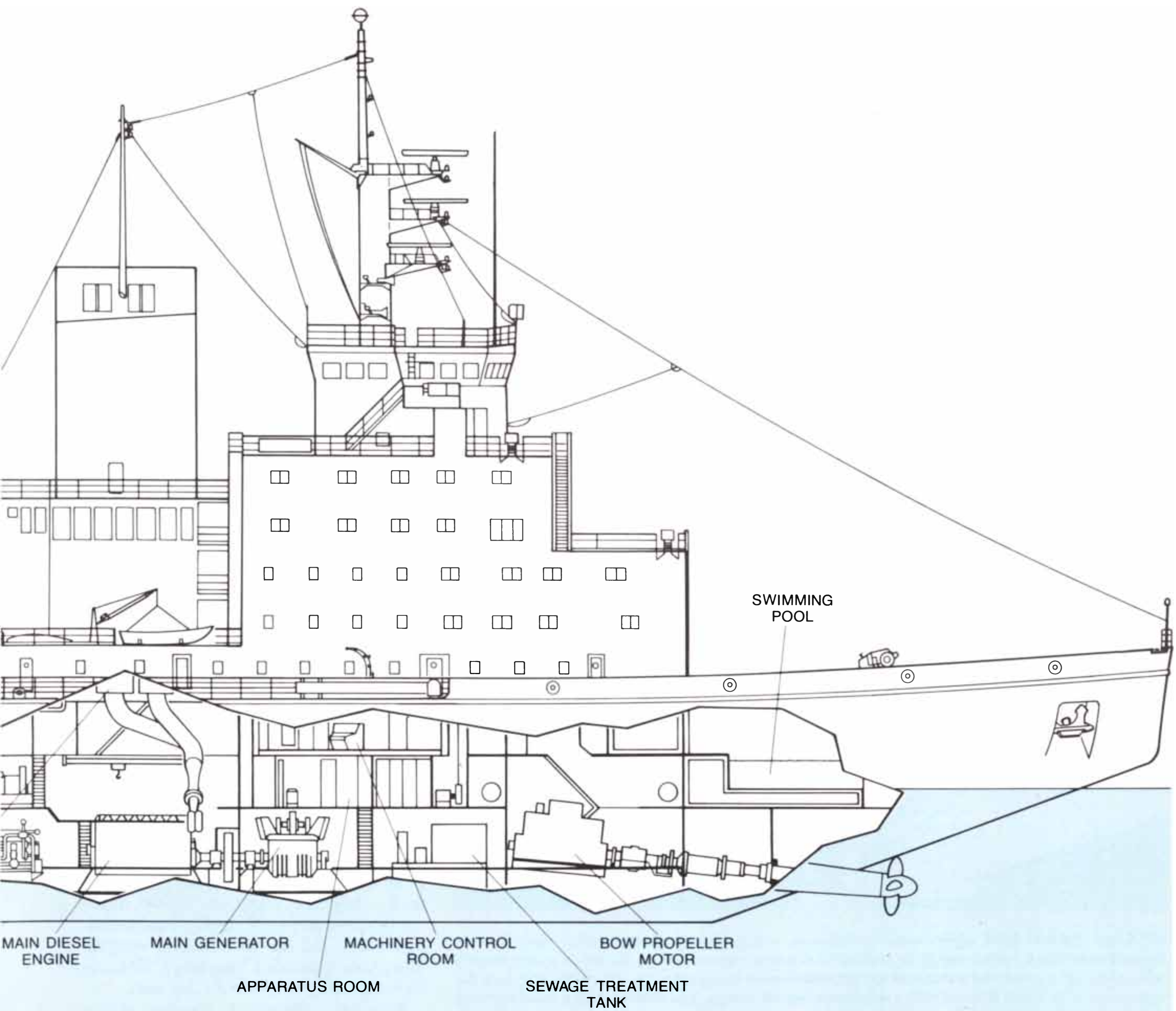
unpredictably in a particular season of icebreaker operations. T. C. Pullen, a retired captain in the Canadian Navy and a specialist on icebreakers, has summarized the situation as follows: "We encounter heavy ice where intelligence has reported ice-free conditions, providing yet another lesson in what Arctic marine operations are like, namely that they are not necessarily what we should assume to be the case."

The variations in the ice give rise to variations in the design of icebreaking vessels. Again the Canadian standards provide a guide. A few years ago the Canadian government published its Canadian Arctic Shipping Pollution Prevention Regulations, which combine a concern for the environmental safety of the Arctic as increasingly larger ships are active in it and an effort to specify the capability of Arctic ships. The regulations define icebreakers and their cousins, ships strengthened for ice, in a range from Arctic Class 1 to Arctic Class 10.

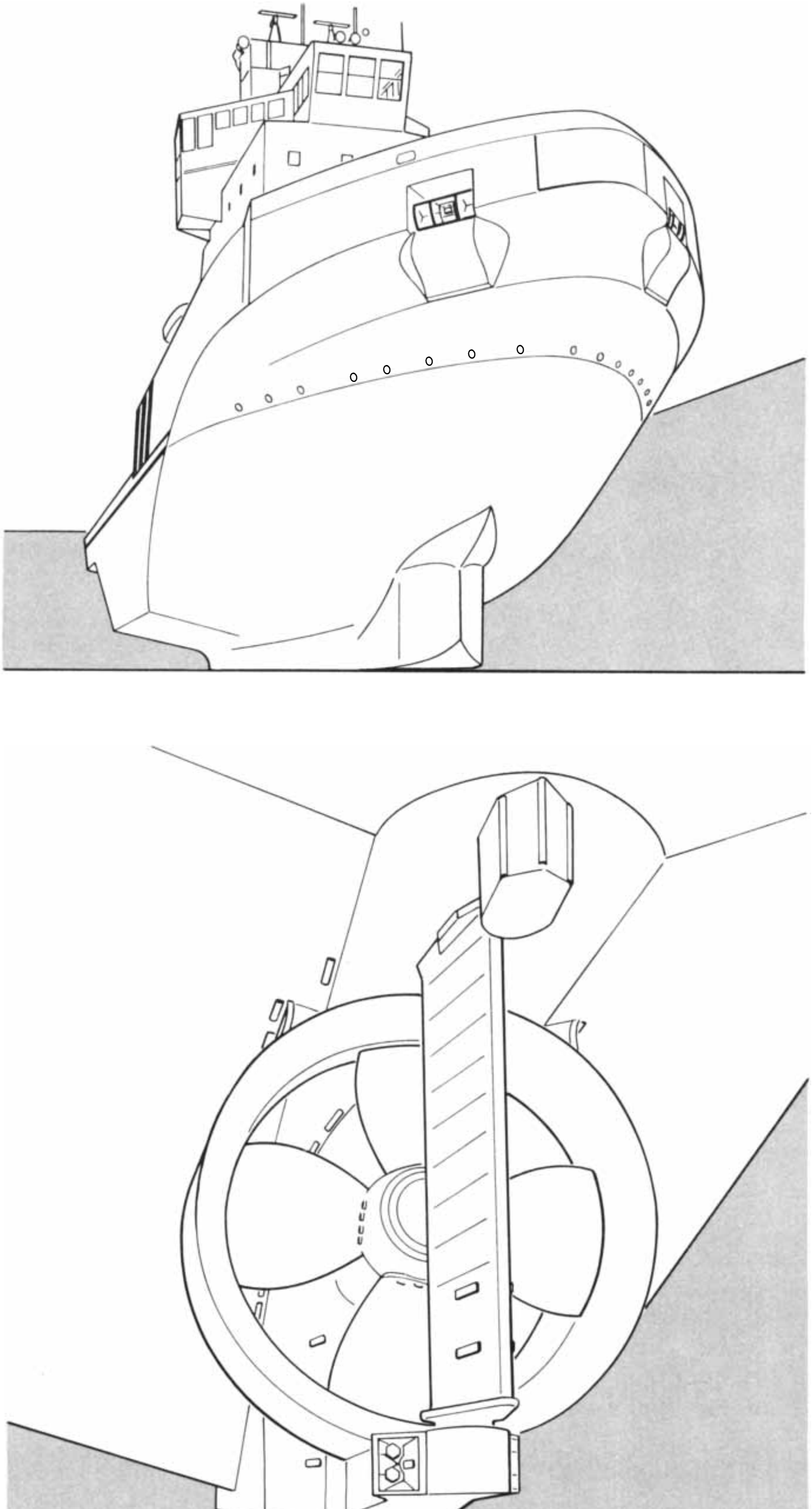
A few examples will indicate the gradient embodied in the regulations. A vessel of Arctic Class 3 can maintain a speed of three knots through ice .9 meter (three feet) thick. A vessel of Class 7 can maintain a speed of three knots in the continuous mode through ice 2.1 meters (seven feet) thick.

The icebreakers of the Baltic type built mainly in Finland and the U.S.S.R. to work in regions with one-year ice can be identified under the regulations as vessels of Class 3 or Class 4. They would not meet the more demanding requirements for a large icebreaker operating throughout the year in the Canadian Arctic. Vessels of this kind are in Arctic Class 8. Canada plans to build its first icebreaker of the class but has not yet begun construction.

The most powerful icebreakers today are the nuclear-powered Russian sister ships *Artika* (recently renamed *Leonid Brezhnev*) and *Sibir,* which can be estimated as in the range from Class 5 to Class 7 under the Canadian regulations. Each is a vessel of 35,000 tons' displacement (equivalent to the weight of the ship and everything that it is carrying) with 75,000 shaft horsepower. Another nuclear icebreaker, the *Rossiya* (40,000 tons and an estimated 100,000 shaft horsepower), is now under construction



appearance. The heeling pipes carry water back and forth between port and starboard tanks when it is necessary to rock the ship from side to side to avoid becoming stuck in ice. The role of the bow propellers is to draw water from under the ice ahead of the ship, weakening the support of the ice sheet and making it easier to break. Such propellers work best in ice that is no more than about a foot thick.



SPECIAL FEATURES of the *Canmar Kigoriak,* **a Canadian-built icebreaker employed by Dome Petroleum Limited mainly to protect oil rigs in the Beaufort Sea, include a spoon-shaped bow, a special reamer and a single propeller surrounded by a protective structure. The bow design is meant to break thin ice with a minimum loss of energy. The reamer digs a short distance into the ice, providing a fulcrum that makes turning easier. The purpose of the propeller arrangement is to avoid the milling of broken ice that floats toward the stern; the milling action reduces the efficiency of the propeller. The** *Kigoriak* **also has an unusually large rudder.**

for the U.S.S.R. at the Wärtsilä shipyard in Helsinki.

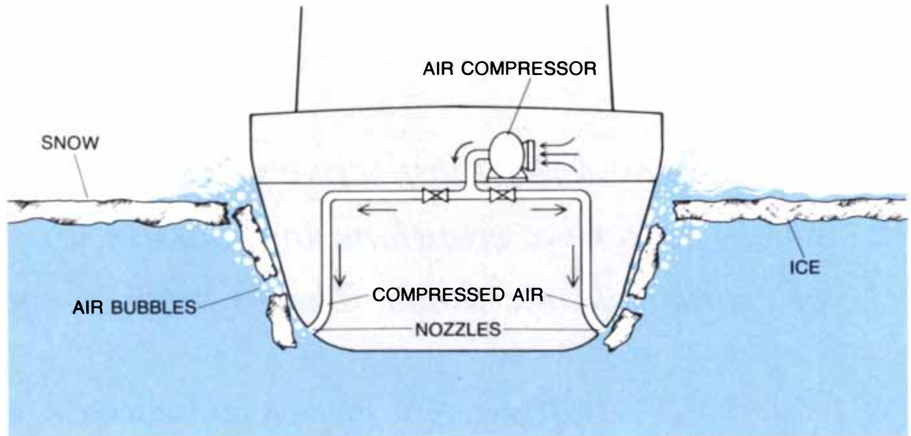
The *Rossiya* will emerge from the shipyard without engines. It will then be towed to Leningrad for the installation of the nuclear power plant. The Russians have not revealed the particulars of these marine nuclear power plants and do not allow foreigners to visit their nuclear icebreakers.

The Antarctic icebreakers of the U.S., Argentina, Japan and West Germany are multipurpose vessels compared with the ones that operate in the Northern Hemisphere. The reason is that they must not only deal with ice but also travel long distances in open water to get to the Antarctic, where they function additionally as observation, research and supply ships. Accordingly they are fitted out with extra cabin space for the scientific workers, landing pads for short-range helicopters that make reconnaissance flights over the ice, heavy deck cranes, cargo elevators and special holds for the supplies they carry to their nation's bases in the Antarctic.

An example of the type is the Argentine *Almirante Irizar,* built by Wärtsilä and commissioned in 1979. One of its special features is a helicopter "touch down" system made in Canada. It is designed to cope with the fact that a shallow-draft, stubby icebreaker can roll as much as 40 degrees in open water when the seas are heavy. A cable is lowered from the helicopter to a locking mechanism on the deck, which slowly draws the aircraft down to the level of the deck. By means of a track set in the deck the helicopter is then towed to its hangar. The vessel also has two 16-ton hydraulic cranes for transferring cargo to the shore, an electric towing winch fitted with tension controls (keeping the cable taut at all times when it is loaded) and capable of exerting a maximum pull of 60 tons, and an air-bubbling system.

The U.S.S.R. is the major operator of icebreakers and the only nation with nuclear-powered icebreakers in regular service. It is also the only country operating significant numbers of merchant ships with a limited capability for icebreaking. (Many countries operate merchant vessels with hulls strengthened for ice.) Among the Russian vessels are a number of multipurpose container ships with icebreaking capability (at about the level of Arctic Class 2) that can proceed on their own in moderate ice conditions if an icebreaker is not available. Typically they are ships of 18,000 deadweight tons (the total carried load), shallow draft and an air-bubbling system. They can penetrate the large Siberian rivers flowing into the Arctic seas.

Russian icebreakers operate across the immense length of the northern sea route and also, during the ice season, in



AIR-BUBBLING SYSTEM was designed by Wärtsilä as an alternative to the heeling technique for preventing an icebreaker from getting stuck in the ice. Compressed air forced through nozzles along the bottom of the ship rises along the hull in the form of bubbles. The result is a strong current of water and air that creates a lubricating layer between the hull and the ice.

such peripheral bodies of water as the Gulf of Finland, the Baltic Sea, the Black Sea, the Caspian Sea and the Yenisei, Ob and Lena rivers. The most impressive icebreaker-convoy operations are in the Russian western Arctic, where fleets of commercial vessels are moving in increasing numbers from Murmansk through the Barents Sea and the Kara Sea to the Yenisei River estuary, where they continue upstream to the interior port of Dudinka that serves the timber- and nickel-exporting Siberian city of Norilsk.

These integrated fleets function in the following way. Nuclear and diesel-electric icebreakers of the *Brezhnev, Kapitan Sorokin* and *Ermak* classes escort the container ships from Murmansk to the Yenisei estuary. The convoy may also include some European cargo ships, mainly British and Scandinavian vessels cleared under Russian ship classifications for navigation behind an icebreaker. At the Yenisei the nuclear icebreakers and the bigger conventional ones stand off the mouth of the river as the shallow-draft icebreakers of the *Kapitan Sorokin* class continue on into the estuary to the point inland where the convoy is taken over by river icebreakers of the *Kapitan Chechkin* class. They are vessels of 4,490 shaft horsepower and extremely shallow draft (3.2 meters, or 10.5 feet). If the ice conditions are too severe for the river icebreakers and the convoy to proceed to Dudinka, the cargoes are unloaded onto "polar utility vehicles," that is, hovercraft that skim over the ice on a cushion of air. The return passage down the Yenisei and the escort back to Murmansk by the heavier icebreakers complete the journey.

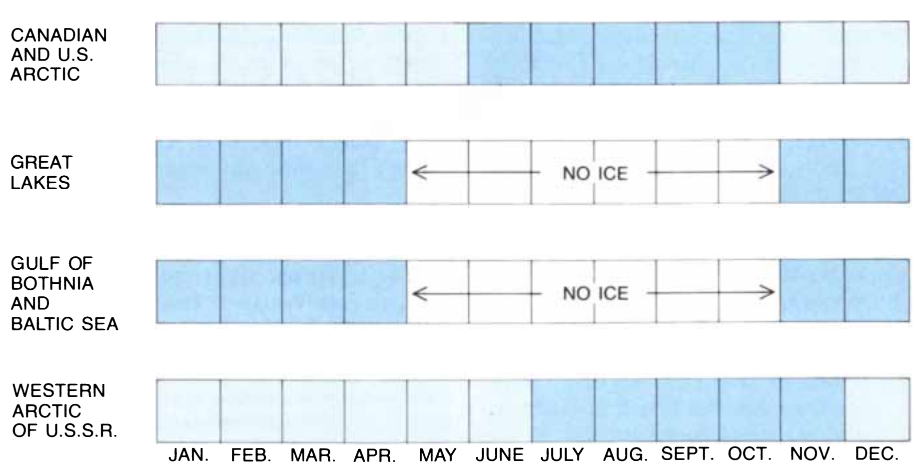
It may be that the characteristics of the icebreaker of the future are being developed by a Canadian vessel recently put in service, the *Canmar Kigoriak* of Dome Petroleum Limited. (The first part of its name is an abbreviation for Canadian Marine Drilling Limited, Dome's shipping subsidiary; the second part is the Inuit name for the northern lights.) This Canadian-built vessel serves mainly in the Beaufort Sea to protect Dome's offshore drilling rigs against damage by ice and to assist the firm's high-Arctic fleet, which includes ice-strengthened drilling ships, tugs and self-propelled barges.

Among the distinctive features the *Canmar Kigoriak* has for coping with the severe ice conditions of Canada's western Arctic is a single propeller, which contrasts with the twin or triple screws of all other Canadian icebreakers. Its designers see this arrangement as a means of avoiding what they call "one of the main perils of twin- and triple-screw icebreakers, the constant milling of ice with the two outboard screws, which reduces propeller efficiency." The propeller is surrounded by a large protective structure that deflects large pieces of ice moving aft and harmlessly mills small ones. The ship also has bow and stern thrusters and an outsize rudder, which increases its maneuverability.

Another feature of the ship is a hull form that distinguishes it from all other icebreakers. Its spoon-shaped bow is meant to break thin ice with a minimum loss of energy. A special reamer is fitted where the bow meets the midsection; it is designed to dig a short distance into the ice, thereby providing a fulcrum that improves the ship's ability to turn. The vessel's capabilities are indicated by the owner's statement that so far it has penetrated first-year ice as thick as 1.73 meters, multiyear ice as thick as 10 meters and first-year ridges grounded in 13 meters of water.

The *Canmar Kigoriak* is also a prototype for a giant ship that Dome may build in the 1990's, an icebreaker-tanker of 200,000 deadweight tons. The plan is based on Dome's estimate that enough offshore oil will be found in the Beaufort Sea to justify marine transporters of this size to move it out. Only ships of such size and enormous power will be equal to the task of operating year-round in the Northwest Passage.

If Canada's Arctic Class 8 icebreaker is built, it will be the largest non-nuclear icebreaker in the world. The *Polar 8,* which is its drawing-board name, would be 194 meters long, 32.2 meters wide and 13 meters deep. At 37,000 displacement tons and 100,000 shaft horsepower, the vessel would represent a substantial change in the standard formula for icebreaker design that relates displacement tonnage to shaft horsepower. In non-nuclear icebreakers a characteristic ratio is 15,000 displacement tons to 20,000 shaft horsepower, as in Wärtsilä's widely employed *Kapitan Sorokin* class now in service in Finland, the U.S.S.R. and Argentina. The *Polar 8* would be the first all-weather icebreaker capable of working in Arctic ice conditions that are too severe for the U.S. and Canadian icebreakers in service today.



SEASON OF OPERATION in the main icebreaking areas is charted. The months when ice is present are indicated in color; the dark color shows the months of icebreaker operation.

Io

The solar system's most active volcanic world is a moon of Jupiter whose geyserlike eruptions and massive lava flows yield evidence of exotic volcanic fluids. Their source of energy is gravitational

by Torrence V. Johnson and Laurence A. Soderblom

In the modern era of planetary exploration Io, a moon of Jupiter discovered by Galileo almost 400 years ago, has emerged as one of the strangest worlds in the solar system. First in the 1960's astronomers discovered that Io somehow modulates the intense bursts of radio waves emitted by Jupiter. In addition sensitive instruments trained on Io from the earth confirmed that the object is as reflective as fresh snow and yet its color is yellow orange. This raised perplexing questions about the composition of Io's surface. Then in the 1970's it was discovered that great clouds of ions and neutral atoms are being injected by Io into the Jovian magnetosphere, the region in which electrically charged particles are trapped by the magnetic field of Jupiter. Finally spacecraft entered the Jovian satellite system. The flybys of *Voyager 1* and *Voyager 2* in 1979 yielded the first close look at Io.

The Io we know today seems even stranger than before. Heated by tidal forces exerted by Jupiter, it appears to be the most volcanically active body in the solar system. It is clearly more active than the earth. Great volcanic geysers on Io send eruptive plumes to heights of hundreds of kilometers, and matter escaping from the tenuous volcanic atmosphere of Io by processes that are not yet well understood does indeed dominate the Jovian magnetosphere, thereby controlling the rate at which Jupiter radiates energy and probably affecting phenomena as diverse as auroras on Jupiter and radio bursts from the planet and its satellites. The discoveries made by the Voyager spacecraft have raised a host of new questions; hence this article cannot be definitive. It is instead a report on the continuing investigation of a remarkable celestial object.

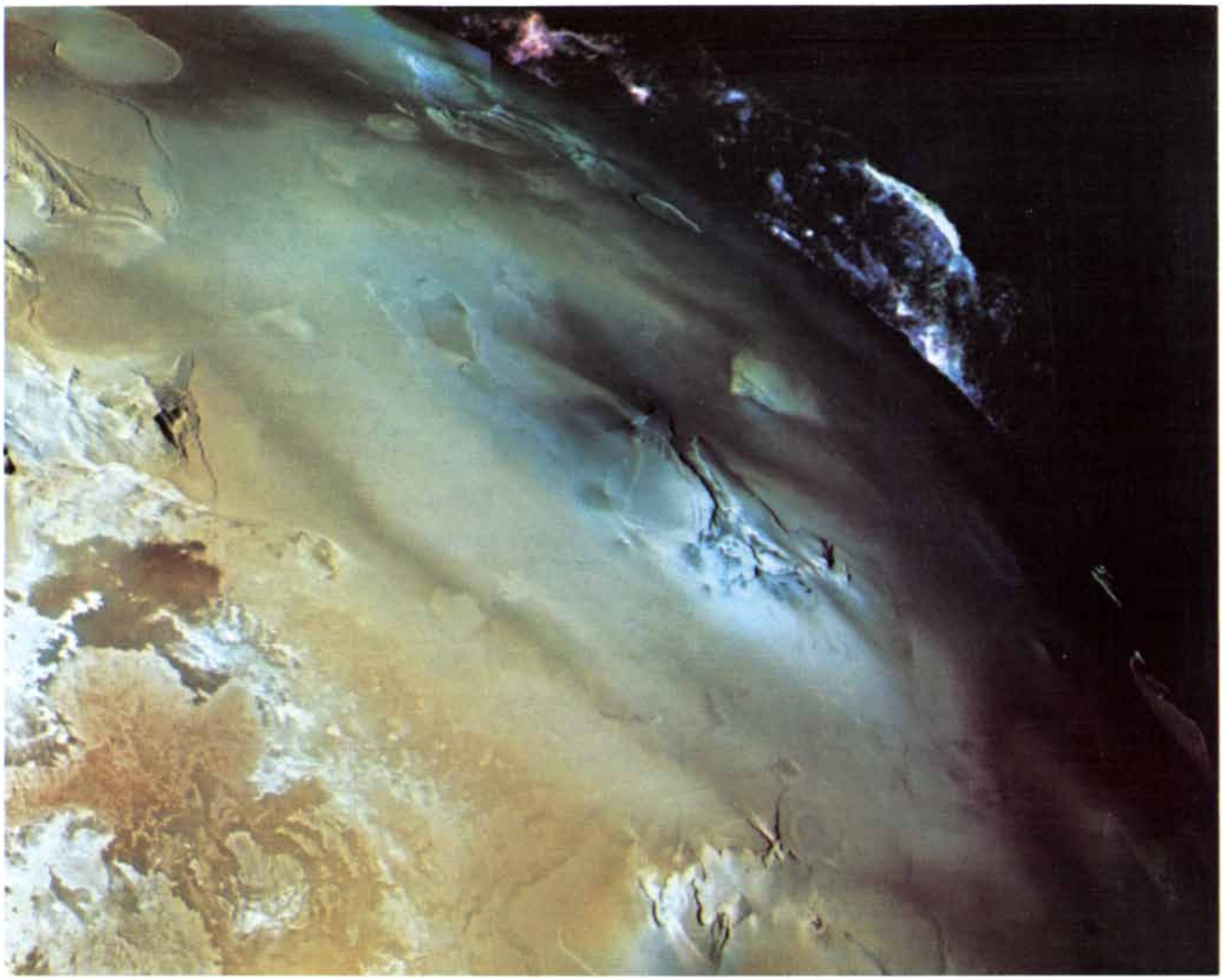
Io is one of Jupiter's "regular" satellites: the ones that travel in essentially circular orbits more or less in the same plane as Jupiter's equator and so are thought to have formed around Jupiter much as the planets formed around the sun. At an orbital distance of 350,000 kilometers (about the same as the distance of the earth's moon from the earth), it is the innermost of the four largest Jovian satellites, which were discovered by Galileo. Io and Europa, the next outward of the four, are about the size of the earth's moon, and both have a rocklike density: respectively 3.5 grams per cubic centimeter and three grams per cubic centimeter. Ganymede and Callisto, the outermost of the four, are larger (they are about the size of Mercury), and they have a significantly lower density, about two grams per cubic centimeter. Apparently they consist half of water ice and half of rock, mostly silicates. The density variation among the four bodies has often been called a smooth trend, but in truth it is more of a bimodal distribution. The inferred mass of the silicates in Ganymede or Callisto is approximately the same as the total mass of Io or Europa; thus the four can be viewed as being similar except that the inner two are missing their share of volatiles, principally water. (Europa is known to be covered by ice, but its bulk density indicates that this covering is a veneer whose thickness is less than one 20th Europa's radius.)

The density variation is usually attributed to the effect Jupiter had on the early conditions in its vicinity. Even today Jupiter radiates at infrared wavelengths about twice as much energy as it receives from the sun; it has clearly been cooling slowly from an early luminous state. Hence models of Jupiter's early history (based on physics much like that of low-mass stars) suggest that the early Jupiter was an important local energy source for several hundred million years. Models developed by James B. Pollack of the Ames Research Center of the National Aeronautics and Space Administration and his colleagues suggest that over a period on the order of 100 million years the temperature and the pressure at Io's current orbital position were dominated by Jupiter, not by the sun. In fact, a body at Io's orbit in the first million years of that period would have received more energy from Jupiter than the earth receives now from the sun.

It is likely, therefore, that the composition of the regular moons of Jupiter was affected by their relative proximity to the planet. In this view Io and Europa are rocky bodies because conditions near Jupiter were too hot to allow those moons as they were forming to retain any substantial quantity of water. In contrast, Ganymede and Callisto, which did retain water, became mixtures of ice and rock and grew larger than Io and Europa. It is notable that the densities of Ganymede and Callisto are about what one would expect if a gas with the composition of the sun were cooled to the temperature at which ice condenses.

This simple view cannot be the entire story. For one thing the bodies in the solar system were intensely cratered by planetesimals in a period called the heavy bombardment, which ended about four billion years ago. Planetesimals falling into Jupiter's gravitational

IO'S SOUTHERN HEMISPHERE was photographed in 1979 by a wide-angle camera on the spacecraft *Voyager 1*. The field of view encompasses about 2.5 million square kilometers, or 6 percent of Io's surface; one centimeter represents about 50 kilometers. Several classes of features are seen. Calderas (volcanic-collapse craters) are the most numerous. Some are surrounded by diffuse surface markings, which probably represent the painting of the surface by sulfurous geysers. Others are surrounded by what are probably eruptive flows. Still others appear to be filled with eruptive material; the lozenge-shaped caldera with red and black patches in it is an example. White splotches dotting the surface are thought to be laid down by mixtures of sulfur dioxide gas and frost erupting along scarps and fractures. A mountain at the lower left is taken to be an outcrop of the silicate crust of Io and is estimated to be nearly 10 kilometers high. The image has been processed by a computer of the U.S. Geological Survey so that one views the various features as if one were hovering directly above the landscape.



PELE was the first volcano discovered on Io; it is also the largest geyserlike eruption observed there so far. The plume of the geyser is visible above the limb of the moon; it ascends to a height of about 300 kilometers and has deposited on the surface a concentric set of yellow and brown rings. The outermost ring averages 1,400 kilometers in diameter. At the center of the deposit is the source of the geyser: a complex of hills with a central valley. Markings and flows are evidence of past eruptions. This view of Pele was made by *Voyager 1;* it is a mosaic produced by a technique developed by Alfred S. McEwen of the Geological Survey in which high-resolution images contribute spatial detail and low-resolution images contribute color data. When *Voyager 2* arrived four months after *Voyager 1,* Pele was inactive.

field would have been greatly accelerated; hence Io and Europa may have received a more thorough battering than the Jovian satellites farther out. In addition the intense volcanic activity on Io has undoubtedly altered Io's content of volatiles. Still, it seems that Io, and more generally the regular moons of Jupiter, resulted from processes similar to those that gave rise to the planets. Jupiter and its moons are thus a miniature solar system in a deeper sense than the merely geometric one.

The surface Io revealed in Voyager images is dominated by volcanic activity to an extent that was almost totally unexpected. The reason is as follows. The normal influences on the thermal evolution (and so on the volcanism) of a planetary body are thought to be, first, the heating of the body by the release of gravitational potential energy in its accretion and also by the decay of short-lived radioactive isotopes such as aluminum 26; next, the gradual buildup of heat due to the decay of long-lived radioactive isotopes, mainly of uranium, thorium and potassium, and finally, a gradual cooling as the radioactive heat sources are depleted and heat is lost through the surface of the body by convection and conduction. It is a sequence favoring volcanic activity in large solid bodies, which have a greater ratio of volume to surface area than small ones. Hence the earth is fairly active today, whereas the moon's volcanic fires were banked long ago. (The ones on Mars are probably active today but are much less intense than the earth's.) Io, which has the size and the mass of the earth's moon, would be expected to exhibit no more than signs of ancient volcanism.

This expectation changed dramatically in March, 1979, just days before *Voyager 1* reached the Jovian system, when Stanton J. Peale of the University of California at Santa Barbara and Ray T. Reynolds and Patrick M. Cassen at the Ames Research Center published their analysis of tidal forces as a heating process for Io. Io and Europa are locked into one of the gravitational resonances first studied early in the 19th century by Pierre Simon de Laplace. In each such resonance the gravitational interaction of two moons adjusts their orbital distances until their orbital periods are common multiples. (The orbital period of Europa is twice that of Io.) One effect of the Io-Europa resonance is that Europa continually forces some eccentricity, or deviation from circularity, into Io's orbit. As a result the sub-Jupiter point on Io (the point on the body closest to Jupiter) oscillates, producing a tidal bulge that migrates back and forth on Io and varies in amplitude, heating the body by friction. Peale and his colleagues calculated that the amount of heating could be surprisingly large, and so they predicted that significant volcanic features might be present.

As *Voyager 1* neared Io it became clear that Peale and his colleagues were right. In images of ever increasing resolution no traces of large-scale impact craters were found. Instead the surface was covered with varicolored red, orange and yellow splotches, which soon resolved themselves into features resembling terrestrial and Martian lava flows and calderas (volcanic-collapse craters). It was obvious that Io has a geologically young, active surface. A few days after the *Voyager 1* encounter a further analysis of the images by the Voyager navigation and imaging-science teams revealed active volcanic eruptions on Io. Since then the nature of the volcanic activity revealed in the Voyager images has been intensively studied, and a preliminary picture of the kinetics and thermodynamics of Io's volcanism has begun to emerge.

Many of the characteristics of Io's volcanoes stem from the exotic chemistry of its surface. Well before *Voyager 1* the absence of certain features in Io's infrared spectrum (specifically absorption lines due to water ice or frost) had shown that Io's surface is extremely dry, and details of its spectrum at visible wavelengths had led to proposals that sulfur in some form is there. Spectral-emission lines due to sulfur ions trapped in Jupiter's magnetosphere were also suggestive. Then a variety of instruments on *Voyager 1* and *Voyager 2* showed that large quantities of sulfur and oxygen ions are spread throughout the magnetosphere but are concentrated near Io's orbit. Ultraviolet emissions from such ions proved to come from a torus around Jupiter centered on Io's orbit. Finally, members of the Voyager infrared-spectrometer team discovered infrared absorption features due to gaseous sulfur dioxide over one of Io's volcanoes. This made it clear that sulfur and sulfur compounds are central to Io's volcanic activity. Spurred on by the Voyager findings, two research groups, one at the Jet Propulsion Laboratory of the California Institute of Technology and one at the University of Hawaii, identified a prominent feature in Io's infrared spectrum near a wavelength of 4.1 micrometers as arising from sulfur dioxide frost.

Why is sulfur so prominent on Io? In the universe sulfur is fairly common: it is only about 40 times less abundant than oxygen. In the crust of the earth it is scarce, but then terrestrial sulfur is thought to be mostly hidden in the core in the form of iron sulfide. On the moon too sulfur is scarce. There the dearth of sulfur is part of a larger problem: the moon's low content of volatiles, which remains one of the greatest difficulties facing the effort to explain how the moon originated. In meteorites, particularly primitive ones, sulfur is abundant, and on the surface of Mars data collected by the Viking landers reveal the abundant presence of sulfur in one form or another.

Most models of how the moons of Jupiter formed suggest that initially Io should have been richer in volatiles than the earth's moon. It must then be presumed that intense tidal heating over the history of the solar system has driven off all the light volatiles on Io, including water, so that sulfur and its various compounds are the only major volatiles left. A competing possibility is that conditions around the incipient Io were hot enough to prevent the inclusion of water but not too hot to prevent the inclusion of sulfur. The nature of Amalthea, the next moon inward from Io, may decide the issue: if the incipient Io was too hot for water, Amalthea should have been even hotter (assuming it formed in its current place with respect to the other Jovian moons), and so it should consist of very refractory matter indeed. It has been suggested, however, on the basis of limited spectral data that the surface of Amalthea is similar to primitive, carbonaceous meteorites, which implies a significant volatile content.

Several types of volcanic activity on Io must be distinguished; all are probably influenced by the chemistry of sulfur. One type, eruptive plumes, accounts for the most dramatic volcanic activity observed on Io so far. Consider Pele, which was active during the *Voyager 1* flyby and was the plume discovered first. It rose about 300 kilometers and distributed its ejecta in an umbrella-shaped pattern about 1,400 kilometers in diameter. The umbrella shape, which typifies many of the plumes, suggests that their ejecta travel ballistically. On the other hand, a set of images of Pele's plume records a swirling structure. The swirls suggest turbulence, and hence that the visible particles are interacting with some invisible propulsive gas. (The Voyager images show primarily solid particles of micrometer size or smaller; the associated gases would not be seen directly.) Computer models of plumes formed by ballistic particle ejection under Io's gravity match the general characteristics of Io's symmetrical plumes quite well.

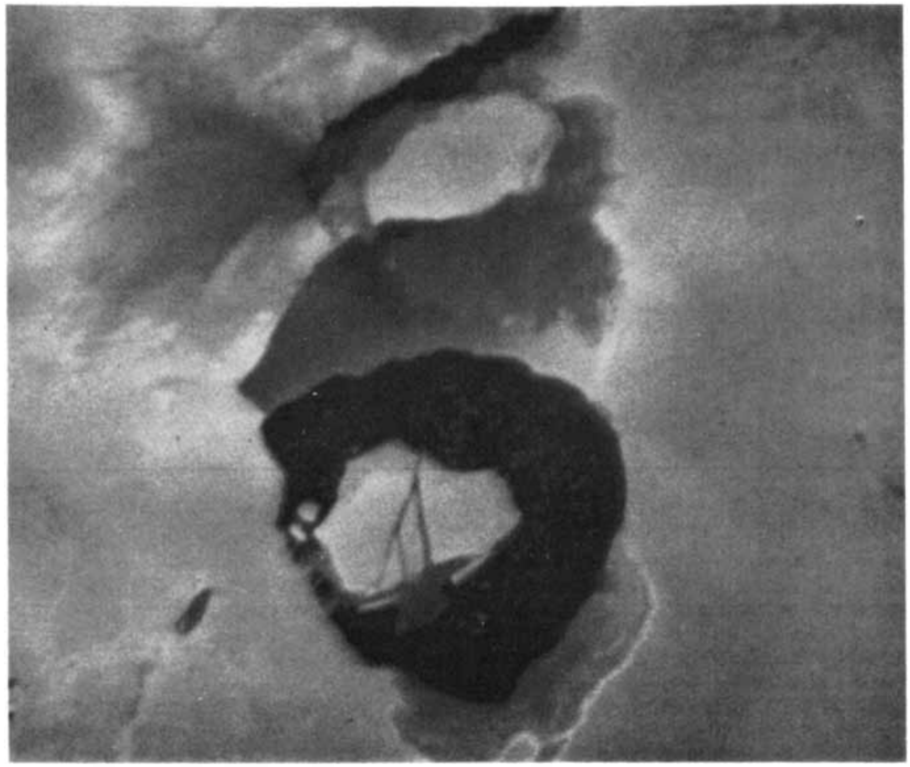
Any model of the mechanisms responsible for the plumes must explain the high exit velocities derived from the ballistic analysis, which range between 500 and 1,000 meters per second. Bradford A. Smith of the University of Arizona and his colleagues Eugene M. Shoemaker and Susan W. Kieffer of the U.S. Geological Survey and Allan F Cook II of the Center for Astrophysics of the Harvard College Observatory and the Smithsonian Astrophysical Observatory have proposed a geyserlike model, taking as their starting point the inference that the great amount of heat being pumped into Io by tides must result in subsurface temperatures that allow liquid sulfur dioxide and molten sulfur to be in contact at relatively shallow depths. Accordingly Kieffer has made an intensive study of the thermodynamics of plume eruptions for a wide

PROMETHEUS represents a class of volcanic plumes on Io that are smaller, cooler and longer-lived than the ones resembling Pele. These three views of Prometheus were made by *Voyager 1*. The top view shows the plume in profile; it is 100 kilometers high and 300 kilometers across. The other two views show the plume as the spacecraft passed overhead. The plume is dark against the surface. Nevertheless, it has laid down a ring of bright material. The brightness probably indicates a surface accumulation of sulfur dioxide frost.

LOKI PLUMES emerge from the ends of a linear black fissure some 200 kilometers long. Each has produced a bright fan and a larger, darker deposit, as if the two plumes mixed the eruptive styles of Prometheus and Pele. Below the Loki fissure is a *D*-shaped black patch that may be a lava lake, although it remains to be established whether the lava is sulfur or silicate. Measurements of Io's infrared radiation made by the Voyager spacecraft indicate that most of the surface of the lake has a temperature of about 300 degrees Kelvin, which is 170 degrees higher than the temperature of the surrounding terrain. The image is a *Voyager 1* mosaic processed by McEwen's technique.

variety of conditions involving these substances. At the low-energy end of a range of possible eruptions is one that starts when liquid sulfur dioxide comes in contact with hot liquid sulfur. The liquid sulfur dioxide begins to boil, so that a mixture of liquid and gas begins to expand up some conduit toward the surface. As the expansion continues, the production of vapor increases. Then the sulfur dioxide attains its triple point: the temperature at which ice, liquid and gas coexist. At this juncture, still at some distance below Io's surface, the remaining liquid freezes, and the further expansion of sulfur dioxide is marked by the condensation of sulfur dioxide vapor into a snow of solid particles. An alternative higher in energy starts with superheated sulfur dioxide vapor. Here the expansion toward the surface is accompanied first by the condensation of liquid and then by the freezing and snow.

In each case the mixture of sulfur dioxide phases is accelerated inside its conduit until the flow reaches the velocity of sound (in that medium). At that point its velocity (a few hundred meters per second) can no longer increase. Nearer the surface the conduit may widen. Moreover, on reaching the surface the flow can expand into the tenuous atmosphere of Io. Hence the flow can attain an impressive velocity just above the vent. By calculating the amount of energy available Kieffer has estimated that exit velocities of more than 1,000 meters per second are possible. She notes, however, that a large range of subsurface conditions can account for the Voyager observations, so that it is not yet possible to specify for individual plumes values of temperature, depth of origin and composition.



CLOSER VIEW of the Loki lake is afforded by this high-resolution image, also made by *Voyager 1*. The large bright raft in the lake is traversed by several cracks, and pieces of bright material appear to have broken from its edges. They dot the surface of the lake. Perhaps the cooling crust of the lake has been disrupted by convective overturn or the addition of fresh eruptive material. Conceivably the raft is made of elemental sulfur. The lake is 250 kilometers across.

Additional information about the plumes can be derived from their distribution on the surface of Io and from their temporal behavior. At least nine plumes were seen to be erupting when *Voyager 1* arrived, and some of them were seen several times over the days of the encounter. The nine were more or less uniformly distributed in longitude but were concentrated at low latitudes: eight of them were within 30 degrees of the equator. Eight of the nine were active four months later, when *Voyager 2* arrived. (Pele, the largest plume, was no longer active.) Two of the nine formed a pair of plumes, one at each end of the fissurelike feature north of a black patch named Loki. They exhibited major variability in the height of their eruption during the encounters and also between them. Moreover, the surface patterns around the Pele and Loki vents changed substantially between the two encounters.

Pele and Loki are fairly near each other. In particular, they are both well inside a hemisphere intermediate between the one that faces Jupiter and the one that trails in Io's orbit. (Io, like the earth's moon, keeps the same face turned toward its planet.) That hemisphere has long been known to be much redder and darker on the average than the opposite hemisphere. The only other major changes on Io's surface between the two encounters also took place there. Two dark reddish ring-shaped deposits, each on the order of 1,400 kilometers in diameter, or about the size of the Pele deposit, developed around two calderas, both at high latitude, one, named Surt, in the north, the other, named Aten Patera, in the south. Neither was found to be erupting at the time of a Voyager encounter, but the activity at Surt may have been responsible for a period of increased thermal emission at a wavelength of five micrometers (in the infrared) observed at the Mauna Kea Observatory by William M. Sinton of the University of Hawaii at Honolulu one night about a month before the *Voyager 2* encounter. The Aten Patera deposit was discovered during a recent analysis of Voyager images by Alfred S. McEwen and one of us (Soderblom) of the U.S. Geological Survey.

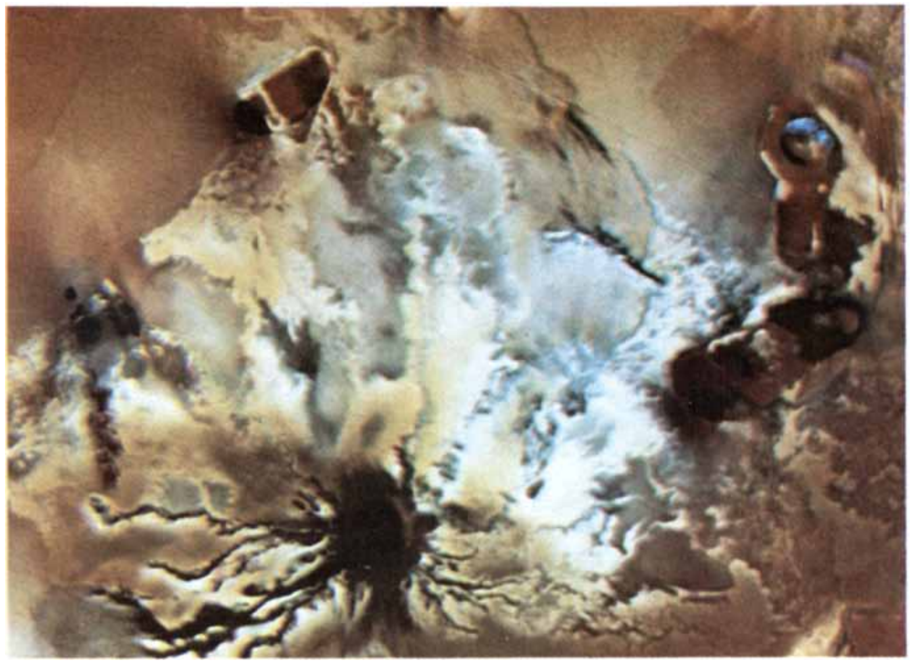
The color pattern and the scale of these two enormous deposits make them closely resemble the one surrounding Pele; therefore McEwen and Soderblom suggest they belong to a class of large, short-lived energetic eruptions of which Pele is the best example so far. Employing measurements made by the infrared spectrometer on *Voyager 1*, John C. Pearl and his colleagues at the Goddard Space Flight Center of the National Aeronautics and Space Administration estimate that the temperature of the Pele vent is about 650 degrees Kelvin; Sinton's observations, presumably of Surt, suggest a similar figure.

Eruptions of a second class are smaller: they deposit rings about 300 kilometers in diameter. They are long-lived: all of them were observed by both Voyager spacecraft, and from that fact the lifetime of such eruptions can be estimated as at least several years. They appear to be cooler: about 400 degrees K. They are restricted to an equatorial band of bright white material thought to be rich in sulfur dioxide frost. (Their ring-shaped deposits consist of the same material.) The best example is the feature called Prometheus, five degrees south of the equator. The pair of plumes at the Loki fissure seemed to oscillate between the two classes. The deposit surrounding them is complex: each has an inner zone consisting of a fan-shaped deposit, probably rich in sulfur dioxide, about 200

MAASAW PATERA and the surface markings surrounding that caldera on Io are much like the ones geologists see on the earth and on Mars. The principal difference is that Maasaw Patera is huge: the caldera alone is 50 kilometers across. Two levels in Maasaw Patera suggest stages of collapse. The greater part of the caldera is .7 kilometer deep, according to calculations based on the shadows in the caldera. The remaining part, at the upper right of the caldera, is two kilometers deep. At some point upwelling lava must have overflowed the rim of the caldera, producing streams of lava hundreds of kilometers long. The image is another of McEwen's *Voyager 1* mosaics.

kilometers long, and an outer zone similar in scale and color to the deposits around the three large, Pele-like plumes.

McEwen and Soderblom suggest that two distinctly different volcanic systems driven by different volatiles are responsible for the two classes. They also suggest that the classes arise from the peculiar relation in sulfur between temperature and viscosity. When sulfur is heated past its melting point, it undergoes several sequential changes. It starts as a yellow solid; then, at a temperature of about 400 degrees K., it melts to a low-viscosity yellow fluid. Further heating changes its color first to orange and then, at 430 degrees, to a clear pink. Next it turns deep red and thickens; by about 500 degrees it turns into a blackish tar. At about 600 degrees its viscosity starts to drop, and by 650 degrees it has become quite fluid again. Finally, at a temperature depending on the pressure, it vaporizes.



RA PATERA and the surrounding surface markings differ markedly from what geologists see on the earth and on Mars. Here Ra Patera is at the bottom of the image. It is a caldera filled with black material that sends out long, sinuous flows, mostly toward the left, over distances as great as 200 kilometers. David C. Pieri of the Jet Propulsion Laboratory of the California Institute of Technology and Carl Sagan of Cornell University propose that the dark center of Ra Patera represents a black, tarry phase of sulfur and that the flows represent cooler, less viscous phases. Two other calderas are notable. Toward the upper left an elongated caldera is banded in yellow, orange and black. The colors may represent different phases of sulfur. At the upper right a circular caldera includes a bluish crescent. In an image made by *Voyager 1* six hours before it made this one, the crescent is missing. Perhaps the crescent signals a low-energy eruption.

Low-viscosity molten sulfur that can readily transport heat occurs, then, in two forms: a clear reddish one at temperatures between about 400 and 430 degrees K. and an opaque black one at temperatures in excess of 650 degrees. The idea, therefore, is that the small plumes such as Prometheus are driven by liquid red sulfur coming in contact with liquid sulfur dioxide, making the sulfur dioxide vaporize and expand, as Smith, Shoemaker, Kieffer and Cook propose, whereas the large plumes such as Pele arise when hot silicates in the crust of Io vaporize sulfur at temperatures between 700 and 1,200 degrees K.

The remarkable difference in the duration of the two classes of plumes can be explained by this model. The small sulfur dioxide plumes persist for years because sulfur dioxide can be a liquid over a great range of depths in Io's crust. The liquid has extremely low viscosity (about that of alcohol at room temperature), and so it migrates easily through the crust to reach vents at the surface. In contrast, the large black sulfur plumes last for only days because the conditions favoring their eruption are delicately balanced. The venting of sulfur could stop, for example, if the temperature fell much below 650 degrees at the vent, in the conduit or deep in the crust where the sulfur and the silicates meet, because the sulfur would quickly cool into a tarry, nearly solid substance.

It remains an open question why the two classes of plumes are distributed as they are. The Voyager images of the dark red hemisphere of Io, where the large plumes are found, show numerous mountains that must consist of silicates protruding above Io's plains. (A sulfur mountain would not be strong enough to sustain its own weight.) Images of the opposite hemisphere have a lower resolution. In some of them, however, the region where the small, Prometheus-type plumes are most abundant is on the terminator (the day-night boundary), where mountains should be easily visible. In other images it is on the bright limb of Io. These views show mostly smooth plains.

Accordingly McEwen and Soderblom suggest there are major variations in Io's crust. Specifically, the crust in the region of the large, Pele-like plumes may be much thinner than elsewhere, so that hot silicates are closer to the surface and the pressure confining sulfur vapor is less. In addition the crust in the region of the small, Prometheus plumes may be enriched in liquid sulfur dioxide because some geophysical process (say a hydrostatic pressure gradient that varies with latitude arising from the rotation of Io) may drive the liquid toward the equator. It is worth noting that the Loki pair of plumes, which are perhaps a hybrid between the two distinct classes of plumes, lie in the area of overlap of the regions where the two classes are found.

A type of volcanic activity quite different from Io's geyserlike plumes is revealed by large volcanic calderas and their associated flows and surface markings, which are among the most prominent features identified in Voyager images. In general the calderas resemble terrestrial calderas except that they are larger: more than 200 of them are at least 20 kilometers across, as opposed to 15 on the earth, which has 3.5 times as much land area. The calderas on Io, unlike the plumes, show no strong concentration with latitude, being nearly as abundant per unit area at high latitudes as they are at the equator.

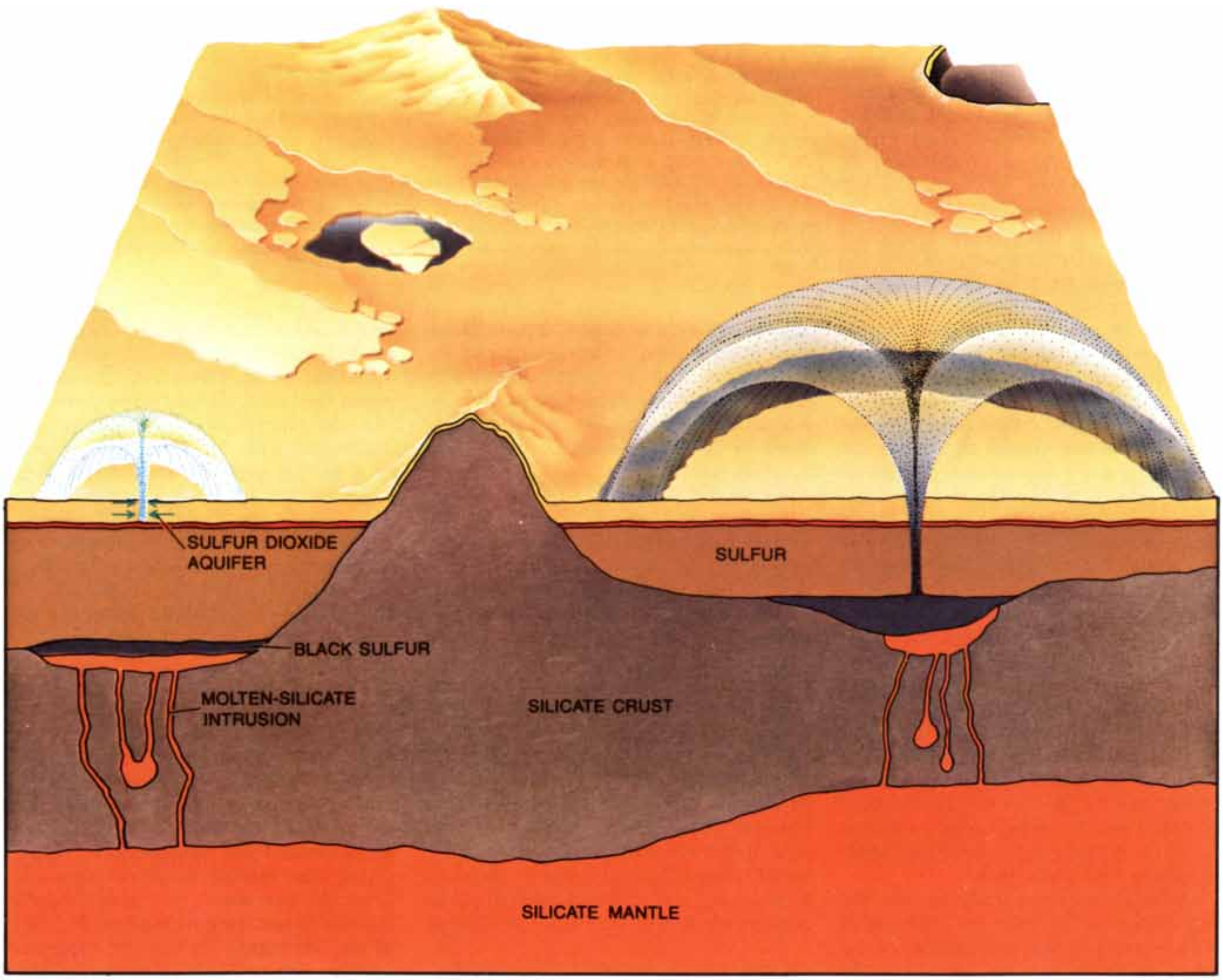
Debate about the nature of the calderas has centered on the issues raised by the plumes, namely the subsurface thermal conditions and the availability of sulfur and sulfur dioxide. Two extreme views of the calderas can be called the sulfur model and the silicate model. In the former, proposed in its purest form by Carl Sagan of Cornell University, the underground pools of molten sulfur postulated by Smith and his colleagues to explain Io's sulfur dioxide plumes also account for the volcanic flows on the surface; the flows around the calderas result from liquid-sulfur "lavas" rather than molten basalt or other silicates. In the latter, first proposed by Harold Masursky and Michael H. Carr of the Geological Survey and their colleagues, the calderas represent silicate volcanism much like that on the earth. The only difference is that the silicate flows on Io are tinted by sulfur.

The sulfur model is supported by the undoubted abundance of sulfur and sulfur compounds on Io, by the detection of sulfur dioxide gas near one of Io's eruptions and by the probability that these substances reach their melting point at shallow depths in Io's crust. In addition it is suggested that sulfur's sequence of changes in color and viscosity on heating or cooling explains the change in color with distance observed around many of Io's calderas. Ra Patera is a good example. David C. Pieri of the Jet Propulsion Laboratory and Sagan propose that the dark regions at the center of Ra Patera represent the high-temperature black phase of molten sulfur. Long, fingerlike flows that radiate from the caldera represent cooler, less viscous streams of the red and orange phases.

One problem with this proposal is that the behavior of the various sulfur phases under Io's conditions is uncertain. To be sure, the colors of the high-temperature phases can be retained in the laboratory by rapidly cooling the liquid. Nevertheless, the phases are only metastable and generally revert to ordinary solid sulfur after some period of time. Moreover, the phases on Io are likely to have impurities, and the impurities are likely to yield further complications.

The case for silicate volcanism rests in part on the measured topography of Io. The measurements are difficult because the spatial variations in the brightness of Voyager images resulting from differences in surface composition tend to be much stronger than the variations in brightness resulting from topographic relief. Still, some measurements have been tried on high-resolution images made under favorable illumination during *Voyager 1*'s closest approach to Io. The measurements suggest that Io has a range of topography far greater than that of the other Galilean moons: its mountains are from five to 10 kilometers high and at least a few of its calderas are from two to three kilometers deep. The maintenance of such a topography would be distinctly difficult on a body whose crust included thick deposits of warm, possibly liquid sulfur and sulfur dioxide. A stronger crust whose skeleton, at least, is silicate seems to be required; hence the broad distribution of calderas on the surface of Io would seem to rule out extensive subsurface sulfur "oceans."

The other main argument advanced by the proponents of silicate volcanism is that the structures seen on Io (calderas, long fingerlike flows, volcanic mounds with central craters and so on) are more or less what geologists are accustomed to seeing on the earth and on Mars. It seems implausible that sulfur, an unusual volcanic material with an exotic temperature-viscosity relation, should so closely mimic the structures produced by basaltic lavas. Unfor-



HYPOTHETICAL CROSS SECTION of Io shows an attempt to account for the body's volcanic activity. The mantle in the section is molten (or partially molten) silicate (*orange*); the crust is solid silicate (*brown*), overlain by layers of silicate debris rich in sulfur. The uppermost sulfur is cool and solid (*yellow*); then comes a thin zone in which the sulfur is molten (*red*) and a thicker zone in which the sulfur is hot and tarry (*yellowish brown*). At the left a plume like Prometheus is modeled. Molten sulfur makes contact with liquid sulfur dioxide (*blue*), which then boils up a conduit, producing at the surface a plume rich in sulfur dioxide snow. At the right a plume like Pele is modeled. Here a silicate intrusion heats sulfur so that it enters a very hot, tarry phase (*black*). Indeed, it vaporizes; hence vaporized sulfur erupts. A lava lake like Loki is at the upper left. A caldera is at the upper right. Its floor is the silicate crust underlying a thin layer of sulfur.

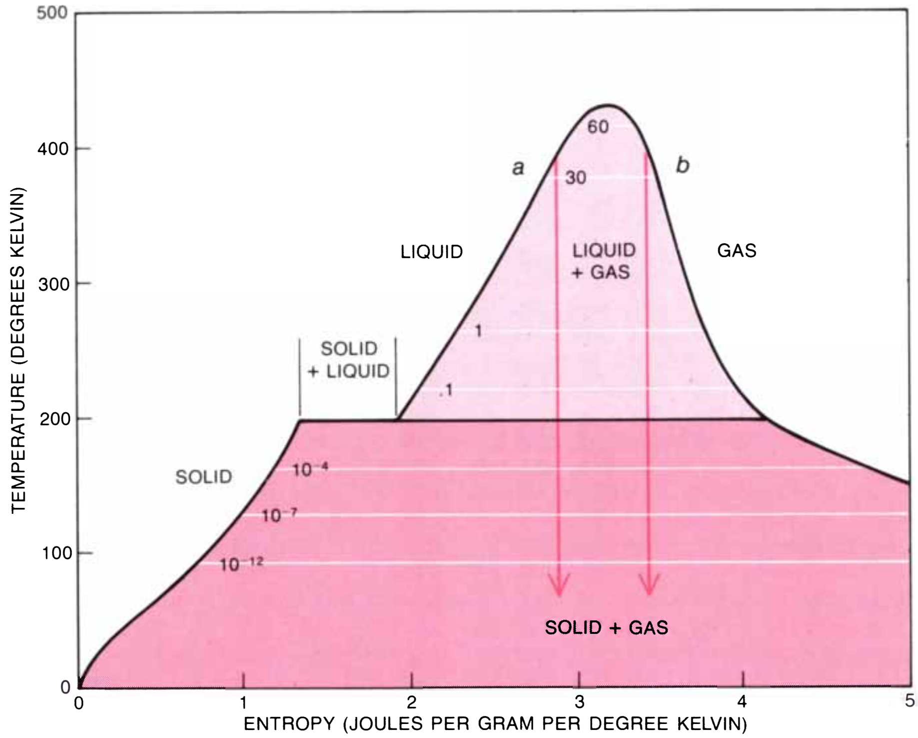
tunately geologists have only limited personal experience with sulfur as a volcanic fluid. There are rare sulfur flows on the earth, but they occur only when silicate volcanism melts sulfur deposits.

What, then, is happening on Io? Each side in the sulfur v. silicate debate now concedes the likelihood that the opposition's type of volcanic activity is occurring in at least some places on Io. Gerald G. Schaber of the Geological Survey observes, for example, that certain calderas are stamped out in mountainous regions of Io. Since the mountains presumably are silicate, the calderas must be silicate structures as well. Moreover, both sides agree that the volcanic activity at Io's surface, whatever its nature, is driven by molten silicates deep in Io's crust. The question is whether one type of volcanic flow, sulfur or silicate, dominates at the surface.

One further volcanic feature remains to be described. A discrete dark feature south of the Loki plumes is likely to be a huge lava lake of the kind that typically fills active calderas during eruptions on the earth. At the time of the *Voyager 1* encounter this feature was the largest "hot spot" on Io: it had a temperature of about 300 degrees K., whereas the local background temperature was only 130 degrees. Moreover, high-resolution images showed that the feature had a "raft" of bright material in it. The raft appeared to be traversed by cracks, and smaller pieces of the same material appear to have broken off from its edges. It is as if the cooling crust of the feature had been broken by convective overturn or the addition of fresh eruptive material. The feature is vastly larger than the lava lakes in Hawaiian calderas. Indeed, its length of about 250 kilometers would span the entire Hawaiian island chain. Is it filled with molten silicate or molten sulfur that now has chilled? We do not know.

The most striking evidence that the surface of Io is geologically young is the total absence of craters of impact origin in any Voyager image. Planetary geologists had found these signs of ancient and modern bombardment everywhere else they were able to look, including Mars and the earth, and the cratered surfaces of Ganymede and Callisto show that the Jovian satellites have had a history of impacts similar to that of the inner planets. Furthermore, calculations by Shoemaker suggest that the current flux of comets and asteroids through the Jovian system should be producing large craters on Io at a rate within an order of magnitude of the rate at which the earth's moon is now being cratered.

On the basis of Shoemaker's finding we have calculated how much volcanic activity is needed to bury Io's craters. We conclude that a one-millimeter thickness of volcanic debris must be spread over Io in an average year. This implies a very active body indeed. Specifically, it implies that Io erupts several thousand metric tons of material per second, or about the total amount of material ejected by the Mount St. Helens eruption of May, 1980, every month. The rate has significant consequences. Among other things, it suggests that at least the upper part of Io's mantle and crust has been recycled many times in the span of Io's history. This recycling is consistent with the idea that Io has been thoroughly "sweated out" by tidal heating, so that it has lost whatever light volatiles it once had.



THERMODYNAMICS OF SULFUR DIOXIDE suggest that a range of geyserlike eruptions is possible on Io. The vertical scale plots the temperature of sulfur dioxide; the horizontal scale plots the entropy. (Pressures are indicated by isobars labeled in units of bars.) On such a plot the horizontal position of a parcel of sulfur dioxide represents proportions of phases (for example liquid and gas). Moreover, a vertical line represents the evolution of a sulfur dioxide eruption that does no work (such as pushing against a barrier) and neither gains nor loses heat. Two such eruptions are shown. One of them (*a*) begins when liquid sulfur dioxide is raised to 393 degrees K., the temperature of molten sulfur deep in the crust of Io. The sulfur dioxide thus boils upward toward the surface. The other eruption (*b*) begins with gaseous sulfur dioxide. In each case the sulfur dioxide explodes as a cloud of ice and gas in the near-vacuum above Io's surface. The thermodynamics were analyzed by Susan W. Kieffer of the U.S. Geological Survey.

The details of the recycling are not well understood. For example, the relative importance of plumes and lava flows for covering the surface with new material is not clear. Estimates of the total amount of material erupting in Io's plumes imply that the resurfacing rate attributable to plumes alone ranges from a few ten-thousandths of a millimeter per year to a tenth of a millimeter or more. This suggests that flows are at least competitive with the plumes. Of course, one must bear in mind about such estimates that Io is highly dynamic and that the Voyager encounters with it have given us, so to speak, only two snapshots.

Is any evidence of Io's volcanic activity apparent from the earth? Remarkably, the answer is yes. In fact, evidence of volcanoes on Io existed in puzzling data gathered years before the Voyager encounters. Telescopic observations of Io at infrared wavelengths, made in the late 1960's and early 1970's, reveal two curious things about the object. First, when Io is eclipsed by Jupiter its temperature falls precipitously, as one would expect of a body with little or no atmosphere whose surface is a layer of fine insulating powder. On Io, however, as opposed to the earth's moon, the minimum temperature, attained when Io is deep in Jupiter's shadow, is too high to be consistent with a homogeneous insulating layer. Second, the temperature of Io inferred from its infrared brightness is not the same for all infrared wavelengths; it is significantly greater at shorter wavelengths.

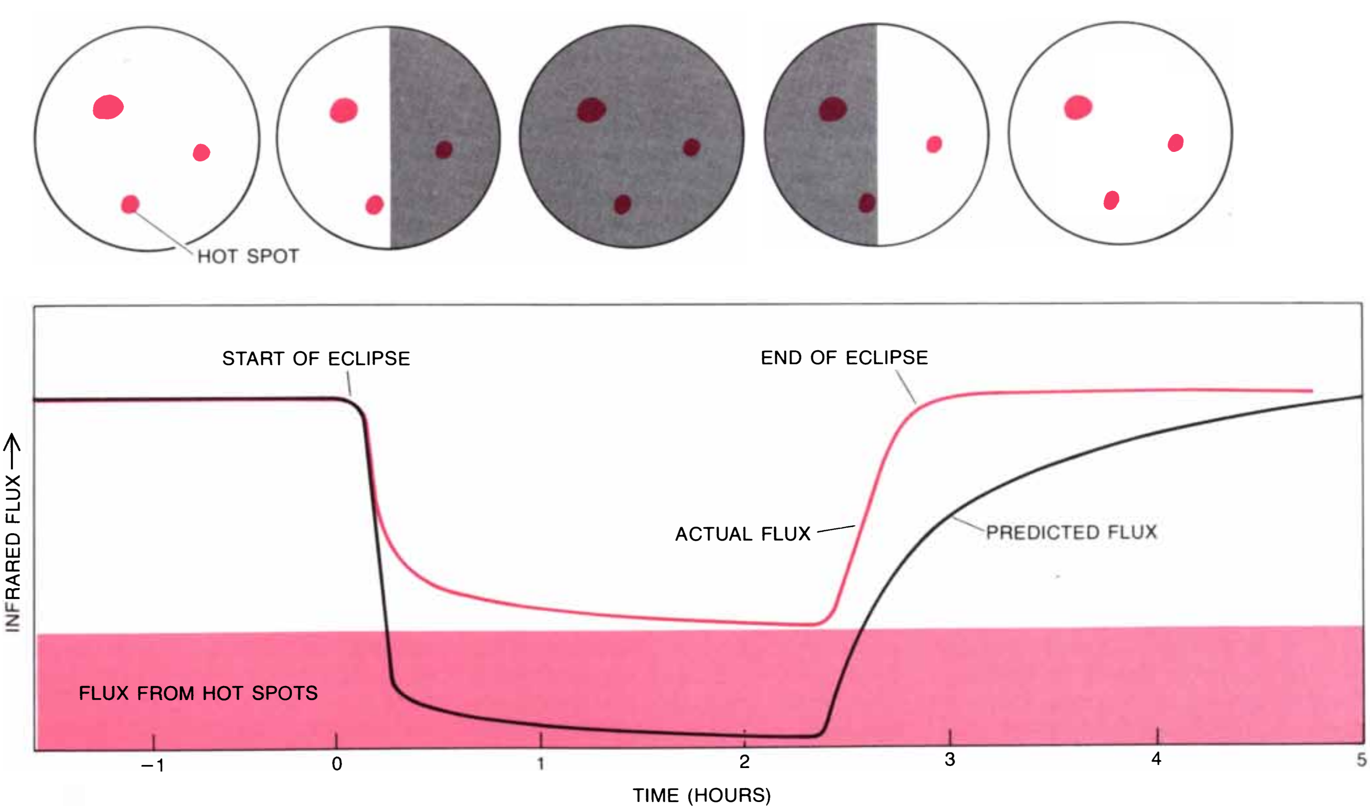
Individually these problems could be handled, although in an ad hoc fashion. Two-layer models of Io were devised to explain the eclipse data, and wavelength-dependent variations in Io's emissivity were proposed to explain the temperature anomalies. After the Voyager encounters and the discovery of volcanoes on Io, three investigators at the Jet Propulsion Laboratory, Dennis L. Matson, Gary A. Ransford and one of us (Johnson), returned to the puzzling data, hoping to solve the old problems in the light of the new information. They devised a simple model in which the insulating surface of Io is dotted with hot spots of a given temperature. It emerged that the problems would disappear if about 1 percent of Io's surface were covered with hot spots whose temperatures are similar to the ones the Voyager spacecraft measured.

The modeling yielded a value for the amount of energy radiated by the hot spots: averaged over the surface of Io, it came to approximately two watts per square meter. New analysis of Voyager data and new infrared observations made from Mauna Kea yield independent estimates of the amount; they agree with the model and with the earlier data. The current best estimate of the average heat flow from Io, drawing on all these sources, is 1.5 watts per square meter, plus or minus .5 watt. In comparison, the earth radiates about .08 watt per square meter, and the moon, with nearly the same size and mass as Io, radiates only .03. The radiation from Io stands out as being completely anomalous. Even without the proposal of tidal heating it would be clear that a source of energy other than internal radioactivity is active on Io.

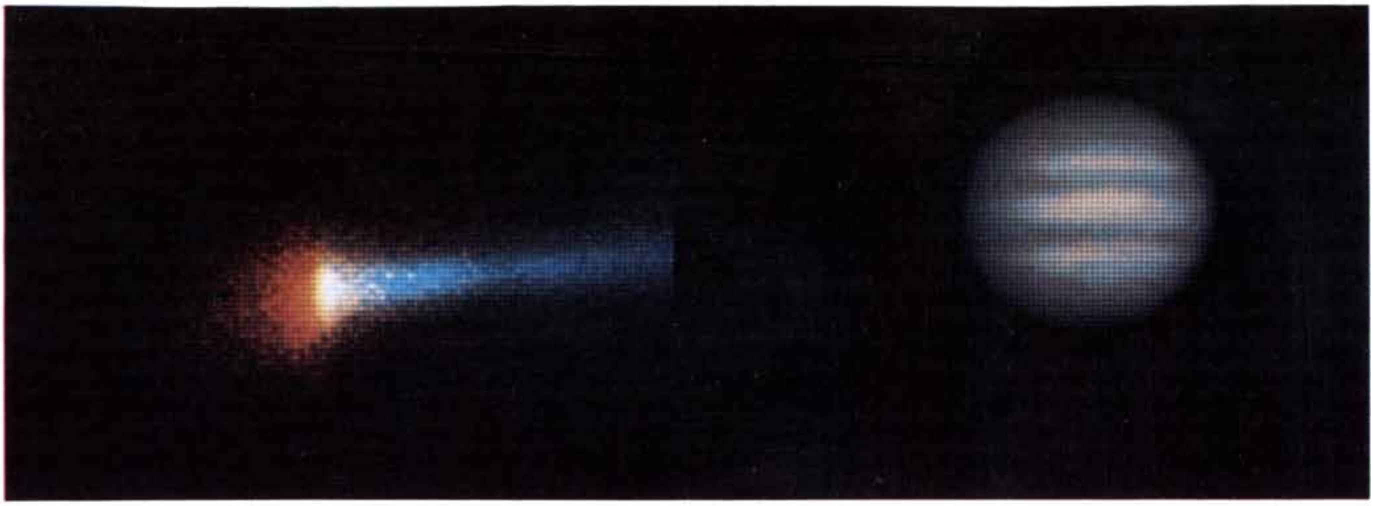
It should be said that the heat flow on Io differs from the heat flow on the earth and the moon not only quantitatively but also qualitatively. On the earth and the moon nearly all the excess internal heat generated by radioactivity reaches the surface by conduction through the crust. (Volcanic eruptions on the earth account for only a small part of the total.) From there it is radiated into space without measurably raising the surface temperature of the body. (The surface is heated primarily by the energy the surface gets from the sun.) This means the excess cannot be measured at a distance as infrared emission. What is required instead is the precise measurement of the thermal gradient in the upper part of the crust. On Io matters are different. The excess internal heat reaches the surface mostly by convection: it is carried by hot fluids rising to the hot spots. From there it is radiated into space at temperatures far higher than Io's surface temperature; hence it is readily measured. The energy output from Io's surface is between 10^{13} and 10^{14} watts.

So far two processes other than internal radioactivity have been proposed as the sources of this energy. One is tidal heating; the other is that Io is heated by its resistance to electric currents driven into the body by its interaction with Jupiter's magnetosphere. Indeed, Thomas Gold of Cornell suggests that Io's plumes arise from such an interaction when ionized matter ascends along the field lines of current flowing into Io's surface like great sustained lightning bolts. He bases his suggestion on the difficulty of driving volcanic gases to velocities exceeding the speed of sound and on data from the Voyager magnetometers implying that a current of about one million amperes flows along magnetic field lines in Io's vicinity. (A current of one million amperes is also what theoretical calculations of the flow of ions in the Jovian magnetosphere would lead one to expect.)

Io's output of 10^{14} watts can be exploited to test the possible importance of electromagnetic effects in the object's overall energy balance. Suppose all the current thought to impinge on Io is converted into heat in the interior by resistive heating. The net amount of heating would range from 10^{11} to 10^{12} watts, no more than a hundredth of the amount needed to account for Io's output. Electromagnetic effects might nonetheless be important for events such as eruptive plumes. In that case, however, the "lightning bolt" giving rise to a plume would produce a spot on Io with an effective temperature on the order of 100,000 degrees K. Searches for such spots on the night side of Io have not been successful.



ECLIPSE COOLING of Io puzzled observers who measured the infrared flux from Io as Io passed through Jupiter's shadow. The observers thought the flux would fall rapidly toward zero at the start of each eclipse and then return to its normal value when sunlight warmed Io's surface (*black curve*). Instead the flux stayed well above zero (*colored curve*). The explanation is that volcanic "hot spots" on Io emit a more or less constant flux of heat throughout the eclipse. Io's volcanic activity can therefore be monitored from the earth.



TORUS OF IONIZED SULFUR circles Jupiter and is centered on Io's orbit. It derives from Io's volcanic activity, which is thought to eject about one metric ton of sulfur and oxygen atoms per second into the space surrounding Jupiter. This image of one limb of the torus was made at the Mount Wilson and Las Campanas Observatories by John T. Trauger of Cal Tech. False colors designate the proportions of singly ionized sulfur atoms (*blue*) and doubly ionized sulfur atoms (*red*). The apparent right edge marks the limit of Trauger's data.

Meanwhile a problem has arisen with the tidal-heating hypothesis. The tidal interaction of Io and Jupiter involves tidal bulges raised on both the satellite and the planet; the bulge on Jupiter applies a gravitational torque to Io, causing the satellite to accelerate and move into an ever higher orbit. (A similar interaction is responsible for the slow but steady retreat of the moon from the earth.) The current best estimates of the strength of the interaction place an upper bound on the average input of energy to Io, and that upper bound is about a factor of two lower than the measured heat flow from Io.

The problem is still unresolved. Perhaps Io's heat flow and volcanic activity are highly variable over time, and the Voyager spacecraft arrived in a period when the activity was unusually pronounced. Perhaps some of the parameters and assumptions involved in the calculations of Io's orbital evolution and tidal heating are mistaken. Or perhaps the hot spots are not uniformly distributed over Io's surface. Recent data gathered at NASA's Infrared Telescope Facility in Hawaii by workers at the Jet Propulsion Laboratory and the University of Hawaii including one of us (Johnson) suggest that the region around Loki may contribute a large fraction to Io's total heat flow. If that is the case, and if Loki and its vicinity was mistakenly taken to be representative of the entire body, the total power actually emitted by Io may be closer to the theoretical expectation.

There seems to be no evidence that Io's heat flow varies much: the eclipse data recorded in the early 1970's give results in agreement with Voyager measurements and with eclipse data recorded in the 1980's. One type of infrared emission does exhibit dramatic short-term changes: the emission near a wavelength of five micrometers shows occasional bursts that probably are related to day-to-day processes on Io. The first such burst was detected in 1978 by Fred C. Witteborn of the Ames Research Center and his colleagues aboard the instrumented high-altitude aircraft named the Kuiper Airborne Observatory. Their brief observation of a five-micrometer flux of more than twice the intensity of what could be due to sunlight reflected from Io was startling.

The flux suggested that a small area amounting to 10^{-4} of Io's surface had a temperature of about 600 degrees K., but volcanic activity seemed almost unthinkable at the time. In the wake of the *Voyager 1* encounter the significance of Witteborn's observation was immediately apparent: the infrared spectrometer on *Voyager 1* had detected small regions on Io that had a similar temperature. William Sinton, working at Mauna Kea, began his systematic search for more "five-micron events." One of his first successes was the detection of the event between the Voyager encounters: the event thought to have been associated with the change around the Surt caldera.

Since then Sinton has accumulated several years' worth of data. He has found few events that rival the early ones, but his data show a continual variation in the flux from Io's very hot "hot spots." The total amount of energy these small, high-temperature areas radiate is much less than the amount more extended areas emit at lower temperatures. Hence the five-micron events do not contribute significantly to Io's overall energy budget. They nonetheless yield insight into the violent events that presumably generate large, Pele-like plumes. In addition they will facilitate the search for evidence of silicate lavas. More broadly, observations of Io made from the earth, including eclipse studies and infrared measurements, are all potential ways of monitoring the volcanic activity on Io and will provide an important link between the Voyager data and the data to be amassed by the next spacecraft sent to the Jupiter system.

The next spacecraft will be *Galileo.* Scheduled for launch in 1986 and arrival near Jupiter in August, 1988, it will include a probe to enter the atmosphere of Jupiter and a long-lived orbiter to allow some 20 months of detailed observations of the planet, its magnetosphere and its satellites. On its way toward Jupiter *Galileo* will pass within 1,000 kilometers of Io, or 20 times closer than the nearest approach of the Voyager spacecraft. Instruments aboard *Galileo* will map Io's surface and make images at a resolution comparable to that of Landsat images of the earth. The interaction of Jupiter's magnetosphere and Io's tenuous atmosphere will be examined, and radio tracking of the spacecraft will place new constraints on models of Io's interior.

After this nearest encounter *Galileo* will have to avoid the vicinity of Io: it would otherwise be disabled by a prolonged dose of radiation. Still, most of the orbiter's 20-month mission will keep it well within the range from which the Voyager spacecraft detected Io's volcanoes. And on about a dozen occasions *Galileo* will be able to scan Io for volcanic changes, make thermal maps that show the location of hot spots and search for evidence of ions passing from Io into Jupiter's magnetosphere. Io had already intrigued astronomers for 400 years when the Voyager spacecraft revealed a world more complex and yet more fascinating than it had been before. The trend will surely be continued by the next round of exploration.

An Early Iron Age Farm Community in Central Europe

Excavations at a site in Bavaria unearth clues to an economy of 1000 to 800 B.C. in which farmers were just beginning to exchange their surplus for the work of specialized craftsmen

by Peter S. Wells

Town life was commonplace in the Near East by 3500 B.C., but most settlements in Europe north of the Alps were little more than hamlets until the end of the late Bronze Age and the beginning of the early Iron Age. In that part of the world the transition to town life first took place in about 800 B.C. It was a time of rapid change in temperate-climate Europe that saw both the growth of trade and the expansion of metal production. Archaeological knowledge of the transition has come mainly from the excavation of cemeteries that number in the thousands and from the chance (but not infrequent) discovery of buried hoards of metal. Only a few European settlements of the period, however, have been investigated systematically, and so the economic developments that led to the formation of the first central European towns remain rather poorly known. Here I shall describe the findings of four seasons' work at a farmstead site of the period in Lower Bavaria and relate those findings to the rise, both there and elsewhere in central Europe, of larger settlements: precursors of the trading towns of the medieval period.

The site, Hascherkeller, is on a sand and gravel terrace that forms the northern border of a narrow river valley on the outskirts of Landshut, the principal city of Lower Bavaria. The terrace is 15 meters above the river (the Isar, a tributary of the Danube), and its waterborne deposit of glacial debris is covered by a thick layer of loess, a pale yellow sedimentary soil attributed to wind deposition at the end of the Ice Age. The loess in turn is covered by a rich humus, the product of millenniums of soil development, and the top 40 centimeters of the humus has been disturbed by modern deep plowing. Hence nothing of the original ground surface of the prehistoric settlement survives. Only those features that were a result of the settlers' own deep digging, such as pits of various kinds and boundary ditches, still remain: dark, humus-filled intrusions in the otherwise undisturbed loess.

Beginning in 1978 my colleagues and I set about dividing the area of intended excavation into mostly adjacent five-by-10-meter plots. The plow-disturbed humus was removed with shovels in two consecutive excavations, each to a depth of 20 centimeters. The archaeological material in the disturbed humus was collected and catalogued according to the level and the plot of origin. Once the bottom of the disturbed humus was reached the work proceeded with hoes and trowels until the top of the underlying loess came to view. Before the intrusive features thereby uncovered—the pits and ditches—were further excavated they were mapped and photographed. The humus fill in each pit was then bisected and the two halves of the fill were examined separately. All the fill we removed was screened through quarter-inch wire mesh.

As we proceeded it became apparent that the settlement had consisted of three enclosures side by side and extending from east to west, each one bounded by a double ditch. The enclosures were well defined on their northern, eastern and western sides, but stream erosion had destroyed their southern side. Our work in the first year and in the seasons that followed concentrated on the contents of the western enclosure and the middle one.

Although plowing had destroyed the top of the double ditches, what remained was as much as three meters wide. Excavation showed that the ditches were *V*-shaped and about 1.5 meters deep. The fill was dark brown and contained small fragments of pottery and animal bone. One section of the inner ditch of the western enclosure had a row of 19 postholes at the bottom; evidently a wood palisade had once stood there. The average distance between the centers of the postholes was 13.7 centimeters. If this figure represents the average diameter of the individual posts, the palisade was a stout one.

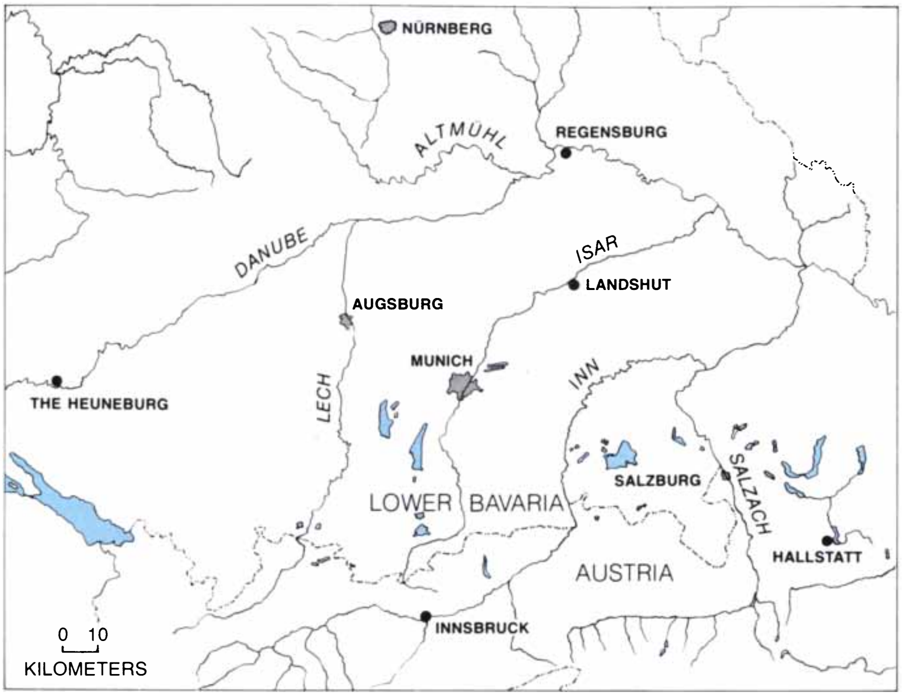
Experiments have indicated that earth-set posts of this kind rot within a few decades. Even though no other postholes have been found so far, many sections of the perimeter ditches show evidence of redigging, which would have been necessary to set up new posts as the old ones decayed. This suggests that all the ditches held palisades. Their purpose was presumably not so much to guard against possible attackers as to keep domestic animals inside the settlement and to keep wild animals out at night.

Most of the information about the community at Hascherkeller has come from the contents of the pits found within the enclosures. Of the 21 largest pits we excavated, 11 could be assigned on the basis of their contents to the three-enclosure settlement; the remaining 10 belonged either to earlier and smaller Bronze Age settlements or to later Roman occupations. Like the ditches, the pits with their dark fill stood out sharply against the loess subsoil. Indeed, the fill of the pits was even darker than that of the ditches, an indication that the pits had held additional organic matter.

The form, size and contents of the 11 pits put them in five functional categories. Three of the pits were long, narrow and cup-shaped, and they were oriented toward the cardinal points of the compass; two were oriented north-south and one was oriented east-west. Such orientations have been found to be prevalent among houses of the same period

RECTANGULAR EXCAVATIONS exposed one of three farmsteads at Hascherkeller, an Iron Age site in Lower Bavaria. The discoloration of the soil just in front of the two-meter scale rod reveals part of the ditch system that separated this farmstead from its neighbor to the west. The farmers erected palisades in the ditches in order to keep livestock from straying and to keep wild animals out at night.

PROFILE OF A DITCH dug at Hascherkeller three millenniums ago is marked by a *V*-shaped intrusion of dark humus into the yellowish loess subsoil covering the gravel of a river terrace. The ditch was one of two concentric ditches surrounding the westernmost of the farmsteads. In the ditches were found fragments of bone, pieces of baked mud plaster from the building walls and sherds of pottery.



SITE OF THE FARMSTEADS is above the Isar River on the outskirts of the city of Landshut. Nearby are Hallstatt, an Iron Age industrial site, and The Heuneburg, a trading town.

excavated elsewhere in Europe. The coincidence suggests that the pits served the function of "cellar holes" for houses. They were probably storage places for large pottery vessels that protected foodstuffs from dampness, temperature fluctuations and animal predation. The potsherds found in the three pits support this conclusion; most of them were fragments of thick-walled vessels.

The mouth of three of the deepest pits was circular and the walls were nearly vertical. Similar pits at other settlement sites of the period have been found to contain carbonized grain; moreover, the pits had been lined with basketry or clay, indicating that they had served for grain storage. The three pits at Hascherkeller did not yield any such positive clues to their use. They contained few potsherds or other kinds of waste, however, and may well have served a similar purpose.

Two more pits seem to have been associated with metallurgical activity. One of them held a hammerstone and a sandstone mold for casting finger rings. Several small scraps of bronze were found nearby. An adjacent pit contained red-stained pebbles and soil, apparently discolored by intense heat. One may conjecture that a hot fire in the second pit melted down bronze for casting. The pit where the mold was found also yielded a loom weight made out of fired clay and five clay spindle whorls. These suggest that the area served for weaving as well as casting.

One large pit, some meters north of the central enclosure, contained much charcoal, and at its bottom were the remains of a boxlike clay structure. More than half of all the potsherds found at the site were taken from this pit. Most of them were from what potters call wasters: pottery that had burst or warped as it was being fired and was therefore discarded. The conclusion seems inescapable that the clay structure in the pit was the firebox of a potter's kiln.

The last two pits, shallow and with gently sloping sides, held small fragments of pottery and animal bone. Apparently these shallow excavations had been borrow pits: places where the Hascherkeller inhabitants dug loess as a raw material for making their pottery or for plastering the walls of their houses with mud. Thereafter the hollows had gradually filled up with settlement debris. In this connection, the 11 pits yielded a total of 198 kilograms of baked mud, fired either when a structure had burned down or when the mud had been plastered on a wall adjacent to some source of heat, such as a hearth. The rest of the mud had simply weathered into silt again after the settlement's buildings had fallen into disuse.

The most abundant evidence of human occupation at Hascherkeller was the broken pottery. The total number of sherds recovered was 14,853. The great majority of them were found in the pits, but 3,828 came from the ditches, from the humus layers overlying the loess subsoil and from other areas. All represent plain, coarse wares typical of ordinary farmers' pottery. Many graves of this period in central Europe contain fine decorated wares, but fewer than 4 percent of the sherds from Hascherkeller had decorations of any kind.

We sorted the sherds on the basis of their thickness and found that they fell naturally into three categories. The first category consisted of relatively thin-walled wares; the sherds were less than 4.5 millimeters thick. The second category consisted of sherds from 4.5 to nine millimeters thick, the third of sherds more than nine millimeters thick. The remnants of small cups, beakers and bowls fell into the first category; they were also the fewest in number and the most frequently decorated. Sherds from larger bowls and from high, wide-mouthed jars with a coarse surface finish made up the second category. Those representative of the third category were chiefly from rough-surfaced jars that probably served mainly for food storage. At other habitation sites of the period many such intact vessels, buried in storage pits and cellar holes, have been found with grain still in them.

Except for the great quantities of fired mud, which provided useful information about the location of structures in the settlement, the most numerous remains at Hascherkeller were fragments of animal bone, 1,435 of them. Brenda Benefit, a doctoral candidate at New York University, has analyzed the bone fragments and finds that 253 of them can be identified as belonging to specific parts of animals of known species. The identifiable bones are predominantly (87 percent) those of domestic animals; the rest are the bones of wild animals. Of the bones of domestic animals, those of pigs predominate (37 percent), those of cattle come next (24 percent) and those of sheep and goats together account for almost all the rest (33 percent). Also found were fragments of a small number of horse and dog bones. The wild animal principally represented among the bone fragments is the red deer (*Cervus elaphus*), but there are also fragments of the bones of hares, hedgehogs and one species of bird (the quail). There are in addition a good number of fishbones but none that allow the identification of the species.

Benefit has analyzed the animal teeth uncovered at the site. She finds that the wear of the pigs' teeth has a bimodal pattern, indicating that the animals were slaughtered at two different times in their life span. Many had been eaten as suckling pigs, not long after birth. The rest were killed at about age two. This is a pattern of pig slaughtering common in Europe down to the present day. It maximizes the yield in meat with respect to the quantity of feed that must be given the animals when in the winter months they cannot forage for themselves.

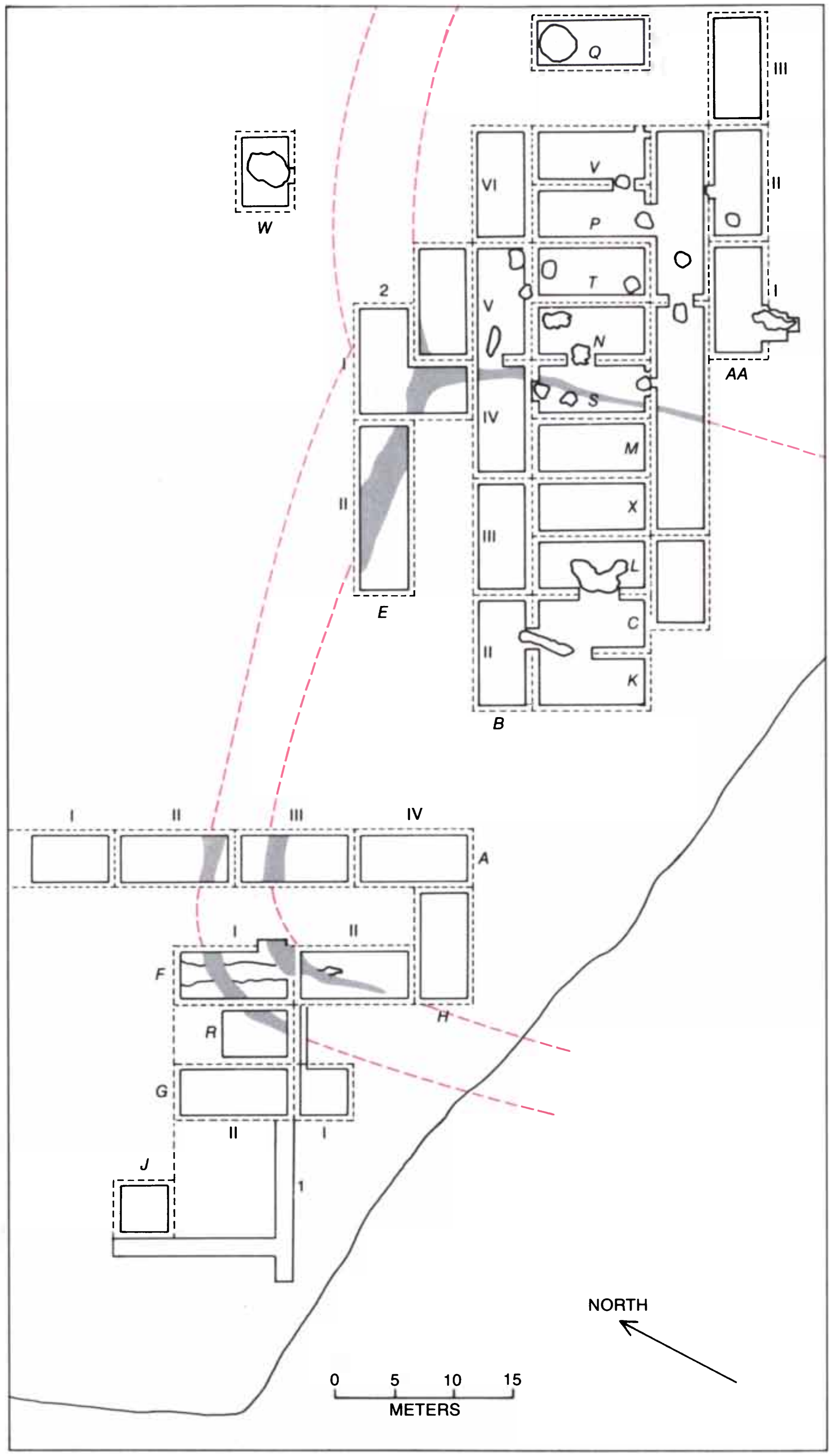
Sheep, goats and cattle led longer lives, regardless of the cost in winter fodder. Presumably the sheep were raised mainly for their wool and the goats and cattle for their dairy products. The cattle may also have been valued as draft animals and, when they were eventually slaughtered, as a source of hide.

In addition to this substantial array of meat and dairy resources the remains of various plants show that the farmers of Hascherkeller relied on cereals, garden crops and certain wild foods to augment their diet. Analysis of the plant remains by Caroline Quillian Stubbs, a doctoral candidate at Harvard University, indicates that the main cereal crops were millet, wheat and barley and that lentils were also cultivated. The inhabitants collected hazelnuts and the products of several other wild plants that are looked on as weeds today but that played an important role in the diet of earlier Europeans. At this settlement the inhabitants collected cleavers (*Galium*), goosefoot (*Chenopodium*) and sorrel (*Rumex*).

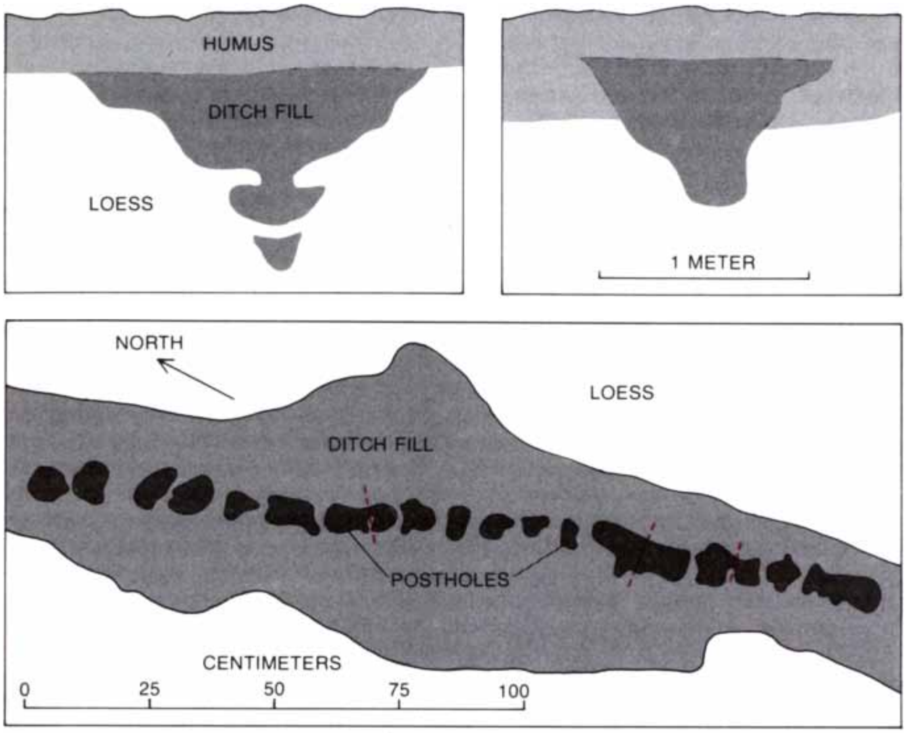
What picture of early Iron Age life can one build from these humble clues? The estimated size of the three ditched enclosures that make up the Hascherkeller farmsteads, each consisting of some 3,000 square meters, corresponds to the size of individual enclosed farmsteads at many other late prehistoric and early historic settlements in Europe. It is likely that each farmstead was inhabited by a family numbering from five to 10 individuals and that each included a dwelling, a barn for the stock and smaller structures such as sheds and workshops. It is evident that all three farmsteads were operating at the same time: their perimeter ditches meet neatly and never cut across each other.

The daily life of the 15 to 30 men, women and children in the settlement can be considered under three interrelated economic headings: subsistence, manufacturing and trade. Under the first heading there is abundant evidence for a self-supporting economic organization: animal husbandry that yielded both meat and dairy products supplemented by hunting and fishing, along with cereal and legume production supplemented by the collection of wild foodstuffs. Whether or not these farm activities yielded a surplus of meat or grain, it is probable that such activities as cheesemaking and hide dressing furnished the settlement with easily preserved commodities in excess of the farm families' own needs.

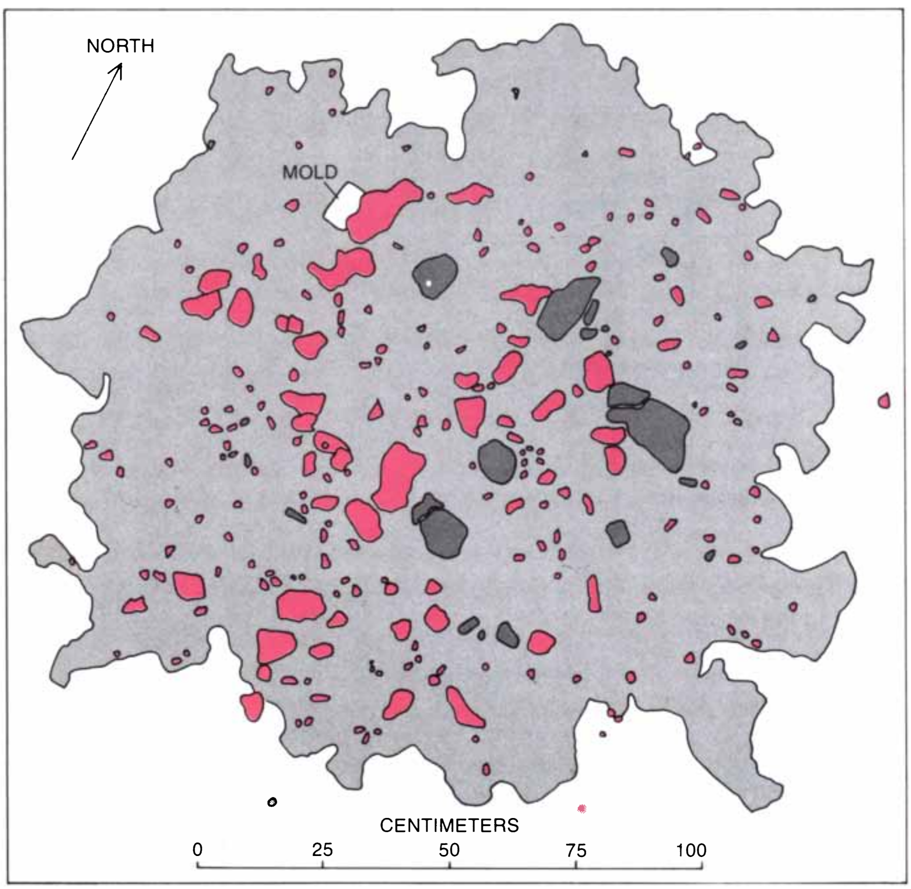
What commodities in addition to cheese, leather and possibly meat may have been produced in surplus quantities? They did not include pottery. At



TWO OF THE THREE FARMSTEADS, the westernmost and central ones, appear in this site plan. The excavated parts of the double ditches that surrounded each farmstead are in gray; their further extent, as indicated by a magnetometer survey, is outlined with broken colored lines. Of the 21 larger pits (*solid outlines*) uncovered by excavators, 11 had been dug during the early Iron Age. More than half of the pottery fragments found at the site came from a single pit (*W*) beyond the ditches of the central enclosure, which also contained the remains of a pottery kiln. Two cellar-hole pits (*C, AA I*) led the others in the quantity of mud-plaster fragments they contained, respectively 75 and 44 kilograms. The west pit in rectangle *N* (*see bottom illustration on next page*) held a quartzite hammerstone, a sandstone mold for casting bronze finger rings, a clay loom weight, five spindle whorls and 20 kilograms of plaster fragments.



DITCHES WITH A *V*-SHAPED CROSS SECTION (*left and right at top*) are distinguished from the lighter loess subsoil by the darkness of their humus fill. The plan view of a section of one ditch (*bottom*) shows 19 postholes over a distance of 1.6 meters. The broken lines in color indicate double holes. This section is a part of the inner ditch of the western enclosure.



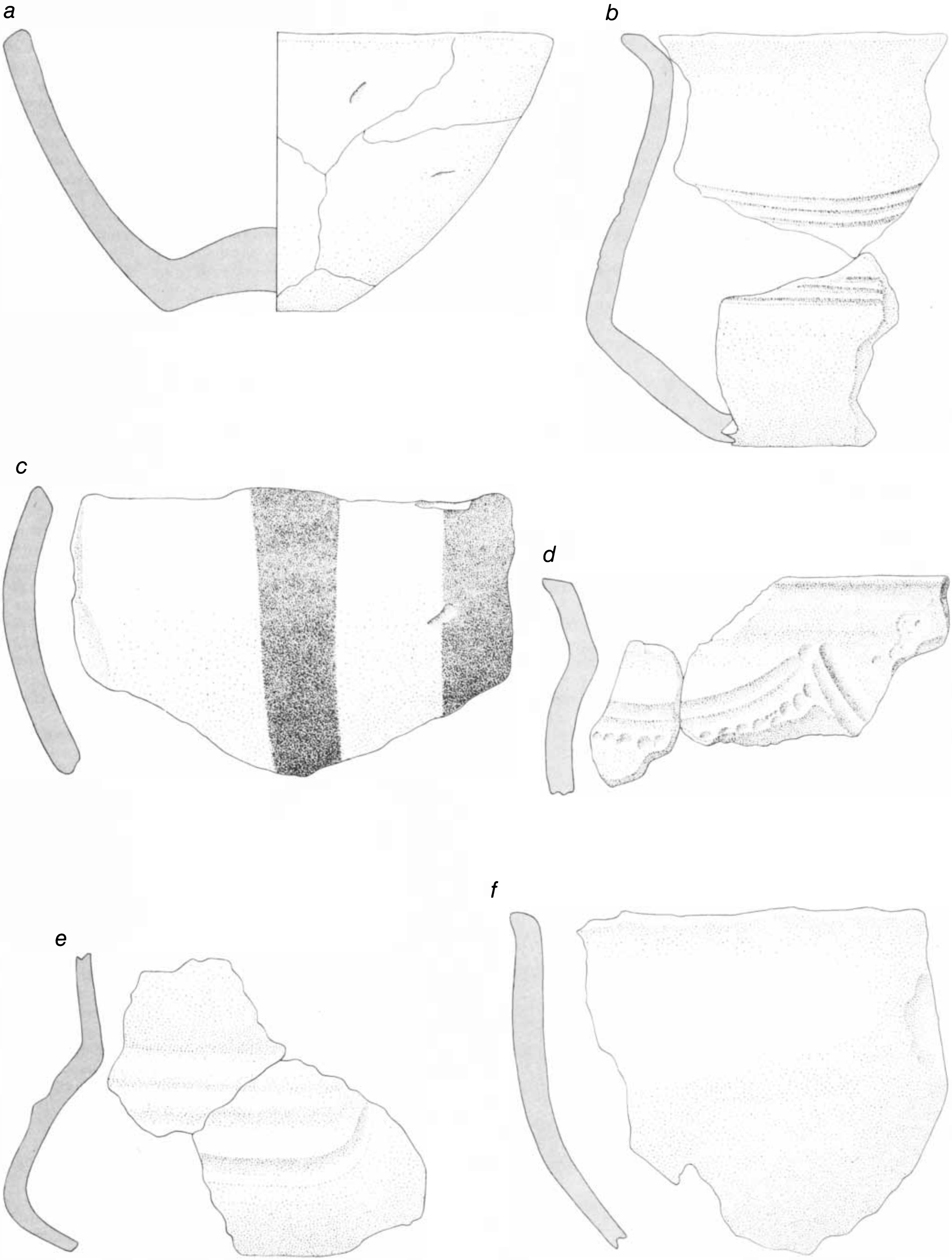
"WORKSHOP" PIT in rectangle *N* contained, in addition to casting and weaving equipment, numerous potsherds (*black*) and plaster fragments (*color*) in sufficient quantities to suggest that the structure housing the metal and textile workers also had mud-plastered walls. The sandstone casting mold is shown in situ to the left of center near the top of this plan view.

Hascherkeller each clay bowl, cup or jar was made not on the potter's wheel but by laborious coiling and paddling, which is unlikely to have given a farm settlement a pottery surplus. Our discovery of the loom weight and spindle whorls, together with the evidence that sheep were kept for a number of years before they were slaughtered, suggests, however, that one commodity perhaps made in excess of local needs was wool cloth.

This brings us to the third economic heading: trade. Evidence that the settlement's inhabitants imported exotic materials includes the fact that bronze artifacts such as finger rings were being made locally out of melted-down scrap bronze. The scrap was certainly imported, and several of the bronze pins unearthed may also have been brought in from the outside. On the other hand, we also found five fragmentary iron objects, too small and corroded to identify. Although they vouch for the settlement's being one of the Iron Age, they do not necessarily indicate that the farmers imported any iron artifacts. Iron ore is nearly ubiquitous, whereas the raw materials for the alloy we call bronze, tin in particular, are not. Evidence for local iron smelting, in the form of one lump of iron slag, supports the conclusion that the people of Hascherkeller were familiar with ironsmithing. There is no evidence, however, for local bronze smelting.

Additional imports include glass beads and graphite, the latter used for the surface decoration of a few of the local pots. If a smear of graphite is applied to the vessel's surface before firing, a shiny black finish is formed. This kind of ornamentation, either as a complete surface coating or as a series of bands, was particularly popular in central Europe beginning in about 1000 B.C. The principal sources of the graphite were deposits at least 100 kilometers from Hascherkeller, east of Passau on the Danube and to the north in Bohemia. As for the glass beads, four of them, blue green in color, were among the artifacts recovered: two from the pit that held the pottery kiln and two from another pit. Their place of manufacture is not known, but no evidence for local glass production has been found either at the settlement or at other sites of the period in central Europe. Judging from the relatively wide distribution of such beads, however, it seems likely that more than one bead-production center was regularly exporting its wares to the farm communities of the region.

What makes this pattern of scrap-bronze, graphite and bead importing particularly significant in relation to the later development of Iron Age towns in Europe north of the Alps is that all three



SHERDS FROM SIX POTS are fitted into the complete or partial profile of the original vessel. The decorated sherds are thinner than one of the two undecorated sherds (*a*). One vessel (*c*) was decorated with black bands, made by rubbing its surface with graphite before it was fired. The nearest sources of graphite were 100 kilometers away, so that the local potter probably obtained the material by trading.

imports were luxury products. The manufacture at Hascherkeller of metal sickles or other agricultural implements, although they can perhaps be regarded as overly stylish, has a utilitarian purpose. No farmer, however, really needs bronze rings and pins, graphite-ornamented pottery or blue green beads in order to raise more pigs or sow more millet. The settlement was trading with the larger world not for the necessities of life but for ornaments.

When was all this going on and for how long? Six charcoal samples taken from three pits yielded carbon-14 dates indicating the time of occupation was a 200-year period between 1000 and 800 B.C. The settlement's pottery and bronze objects and the imported glass beads in turn fit with a central European archaeological phase known as Hallstatt *B*. When material from this phase is cross-dated with material from Mediterranean cultures for which historical records exist, the Hallstatt *B* phase is assigned to the three centuries between 1000 and 700 B.C. Thus both the absolute chronology and the relative chronology are in close agreement.

As for how long the settlement endured, here the evidence is less direct. The ditches were renewed, and I have suggested that the renewal was needed in order to rebuild rotted palisades. Since the timbers of the palisades would have lasted for a few decades, the settlement could have lasted no less. Another indication of the length of occupation comes from the mud plaster preserved by baking. Many pieces of baked mud give evidence that the walls of some structures had been replastered and repainted two or three times. This suggests occupation for as long as two or three generations. Could the occupation have been even longer? Both the carbon-14 determinations and the analysis of the pottery indicate an upper limit of no more than two centuries of occupation. The pottery is completely homogeneous in style. It is difficult to believe the settlement would have remained unaffected by the winds of change over any period longer than 200 years.

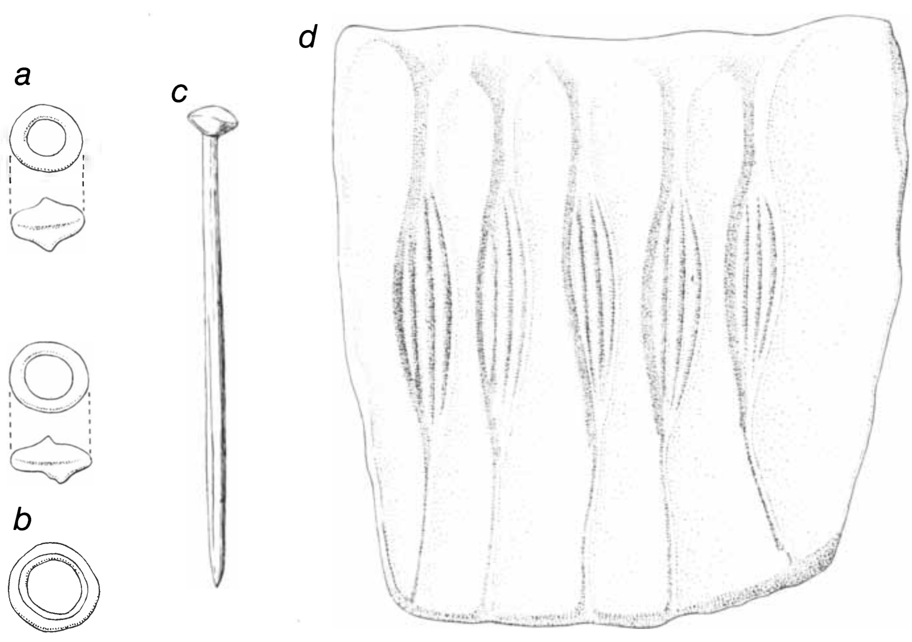
The picture of Hascherkeller that emerged after four seasons of excavation is one of a community economically self-sufficient for its necessities but at the same time tied into the larger world of central Europe by the desire for small luxuries. To satisfy these desires only one path could be followed: trade. Trade, however, is a two-way path. What the farmers could offer their trading partners was farm produce: wool cloth or yarn, cheese, possibly butter and salted meat and hides (or finished leather goods).

At this time a general intensification of agricultural production was under way in central Europe. Whether or not the Hascherkeller settlement remained active for a full two centuries, other farm settlements were then being occupied for as long and much longer. Manuring, fallowing and crop rotation were preserving soil fertility. Many more metal tools were being manufactured, not only sickles but also axes, saws, chisels and hammers, all implements that increased human efficiency.

Imagine for a moment that expanded metallurgy had led to the development of associations of smiths who did no farming of their own. As the graves of the period reveal, more than metal implements were being cast and forged. The dead were accompanied by ornate weapons and household goods: swords, helmets, large bronze vessels and gold ornaments. How would the smiths have fed themselves? Probably by trading their manufactured goods for surplus agricultural produce from the increasingly productive rural farmsteads. One can even imagine traveling middlemen finding their place in such trade networks.

The archaeological record provides an actual example of such an association that arose in what is now Austria no more than 160 kilometers from Hascherkeller. At the salt mines of Hallstatt a community with a population of some 200 between 800 and 400 B.C. devoted its energies exclusively to the extraction and trading of salt. The exceptionally rich assemblages of trade goods that accompanied the Hallstatt dead to the grave are eloquent evidence of the success of their experiment in a one-commodity communal venture. The first salt miners at Hallstatt began their work in about 1000 B.C., but they may have been farmers who mined salt chiefly for themselves, as the Hascherkeller bronze casters, weavers, shepherds and cheesemakers were to do in their own small settlement in the centuries that followed.

In any event the rise of the mining town of Hallstatt was not unique. Towns with populations in the hundreds, busily devoted to smelting and forging iron, sprang up in the Alpine border region of what is now Slovenia. Closer to Hascherkeller, in southwestern Germany, one well-studied site, The Heuneburg, grew into a commercial center: a densely packed cluster of substantial wood buildings not unlike the early medieval trading towns of 14 centuries later. The craftsmen of The Heuneburg engaged in different kinds of primary manufacturing rather than in some single specialty, and the same was true in similar commercial towns in central Europe. The principal function of the towns, however, was to engage in trade. None of these population clusters could have arisen without the support of thousands of small agricultural settlements such as Hascherkeller, able and willing to produce ever greater agricultural surpluses to exchange for the townsmen's goods. In Shakespeare's *The Tempest* Antonio says, "What's past is prologue." At Hascherkeller we can see the early Iron Age prologue to the urbanism of medieval and Renaissance times that ultimately shaped our modern world.



LUXURY PRODUCTS unearthed at Hascherkeller included glass beads (*a*), a bronze ring (*b*) and a bronze pin (*c*). Although the mold for casting the strips that were fashioned into finger rings (*d*) was found, no rings themselves were uncovered. The bronze pin and ring and the beads, like the graphite used to decorate the pottery, were probably traded for farm surplus.

Museum piece, circa 1987.

The clapstick may be taking an early retirement. This and other conventional tools of the filmmaking trade may find themselves relegated to crates and carted off to museums.

What will replace them? Kodak film with Datakode magnetic control surface! This radical advance in film manufacture, which unites chemical and magnetic imaging technologies, gives film the ability to "converse" with computers. This achievement could substantially reduce the time and costs associated with film postproduction.

The Datakode magnetic control surface is a thin, transparent layer of iron oxide particles (approximately 9 billion per square inch) coated across the entire back of the film during manufacture. Less than 8 microns thick, this transparent layer provides the means to record machine-readable information (100 bits of binary data in a single track on each frame of film), and makes possible a uniform frame-indexing code that can be used with both film and videotape. Best of all, it accomplishes this interface without altering the quality or characteristics of the final image.

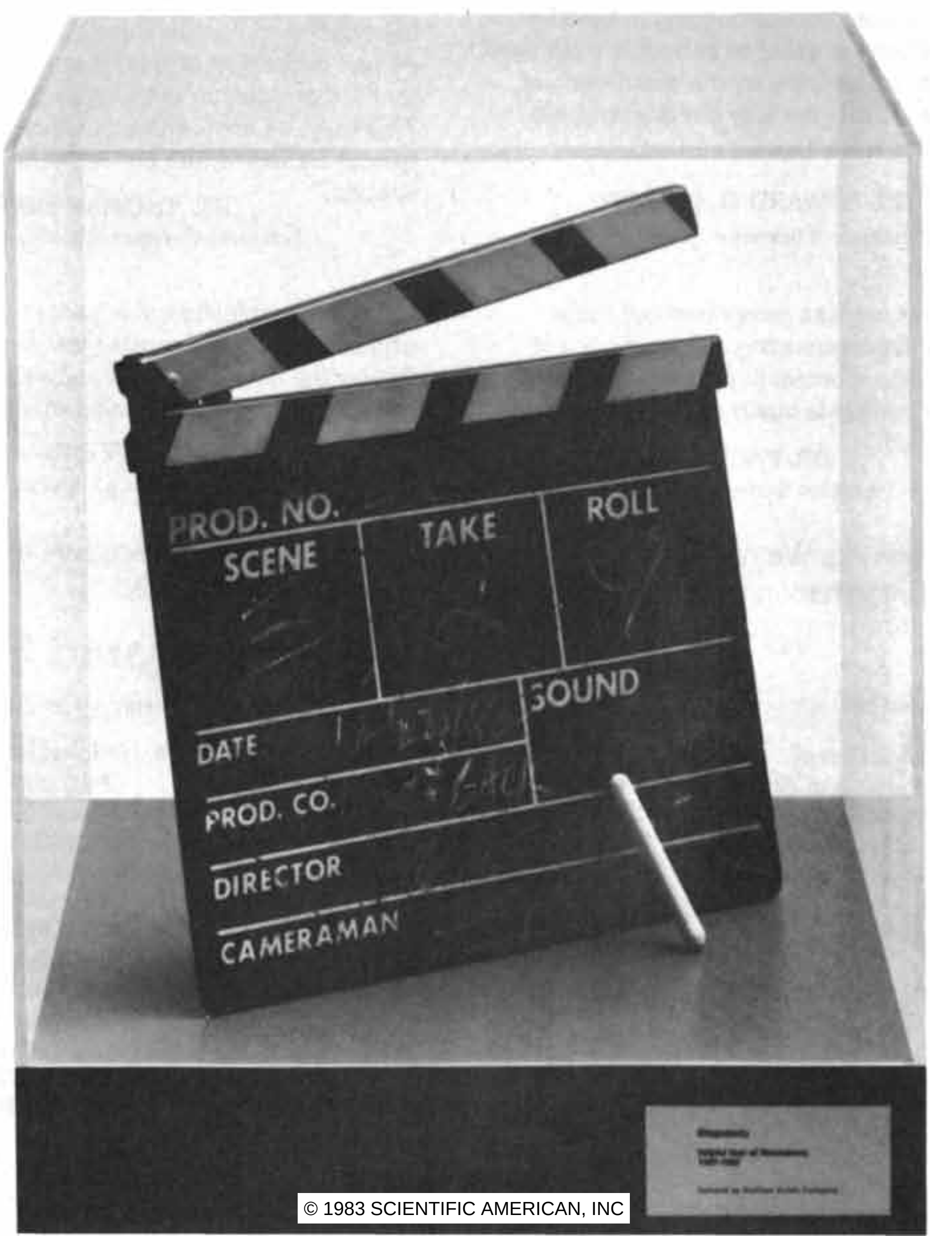
In the not-too-distant future, the use of discs, video displays, time-code synchronization, and automated printing equipment will speed filmmakers through all the noncreative, repetitive steps associated with film postproduction.

The new Datakode magnetic surface is more than just another milestone for Kodak and a boon to filmmakers. It represents the most significant development in motion picture technology in the last 50 years. It's a victory for science. And for the chemists and researchers who met the challenge to add to film's capabilities.

For more information on the new Datakode surface, send your request to: Eastman Kodak Company, Department GBSA-7, 343 State Street, Rochester, NY 14650.

Kodak

Kodak. Where technology anticipates need.



EXTINCTION. IT'S FOREVER.

One quarter of all species of animals and plants on Earth may disappear in the next 30 years because of man's destruction of their habitat. The rate of extinctions is increasing enormously as forests are destroyed and other wild areas are lost. Organisms that evolved over hundreds of millions of years will be gone forever. The complex interdependence of all creatures, from the largest mammals to the smallest plants, is being shattered. It is a crisis with profound implications for the survival of all life. Unfortunately, little is being done to save our planet's natural heritage. Here are some warnings by leading scientists:

"The worst thing that can happen—will happen—in the 1980's is not energy depletion, economic collapse, limited nuclear war, or conquest by a totalitarian government. As terrible as these catastrophes would be for us, they can be repaired within a few generations. The one process ongoing in the 1980's that will take millions of years to correct is the loss of genetic and species diversity by the destruction of natural habitats. This is the folly our descendants are least likely to forgive us."

DR. EDWARD O. WILSON
Baird Professor of Science • Harvard University

"We are encroaching on nature, in the U.S. and around the world, at an unprecedented rate. A large proportion of the chemicals in use in our present-day civilization were 'invented' by nature, not by the chemist in the laboratory. An estimated 40% of all drug prescriptions in the U.S. contain as their chief ingredients compounds derived from plants. There is no end to the potential for discovery in nature, because we have only begun the chemical exploration of nature. Tragically, we are burning our library of priceless genetic treasures with our reckless destruction of species."

DR. THOMAS EISNER
Schurman Professor of Biology • Cornell University

"Few problems are less recognized but more important than the accelerating disappearance of Earth's biological resources. In pushing other species to extinction, humanity is busily sawing off the limb on which it is perched."

DR. PAUL EHRLICH
Bing Professor of Population Studies • Stanford University

"The extermination of a quarter of the plant species on Earth during the coming few decades could lose forever the medicines and food sources we need to cure the diseases and hunger that plague mankind."

DR. PETER RAVEN
Director • Missouri Botanical Garden

The Ark is sinking. We need the help of every concerned citizen to conserve the diversity of life on Earth. For information about how you can help, please write to us at:

Save Endangered Species

P.O. Box 50771 • Washington, D.C. 20004

Sponsored by:

- American Association of Zoological Parks and Aquariums
- American Cetacean Society
- American Humane Association
- American Institute of Biological Sciences
- Animal Protection Institute of America
- California Native Plant Society
- Center for Action on Endangered Species
- Center for Environmental Education
- Defenders of Wildlife
- Desert Fishes Council
- Environmental Defense Fund
- Florida Audubon Society
- Friends of the Earth
- Friends of the Sea Otter
- Fund for Animals
- Garden Club of America
- Greenpeace U.S.A.
- Humane Society of the United States
- International Council for Bird Preservation
- Massachusetts Audubon Society
- National Audubon Society
- National Parks and Conservation Association
- National Shell Fisheries Association
- Natural Resources Defense Council
- Ocean Research and Education Society
- Sierra Club
- Society for Animal Protective Legislation
- Wilderness Society
- World Wildlife Fund - U.S.

Medicine is always changing, now there's a text that changes right along with it.

SCIENTIFIC AMERICAN *Medicine.*

The answers that you get from SCIENTIFIC AMERICAN *Medicine* are current, because each month new chapters replace the old. You will know where to look for the answers, because this advanced text follows the familiar structure of medical knowledge, changing as medicine changes.

Comprehensive. Leading authorities from Harvard and Stanford cover the full range of medicine for you the seasoned clinician.

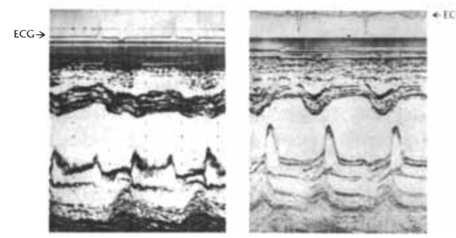
Complimentary CME. At no extra cost, we offer a CME program in which you can earn 32 top credits* per year.

Trial Offer. We invite you to try SCIENTIFIC AMERICAN *Medicine* for two months with our compliments. Send us the coupon and you will receive the two-volume text and two monthly updates. You may also take a CME test for credit. At the end of 60 days, if you decide to continue the subscription, you will owe us $220 for the first 12 months (current renewal is $170); otherwise, return the books; there is no charge.

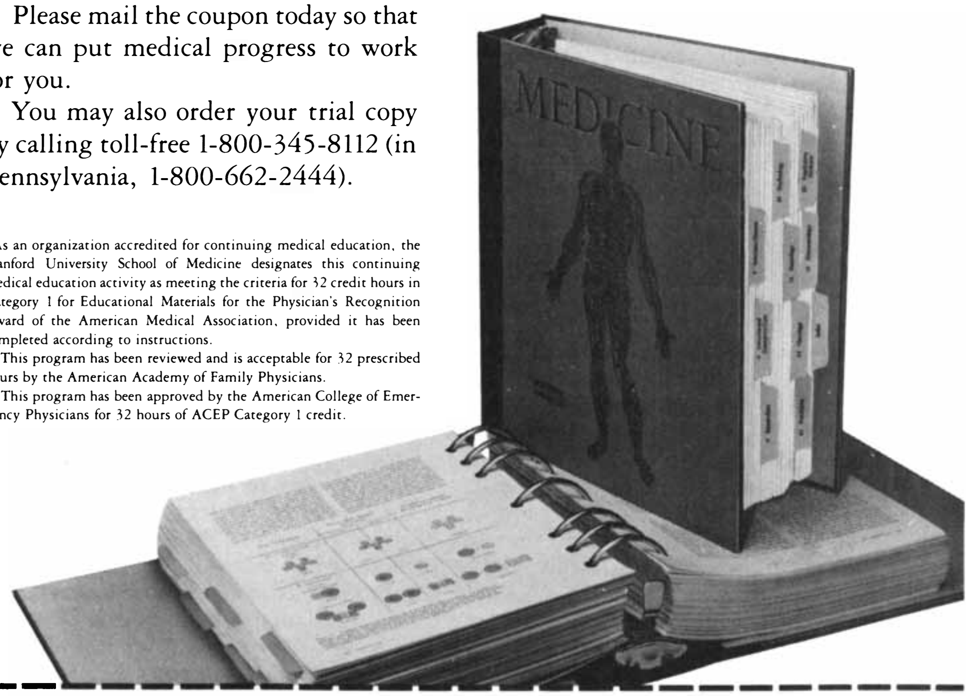
Please mail the coupon today so that we can put medical progress to work for you.

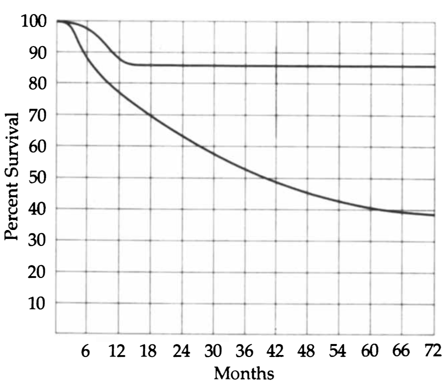
You may also order your trial copy by calling toll-free 1-800-345-8112 (in Pennsylvania, 1-800-662-2444).

*As an organization accredited for continuing medical education, the Stanford University School of Medicine designates this continuing medical education activity as meeting the criteria for 32 credit hours in Category 1 for Educational Materials for the Physician's Recognition Award of the American Medical Association, provided it has been completed according to instructions.
This program has been reviewed and is acceptable for 32 prescribed hours by the American Academy of Family Physicians.
This program has been approved by the American College of Emergency Physicians for 32 hours of ACEP Category 1 credit.

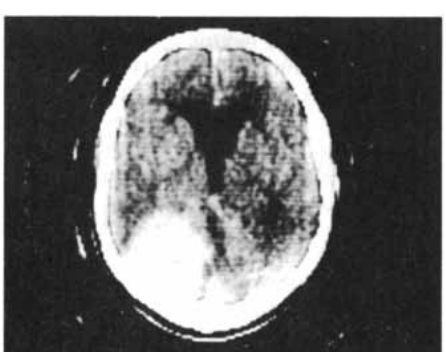


Echocardiograms from patients with aortic regurgitation.





Top curve represents survival in chronic hepatitis patients treated with steroids.



CT scan reveals large meningioma and mild hydrocephalus.

SCIENTIFIC AMERICAN MEDICINE

415 MADISON AVENUE, NEW YORK, NEW YORK 10017

Please enroll me as a subscriber to SCIENTIFIC AMERICAN *Medicine.* On receipt of this coupon you will send me the advanced two-volume text described in your announcement and update it regularly by sending me new monthly subsections. I understand that the annual subscription of $220 for SCIENTIFIC AMERICAN *Medicine* is tax deductible. If I am not entirely satisfied, I may cancel at any time during the first 60 days, returning all materials for a *complete refund.*

☐ Please enter my subscription for SCIENTIFIC AMERICAN *Medicine*
☐ I shall also enroll in the CME program
☐ I enclose a check made out to SCIENTIFIC AMERICAN *Medicine* for $220*
☐ Bill me ☐ VISA ☐ MasterCard

Expiration Date________ Account Number________

*Please add sales tax for California, Illinois, Massachusetts, Michigan, Ohio and New York

Name________ MD Specialty________

Address________

City________ State________ Zip________

Signature________

Please allow six to eight weeks for delivery. All payments must be in U.S. dollars. CME is available outside the U.S. and its possessions for an additional $65. Surface routes used unless airmail delivery is prepaid. 7G

SCIENCE AND THE CITIZEN

Thinking Nuclear

Are U.S. nuclear weapons intended primarily to prevent a war with the U.S.S.R. or to fight one? That question, never far from the minds of professional military planners since the beginning of the nuclear-arms race, has come into increasing public prominence lately in the wake of the Reagan Administration's apparent efforts to shift the emphasis of U.S. policy from the former purpose to the latter. Since the early 1960's the prevailing strategic doctrine in Washington, if not in all corners of the military establishment, has been that the overriding objective of the nation's nuclear forces is to deter a preemptive nuclear attack by the U.S.S.R. on the U.S. or its allies. Indeed, notwithstanding the thousands of tactical nuclear weapons stationed with U.S. forces in Europe and the repeated assurances to Western European leaders that these weapons are meant in part to deter a conventional, or non-nuclear, invasion of their countries from the east, there is good reason to believe that for at least the past two decades no U.S. president would have ordered the use of nuclear weapons for any purpose other than to retaliate for a nuclear first strike by the U.S.S.R.

Support for this conclusion can be found in an extraordinary statement made recently by Robert S. McNamara, Secretary of Defense under presidents Kennedy and Johnson and one of the chief architects of the doctrine of nuclear deterrence. Writing in *Foreign Affairs,* McNamara reports that during the early 1960's "in long private conversations with successive Presidents—Kennedy and Johnson—I recommended, without qualification, that they never initiate, under any circumstances, the use of nuclear weapons. I believe they accepted my recommendation." Although careful to avoid suggesting that all U.S. presidents would behave in the same manner, McNamara quotes approvingly a revealing statement by another influential presidential adviser: Henry A. Kissinger. "The European allies," said Kissinger in a 1979 speech "should not keep asking us to multiply strategic assurances that we cannot possibly mean, or if we do mean, we should not want to execute because if we execute, we risk the destruction of civilization."

The currently favored alternative view, enunciated most notably by Secretary of Defense Caspar W. Weinberger, stresses instead the utility of nuclear weapons for waging a wide range of "limited" or "protracted" nuclear wars and the concomitant undesirability of a "no first use" policy. Although Weinberger and other Administration officials have tended to mute their comments on this issue somewhat since the early days of President Reagan's term in office, the material evidence of their planned restructuring of the nation's military priorities is clearly discernible in the number, size and variety of the nuclear-weapons programs included in the Administration's past and future military-budget requests.

According to a recent study by budget analysts at the Center for Defense Information, spending on nuclear weapons has more than doubled since President Reagan took office three years ago. Over the next six years, they report, the share of the military budget devoted to preparations for nuclear war is projected to continue its rapid growth, accounting for some $450 billion, or 22 percent of the total military outlay for that period. By 1992, the report says, some 17,000 new nuclear weapons will have been made in the U.S.—an average of five per day for the next decade. Assuming the retirement of approximately 11,000 older weapons over the same period, the analysts remark, "the U.S. nuclear stockpile could grow by an estimated 6,000 bombs and warheads."

Included in the Administration's military-budget projections for the next 10 years are funds for 15 different types of nuclear warheads, aerial bombs and artillery shells. Nine of these new explosive devices are already under construction and work on several others is scheduled to begin soon. In general, the Center for Defense Information analysts observe, "the delivery systems these weapons will be fitted to are more sophisticated, more accurate and more destructive than those they are replacing." More to the point, they add, "many of the weapons being sought and the reasons given for their necessity fit a nuclear-war-fighting strategy."

Not surprisingly, the increasing reliance of the nation's military services on nuclear weapons is also reflected in a progressive "nuclearization" of training exercises and other war preparations. The effect of this trend is evident in testimony by Admiral Powell Carter before the Senate Armed Services Committee. "We have revamped the training command in order to teach our people how to think nuclear," he is reported to have said. "We are extending to where all our war exercises are into a nuclear phase."

It is precisely such thinking on the part of many of today's military and political leaders that former high officials such as McNamara evidently are beginning to find unsettling. "Having spent seven years as Secretary of Defense dealing with the problems unleashed by the initial nuclear chain reaction 40 years ago," he writes, "I do not believe we can avoid serious and unacceptable risk of nuclear war until we recognize—and until we base all our military plans, defense budgets, weapon deployments and arms negotiations on the recognition—that *nuclear weapons serve no military purpose whatsoever. They are totally useless—except only to deter one's opponent from using them.*"

Whirl Pool

The discovery of the three long-sought intermediate vector bosons at the European Laboratory for Particle Physics (CERN) this year has quickened the efforts of particle physicists around the world who seek a better understanding of the unified nature of the weak force and the electromagnetic force. The strong experimental evidence favoring the existence of the W^+, the W^- and the Z^0 particles confirms one of the main predictions of the electroweak theory; the theory regards the weak force and the electromagnetic force as two aspects of a single underlying phenomenon. The success of the theory suggests to many physicists that its predictions about interactions of still higher energy may well be true; confidence in the theory now seems high enough to justify the considerable risk in time, effort and money that will be needed to continue its verification. In the next 10 to 15 years a substantial part of the resources of other major laboratories will be devoted to testing and probing the predictions of the electroweak theory and, not incidentally, to seizing the mantle of leadership from CERN.

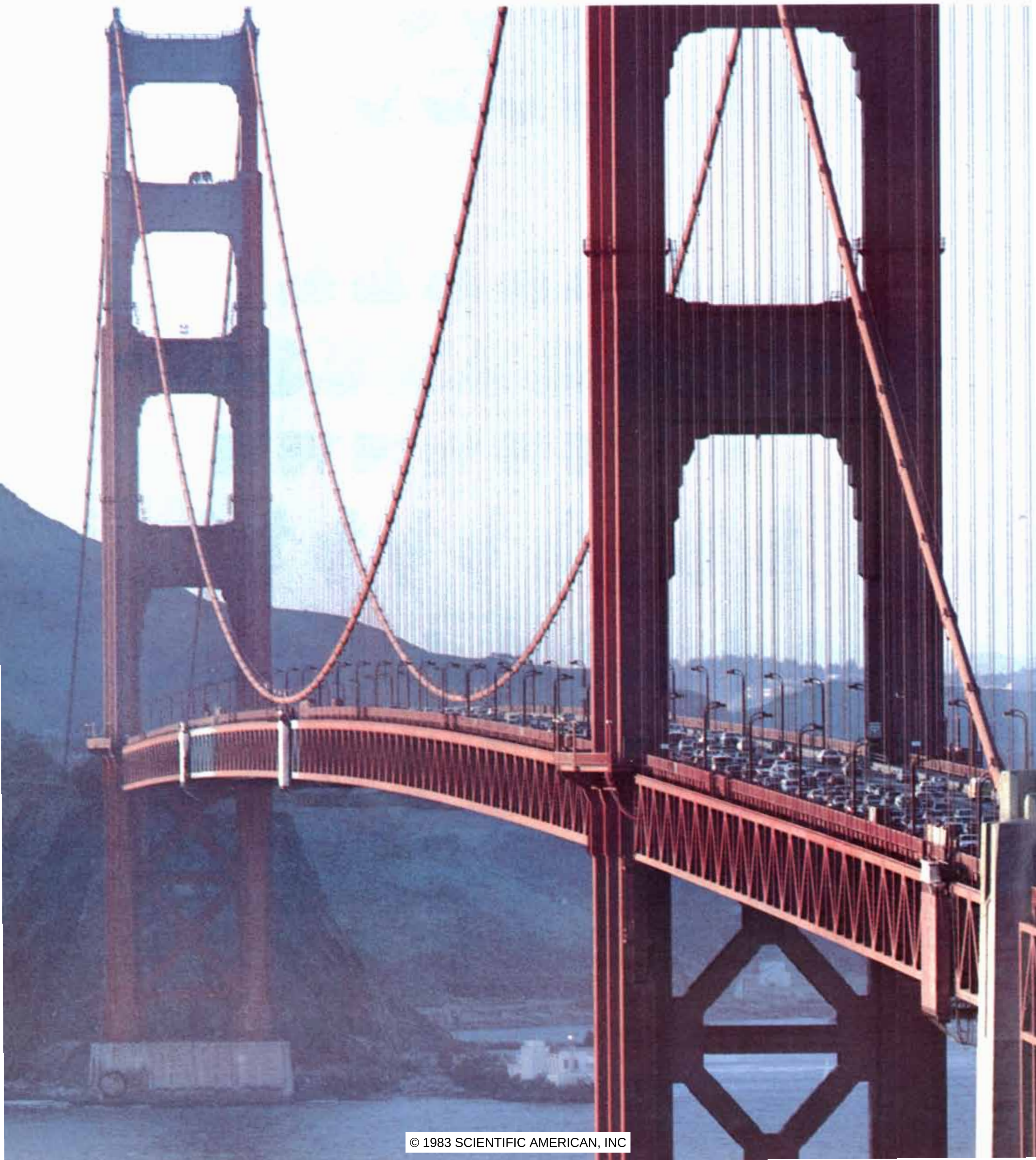
In the U.S. the results at CERN have emboldened the physics community to begin serious work on the design of a machine nicknamed the Desertron, so called for one of its likely sites. The Desertron, more formally called the Superconducting Super Collider, would accelerate two beams of protons in opposite directions around a track perhaps 100 kilometers in circumference. Protons in each of the two beams would reach an energy of about 10 TeV (10 trillion electron volts) before colliding, thereby liberating twice that energy in the collision. The net energy available for the materialization of new particles would be on the order of 1 TeV, or roughly 40 times the energy that can be extracted from the present machine at CERN. Many physicists believe that at such energies a new particle called the Higgs particle, required by the structure of the electroweak theory, could be created.

The detection of a Higgs particle would be a discovery of the first rank,


A partridge in a pear tree
Two turtledoves
Three French hens
Four calling birds
Five gold rings
Six geese a-laying
Seven swans a-swimming
Eight maids a-milking
Nine ladies dancing
Ten lords a-leaping
Eleven pipers piping
Twelve drummers drumming

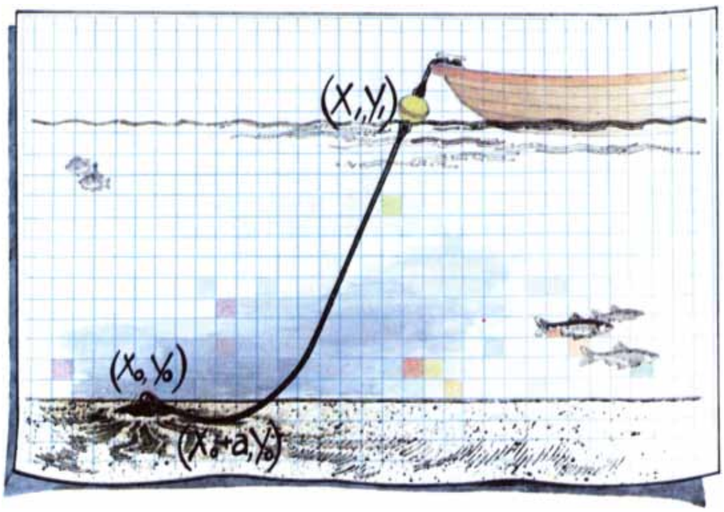
The Texas Instruments

Because y(x)=(1/k) co



Home Computer. sh [k(x+a)]+b.

Whether it's beginning math or this catenary equation, the TI-99/4A Home Computer gives you real power to solve real problems.



The problem: A 100-foot-long cable, whose weight is 0.67 pounds/foot (corrected for buoyancy) is used to secure a mooring buoy to an anchor in 30 feet of water in a flat-bottomed lake. Determine the horizontal and vertical force components at the anchor and buoy when the buoy is displaced horizontally 92 feet from the anchor. If any of the cable lies on the bottom of the lake, find the distance from the anchor at which it breaks away from the bottom. When the cable is unconstrained (i.e., not touching bottom), the equation of the catenary is $y(x) = (1/k)\cosh[k(x+a)]+b$. See the computer screen below for the solution.

A true home computer should be user friendly to every family member, novice or professional. It takes power to do it.

That's what makes the TI-99/4A Home Computer the one to choose. It's a powerful package designed around our 16-bit TMS 9900 microprocessor. And it contains a lot of fully usable memory – 16K RAM and 26K ROM, plus up to 38K more ROM in each solid state cartridge.

With TI's BASIC system built into ROM, it doesn't use up Random Access Memory (RAM). Others, who load their operating systems into RAM, can use up half their memory before you even get started.

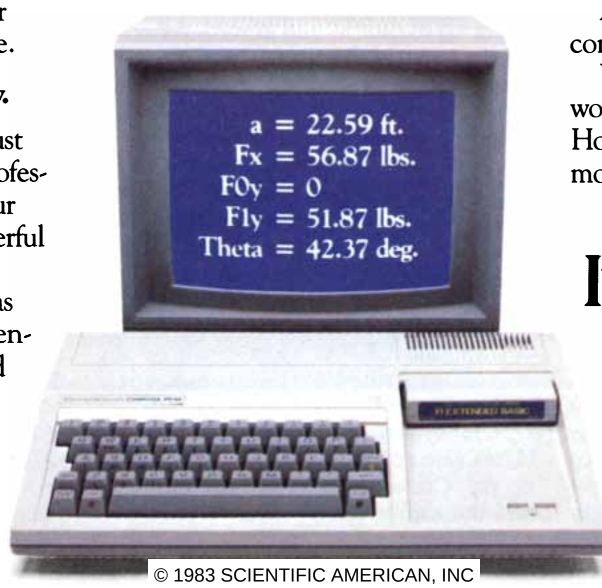
The result is that the TI Home Computer performs highly sophisticated computing tasks as easily as it teaches third grade math. For the money, no other computer is better suited for technical or scientific use.

Greater arithmetic accuracy.

Our unique 16-bit processor is just like the one found in expensive professional computers. Coupled with our rich TI BASIC, or still more powerful TI Extended BASIC, it makes our scientific programming capability as good as that provided by many scientific languages for mainframes, and better than some.

We give you 13-digit accuracy, and number range spanning $1\text{x}10^{\pm 127}$. All other home computers, using 8-bit processors, limit you to 6 to 9 digits, and only $1\text{x}10^{\pm 38}$ range. It's a big difference. And it means you can solve problems on the TI Home Computer which aren't feasible on the others.

TI BASIC also gives you more powerful commands and better flexibility for naming variables, since it recognizes up to 15 character descriptive names. Some others recognize only the first two significant characters in names, which means they can't tell the difference between "TAX" and "TARIFFS." So you have to resort to confusing acronyms.



The TI Home Computer is fluent in seven languages.

You can choose languages for all programming levels, from built-in TI BASIC to TI Extended BASIC, TI LOGO II, UCSD Pascal,* TI PILOT, TI FORTH and our powerful TMS 9900 Assembly Language.

You can also write useful non-scientific programs with the TI Home Computer. It's the only one in its price range that lets you combine 16-color graphics, single-command animation, sound effects, music, and speech with unlimited vocabulary into the same TI BASIC program.

And only TI's home computer comes with a 1-year warranty.

Whatever kind of problem you're working on, work it out with the TI Home Computer. You can't find more computer for the money.

TEXAS INSTRUMENTS

Creating useful products and services for you.

*UCSD Pascal is a trademark of the Regents of the University of California.

and the possibility that the Desertron could find it has mobilized the American physics community behind the machine. In June, citing the need to consolidate funds and talent, a subpanel of the High Energy Physics Advisory Panel (HEPAP) of the Department of Energy voted 10 to seven to abandon construction of the half-finished Colliding Beam Accelerator at the Brookhaven National Laboratory in favor of the Desertron. Soon after the vote the members of HEPAP unanimously agreed to make the Desertron their top priority.

In October the Department of Energy redirected $18 million originally designated for the Colliding Beam Accelerator to a feasibility study for an ultrahigh-energy accelerator. Research and development over the next three to four years is expected to cost $150 to $200 million. The total cost of the machine is estimated to be at least $2 billion, and the length of time needed to design and build it would probably be 10 to 12 years. Nevertheless, many physicists are convinced that whether or not the Higgs particle is found, the Desertron would reach energy regions of great importance to physics and must be built if the most talented young physicists are to be attracted to the U.S. program.

There has been a major technological advance that makes it feasible to contemplate the Desertron, namely the development of reliable superconducting magnets. Such magnets can generate enormous magnetic fields for steering and confining a beam of particles, and they are far less costly to operate than nonsuperconducting electromagnets because superconducting materials have no resistance to the flow of electric current. Superconducting magnets were to have been installed at the aborted Brookhaven accelerator, and the difficulties encountered at Brookhaven in their manufacture caused the delays in construction that finally led to the cancellation of the project. The problems with the magnets have now been solved. According to Nicholas P. Samios, director of Brookhaven, the Desertron might be a version of the Brookhaven accelerator scaled up by a factor of 50.

Superconducting magnets have been installed at the Tevatron I, the main accelerator ring at the Fermi National Accelerator Laboratory (Fermilab); the ring will eventually accommodate countercirculating beams of protons and antiprotons. According to Leon M. Lederman, director of Fermilab, a single beam of protons reached a world-record energy of 700 GeV (700 billion electron volts) in August, and work is now focused on raising the luminosity, or intensity, of the beam and on extracting it for bombardment of a fixed target. The Tevatron I is to be tested and operated as a fixed-target machine until the summer of 1985; in other words, the high-energy protons will be directed at matter that is at rest instead of against particles in a countercirculating beam. Thereafter the machine is to be run as a colliding-beam accelerator. The energy released by the collision of protons and antiprotons is expected to reach 2 TeV.

The Tevatron I is the major competitor of the Super Proton Synchrotron, the proton-antiproton collider at CERN that was responsible for the discovery of the *W* and *Z* particles. Indeed, once the Tevatron I begins to operate at full power as a colliding-beam machine, which Lederman expects by the summer of 1986, it will surpass the energy and luminosity capabilities of the CERN accelerator.

Accordingly there is much interest at CERN in carrying out as many new experiments as possible before the Tevatron is ready. The energy capability of the CERN machine will be raised slightly to 310 GeV for each beam, and a major effort will be made to increase the luminosity of the beam. The current plan calls for the construction of a new collector for antiprotons, which would cost about $20 million and would increase the luminosity of the antiproton beam by a factor of 10. These improvements would guarantee a much higher production rate of the *W* and *Z* particles than is now possible.

In September ground-breaking ceremonies were held for another machine at CERN called the Large Electron-Positron collider (LEP); the LEP should generate about 10,000 vector bosons per day when it begins operation in 1988. The main tunnel in the LEP collider will be 27 kilometers long, and in its first phase of operation the machine will enable electrons and positrons to collide at energies of up to 60 GeV per particle. Eventually that energy will be raised to 100 GeV. Although the energy generated by the collisions is far less than that generated by the heavier protons and antiprotons, a large fraction of the electron-positron energy becomes available for the creation of new particles. (Because the proton and the antiproton are made up of quarks and gluons, the energy supplied to the proton and the antiproton must be divided among several elementary particles.) There is no firm theoretical prediction for the mass of the Higgs particle, and so the LEP could well find the particle before the Desertron is built. Moreover, there is discussion at CERN that the LEP tunnel could later be converted into a proton ring, where energies on the order of 8 TeV could be achieved.

Until the LEP is built much of the physics generated by electron-positron interactions will be carried on at two U.S. laboratories. The HEPAP subpanel has also recommended the upgrading of the Cornell Electron Storage Ring and the rapid construction of the Stanford Linear Collider. The latter, to be built at the Stanford Linear Accelerator Center (SLAC), will accelerate electrons and positrons along the track of the existing three-kilometer linear accelerator to 50 GeV. The two beams are then to be split by an electromagnet and sent in opposite directions halfway around a ring, where they will collide. Construction is now under way on the $113 million project and should be completed by the end of 1986. The machine will generate copious numbers of *W* and *Z* particles, and it will also test the linear accelerator as a cheaper alternative to the well-established storage ring.

At the Deutsches Elektronen Synchrotron (DESY) in Hamburg construction is to begin early next year on a device called the Hadron-Electron Ring Accelerator (HERA). The name points to the distinctive feature of the accelerator: it will be the only machine in which a hadron, such as the proton, is made to collide with a lepton, such as the electron. The design will enable physicists to closely examine the effects of the weak nuclear force without having to account for large effects of the electromagnetic force or the strong nuclear force, which can obscure the data collected from other machines. One major goal of the HERA machine is to penetrate the field surrounding the quark to within 10^{-17} centimeter. If the quark has a finite size, the size may become evident at that scale. It is left to the more powerful Desertron to determine the shape and structure of the quark, if they exist. The HERA is expected to go into experimental service in 1990.

At the 12th International Conference on High Energy Accelerators held at Fermilab in August physicists heard several reports on the development of accelerators in China, in the U.S.S.R. and in Japan. Outside Beijing a 3-GeV electron-positron collider is being constructed that should go into operation in 1987. The Russian laboratory at Serpukhov is building a 400-GeV proton accelerator called UNK, scheduled for completion in 1988. The accelerator is to be upgraded in energy to 3 TeV in the 1990's, and a second beam of 3-TeV protons is eventually to collide with the first. The Japanese project Tristan, started in 1981, is to be finished by 1986. A beam of 30-GeV electrons will collide with a beam of 30-GeV positrons.

Unruly Child

An El Niño is an appearance in the central and eastern equatorial Pacific of anomalously warm water. El Niños occur at irregular intervals of three to 10 years in conjunction with a large-scale fluctuation in atmospheric pressures known as the Southern Oscillation. The consequences of these coupled phenomena are often severe, ranging

Ford Escort Diesel: Better mileage than this leading import.

We didn't believe it at first, either.

But EPA testing figures established it. Our new Escort Diesel is rated approximately four miles per gallon higher than a Honda 750.

Just take a look at our numbers:

46 EPA EST. MPG.

68 EST. HWY.

And because this diesel is an Escort, there's a lot more to talk about than great economy.

Like the fact that Escort's the best-selling car in the world.**

Or that it comes with more total passenger room and more total cargo room than a Honda Accord.†

More standard features than a Toyota Tercel.††

There's even a fully-independent suspension system for a smoother ride than a Nissan Sentra.

All of which means Ford Escort not only gives you a big advantage over that motorcycle pictured above.

It also beats more than its share of cars.

THE BEST-BUILT AMERICAN CARS.

When we say "Quality is Job 1," we are talking about more than a commitment. We are talking about results. An independent survey concluded Ford makes the best-built American cars. The survey measured owner-reported problems during the first three months of ownership of 1983 cars designed and built in the U.S.

And that commitment continues in 1984.

* For comparison. Honda 750 mileage is obtained from EPA emissions testing and is *not* an official rating. Your mileage may vary depending on speed, trip length, weather. Actual highway mileage lower. Escort Diesel mileage applicable to sedans with FS engine and without power steering and A/C. Not available in California.

** Sales estimates based on world-wide production figures.

† Based on EPA Interior Volume Index.

†† Escort GL (shown) compared to Toyota Tercel 3-door deluxe liftback.

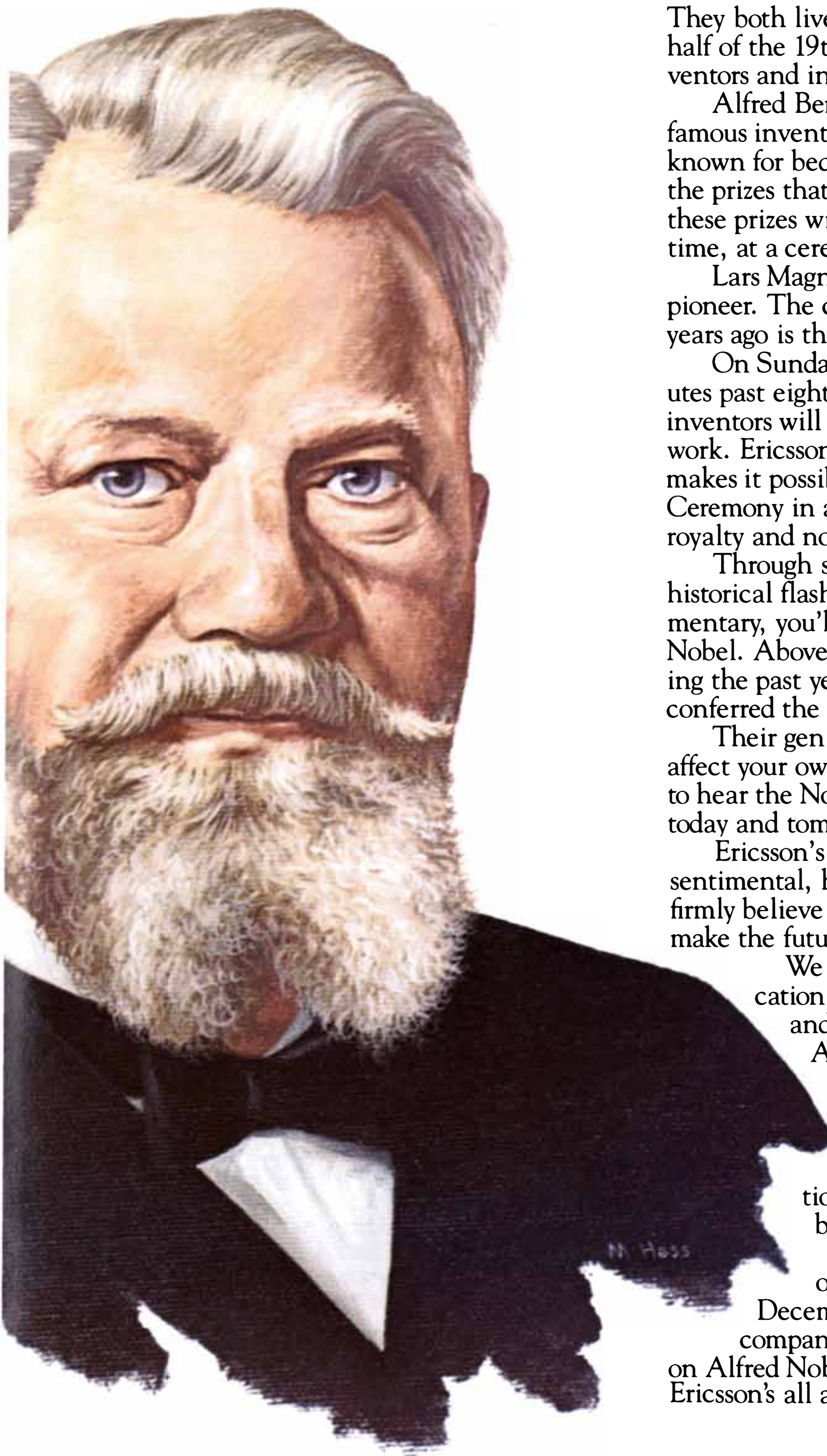
Get it together – Buckle up.

Have You Driven A Ford... Lately?

Ford

A. B. Nobel, 1833–1896

ONE MAN'S NAME IS ON EVERYBODY'S LIPS. THE OTHER ONE'S SOON WILL BE.



L.M. Ericsson, 1846–1926

They both lived in Stockholm during the second half of the 19th century. Both were prominent inventors and industrialists.

Alfred Bernhard Nobel was a chemist. His most famous invention was dynamite. But he is best known for bequeathing his vast fortune to institute the prizes that bear his name. On December 10, these prizes will be awarded again, for the 82nd time, at a ceremony in Stockholm's City Hall.

Lars Magnus Ericsson was a telecommunications pioneer. The company he started over a hundred years ago is the world's fourth largest in its field.

On Sunday, December 11, at exactly five minutes past eight in the evening (EST), these two inventors will meet again—on WTBS Cable Network. Ericsson's role as a Nobel Fund contributor makes it possible for you to view the Nobel Awards Ceremony in all its pomp and circumstance, with royalty and nobility in attendance.

Through satellite coverage, interspersed with historical flashbacks and up-to-the-minute commentary, you'll get to know the enigmatic Alfred Nobel. Above all, you'll meet the people who, during the past year, to quote Nobel himself, "have conferred the greatest benefit on humankind."

Their genius and their insights will profoundly affect your own future. Don't miss this opportunity to hear the Nobel Laureates' views on the issues of today and tomorrow.

Ericsson's contribution to this event is not for sentimental, historical reasons, but because we firmly believe that science and technology can help make the future better for all of us.

We have been helping to make communication between individuals, organizations and nations easier ever since the days of Alfred Nobel. Today, we're known as the pioneer in the digitalization of the global telephone network.

And the more we innovate in areas like data processing, information systems and cellular radio, the better you, too, will get to know us.

Watch the Nobel Prize Ceremony on WTBS Cable Network on Sunday, December 11. And write to us, on your company letterhead, for a free 12-page booklet on Alfred Nobel and his work—and a little on what Ericsson's all about.

ERICSSON

ERICSSON, INC., GREENWICH OFFICE PARK 3, GREENWICH, CT 06836-3110 from droughts in Australia and South Africa to devastating rains and sharply reduced fish catches in Ecuador and Peru. The 1982–83 El Niño has been the strongest in this century: typically, sea-surface temperatures off the South American coast rise by two to three degrees Celsius during an El Niño, but the anomaly reached a peak of seven degrees in June. The anomaly was also anomalous in that it began later in the year than previous events and did not announce itself with the usual atmospheric signal.

During a Southern Oscillation sea-level air pressure subsides in the high-pressure system of the southeastern Pacific and rises in the low-pressure area normally over Indonesia and northern Australia. The result is a decrease in the pressure gradient that drives the easterly trade winds along the Equator. Under normal conditions these winds and the westward equatorial current pile up very warm water (29 degrees C.) in the western Pacific. When the trade winds slacken, the warm water can slosh back eastward, leading to anomalously warm sea-surface temperatures in the central Pacific. The wind shift also appears to trigger a series of subsurface waves, called Kelvin waves, that propagate along the Equator as vertical displacements of the thermocline (the boundary between the warm upper water of the ocean and the cold water below it). The Kelvin waves depress the thermocline and thereby thicken the warm surface layer.

It is the latter phenomenon on which oceanographers now rely to account for the appearance, sometimes with disastrous ecological and economic consequences, of unusually warm water off the South American coast. Normally the southeast trade winds blowing along the coast push surface water out to sea, allowing cold water from below the thermocline to rise to the surface. After the Kelvin waves hit the coast, however, the sea level rises and the warm surface layer gets so thick that the water welling up is warm, not cold. Moreover, it does not contain the rich supply of nutrient salts that the cold water normally brings to the surface and that phytoplankton require for photosynthesis. The resulting decline in plankton levels is felt all the way up the food chain, ending with the sea birds and fishermen who work the usually bountiful sea.

Previous El Niños had been preceded by a buildup of the easterly trade winds. Following the drop in the winds in the fall, warm water tended to appear first off South America just after Christmas, in January or February (hence the term El Niño, literally "the little one" but referring to the Christ child). The peak of the anomaly came in May or June. This event was different: the trades never peaked before slackening, and they did not slacken until April, 1982. Warming of the sea surface began in the central Pacific in the summer and in the eastern Pacific in September.

Investigators had come to regard a buildup of the trade winds as being essential for the piling up in the western Pacific of warm water that would then slosh back eastward during the El Niño. They are now reconsidering this concept. The odd timing of this year's event is also a puzzle, but according to Mark A. Cane of the Massachusetts Institute of Technology it need not be. Cane suggests that one of the first things to happen in an El Niño is a movement of the massive Indonesian low-pressure system eastward along the Equator toward the central Pacific. In previous events such movement has tended to come in the fall, but there is no reason it should not come in the spring, as it did last year; in both seasons the low-pressure zone is less powerfully centered, more unstable and therefore more mobile.

Why the Indonesian low should move in some years and not in others is not known. Once it starts moving, however, a feedback mechanism could keep it going and eventually produce an El Niño, according to Eugene Rasmusson of the National Meteorological Center. Like other lows, the Indonesian low is a convergence zone, where air is heated and rises and is replaced by winds flowing into the zone. Convergence zones are generally over areas where the surface is warmest and evaporation and upward motion of the air are therefore greatest.

When the Indonesian low is in its normal position over Indonesia, the inflow into the system over the western Pacific is provided by the easterly trade winds. As the low moves eastward toward the International Dateline in the middle of the Pacific, however, it draws air in behind it from the west, creating anomalous westerlies that damp the trade winds. Now the warm pool of water in the western Pacific can begin to move eastward, taking with it an area of heavy rainfall. The precipitation warms the atmosphere, intensifies the westerlies (in 1982–83 the trade winds actually reversed direction in the western Pacific) and draws the low-pressure zone farther east. Meanwhile the signal that the winds have shifted in the west is propagating along the Equator toward South America, at a speed of several hundred kilometers per day, as a series of Kelvin waves.

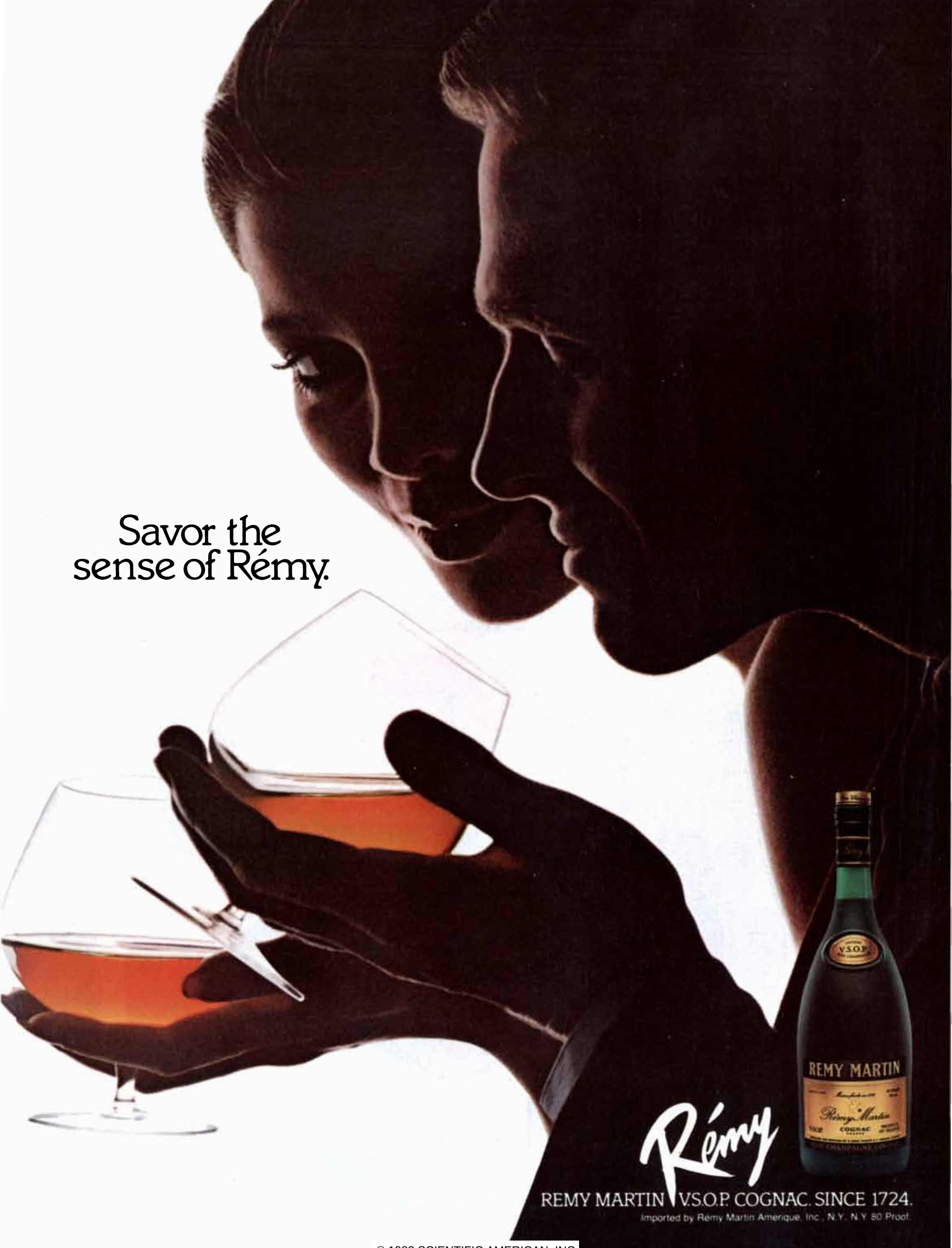
How is this signal transmitted to the mid-latitudes of the Northern Hemisphere? When an equatorial El Niño occurs, almost invariably a mid-latitude El Niño produces smaller sea-surface temperature anomalies (two to three degrees C. in 1982–83) off the North American coast. Although the present equatorial El Niño appeared to be fading away this past fall, the mid-latitude anomaly, which has reached as far north as British Columbia and Alaska, is if anything showing signs of getting stronger. That the two events are connected seems obvious, but the nature of the connection is not.

According to one hypothesis, the equatorial Kelvin waves split in two when they hit South America and travel north and south along the coast, carrying warm water into the mid-latitudes and depressing the thermocline as they go. James J. Simpson of the Scripps Institution of Oceanography reports, however, that the warm water reaching the California coast during the present El Niño contains too little salt and too much dissolved oxygen for it to be coming from the south. Detailed measurements made by Simpson and his colleagues along a line extending southwest from San Diego indicate instead that the warm water comes primarily from the west and is being driven onshore by the same anomalous southwesterly winds that battered California with storms all last winter. These results suggest a primarily atmospheric rather than oceanic "teleconnection" between the equatorial and mid-latitude El Niños.

Arthur Douglas of Creighton University offers a possible scenario for such a teleconnection. In normal years the low-pressure zone near the Equator where the northeast and southeast trade winds meet, called the Intertropical Convergence Zone (ITCZ), migrates north in May or June. During an El Niño, for reasons that are not entirely clear, the ITCZ stays south. As a result the upper-level high engendered in the subtropics by excessive air rising from the ITCZ stayed over Mexico in 1982, producing hot, dry weather that allowed the Pacific to warm up as far as 700 miles offshore.

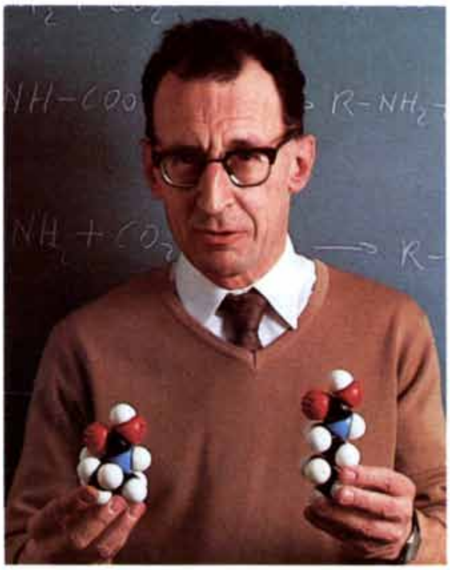
Heat energy from this pool of warm water then intensified the southwest winds flowing into the upper-level low-pressure trough that often lodges over California in September. These winds drove warm water onto the coast and prevented the normal upwelling of cool water. By winter the warm anomaly had spread to British Columbia as the onshore winds associated with a low-pressure system over the Gulf of Alaska intensified. The forces driving this last process are also not well understood, but they are assumed to be related to the large temperature contrast between the wide band of El Niño water at the Equator and a pool of anomalously cold water observed this year in the central Pacific.

Douglas thinks a similar sequence of events has taken place in previous El Niños, but with differences that account for the vastly different effects the events have had on North American weather. During the El Niño of 1976–77 California had a drought and the winter in the U.S. was unusually cold. The current


Savor the
sense of Rémy.
REMY MARTIN
Rémy
REMY MARTIN V.S.O.P. COGNAC. SINCE 1724.
Imported by Remy Martin Amerique, Inc., N.Y., N.Y. 80 Proof.

How Exxon developed can double the productivity

Guido Sartori's work on hindered amines may impact an entire industry.



Removing impurities such as carbon dioxide and hydrogen sulfide from natural, refinery, and synthesis gases is an expensive, energy-consuming process.

But at Exxon Research and Engineering Company a new chemistry discovery, and cross functional teamwork, have led to the development of a new technology—one that significantly decreases the cost and increases the capacity of commercial gas treating processes.

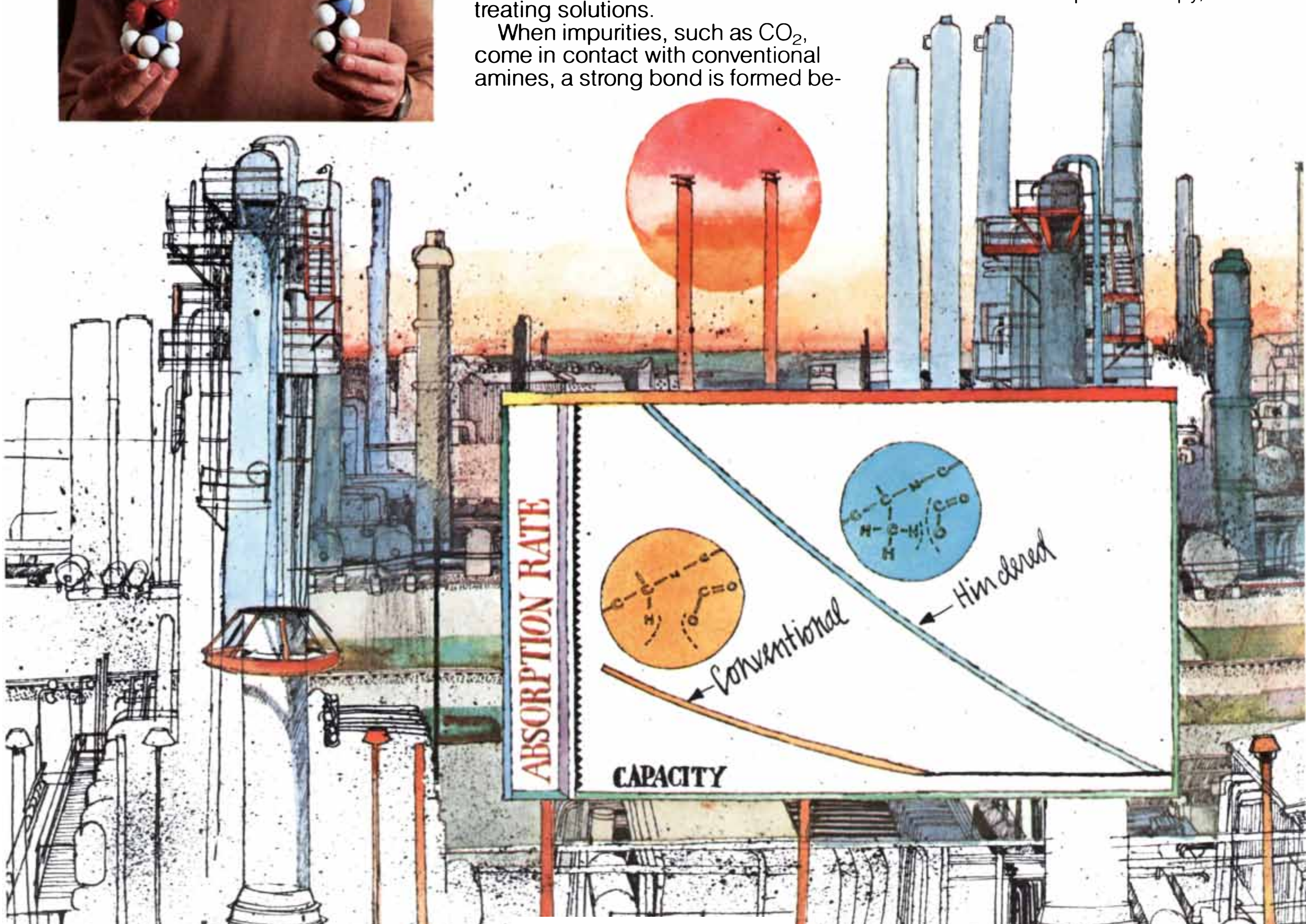
Research Led to a Discovery

Guido Sartori, a chemist in Exxon Research and Engineering Company, had been conducting research on amines—organic nitrogen-containing molecules—to increase both the absorption rate and capacity of gas treating solutions.

When impurities, such as CO_2, come in contact with conventional amines, a strong bond is formed between the CO_2 and the nitrogen atom of the amine. This strong bond ties up a disproportionate amount of useful amine. Sartori theorized that both the absorption rate and capacity of the amine would be improved if the bond at the nitrogen site could be weakened. Continuing research revealed the advantages of a whole new class of amines, which he called hindered amines.

Observing Molecular Behavior

Sartori and others began a comprehensive evaluation of the discovery, utilizing the company's advanced analytical capabilities. To understand the behavior of hindered amines, and to monitor reactions, Sartori employed the results of carbon-13 nuclear magnetic resonance spectroscopy, a



new molecules that of gas treating plants.

state-of-the-art technique not previously used for this purpose.

Further research confirmed the hindered amines' capability to substantially increase the rate and capacity of carbon dioxide absorption through the formation of low stability bonds. Low stability was achieved by placing a bulky substituent next to the nitrogen sites, thereby hindering bond formation with CO_2. Building on this new understanding, he synthesized new molecules to meet the performance requirements for specific applications.

Integrated Innovation

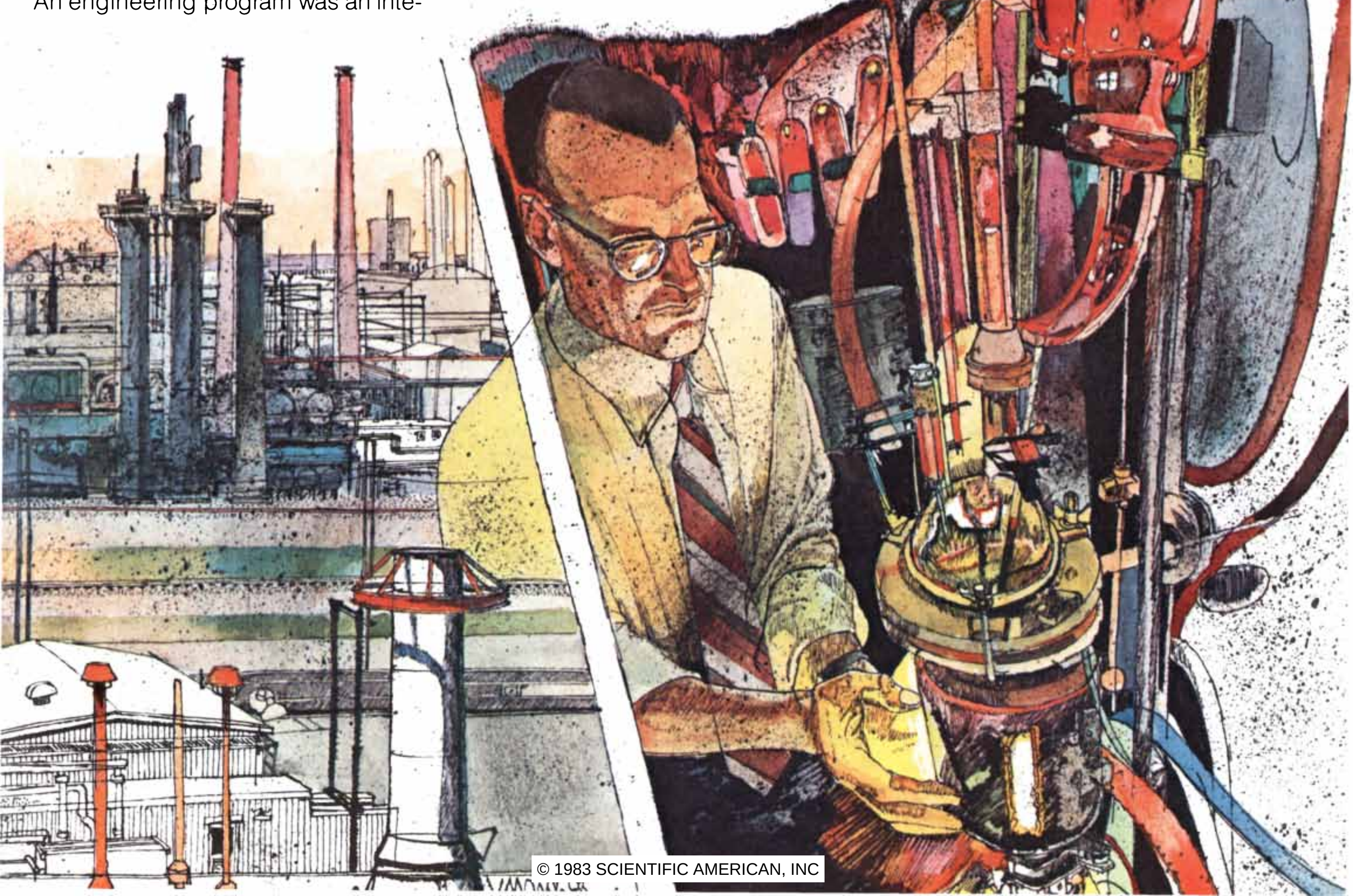
Other Exxon organizations joined the effort to develop improved gas treating technology. After the hindered amines had been evaluated at the laboratory bench, process development was required on a larger scale. A major pilot plant program confirmed, broadened and extended the bench scale results and helped to define the capabilities of the hindered amines. An engineering program was an integral part of the research and development required to convert these laboratory discoveries into commercially feasible technologies. Capacity increases of 50% have been achieved commercially using this technology with no added facilities.

Through integrated innovation–the combined efforts of the company's basic research, process development, and engineering staffs–hindered amine technologies advanced from scientific discovery through commercial use in less than three years. Further research has enabled ER&E to identify or synthesize other practical hindered amines.

Exxon Research and Engineering Company

Research on hindered amines is just one example of the numerous programs underway at ER&E. A wholly owned subsidiary of Exxon Corporation, ER&E employs some 2,000 scientists and engineers working on petroleum products and processing, synthetic fuels, pioneering science and the engineering required to develop and apply new technology in the manufacture of fuels and other products. For more information on Exxon's hindered amine technology or ER&E, write Dr. E. E. David, President, Exxon Research and Engineering Company, Room 705, P.O. Box 101, Florham Park, New Jersey 07932.

EXXON



event produced downpours in California and may account for the mild winter of 1982–83 in most of the U.S.

Earlier Americans

When did men out of Asia first enter the New World? One position on the question, held by many conservative-minded archaeologists, is doubtful of any date earlier than some 12,000 years ago. Another position, held by a few venturesome workers, rests on a single California site: Calico, an ancient alluvial fan in the Mojave Desert. Geomorphologists consider the fan to be from one to two million years old, but proponents of Calico as an early-man site suggest that fractured stones from the lower part of the fan are manmade artifacts some 200,000 years old.

Richard S. MacNeish of Boston University reviews the question in the light of archaeological findings, including his own near Ayacucho in Peru. Writing in *The L. S. B. Leakey Foundation News*, he addresses his main argument to his conservative colleagues: the proposal that man pushed eastward across the Bering Strait as recently as 12,000 years ago must be dismissed if earlier evidence of immigration exists in South America.

MacNeish points out that just such dates have been established by carbon-14 analyses of material from the lower levels of one of his Ayacucho sites, Pikimachay Cave. Two separate analyses from the cave's lowest man-occupied stratum yielded respective dates of 19,600 and 20,200 B.C. The dates of three overlying strata are 16,100, 14,750 and 14,100 B.C. MacNeish goes on to list other South American findings: (1) at Los Toldos in southern Argentina stone tools have been found in levels older than 10,650 ± 600 B.C., (2) at the Piaui Caves in Brazil flake tools older than 16,650 ± 600 B.C. have been found in lower levels, (3) at Monte Verde in Chile tools have been found with mastodon bones dating to about 12,000 B.C. and (4) at Taima-Taima in Venezuela a projectile point has been found embedded in a mastodon pelvis; although its carbon-14 age has not been published, the date is estimated at 14,000 B.C.

Turning to North America, MacNeish points to Valsequillo and Tlapacoya in Mexico. At Valsequillo tools in association with the bones of extinct animals are estimated to range in date from 20,000 to nearly 40,000 B.C. At Tlapacoya hearth charcoal with a carbon-14 age of 23,150 ± 500 years before the present is associated with both artifacts and the bones of extinct animals.

Commenting on the Los Toldos site in Argentina, the one with a date closest to 12,000 years ago, MacNeish remarks: "No matter how fast man ran the length of [South America] he had to come across Bering Strait well before 12,000 years ago in order to drop these tools at this date." He goes on to report a number of recent defections from the conservative ranks, perhaps most notably that of C. Vance Haynes of the University of Arizona. Haynes now places the date of the first New World immigration at more than 30,000 years ago.

As for the Calico site enthusiasts, either of two factors may weaken their position. If, as geomorphologists assert, the alluvium was deposited more than a million years ago, no archaeologist would suggest that the deposit contains human handiwork. If, on the other hand, the deposit is only a fraction as old, it still remains to be established whether the Calico objects are crude stone tools worked by man or stones worked by natural agencies. At present arguments are offered on both sides of the question.

Forests of the Night

The hallmark of a quasar is a set of spectral lines prodigiously out of position. The Lyman lines are the best example. They consist of the photons (the quanta of light) emitted by electrons in hydrogen atoms when they "fall" from excited states to the ground state: their state of minimal energy. In the laboratory the Lyman lines are found in the ultraviolet; Lyman alpha, the most prominent line in the set, has a wavelength of 1,216 angstrom units. Yet in the light from a typical quasar the wavelengths of the Lyman lines are several times longer: they lie in the visible part of the spectrum. In the most widely accepted hypothesis quasars are taken to be "engines" at the core of newly formed galaxies extremely distant in space and in time, and the systematic displacement of the lines in the light each one emits is taken to be a red shift: an increase in wavelength indicative of the quasar's recession from the earth as part of the general expansion of the universe.

In an effort undertaken by Wallace L. W. Sargent of the California Institute of Technology and Alec Boksenberg of University College London some Lyman-alpha lines in the light emitted by quasars are significant in quite a different way. They are not emission lines. They are absorption lines, that is, they are dips, not peaks, in the spectrum. They are caused when the electrons in hydrogen atoms pluck light from a quasar's broadband radiation and are excited from the ground state into a state of greater energy. In the light from a single quasar these Lyman-alpha absorptions can number more than 100. Each has a different wavelength longer than 1,216 angstroms but shorter than the red-shifted Lyman-alpha emission of the quasar itself. Sargent and Boksenberg propose they represent the absorption of quasar light by clouds of hydrogen at different positions along the line of sight to the quasar. Each cloud would be receding from the earth at some velocity less than that of the quasar, and so its hydrogen atoms would absorb the quasar's light at a distinctive red-shifted value. The resulting "forest" of Lyman-alpha absorptions means that the light of the quasars can serve as a probe of the hydrogen clouds in intergalactic space.

Sargent and Boksenberg approach the Lyman-alpha forest by removing everything else. Some of the absorption lines they encounter in the light of a typical quasar are relatively prominent and are red-shifted from their laboratory wavelengths by only a small amount. Such lines are attributed to the absorption of quasar light by elements such as carbon, oxygen, silicon, iron and magnesium in the gas and dust that permeate our own galaxy. Other lines are less prominent; they turn out to be absorption lines due to hydrogen accompanied by absorption lines due to heavier elements, all moved from their laboratory wavelengths by an identical red shift. They are attributed to the absorption of quasar light by galaxies other than our own along the line of sight to the quasar. Still other lines are hydrogen lines at a multitude of red shifts. They make up the Lyman-alpha forest. They prove not to be accompanied by lines due to heavier elements; thus Sargent and Boksenberg calculate that the abundance of heavy elements relative to the abundance of the hydrogen can be no greater than a thousandth the abundance of heavy elements in the sun. Evidently the clouds consist of primordial hydrogen that has never condensed into stars.

The properties of the clouds are beginning to emerge. For one thing, Sargent and Boksenberg note that the light from quasars that have small angular separations in the sky can reveal the size of the clouds. They have already found that the Lyman-alpha forests in the spectra of 10-billion-year-old quasars one minute of arc apart are different from one another. They conclude that the clouds are probably smaller than .5 megaparsec, roughly the size of a cluster of galaxies. More generally, Sargent and Boksenberg report that the lines in a Lyman-alpha forest show no grouping at particular red shifts; hence the clouds do not group together in space. In this they differ from galaxies. On the other hand, the lines are more numerous at greater red shifts; hence the ratio of clouds to galaxies increases with increasing distance in space and in time. That is, the clouds were more numerous in the early universe. What made their numbers decrease? Not gravitational collapse, or else stars would have formed. Sargent and Boksenberg suspect the clouds are held together by the pressure of the surrounding intergalactic medium, a hydrogen gas whose atoms are ionized by the photons streaming from quasars.

BMW'S ANSWER TO PROGRAMMED OBSOLESCENCE: PROGRAMMED RELEVANCE.

There's a simple reason why some cars refuse to become outdated, while others do it all too willingly.

The difference lies in how an automobile manufacturer applies technology—whether for purposes that are substantial and enduring, or simply to bestow upon cars an annual patina of newness.

Nowhere is this distinction made more vivid than in the BMW 528e.

It is built by a company that developed its first independent suspension in 1932. That began exploring aerodynamics as early as 1923. And that has been incorporating fuel injection into its cars for over a decade.

These and other indications of BMW's foresight are abundantly apparent in the 528e.

Its computer-governed Eta engine was designed to reconcile performance and efficiency. It won't have to be rethought the next time the world convulses between energy shortage and energy glut.

Its Service Indicator tells the driver when routine service is needed. It will remain relevant for as long as cars require maintenance.

And its easily-readable instrumentation provides a wealth of vital data, including seven different measures of the car's operational readiness, along with accurate fuel-mileage figures. None of it was inspired by a passing infatuation with video games, but by the timeless need of drivers to receive and instantly comprehend information.

In short, the BMW 528e has been programmed to provide a service that most other cars can't:

The assurance that the future will come as a fulfillment, not a shock.

BMW

THE ULTIMATE DRIVING MACHINE.

Who on earth will install enough lightguide fiber to circle the earth 10 times?

Western Electric, an AT&T company.

In just five years, Western Electric has already created and installed enough lightguide fiber to circle the earth. That's 25,000 miles of fiber in 60 lightwave communications routes.

Those totals should increase at least tenfold by the year 2000, as Western Electric continues to install lightwave systems to transmit the high-speed voice and data services of the Information Age.

Far more efficient and economical than copper cable, lightwave communications speeds your messages over 90 million light pulses per second through hair-thin glass fibers. A single lightguide cable can carry 240,000 telephone calls at once.

Right now Western Electric is installing the world's largest lightwave system in the busy Northeast Corridor. Another will transmit the entire 1984 Los Angeles Summer Olympics to ABC television studios.

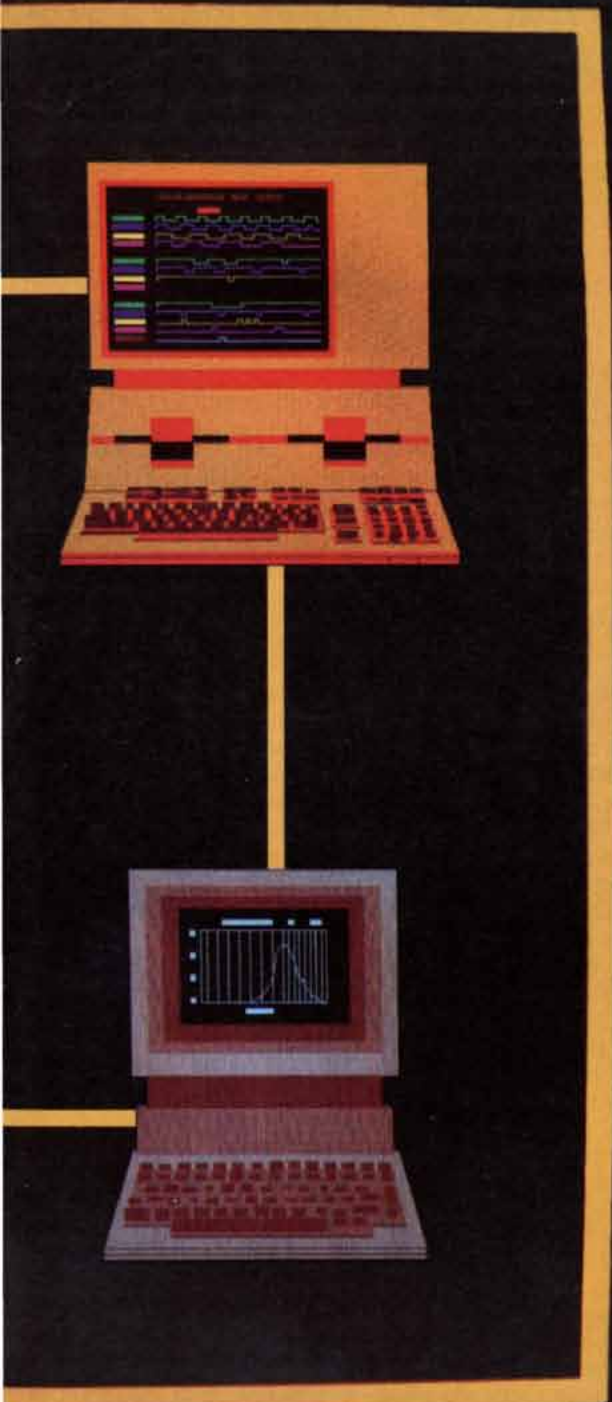
And our research and development arm, Bell Labs, is working on a transatlantic lightwave cable that will triple calling capacity and enhance transmission quality.

These systems are forerunners of a continually expanding lightwave network that one day will literally circle the earth, bringing everyone exciting new information services.

Western Electric, an AT&T company. Applying the technologies of lightwave, microelectronics, and software to make the dream of the Information Age a reality.

AT&T

oductivity Network.



Presenting your broadest choice of engineering workstations anywhere. From a low-cost personal computer to a powerful 32-bit system.

No doubt about it; when you put a computer on every engineer's desk, it helps them be more productive.

In fact, at Hewlett-Packard, the workstation approach has worked so well for our own engineers that we've taken it to an entirely new dimension.

It's called the Engineering Productivity Network. As you can see, it includes the broadest range of workstations available anywhere. From the recently-introduced 32-bit HP 9000 desktop mainframe to our personal computer designed just for engineers.

With such a wide range to choose from, you can match the computing power to the individual task. And give every engineer the tool it takes to do the best possible job.

To help you tailor every workstation to a task even more closely, we have over 200 specific application software packages. And since our workstations use HP-UX—our enhanced version of the UNIX* operating system—we'll be able to bring you a lot more in the future.

Of course, when you tie all these individually-optimized workstations together into a network, you can look for an even higher level of performance. Whether you choose our HP-UX networking, Ethernet local area network, or Shared Resource Manager, your engineers can share data and peripherals, communicate much better, and get more done than ever before.

The HP Engineering Productivity Network. You can start with one workstation, or with dozens. But no matter where you start, you're actually making your whole department more productive. One step at a time.

To find out more, just call the nearest HP office listed in the white pages of your telephone book, and ask for a Technical Computer Representative. Or write to Pat Welch, Dept. 41189, Hewlett-Packard, 19447 Pruneridge Avenue, Cupertino, CA 95014. In Europe, write Henk van Lammeren, Hewlett-Packard, Dept. 41189, P.O. Box 529, 1180 AM, Amstelveen, the Netherlands.

We'll be happy to show you how far you can really go with a good idea.

*UNIX is a trademark of Bell Laboratories.

HEWLETT PACKARD

BD-02316

The DNA Helix and How It Is Read

X-ray analysis of crystals of three types of double-helical DNA molecule leads to the realization that base-sequence information can be stored in the local structure of the helix

by Richard E. Dickerson

Two kinds of genetic information are carried in the double helix of DNA, and they are encoded and interpreted in very different ways. The genetic message itself—the information specifying the structure of proteins—is written in the familiar triplet genetic code. That code is linear, and its interpretation is extrinsic. The sequence in which successive triplets of the chemical groups called bases are arrayed along a single strand of the helix encodes the sequence in which successive amino acids are linked to form a protein chain. There seems to be no innate structural relation between a particular base triplet and the amino acid it specifies. Rather, information transfer is mediated indirectly by complex machinery extrinsic to the DNA: messenger RNA, ribosomes, transfer RNA and a battery of attendant enzymes.

DNA encodes not only the genetic message itself but also instructions for the selective expression of the message. What little is known about this kind of information comes from studies of genetic control in bacteria. Blocks of genes on the bacterial chromosome are turned off by the binding of a repressor protein to a region of DNA (the operator) that has a special base sequence the repressor recognizes; the genes become active when various small molecules bind to the repressor and knock it off the operator. Although much remains to be learned, most molecular biologists expect comparable mechanisms to be the basis of genetic control in life forms higher than bacteria. A bacterial repressor recognizes its operator directly, probably by forming hydrogen bonds between nitrogen and oxygen atoms on its amino acids and on the edges of the complementary pairs of bases in the double helix of DNA. The helix fits snugly against a particular conformation of the protein. The information specifying genetic control, then, is three-dimensional, and its interpretation is intrinsic: it depends on the implicit structural properties of the protein and the helix.

Does DNA play only a passive role in the recognition process, with amino acid side chains inserting themselves into the grooves of a static double helix, or does the base sequence itself modify the structure of the helix in a way that contributes to the recognition of specific sites by the control proteins? As recently as four years ago the question would have been pointless because it was totally unanswerable from the type of data then available.

James D. Watson and Francis Crick proposed their double-helix structure for DNA in 1953 on the basis of X-ray-diffraction photographs made by Rosalind Franklin and Maurice Wilkins. The photographs were of fibers of DNA. Further analysis of such diffraction patterns led to the realization that fibers of DNA can exist in two forms: as *B* DNA under conditions of high humidity and as *A* DNA when the humidity is lower. Molecular models were developed for both forms. The amount of structural information that could be extracted from fiber diffraction patterns was inherently limited, however, by the disorder of the DNA strands around and along the fiber axis. The most that could be established was an averaged, overall helix structure. Any local variations in structure, such as might be induced by a particular base sequence, could not be detected.

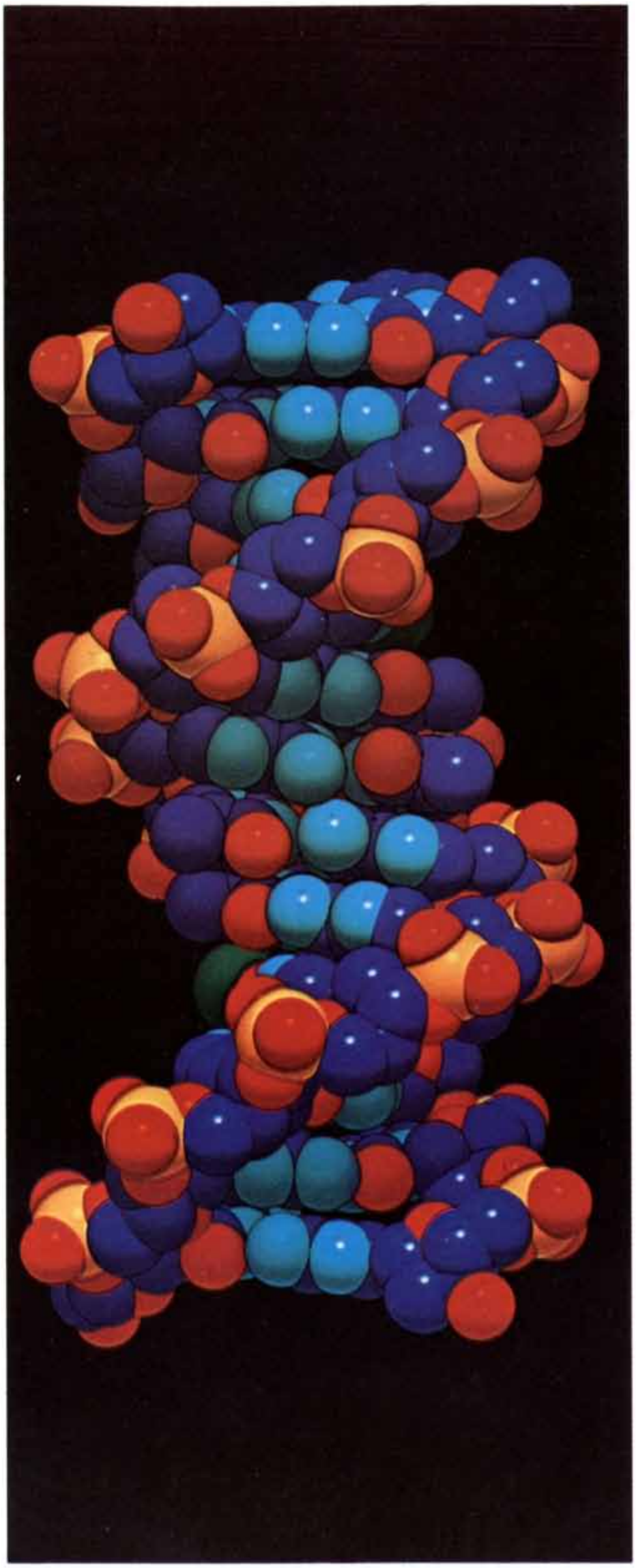
Now they can be detected and measured precisely. Improved methods for the organic synthesis of short DNA molecules of any desired sequence have made it possible for the first time to make DNA oligomers (short DNA chains) with from four to 24 bases in sufficient quantity and pure enough to be crystallized and studied by conventional single-crystal X-ray-diffraction methods. This article is intended as a window on some of the exciting new results beginning to come from the structural analysis of short double-helical DNA molecules.

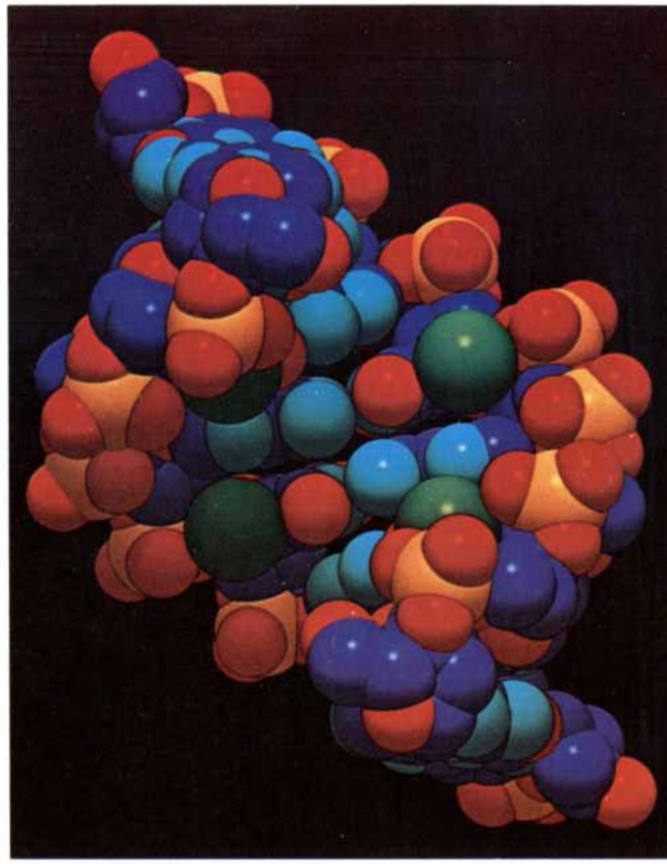
Fiber Diffraction

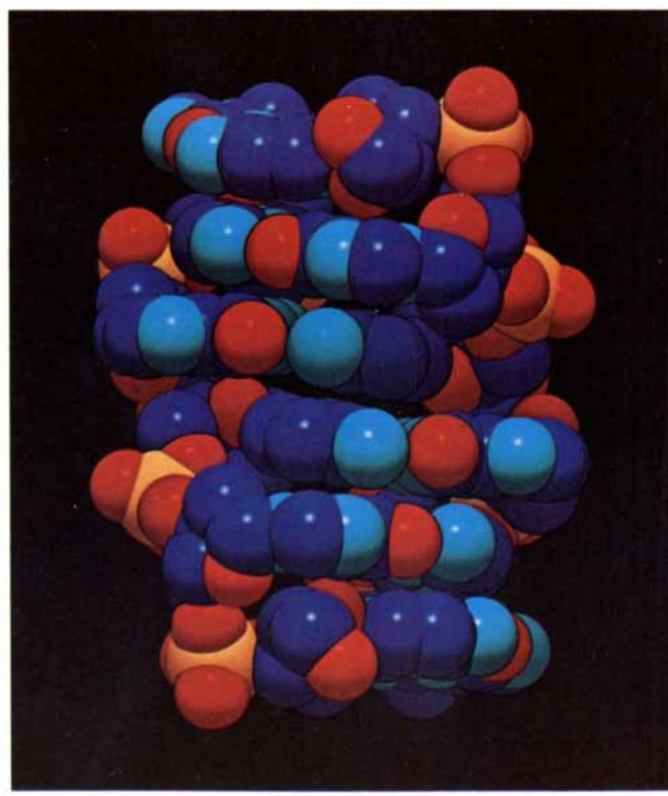
Examination of X-ray patterns from stretched fibers and thin films of natural DNA revealed the basic structure of the two fundamental types of double helix: the *B* configuration, which is stable at a relative humidity of about 92 percent, and the *A* configuration, which most base sequences assume when the humidity falls to about 75 percent. Both have the shape of a flexible DNA ladder wrapped helically around a central axis. The two rails of the ladder are chains of alternating deoxyribose sugar rings and phosphate groups. The rungs are purine-pyrimidine base pairs. There are two kinds of double-ring purine bases, adenine (*A*) and guanine (*G*), and two kinds of single-ring pyrimidines, thymine (*T*) and cytosine (*C*). The base pairs are held together by hydrogen bonds. Adenine normally is paired with thymine by two hydrogen bonds and guanine is paired with cytosine by three bonds.

The two points of attachment of a base pair to its sugar rings do not lie directly opposite each other across the pair [*see bottom illustration on page 97*], and this is important for the geometry of a DNA double helix. The edge of the base pairs along which the angle between attachments is less than 180 degrees is called the minor-groove edge, and the one forming an angle larger than 180 degrees is the major-groove edge. When base pairs are stacked into a helix, the phosphate backbones build the two walls of a major and a minor groove, which wind around the helix with these base edges as the floor of the grooves.

The floor of the major groove is paved with nitrogen and oxygen atoms that can make hydrogen bonds with the side chains of a protein's amino acids and therefore can have a significant role in







THREE TYPES OF DOUBLE HELIX are displayed in these space-filling models, which are computer images generated by Nelson Max of the Lawrence Livermore National Laboratory. They show three of the short double-helical molecules whose structures have been solved by single-crystal X-ray-diffraction analysis. Carbon atoms are dark blue, nitrogen atoms light blue, oxygen atoms red and phosphorus atoms yellow. Bromine atoms on some of the bases are green. *B* DNA (*left*) is represented by a 12-base molecule whose structure was solved by the author and his colleagues. The view is directly into the major groove of the helix. The phosphate backbone gives a strong sense of the right-handed twist. The eight-base molecule of *A* DNA at the top right is one solved by Olga Kennard, Zippora Shakked and M. A. Viswamitra. It too is right-handed. The view, into the very deep major groove, shows how the base pairs are tilted with respect to the vertical axis of the helix. The molecule of *Z* DNA at the bottom right is one whose structure was solved by Andrew H.-J. Wang, Alexander Rich and their colleagues. As the phosphate groups at the upper right and the lower left of the image indicate, the *Z*-DNA helix is left-handed. The backbone zigzags, and the base pairs are stacked by twos rather than singly, giving the helix an alternating structure.

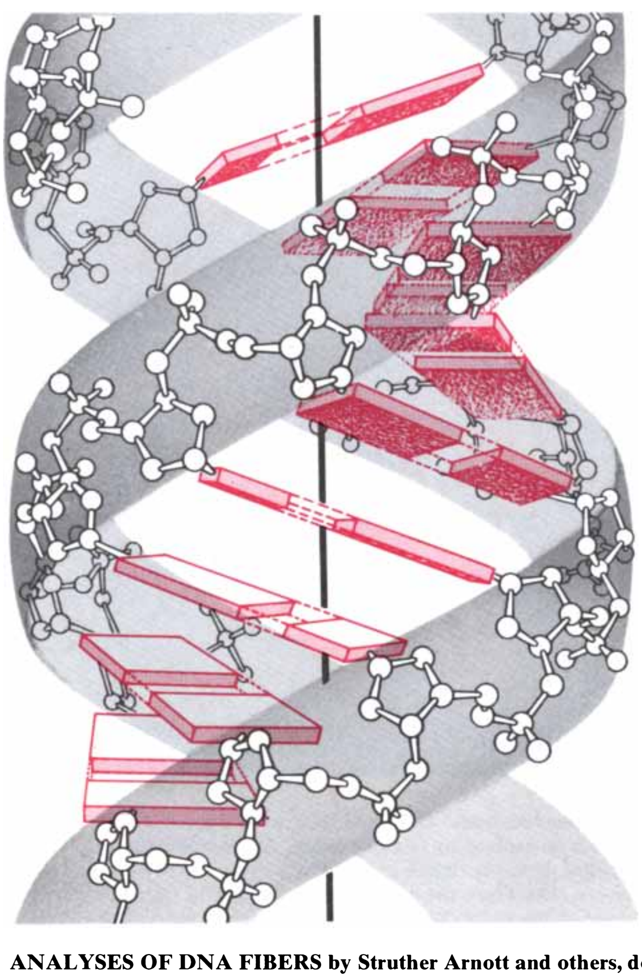
intrinsic coding. The pattern of these hydrogen-bonding groups is different for the two kinds of base pair. Reading from purine to pyrimidine, an *A–T* pair offers a nitrogen atom (a hydrogen acceptor), an NH_2 group (a donor) and an oxygen atom (another acceptor). In contrast, the *G–C* pair offers the same groups in a different order: first a nitrogen (an acceptor), then an oxygen (an acceptor) and finally an NH_2 (a donor). Since each base pair (*A–T* and *G–C*) can also be turned around (to become *T–A* and *C–G*), four different patterns can be exhibited to a repressor or another control protein at each step of the helix. The bottom of the major groove therefore carries a message, the base sequence of the DNA, in a form that can be read by other large molecules.

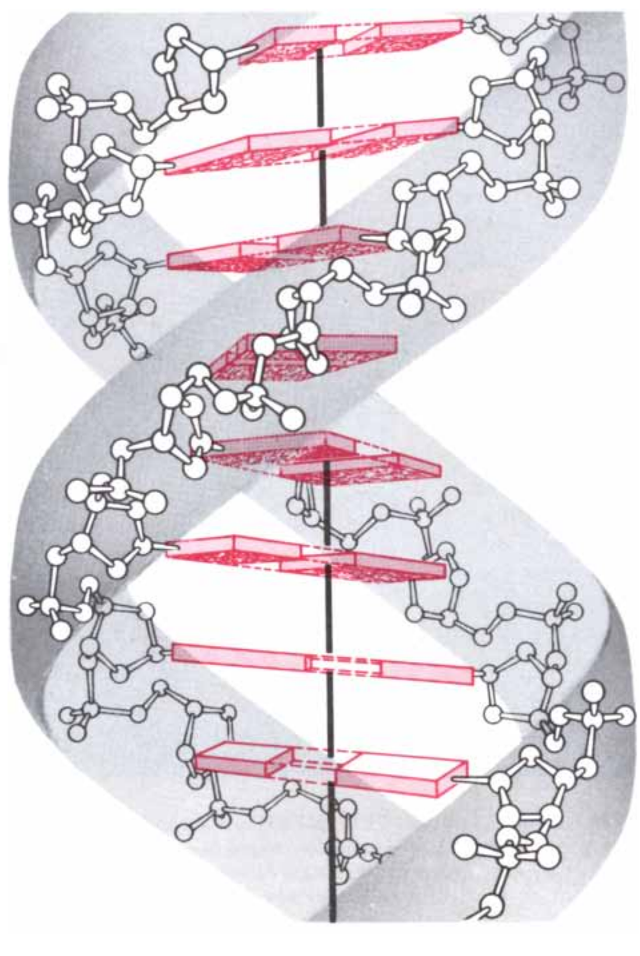
The minor groove is less informative: the hydrogen-bonding pattern in *A–T* base pairs is simply acceptor-acceptor no matter which way the base pair is flipped, and a *G–C* pair differs only by the intrusion of an NH_2 donor between the acceptors. This makes the minor groove a poorer candidate for information readout. As will emerge below, the minor groove has another important function in *B* DNA.

A DNA and *B* DNA differ mainly in the positioning of the base pairs around the helix axis and in base inclination: the tilt of the bases with respect to the axis. In *B* DNA the inclination is close to zero (the base pairs are stacked nearly perpendicular to the helix axis) and the axis runs through the center of each base pair. Although the minor groove is narrower than the major groove because of the base pairs' asymmetric attachment to the sugar rings, the grooves are of similar depth: roughly the same distance must be traversed in from the surface of the helix cylinder before the edge of a base pair is encountered. In *A* DNA, on the other hand, the base pairs are tilted between 13 and 19 degrees from the perpendicular. Moreover, they are shifted toward the outside of the helix, whose axis lies in the major groove, bypassing the bases. Hence the minor groove is shallow, scarcely more than a helical depression spiraling around the outside of the cylinder; the major groove is cavernously deep, extending from the surface all the way past the central axis and part of the way out toward the opposite side.

B DNA has an average of 10 base pairs per turn of the helix, with a spacing of 3.4 angstrom units along the helix axis from one base pair to the next. The *A* helix has closer to 11 base pairs per turn, and because of the way the tilted base pairs are stacked, the distance along the axis is only 2.9 angstroms per base pair. These distances are averages, derived from fiber data; single-crystal analysis eventually revealed large local deviations from the average values. (The original Watson-Crick model had the 10 base pairs per turn and the perpendicular stacking of bases that is characteristic of *B* DNA, but the bases were pulled away from the center to leave a central hole more characteristic of the *A* form. The positioning of sugar rings in the original model also was more like that of the *A* form.)

This was pretty much the state of knowledge about double-helical DNA structure for 25 years after Watson and Crick. Many questions remained unanswered and unanswerable. Why is the *B* helix the high-humidity form? Which form is prevalent in chromosomal DNA, and can one form be converted into another in a living organism? Do recognition proteins such as repressors actually read information from the bottom of the grooves when they bind to specific DNA sequences?





ANALYSES OF DNA FIBERS by Struther Arnott and others, done before the single-crystal studies reported in this article, were the basis of these schematic drawings of *A* DNA (*left*) and *B* DNA (*right*) made by Irving Geis. The sugar-phosphate backbones of the double helix are represented as ribbons and the runglike base pairs connecting them as planks. In *A* DNA the base pairs are tilted and are pulled away from the axis of the double helix. In *B* DNA, on the other hand, the base pairs sit astride the helix axis and are perpendicular to it.

In the late 1970's, when improvements in DNA synthesis made it possible for the first time to study single crystals of short molecules of any selected sequence, both Alexander Rich's group at the Massachusetts Institute of Technology and our own group, which was then at the California Institute of Technology, independently undertook to synthesize *C–G* copolymers: short molecules composed solely of alternating cytosines and guanines. There was a reason for the choice of this sequence. Experiments in solution suggested that the *C–G* copolymer undergoes some kind of structural transition under high-salt or high-alcohol conditions, and so it was the most likely candidate for studying transitions between helix types.

Single-Crystal Studies

Andrew H.-J. Wang of M.I.T. solved the structure of the hexamer *CGCGCG* in mid-1979 and Horace R. Drew in our laboratory determined the structure of the tetramer *CGCG* a few months later. To everyone's surprise the two molecules proved to be neither *A* DNA nor *B* DNA. They were not even right-handed helixes. Instead they were left-handed helixes of an entirely new type, with a peculiar zigzag backbone that led them to be named the *Z* helix. Fiber studies had prepared us to ask a well-defined question: Are these molecules in the *A* form or the *B* form? Nature gave us the chastening and almost rude reply: No.

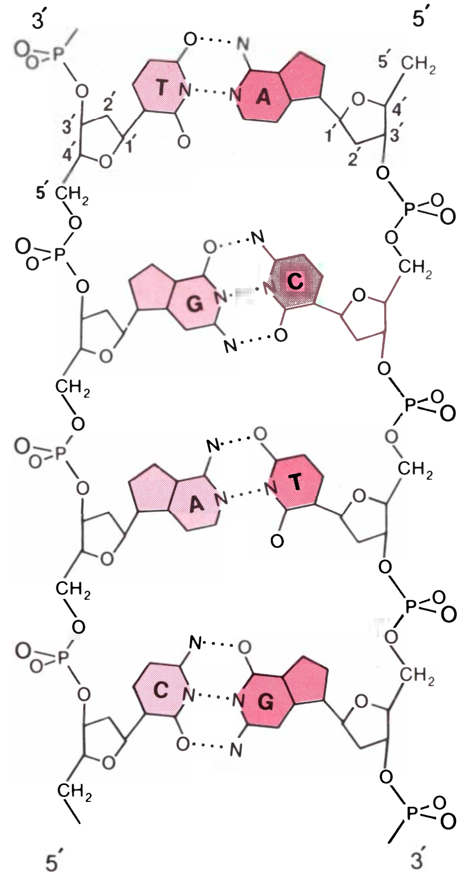
Although the first two double-helical DNA molecules solved by single-crystal X-ray methods broke with all expectations, the next two provided one example each of *B* and *A* DNA. Drew synthesized the dodecamer *CGCGAATTCGCG,* a polymer we chose for two reasons. It coupled *Z*-compatible *CGCG* ends with a *Z*-incompatible *AATT* center, providing a test of the *Z*-forming power of *CGCG* in alien surroundings. And it included the sequence *GAATTC,* which was of interest because it is the recognition and cleavage site for *Eco* RI, one of the restriction endonucleases: enzymes that cut DNA molecules at specific sites. (The reader may note that the oligomer *CGCGAATTCGCG,* like others synthesized for these studies, is self-complementary: the sequence read from left to right is complementary to the sequence read from right to left, so that any two such chains can combine to form a double helix. The result is a major saving of time and energy, since only one strand of a double helix need be synthesized.) The structure of the *CGCGAATTCGCG* molecule turned out to have nothing to do with *Z* DNA but rather to provide a classical example of a *B* helix. Benjamin N. Conner in our laboratory also synthesized the tetramer *CCGG,* solved its structure and showed it to be a fragment of *A* DNA.

Other double-helical DNA molecules followed in quick succession from several laboratories. The *A* helix is represented by *CCGG,* by *GGCCGGCC* from the M.I.T. group and by *GGTATACC* from a group that includes Olga Kennard at the University of Cambridge and Zippora Shakked at the Weizmann Institute of Science in Israel. Wang has also solved the structure of an *A* DNA molecule that is a mixed RNA-DNA hybrid double helix: (*GCG*)*TATACGC*. (The letters enclosed in parentheses represent RNA bases, so that the outer three base pairs, at each end of the double helix, are RNA-DNA hybrids and the central four base pairs are pure DNA.) *B* DNA is represented so far only by *CGCGAATTCGCG* and its derivatives. *Z* DNA is found in the above-mentioned *CGCG* and *CGCGCG* and in a variant of the latter in which methyl groups (CH_3) are added to cytosines, a modification favoring the *Z* structure. These molecules now can be analyzed to see how well they match predictions for *A* and *B* DNA from fiber diffraction, how much variation there is from average helix properties and to what extent the variation can be attributed to the base sequence of the DNA itself.

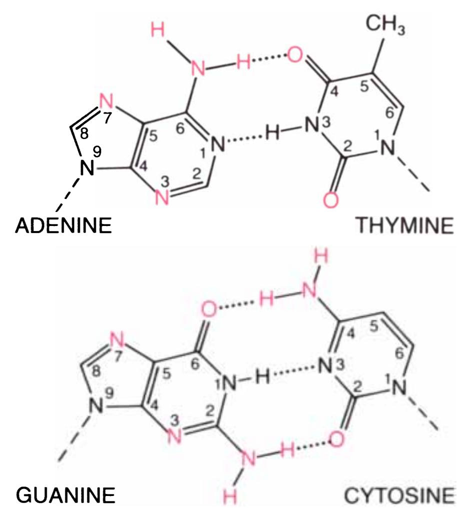
Structural Features

Average values for structural features of each kind of helix have been determined from single-crystal analyses [*see table on next page*]. *A* and *B* DNA are right-handed, with average helical-twist angles from one base pair to the next of 33.1 and 35.9 degrees respectively, corresponding to 10.9 and 10.0 base pairs per 360-degree turn, in close agreement with predictions based on fiber studies. The rise per base pair along the helix axis also has no surprises. What is surprising is the extent of variation in helical-twist angles around these average values: a standard deviation of ± 6 degrees for *A* DNA and of ± 4 degrees for the *B* form. Individual twist angles are as small as 16 degrees and as large as 44 in *A* DNA, and they range from 28 to 42 degrees in *B* DNA. This variation in local helix rotation can be predicted directly from the base sequence by a method I shall describe below. (Is this variable one element by which repressor and other control molecules recognize particular base sequences? The idea is an attractive one.) The bases are inclined away from perpendicularity to the helix axis roughly as was expected from fiber analyses: 13 degrees in *A* DNA and about two degrees in *B* DNA.

A greater departure from fiber-based expectations is found in the "propeller twist" of individual base pairs. Propeller twist is a rotation of the two bases of a pair in opposite directions about their long axis; the sense of the twist is said to be positive when the nearer base (look-



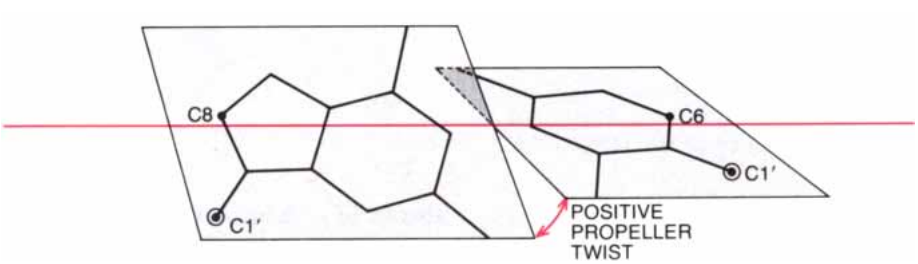
HELIX IS UNROLLED to diagram DNA's sugar-phosphate backbones and base-pair rungs. The backbones run in opposite directions, with 5′ and 3′ ends named for the orientation of the 5′ and 3′ carbon atoms of the sugar rings. Each base pair has one purine base, adenine (*A*) or guanine (*G*), and one pyrimidine base, thymine (*T*) or cytosine (*C*), connected by hydrogen bonds (*dotted lines*).



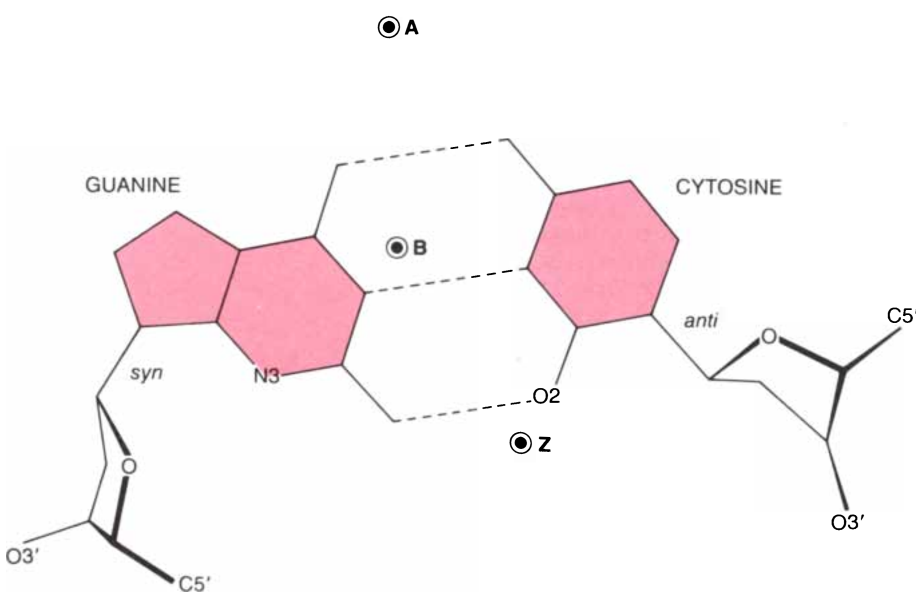
BASE PAIRS are shown in detail, with broken lines indicating bonds to sugar rings of the backbone. Each base pair has a major-groove (upper) edge and a minor-groove (lower) edge. Amine groups (NH_2) are potential hydrogen donors in forming hydrogen bonds with repressors or other regulatory molecules; the nitrogen and oxygen atoms indicated in color are potential hydrogen acceptors. A base pair's characteristic pattern of donors and acceptors is probably "read" by control proteins.

	A DNA	B DNA	Z DNA
HANDEDNESS	RIGHT	RIGHT	LEFT
HELICAL TWIST (DEGREES)			
MEAN AND STANDARD DEVIATION	33.1 ± 5.9	35.9 ± 4.3	*G–C:* −51.3 ± 1.6 *C–G:* −8.5 ± 1.1
OBSERVED RANGE	16.1 to 44.1	27.7 to 42.0	
BASE PAIRS PER TURN	10.9	10.0	12.0
HELIX RISE PER BASE PAIR (ANGSTROM UNITS)	2.92 ± .39	3.36 ± .42	*G–C:* 3.52 ± .22 *C–G:* 4.13 ± .18
BASE INCLINATION (DEGREES)	13.0 ± 1.9	−2.0 ± 4.6	8.8 ± .7
PROPELLER TWIST (DEGREES)	15.4 ± 6.2	11.7 ± 4.8	4.4 ± 2.8
BASE ROLL (DEGREES)	5.9 ± 4.7	−1.0 ± 5.5	3.4 ± 2.1

MEAN HELIX PARAMETERS and standard deviations are listed for three kinds of DNA. The data are based on single-crystal X-ray analysis of molecules mentioned in this article.



PROPELLER TWIST is defined by a schematic drawing of a purine-pyrimidine base pair. A clockwise rotation of the nearer base as one sights down the long axis (*color*) of the base pair in either direction is considered positive propeller twist. Note that the C1′ atoms, at which bases are attached to sugar-phosphate backbones, are shifted up and down by the propeller twist.



TWO CONFORMATIONS of the C–N bond connecting each base to its sugar ring are shown. The *anti* conformation (*right*) appears in all *A* and *B* DNA and at the cytosines in *Z* DNA; the *syn* conformation (*left*) appears at the guanines in *Z* DNA. The alternation of a purine and a pyrimidine (*G, C*) along a strand of *Z* DNA makes possible the *syn-anti* alternation that causes the *Z*-helix backbone to zigzag. The three points labeled *A, B* and *Z* indicate for each type of helix the location of the helix axis with respect to a base pair. Axis is in the major groove in *A* DNA, passes through the base pair in *B* DNA and is in the minor groove in *Z* DNA.

ing along that axis) is rotated clockwise [*see middle illustration at left*]. Little attention had been paid to propeller twist in the fiber studies. A set of coordinates for *A* DNA published in 1972 implied negative propeller twist, and an unpublished revision in 1981 had the proper twist direction but a magnitude of only eight degrees.

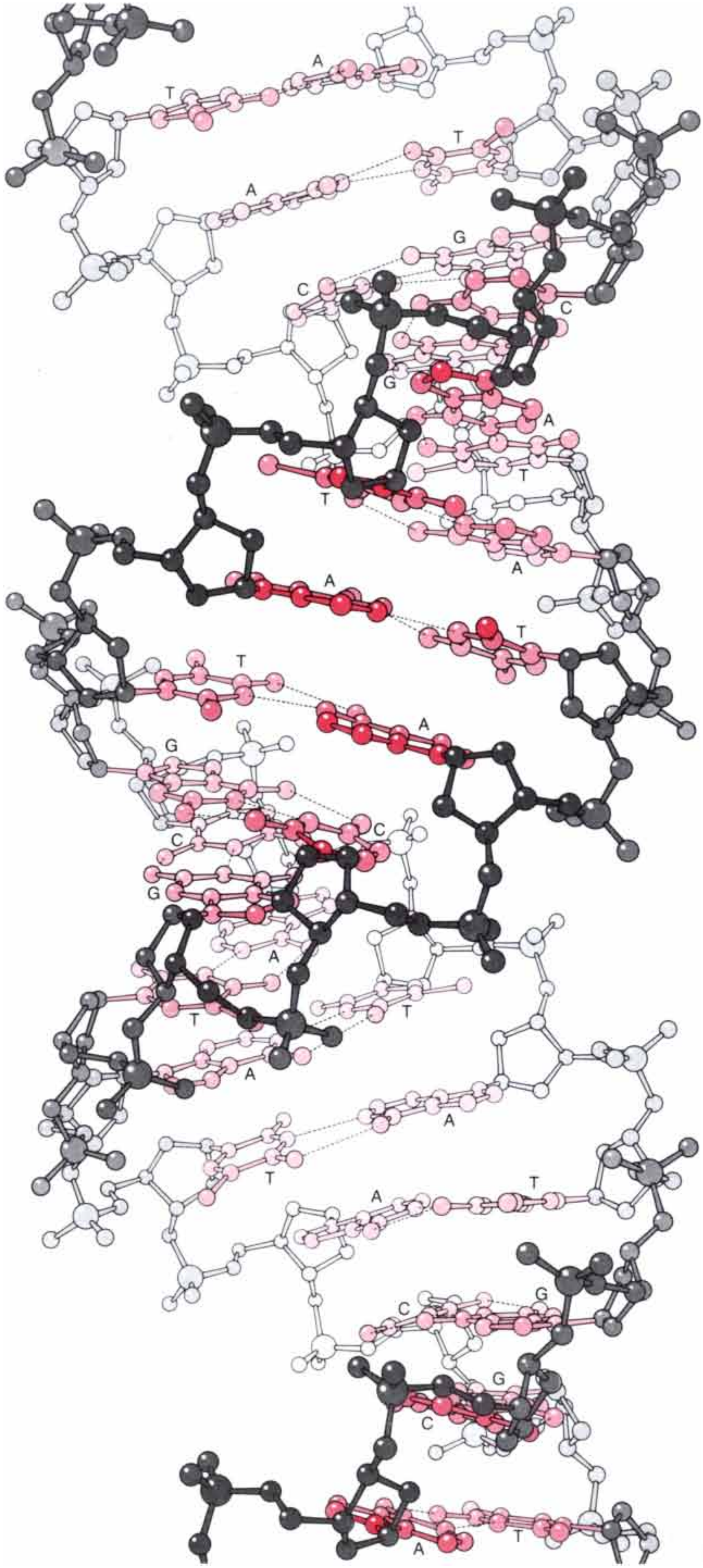
Single-crystal studies show that the twist always is positive, with mean values of 15 degrees for *A* DNA and 12 degrees for *B;* individual values range from as little as three degrees to as much as 25. In a right-handed helix positive propeller twist improves the stacking of bases along each individual backbone strand, making each base overlap more with its neighbors up and down the chain. On the other hand, positive propeller twist also leads to uncomfortably close contact between purines (*G* and *A*) on opposite backbone strands at adjacent base pairs. This, as C. R. Calladine of Cambridge has suggested, is chiefly responsible for several important sequence-dependent variations in helix structure, including propeller twist itself, local helical twist and what is called the base-roll angle.

Base roll measures the orientation of the base pair as a whole (the best mean plane through the purine and the pyrimidine) about its long axis. If two successive base pairs along the helix are rolled in opposite directions, they open up an angle between them toward either the major or the minor groove. The roll angle from one base pair to the next is defined as positive if the angle between them opens toward the minor groove and as negative if it opens toward the major groove. There is no way to measure local base-roll angles from fiber-diffraction data, but single-crystal studies show that base roll is one of the important helix parameters. Since in *B* DNA the base pairs are nearly perpendicular to the helix axis, there need be no systematic mean roll angle from one base pair to the next. In the *A* form, however, the DNA ladder is wrapped around the helix axis in a way that requires an accordionlike opening of base pairs toward the minor groove, and the mean roll angle in *A* DNA is six degrees. Even more significant is the variation, in both the *B* and the *A* helix, of about ±5 degrees from these mean values.

Expansion to Infinite Helixes

The longest DNA double helix that has yet been examined by single-crystal X-ray methods is the dodecamer *CGCGAATTCGCG.* An impression of what longer stretches of helix would look like can be gained by extending three typical molecules [*see illustration on page 95*] to produce long helixes [*see illustrations on pages 100 through 105*].


GM
CIMARRON '84
THIS ONE'S GOT THE TOUCH.
The Cadillac touch. It's everywhere. You can feel it when you put Cadillac's road-hugging Touring Suspension to the test. You can see it ...inside and out. From leather-faced front buckets with lumbar support to aluminum alloy wheels. It's attention to detail, too. Like push-button air. And tungsten halogen fog lamps. You've got to drive this car...and experience the Cadillac touch. Cimarron '84.
BEST OF ALL...IT'S A CADILLAC.
Let's Get It Together...Buckle Up.



This is done by programming a computer to rotate an image of a given helical molecule along its axis until atoms near the beginning of the shifted image coincide with equivalent atoms near the end of the original image, and continuing the process for as long as desired. In this way, for example, the central six base pairs of *A*-helical *GGTATACC* have been repeated four times to produce a 24-base-pair helix with the sequence *GTATACGTATACGTATACGTATAC*. For the first time it has been possible to generate a long double helix from molecules obtained by single-crystal X-ray analysis, without averaging out or idealizing the local, sequence-induced fluctuations in helix structure.

As was expected from fiber studies, the *A* helix is short and fat, with a deep major groove and a broad and shallow minor groove. The *B* helix is slimmer (and taller for the same number of base pairs), with a wide major groove and a narrow minor groove of comparable depth. In contrast, the new left-handed *Z* helix is quite thin and elongated, with a deep, narrow minor groove and a major "groove" that is pushed to the surface so that it no longer is a true groove at all. Because the sugar-phosphate helix backbone in *Z* DNA follows a zigzag path, the repeating unit of the helix is not a single base pair, as it is in *A* and *B* DNA, but rather two successive base pairs: *G–C* followed by *C–G*. This arises because of a difference in the way cytosines and guanines are attached to their sugar rings in *Z* DNA.

At every cytosine the sugar is rotated about its bond to the base so that the puckered sugar ring swings away from the minor groove [*see bottom illustration on page 98*]. This conformation, designated *anti,* also is present in the case of all four bases, and therefore at every step, in *A* and *B* DNA. Each guanine in *Z* DNA, however, has its sugar ring rotated 180 degrees, so that it bends inward toward the minor groove. This *syn* conformation is stereochemically possible only when the sugar is attached to the smaller, five-member ring of a purine rather than to the six-member ring of a pyrimidine; the distance between a *syn* sugar ring and the oxygen atom at position O2 of a cytosine or thymine ring would be too short to be stereochemically allowable. It is the alterna-

A-DNA HELIX is seen in this perspective drawing based on the computer-generated stereoscopic pair on the opposite page. The stereo pair was generated by extending the central six bases of the octamer *GGTATA-CC,* the structure of which was determined by Kennard, Shakked and Viswamitra. The base pairs are not idealized; they are as they are observed in the crystal-structure analysis. Note how phosphate groups on opposite chains face each other across major groove.

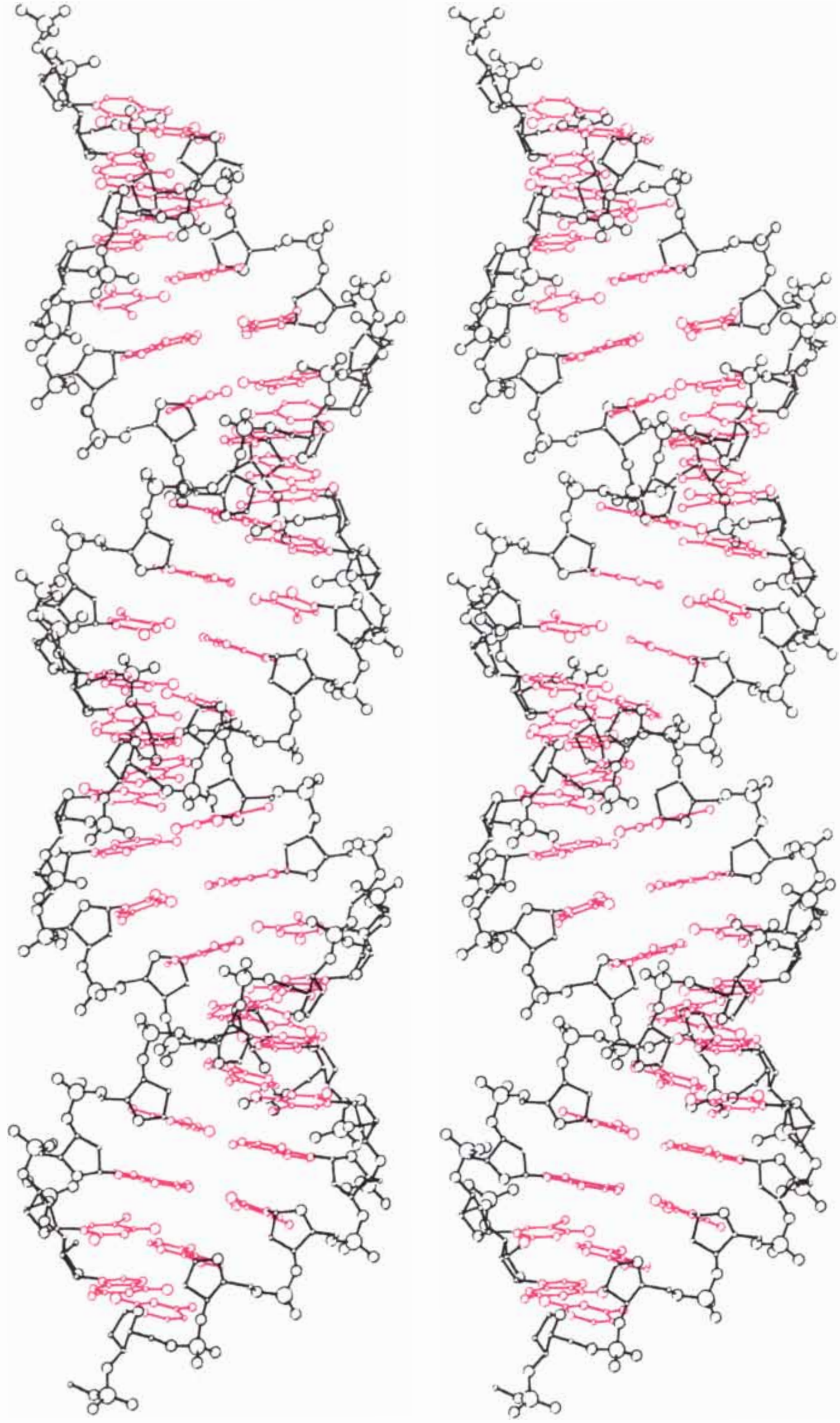
tion of a purine (*G*) and a pyrimidine (*C*) along each chain in *Z* DNA that allows a *syn-anti* alternation of bonds between sugars and bases and produces the zigzag backbone chain. The path of the chain from one phosphorus atom past a guanine to the next phosphorus lies nearly parallel to the helix axis; the path from phosphorus to phosphorus past a cytosine, on the other hand, is tangential to the cylindrical envelope of the helix and perpendicular to its axis. As a result the true repeat unit along the helix is two successive base pairs; although the *Z* helix has 12 base pairs per turn, it formally is a sixfold left-handed helix with six sets of two base pairs per turn.

RNA

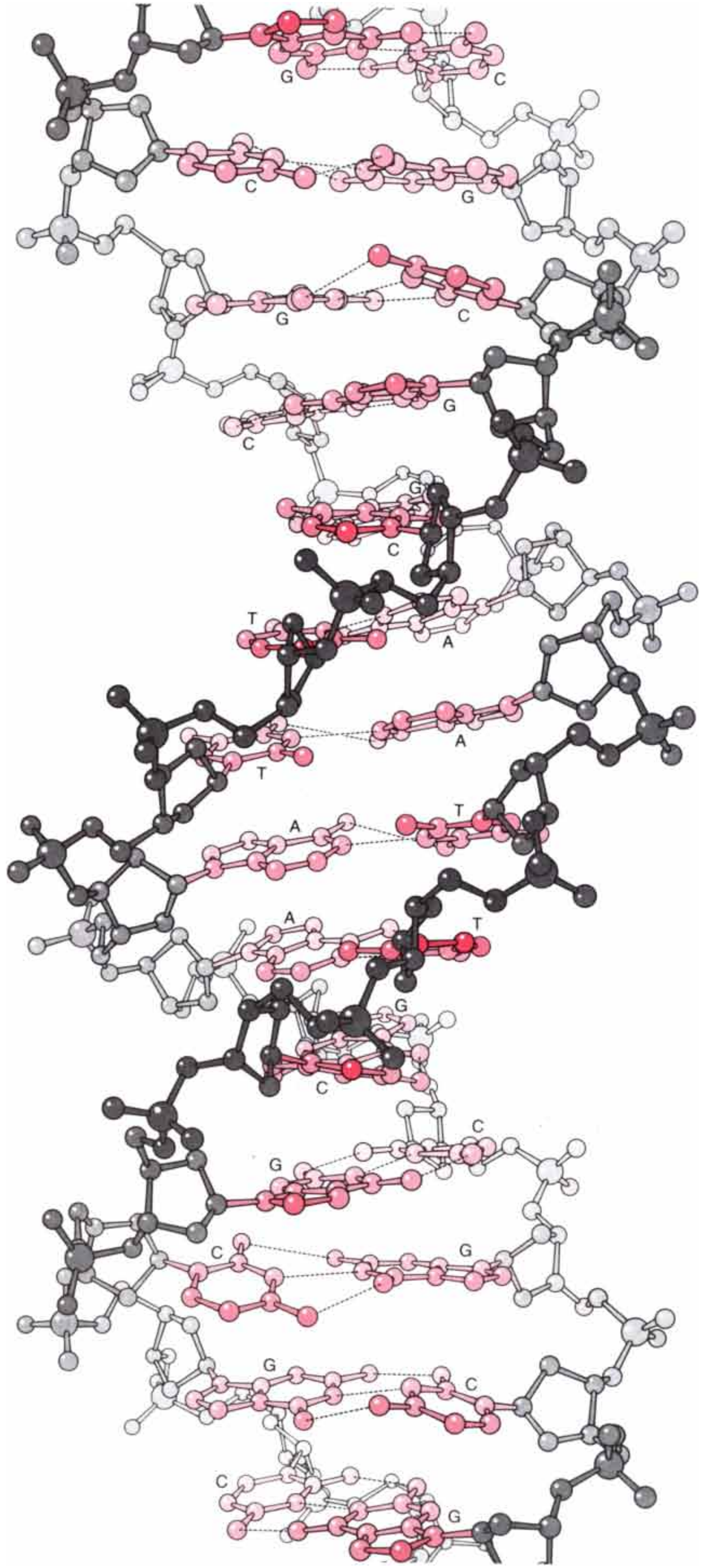
So far I have dealt only with DNA (deoxyribonucleic acid), the archival storage material for genetic information in the nucleus of the cell. The expression of that information depends in large part on the closely related nucleic acid RNA (ribonucleic acid). The base sequence on one strand of the DNA helix is transcribed into a single strand of messenger RNA; transfer-RNA molecules read the sequence and bring the designated amino acids to the ribosomes for assembly into a protein chain; the ribosomes themselves are composed partly of RNA. There are just two differences between DNA and RNA. DNA's thymine is replaced in RNA by uracil, which lacks thymine's methyl side chain, and instead of the deoxyribose sugar rings of DNA's backbone the RNA backbone has ribose rings. In deoxyribose the hydroxyl (OH) group attached to one carbon atom (C2′) of the ribose ring is replaced by a hydrogen.

An important consequence of the extra hydroxyl in RNA is that RNA apparently cannot coil into a *B* double helix. In a hypothetical *B* RNA the OH group would sit in the middle of a cage of atoms of the following phosphate group, sugar ring and base. The distance between the extra oxygen atom and several other atoms would be unacceptably short, making the structure stereochemically unfavorable. In *A* DNA, on the other hand, the 2′ OH group projects out from the helix surface, away from any nearby atoms.

The self-complementary double-helical loops and coils in the transfer-RNA molecule must therefore be variations of an *A* helix. If, as has been suggested, short self-complementary regions of single-strand messenger RNA can form hydrogen-bonded hairpin loops, then these too must be *A* RNA rather than *B*. (In the only example known of RNA participation in a *B* helix, a synthetic hybrid with adenines on the RNA strand and thymines on the DNA strand, the intrusive OH is accommodated by a hy-



STEREO PAIR was generated by manipulating the three-dimensional set of coordinates representing an image of the double-helical *A*-DNA hexamer *GTATAC*, which was derived by deleting the base pairs at the ends of the octamer (because they lacked phosphate groups). The image was rotated and translated along the helix axis until the bottom pair of phosphates on the image coincided with the top pair on an unshifted image; the process was repeated several times to yield what is in effect an infinite *A* helix. The method makes it possible to generate a long stretch of a specific DNA without sacrificing sequence-determined local variations in helix structure. A stereo pair is most easily seen in perspective through a magnifying-lens stereo viewer. With practice a three-dimensional image can be seen without a viewer if one develops the ability to decouple two reflexes that are normally associated: to deconverge one's eyes as if looking at a distant object while focusing the eyes for near vision. Note the depth of the major groove. The helix resembles a double-strand ribbon wound around the outside of a cylinder.



drogen bond to the O4′ atom of the following ribose.) No one has seriously suggested the possibility of a *Z* RNA, but a close examination of the structure indicates that an OH added to a cytosine sugar probably would lie too close to the projecting O2 atom of the cytosine ring.

Stabilization by Hydration

Single-crystal X-ray structure analysis has defined three basic types of DNA double helix rather than two, and one type of RNA double helix. For DNA the *B* form is the stable one under conditions of high humidity, and this makes it seem that the *B* helix is most likely to be present in the nucleus of the cell. If a transient hybrid DNA-RNA helix is formed in the course of the transcription of DNA into messenger RNA, it probably adopts the *A* structure. Under special conditions of stress, right-handed DNA with the purine-pyrimidine alternating sequence can be flipped over into the left-handed *Z* state. For general sequences and under aqueous conditions, however, *B* DNA seems to be the norm. How does water stabilize the *B* helix?

After a DNA molecule has been refined by X-ray analysis one can make a cautious search for solvent (water) molecules around it that happen to be highly ordered, that is, so strongly bound to the DNA that they occupy the same places around every molecule in the crystal. (Less tightly bound water molecules that are at different sites from one molecule to the next in the crystal are blurred in the diffraction image and are barely distinguishable from the background noise.) The water structure has been examined in molecules of all three forms of the DNA double helix, but detailed results have been published only for *B* DNA. Water molecules are observed in *B* DNA in the vicinity of nearly every atom that could make hydrogen bonds with them: free phosphate oxygens, nitrogen and oxygen atoms on the edges of base pairs and, to a lesser extent, the phosphate oxygens that are part of the backbone chain. In effect the helix is coated with a layer of water one molecule deep.

A more extended hydration structure is observed within the narrow minor groove of *B* DNA [*see bottom illustration on page 109*]. The combination of the propeller twist of individual base pairs and the helical twist from one base pair to the next brings an oxygen or nitrogen of one chain into close proximity to an

***B*-DNA HELIX depicted in this drawing was generated from repetition of the central 10 base pairs of the dodecamer *CGCGAATT-CGCG*. The drawing shows the large propeller twist and how the twist improves the stacking of the bases along each backbone chain.**

oxygen or nitrogen of the other chain on an adjacent base pair. The two hydrogen-bond acceptors are bridged by a water molecule. These first-layer molecules are bridged in turn by a second layer of water molecules. Together the two sets of molecules build a zigzag spine of hydration running down the minor groove as it spirals around the *B* helix. The spine of hydration is well developed in regions of *A–T* base pairs but is broken up in *G–C* regions, probably because of the intrusion into the minor groove of NH_2 groups on guanines. In synthetic DNA fibers any base pairs in which guanine is replaced by the modified base inosine, which lacks guanine's NH_2 group, behave like *T–A* pairs rather than *C–G* pairs, in that they help to stabilize the *B* structure.

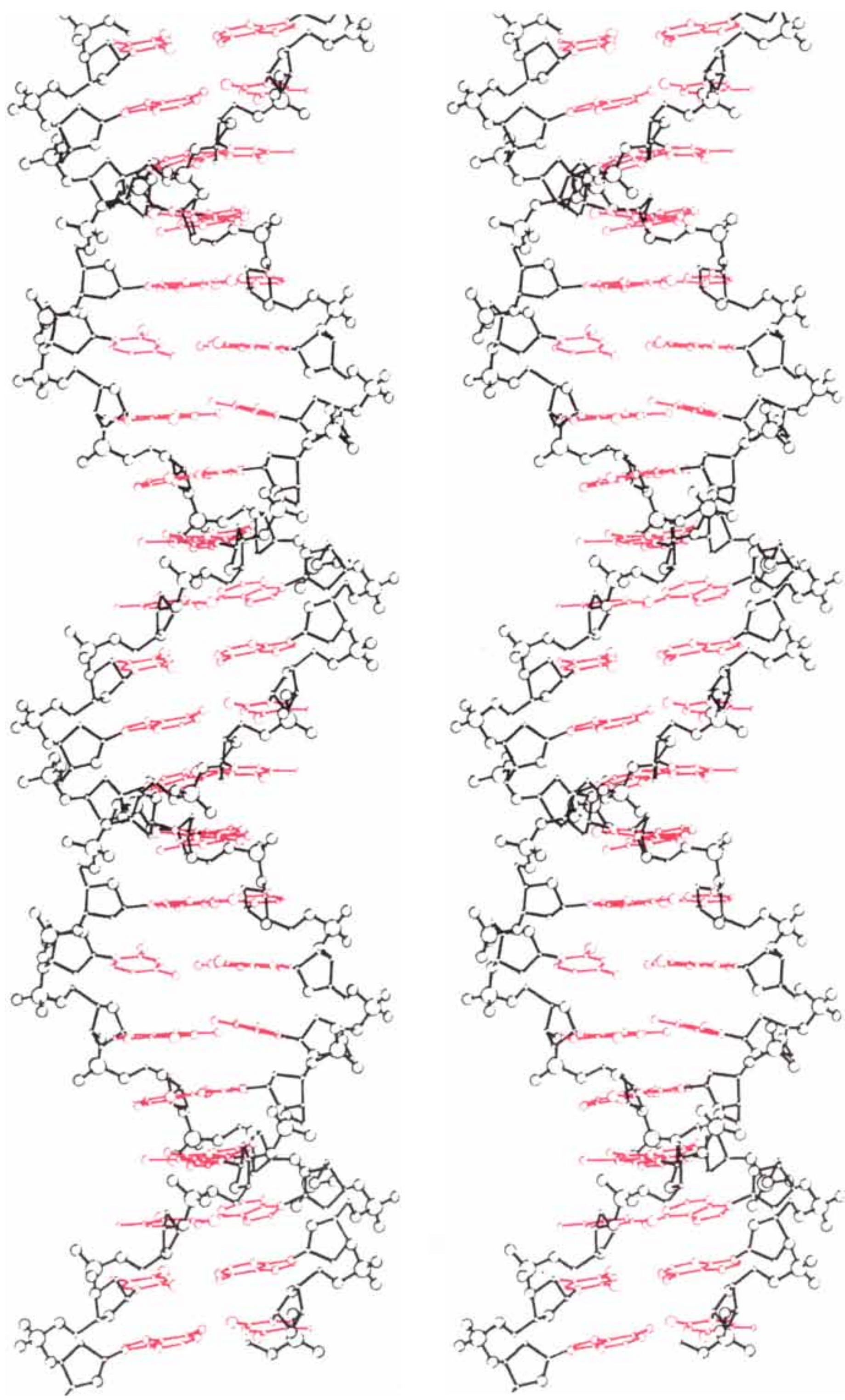
We think that this minor-groove spine of hydration is a prime stabilizing influence on *B* DNA, and that it must be physically disrupted before the minor groove can open up and the helix can slip over into the *A* form. One way to disrupt the stabilizing spine in the laboratory is to remove water molecules by dehydrating the fibers. Another way is to introduce more NH_2 groups into the minor groove; fiber-diffraction studies by Struther Arnott of Purdue University have shown that increasing the *G–C* content in a fiber makes it easier to bring about the transition from the *B* to the *A* form by drying.

No comparable systematic water structure has been found around either the *A* or the *Z* helix. Although water molecules are distributed liberally around atoms on the DNA that could take part in hydrogen bonds, there is nothing that is comparable to the minor-groove spine of hydration in *B* DNA and could be expected to have structural integrity; indeed, the broad minor groove of *A* DNA seems to be even less hydrated than other parts of the *A* helix. It is hard at present to be sure about *A* DNA hydration, however, since in all three *A* DNA crystals—*CCGG/CCGG* (two tetramers stacked to make what is in effect an octamer), *GGCCGGCC* and *GGTATACC*—the end base pairs of one octamer pack against the minor grooves of neighbors, blocking some of the normal hydration that would be expected if the molecules were isolated in solution. Even so, one now has a clear molecular picture of the reasons for the *B* form's prevalence under most general aqueous conditions and for the transition to the *A* form as a result of dehydration.

Genetic Control

Let me now turn to what may be the most significant biological implications of single-crystal X-ray studies of DNA: the influence of base sequence on local helix structure and the role this may play in controlling the expression of genetic information. If the *Z* helix is indeed restricted to only a small number of base sequences with alternating purines and pyrimidines, it is unlikely to have much significance in regions of the gene that encode proteins. As Rich and others have suggested, however, the ability of particular regulatory regions of DNA to reverse the sense of their helical twist could be critically important in controlling the accessibility and readout of genetic information, for example by changing the degree of supercoiling of closed circular DNA.

Such a mechanism would be one example of the influence of base sequence not on the content of the genetic message but rather on its expression. The selective expression or repression of genetic information in particular cells at



STEREO PAIR was prepared by deleting the outermost bases of the dodecamer and reiterating the sequence *GCGAATTCGC*. Here one can see that in a *B* helix the major and minor grooves are about the same depth, although the minor groove is narrower. It is particularly narrow in the *AATT* region at the front and is wider at the back in regions rich in *C–G* base pairs.

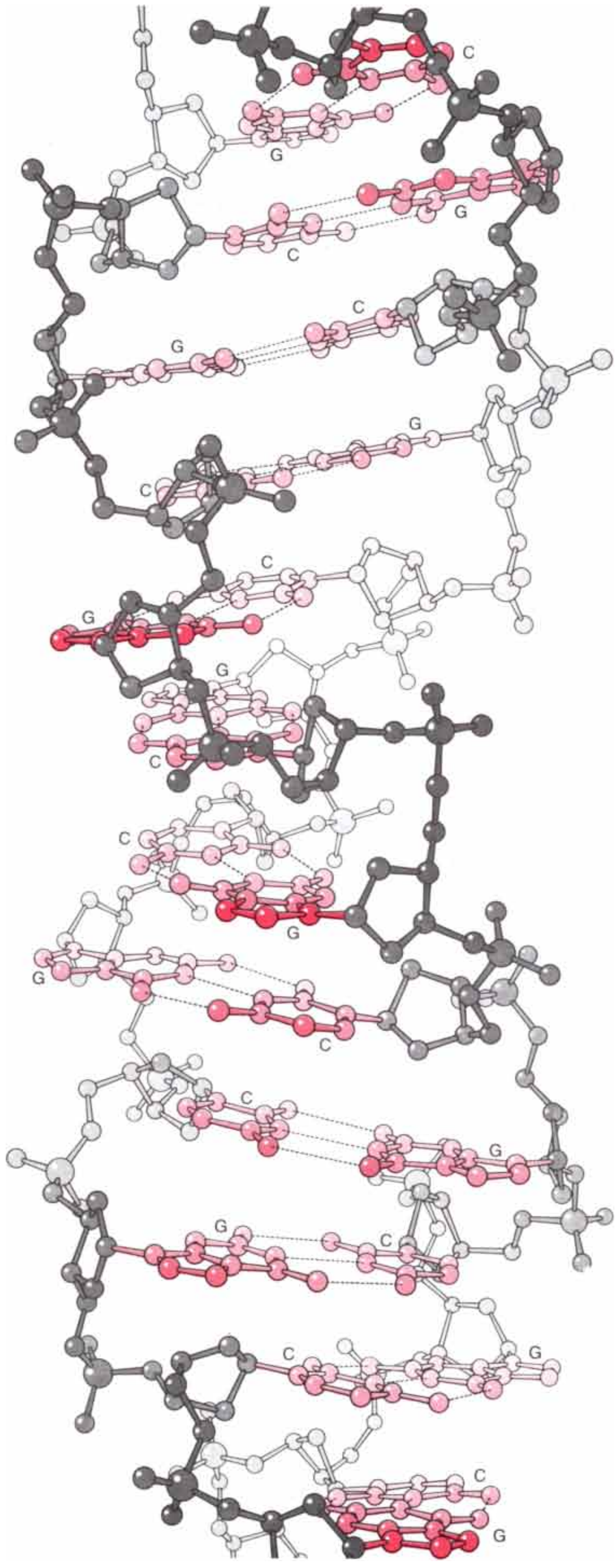
particular times, however it may be accomplished, is the key to development in higher, multicelled life forms. Moreover, a breakdown in the regulation of genetic information can be one trigger for cancer, and it may be involved in the aging process as well. The current level of understanding of the control of genetic information in higher life forms is approximately at the level of understanding of the coding of that information in 1953, when Watson and Crick first proposed their double helix.

The right-hand-to-left-hand transition between *B* and *Z* DNA represents a gross sequence-dependent conformational change. A different kind of structural influence on the readout of genetic information is being revealed by detailed study of local helix variation in *A* and *B* DNA. As I have indicated, the average values of helix parameters are generally about what had been expected from fiber studies; what was not expected are the large local deviations from these averages.

The mean helical-twist angle for *A* DNA is 33.1 degrees, corresponding almost exactly to 11 base pairs per turn, but who would have predicted that individual values could vary from as little as 16 degrees to as much as 44? The mean base-plane roll angles of six degrees for *A* DNA and nearly zero for *B* DNA might have been expected from the way base pairs are wrapped around the helix axis, but no one expected to find standard deviations from these means of about ± 5 degrees in each case. Such local fluctuations at first caused the crystallographers to reexamine their analyses carefully to ensure that the observed variations were not simply artifacts of the refinement process. With this concern satisfied (and in fact with comparable variations observed in different structures refined in three different ways), the obvious question was whether or not the fluctuations have some predictable basis.

From Sequence to Structure

Several attempts were made to find an explanation for structure variation in terms of the sequence of bases along the DNA. The most successful by far was set in motion by Calladine, who applied principles of elastic-beam mechanics to the double helix. According to Calladine's analysis, local variations in helical twist and base-plane roll arise because of steric hindrance between the



Z-DNA HELIX in this drawing is a left-handed helix of alternating guanines and cytosines, generated from the central four base pairs of *CGCGCG* studied by Wang, Rich and their colleagues. Propeller twist is smaller than it is in the two right-handed helixes. Phosphate groups on different chains now face each other across the deep minor groove.

large purine bases, a consequence of the propeller twist in individual base pairs. (Remember that propeller twist itself improves the stacking of bases along each strand of the double helix.)

Because a double-ring purine (*G* or *A*) is larger than a single-ring pyrimidine (*C* or *T*), it extends past the helix axis in any one base pair. If two purines are on opposite backbone strands at adjacent base pairs, they overlap slightly when they are viewed in projection down the helix axis. If each base pair now is given a positive propeller twist, the edges of the purines are brought into unacceptably close contact and a stereochemical clash results. For a purine-to-pyrimidine sequence progressing in the normal 5′-to-3′ direction along a backbone chain, the close contacts are in the major groove; for a pyrimidine-to-purine step the clash is in the minor groove [*see illustration on page 110*]. How might these close contacts between the edges of purine bases be relieved?

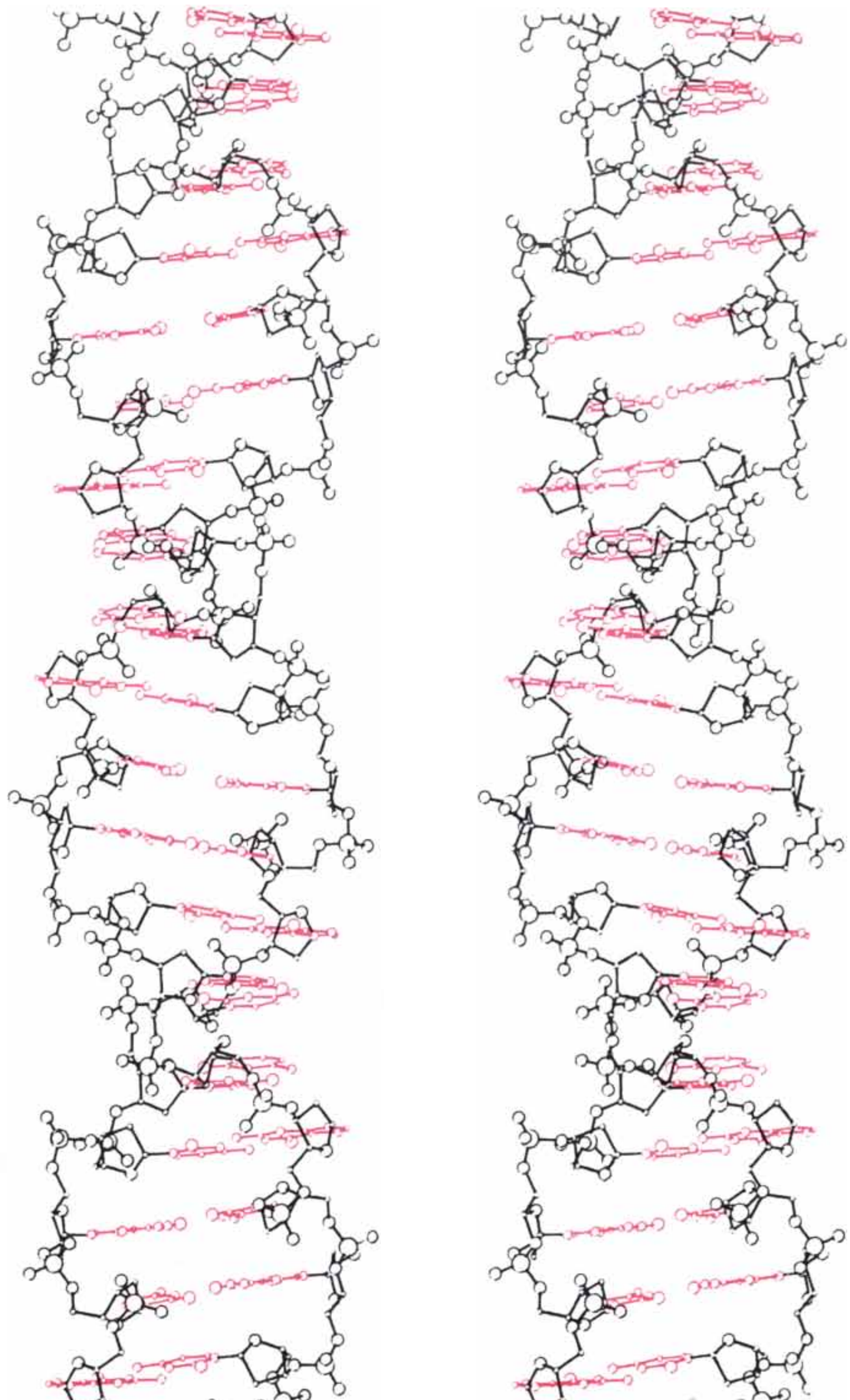
One way is simply to open the base-roll angle on the side where the close contacts take place. For the minor-groove clash of a pyrimidine-purine step the proper response would be to open the angle between the mean plane of the base pairs toward the minor groove, making the roll angle more positive. At purine-pyrimidine steps the roll angle should be opened toward the major groove, making the angle more negative. Drew and I had observed this sequence-dependent behavior of roll angle in the *CGCGAATTCGCG* structure, but our explanation was more convoluted and less plausible than Calladine's.

No matter on which side of a base pair the clash with a neighbor takes place, it always can be lessened by decreasing the local helical-twist angle between the two base pairs. This suggests that helical-twist variation might have the same simple stereochemical explanation as roll angle has. A purine-purine clash across the helix also can be relieved directly by damping down the propeller twist in one or both base pairs, or indirectly by sliding a base pair along its long axis so that the purine is partially removed from the helix stack. Calladine concluded from his analysis of the latter effect that steric clash in the minor groove is roughly twice as pronounced as the clash in the major groove and therefore requires twice the adjustment in local helix parameters.

Predicting the Variation

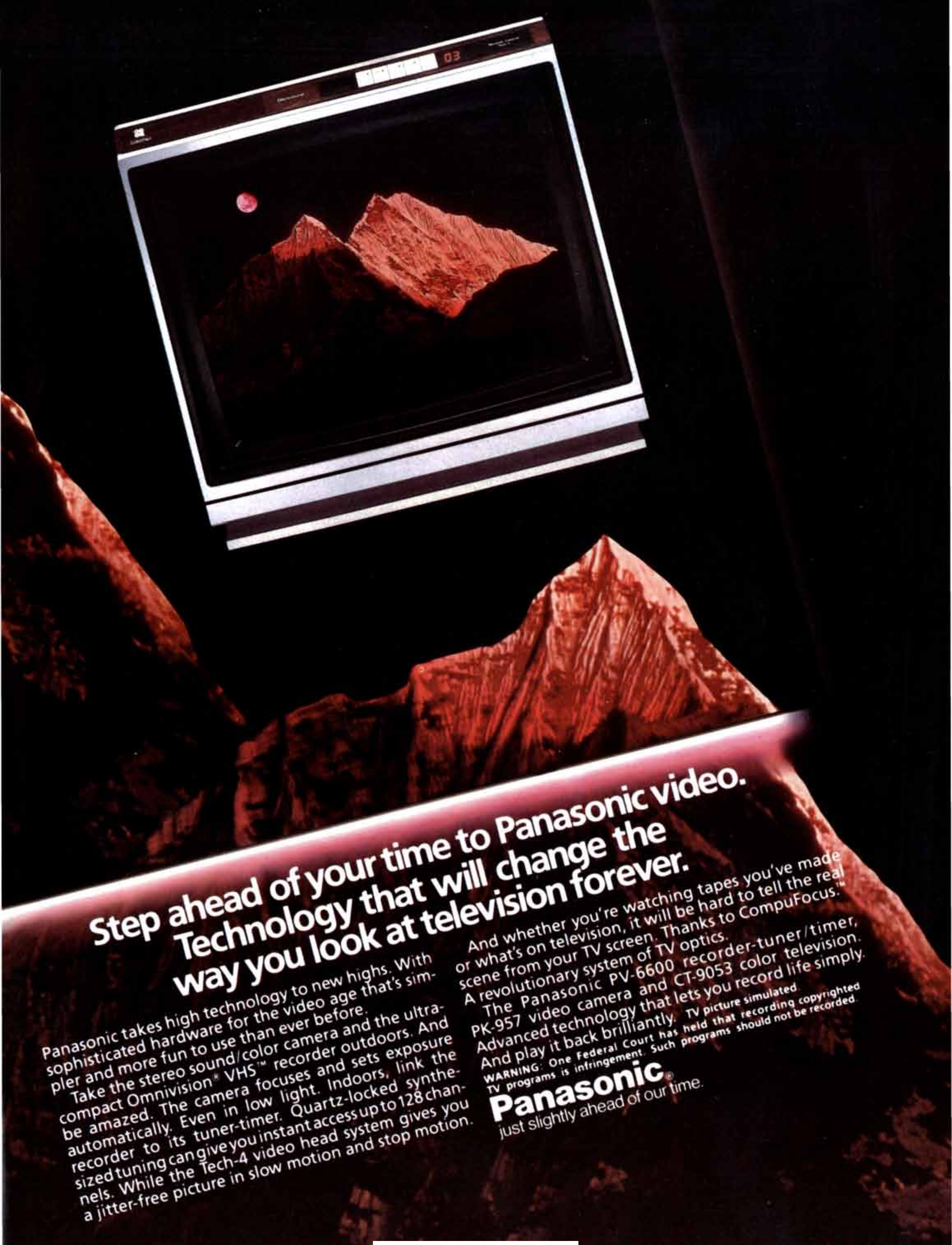
Starting from Calladine's premises, our group (now working at the University of California at Los Angeles) has developed four simple sum functions by which to predict, from a DNA's base sequence, the local variations in helical twist, in roll angle, in propeller twist and also in torsion angle (a measure of the conformation of the backbone chain around the C4′–C3′ bond). The functions for propeller twist and torsion angle seem to apply only to *B* DNA, but the first two functions are valid for both *B* and *A* helixes. As an example of the prediction process, the sum function Σ_1 for helical-twist angle is generated as follows.

1. At each interval from one base pair to the next, assign a number that reflects the relative tendency at that step toward a reduction in helical twist in order to



STEREO PAIR of the extended *Z* helix is a reiteration of *GCGC*. Note that the structure of *Z* DNA and that of *A* DNA are in many ways the inverse of each other. The *Z* helix is tall and slim, with a deep minor groove, a flattened major groove and small propeller twist. The *A* helix is short and broad, with a deep major groove, a very shallow minor groove and large propeller twist.


Panasonic
Panasonic
STEREO
Panasonic
Programmable Tuner/Timer



Step ahead of your time to Panasonic video. Technology that will change the way you look at television forever.
Panasonic takes high technology to new highs. With sophisticated hardware for the video age that's simpler and more fun to use than ever before.
Take the stereo sound/color camera and the ultra-compact Omnivision® VHS™ recorder outdoors. And be amazed. The camera focuses and sets exposure automatically. Even in low light. Indoors, link the recorder to its tuner-timer. Quartz-locked synthesized tuning can give you instant access up to 128 channels. While the Tech-4 video head system gives you a jitter-free picture in slow motion and stop motion.
And whether you're watching tapes you've made or what's on television, it will be hard to tell the real scene from your TV screen. Thanks to CompuFocus.™ A revolutionary system of TV optics.
The Panasonic PV-6600 recorder-tuner/timer, PK-957 video camera and CT-9053 color television. Advanced technology that lets you record life simply. And play it back brilliantly. TV picture simulated.
WARNING: One Federal Court has held that recording copyrighted TV programs is infringement. Such programs should not be recorded.
Panasonic®
just slightly ahead of our time.

VISTA

THE IMPORT WAGON WITH FLEXIBILITY AND ROOM FOR SEVEN.

"Now that's a purpose!"

Purpose: to satisfy the ever-changing transportation needs of people on the move.

Car in point: Vista, a new kind of wagon, imported for Dodge and Plymouth, built by Mitsubishi Motors Corp.

Vista is the wagon reinvented for today. Vista seats 7 passengers and those seats flip up or down, to handle just about any combination of people and things.

36 EST. HWY **28** EPA EST. MPG*

* Use EPA EST. MPG for comparison. Your mileage may vary. Hwy. mileage probably lower. Calif. est. lower.

Vista is quick and agile, easy to park. It has front-wheel drive for sure-footed handling in all kinds of weather or road conditions. It also gets excellent gas mileage, and has a very low base sticker price—yet Vista is not a bare-bones purely functional economy wagon at all. Vista's interior boasts cloth and vinyl seats and heat ducts for rear-seat passengers. And Vista comes with power brakes and electric rear window defroster.

But above all, you just have to love a wagon with all this room and flexibility. Vista. See it at your Dodge or Plymouth dealer.

$8,115

Base Sticker Price excluding title, taxes and destination charges as of October 15. Price subject to change.

IMPORTS WITH A PURPOSE.

Dodge Plymouth

BUCKLE UP FOR SAFETY

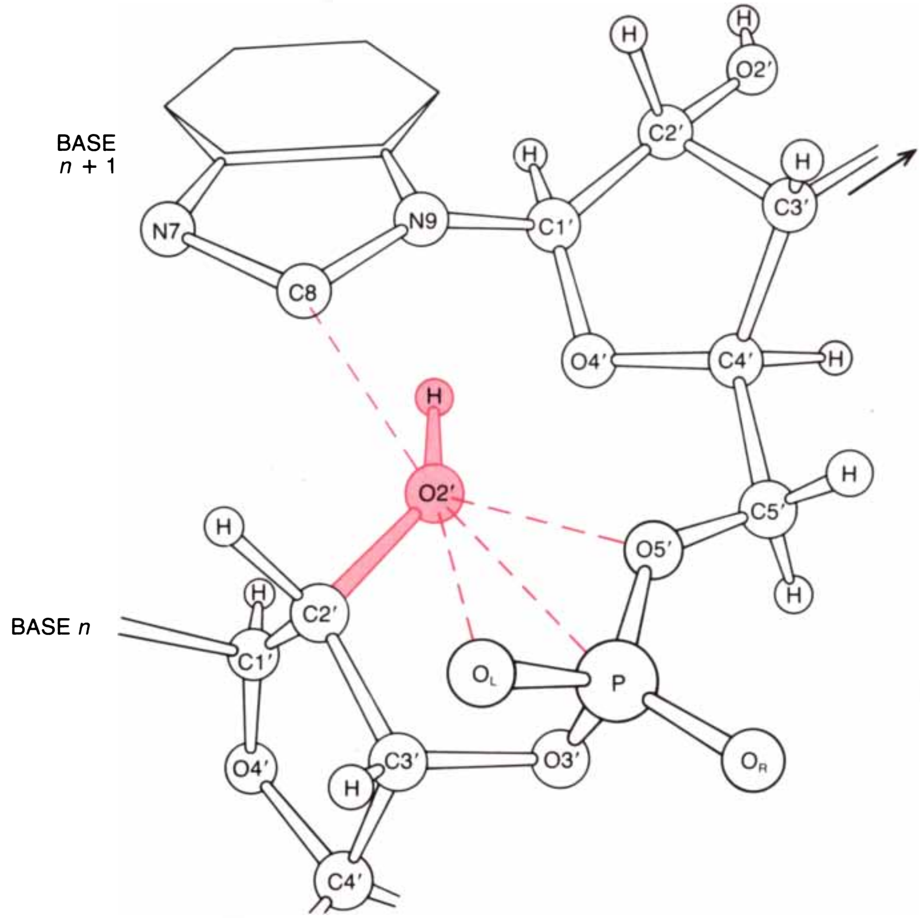
lessen stereochemical clash. Let this number be −2 for purine-pyrimidine steps with their major-groove clash, let it be −4 for pyrimidine-purine steps with their more pronounced minor-groove contact and let it be zero for purine-purine and pyrimidine-pyrimidine steps, which lack steric hindrance.

2. Assign numbers of opposite sign but half the magnitude to the flanking steps, since closing down a central interval by rotating both base pairs will open up the intervals on both sides by half as much. Hence the total contribution to the sum function at each purine-pyrimidine step and its two flanking steps will be +1, −2, +1, and the values for each pyrimidine-purine step and its neighbors will be +2, −4, +2.

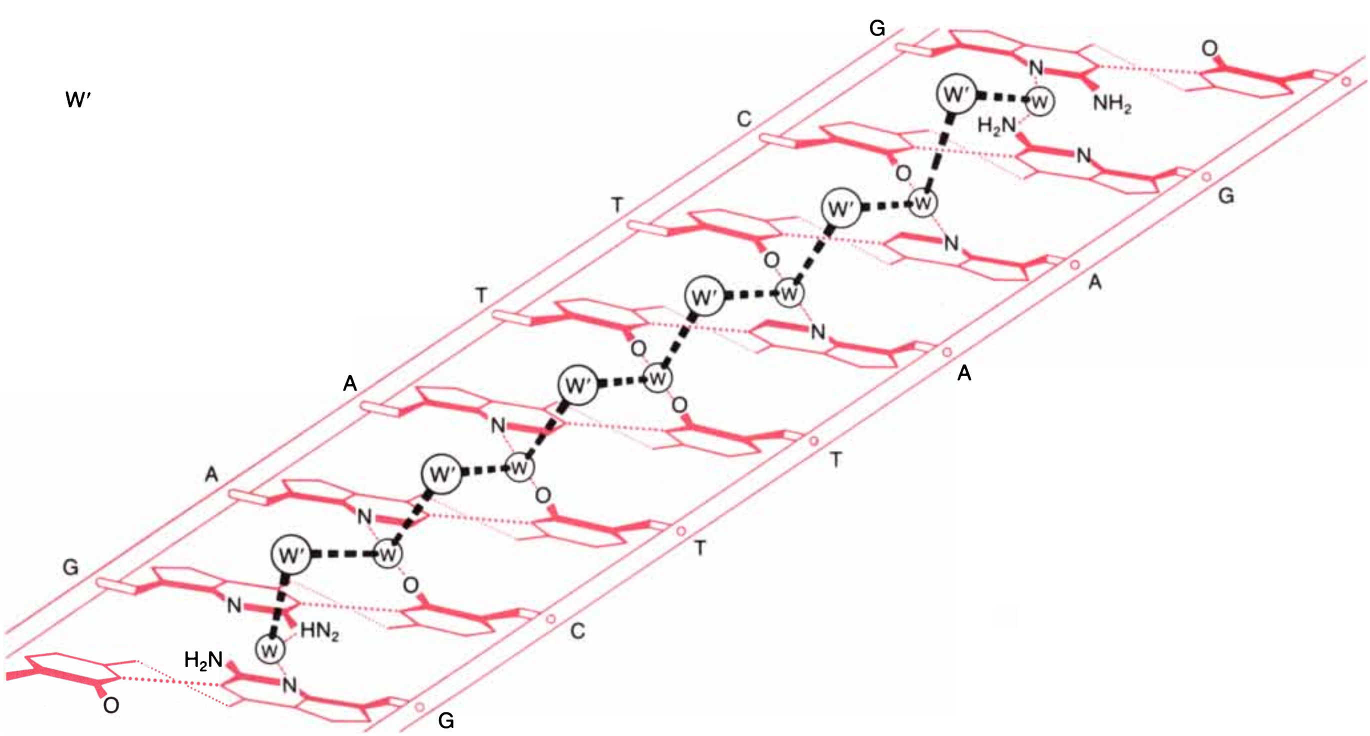
3. Add the contributions from all steps to obtain the full sum function, Σ_1, which measures the local sequence-induced variation from the average helical-twist angle.

The derivation of sum function Σ_1 is illustrated for the *B*-helical dodecamer *CGCGAATTCGCG,* and the predicted and observed values are compared, in the illustration on page 111. A linear-regression comparison of the observed angles with Σ_1 yields a correlation coefficient of .994, or a highly significant correlation between the two quantities. The sum function faithfully reproduces the ups and downs of the observed angles except at the two ends of the sequence, where intermolecular overlap in the crystal distorts the helix.

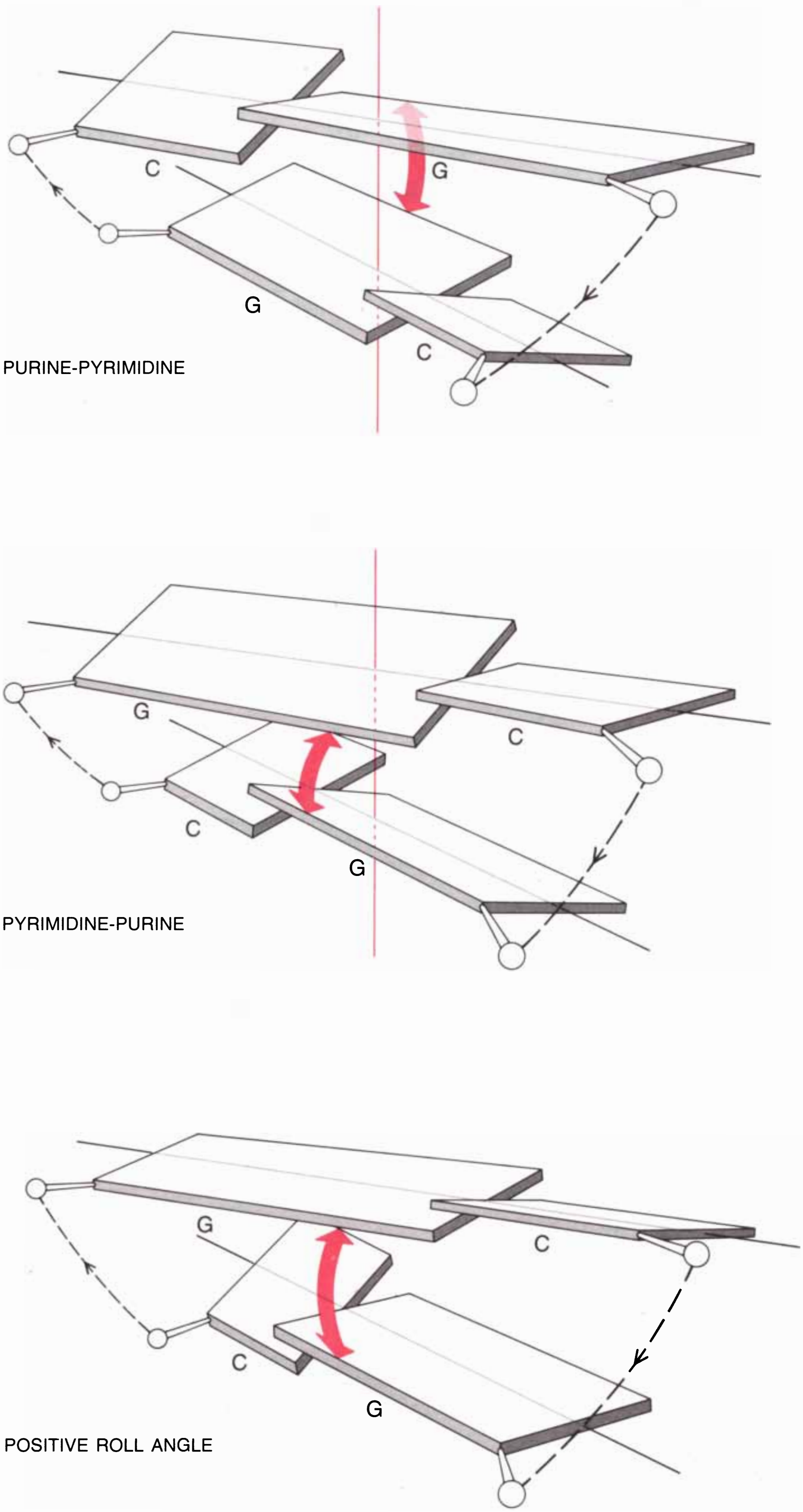
A similar procedure can be followed



RNA does not adopt the *B*-helical structure when it forms a self-complementary double helix. The reason seems to be that RNA has ribose sugar rings, with a hydroxyl group (OH) where the deoxyribose ring of DNA has only a hydrogen atom. Here an OH has been added to a *B*-DNA backbone to simulate a *B* RNA. There is an uncomfortably close contact between the added oxygen atom and four atoms of the adjacent phosphate, sugar and base (*broken lines*).



SPINE OF HYDRATION (*black*) in *B* DNA is formed by a zigzag array of water molecules. This is a view into the minor groove of the unrolled helix of the *B*-DNA dodecamer. The backbones are highly schematized and propeller twist is exaggerated. A bottom layer of water molecules (*W*) is hydrogen-bonded to nitrogen and oxygen atoms on adjacent base pairs; these molecules are linked by a second layer of water molecules (*W′*) to build the spine. The spine keeps the minor groove from opening up to form the shallow depression seen in *A* DNA, and it is therefore thought to be a major factor in making *B* DNA the stable helical structure under conditions of high hydration.



PROPELLER TWIST affects contacts between purines on opposite strands at adjacent base pairs. Here purines and pyrimidines are schematized as planks, the C1′ atoms of sugar rings are designated by small spheres and the backbones are reduced to broken lines, with arrows pointing in the 5′-to-3′ direction. For a purine-pyrimidine step (*top*) positive propeller twist brings the edges of two purines too close together in the major groove (*colored arrow*). For a pyrimidine-purine step (*middle*) the unacceptably close contact is in the minor groove. One strategy (*bottom*) for relieving this clash is to open up the base-roll angle (formed by the tilt of adjacent base pairs as a whole) where the clash takes place, in this case in the minor groove.

to calculate sum function Σ_2, which measures the local variation in roll angle, except that the corrections at pyrimidine-purine steps have reversed signs (-2, $+4$, -2) because at such steps steric hindrance is relieved by opening the angle on the minor-groove side, making the angle more positive. The derivation of Σ_2 and its comparison with the observed roll angles are illustrated for *CGCGAATTCGCG*. Again the increase and decrease of the roll angles is predicted accurately except at the extreme ends of the helix.

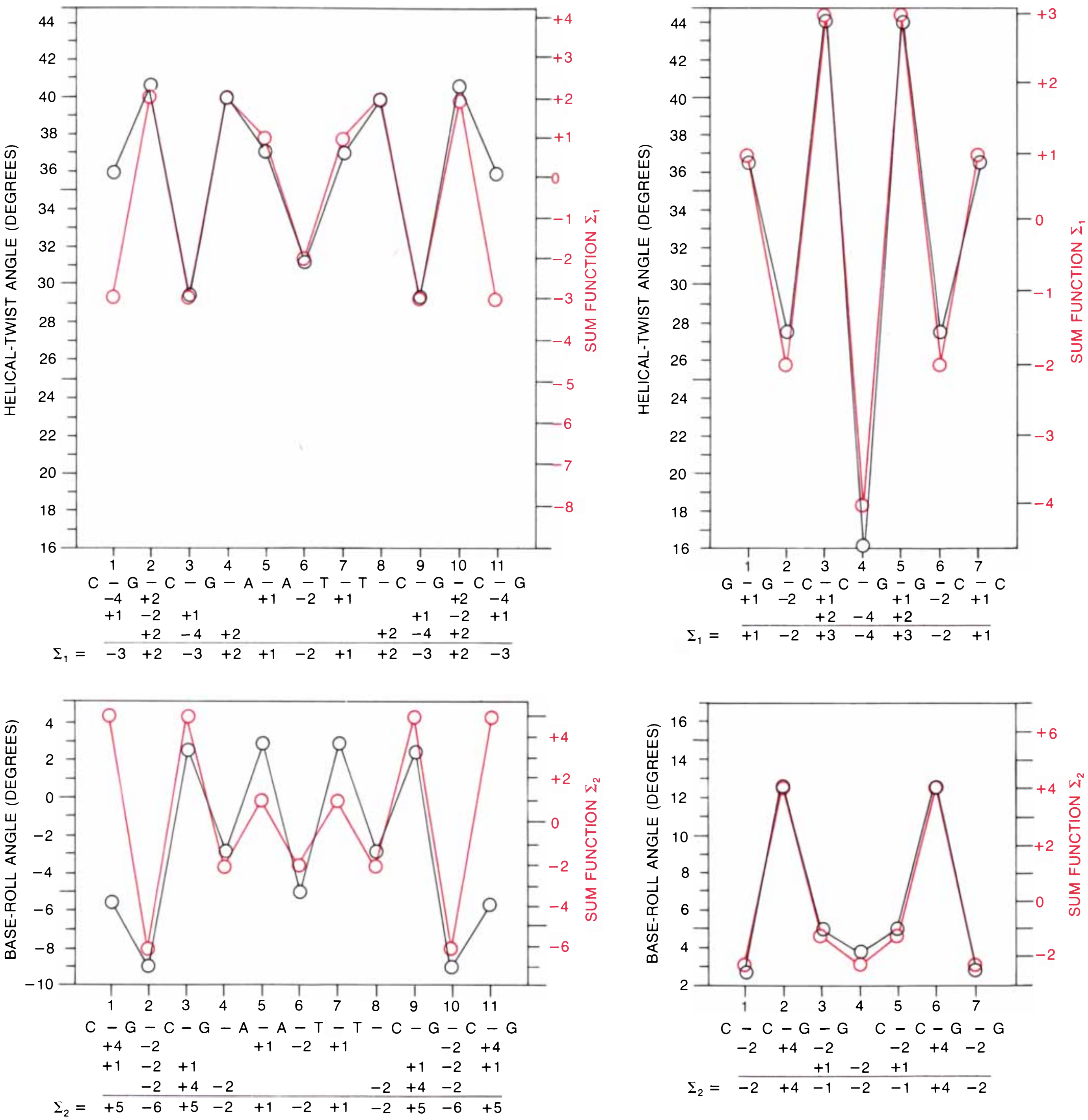
If these two sum functions worked for only one helix of one type, the method would be like many scientific bright ideas that work once in the hands of their inventor but never again for anyone else. It turns out, however, that when the same rules are applied to the four known *A* helixes, the results are fully as good as they are for the *B* sequence. One of the *A* structures offered a particularly stringent test. The two tetramer molecules of *CCGG/CCGG* stack on top of each other in the crystal to form what amounts to an octamer even though the central two base pairs are not bridged by phosphates. In spite of this lack of covalent connection, both the helical-twist angle and the base-plane roll angle across the central step are exactly what would have been predicted by the sum functions for an intact, covalent octamer. Stacking forces between base pairs thus appear to dominate over the influence of the sugar-phosphate backbone in determining double-helical DNA structure. Both the twist function and the roll function also hold for *A*-helical *GGTATACC*, and only the twist function begins to fail in the case of the RNA-DNA hybrid base pairs in the mixed helix (*GCG*)*TATACGC*.

From Structure to Control

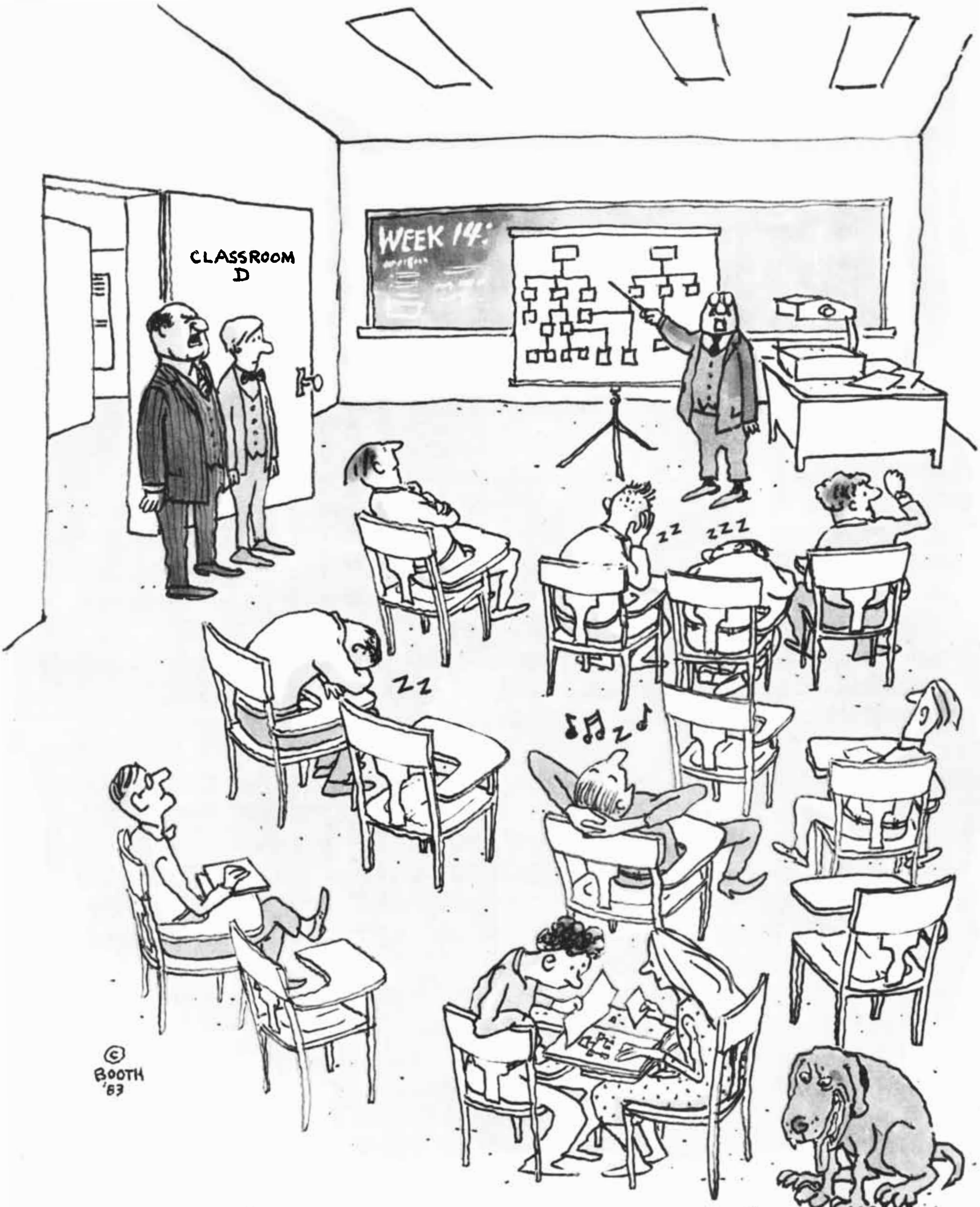
Do these local fluctuations in helix structure actually influence the reading of DNA base sequences by repressors and similar control proteins? If one accepts the idea that the readout process involves hydrogen-bond formation between protein side chains and nitrogen and oxygen atoms on the edges of the bases, it is hard to imagine that these variations of as much as 15 degrees from the mean would not have a major effect on the recognition process. Perhaps the sequence-induced fluctuations represent fine tuning of the fit between DNA and protein, adding to the readout process a dimension that goes beyond the mere arrangement of hydrogen-bond donors and acceptors along the edge of each base pair.

Whether that is the case can only be established by X-ray crystal-structure analysis of a recognition protein in a complex with its particular DNA sequence. Several studies of such complexes are in progress, involving the *lac*, *lambda* and *cro* repressors, catabolite-activator protein (CAP) and various restriction enzymes. John Rosenberg of the University of Pittsburgh has recently calculated an electron-density map of the *Eco* RI restriction enzyme bound to its recognition site on the dodecamer *CGCGAATTCGCG*. He reports that the bound DNA remains in the *B* form but that the helix is distorted.

The questions one can sensibly ask are always influenced by what one already knows. A new advance in knowledge can remove an issue from the realm of idle speculation and place it at the center of scientific inquiry. This is what the recent single-crystal X-ray analyses have done for such questions as: How does base sequence in double-helical DNA affect the structure and behavior of the helix? To what extent is this effect significant in the readout of information? The answers are only beginning to come in, but clearly the questions are well worth asking and will lead eventually to better understanding of genetic control.



LOCAL VARIATIONS in helical-twist angle and base-roll angle are predicted from the base sequence by the sum functions Σ_1 and Σ_2. Here the method is demonstrated for the *B*-helical dodecamer (*top left and bottom left*) and for an *A*-helical octamer (*top right*) and pseudo-octamer (*bottom right*). In each case observed values are shown by the black curves and the scales at the left, predictions by the colored curves and the scales at the right. The derivation of the sum function for each step between adjacent base pairs is shown below each graph. The curves for the observed values and the predicted values are quite congruent: the correlation coefficients between the two sets of values are .994 for the *B*-helix twist angle, .917 for the *B*-helix roll angle, .991 for the *A*-octamer twist angle and .995 for the *A*-pseudo-octamer roll angle. The pseudo-octamer (actually two stacked tetramers, unbridged by phosphates) behaves in this respect like a true octamer; base-pair stacking seems to be more important than the connecting phosphate backbone is for local helix structure.



"It's a fast track here, Herndon. Once you get through our training program."

Quantum Gravity

In a quantum-mechanical theory of gravitation the very geometry of space and time would be subject to continual fluctuations, and even the distinction between past and future might become blurred

by Bryce S. DeWitt

Among the forces of nature gravity seems to have a special status. Other forces, such as electromagnetism, act *in* spacetime, which serves merely as the setting for physical events. Gravity is quite different. It is not a force imposed on the passive background of space and time; rather it constitutes a distortion of spacetime itself. A gravitational field is a "curvature" of spacetime. This is the conception of gravity achieved by Einstein in what he described as the hardest work of his life.

The qualitative distinction between gravity and other forces becomes all the more apparent when one tries to formulate a theory of gravitation consistent with the precepts of quantum mechanics. The quantum world is never still. In the quantum field theory of electromagnetism, for example, the value of the electromagnetic field is continually fluctuating. In a universe governed by quantum gravity the curvature of spacetime and even its very structure would be subject to fluctuations. Indeed, it is possible that the sequence of events in the world and the meaning of past and future would be susceptible to change.

One might suppose that if such phenomena existed, they would surely have been noticed by now. As it happens, however, any explicitly quantum-mechanical effects of gravitation would be confined to an exceedingly small scale, a scale to which attention was first drawn by Max Planck in 1899. In that year Planck introduced his famous constant, called the quantum of action and denoted $\hbar$. He was trying to make sense of the spectrum of black-body radiation, the light that escapes by way of a small opening from a hot cavity. In a curious aside he noted that his constant, when it is combined with the velocity of light and Newton's constant of gravitation, establishes an absolute system of units. These units set the scale for quantum gravity.

Planck's units are totally unrelated to everyday physics. His unit of length, for example, is 1.61×10^{-33} centimeter. This is 21 powers of 10 smaller than the diameter of an atomic nucleus. It bears roughly the same relation to nuclear dimensions as the size of a human being bears to that of our galaxy. The Planck unit of time is even more fantastic: 5.36×10^{-44} second. To probe these scales of distance and time experimentally, using instruments built with present technology, one would need a particle accelerator the size of the galaxy!

Because experiment can offer no guidance, quantum gravity is unusually speculative. Nevertheless, it is basically conservative in spirit. It takes well-established existing theory and merely pushes it to its extreme logical conclusions. In its barest essentials quantum gravity aims to combine three component theories: special relativity, the Einstein theory of gravitation and quantum mechanics, nothing more. Such a synthesis has not yet been fully achieved, but much has been learned from the effort. Furthermore, the development of a working theory of quantum gravity offers the only known path toward understanding the origin of the big bang and the ultimate fate of black holes, events that can be taken as marking the beginning and the end of the universe.

Of the three component theories of quantum gravity special relativity came first historically. It is the theory that unites space and time through the (experimentally confirmed) postulate that the velocity of light is the same for all observers moving in regions of empty space free of external forces. The consequences of this postulate, which was introduced by Einstein in 1905, can be described with the aid of a spacetime diagram, a graph bearing curves that represent the positions of objects in space as a function of time. The curves are called world lines.

For the sake of simplicity I shall ignore two of the spatial dimensions. A world line can then be drawn on a two-dimensional graph where spatial distances are measured horizontally and time intervals are measured vertically. A vertical straight line is the world line of an object at rest in the frame of reference chosen for the measurement. A tilted straight line is the world line of an object moving at a constant velocity in the chosen frame of reference. A curved world line represents an object undergoing acceleration.

A point on the spacetime diagram specifies both a position in space and a moment in time and is called an event. The spatial distance between two events depends on the frame chosen, and so does the time interval. Indeed, the concept of simultaneity is frame-dependent. Two events that are connected by a horizontal line in a chosen frame are simultaneous in that frame but not in other frames.

To establish a relation between frames in relative motion one must introduce a common unit for the measurement of space and time. The velocity of light serves as a conversion factor, relating a given distance to the time needed for light to traverse it. Here I shall adopt the meter as the unit of both space and time. One meter of time is about $3\frac{1}{3}$ nanoseconds (billionths of a second).

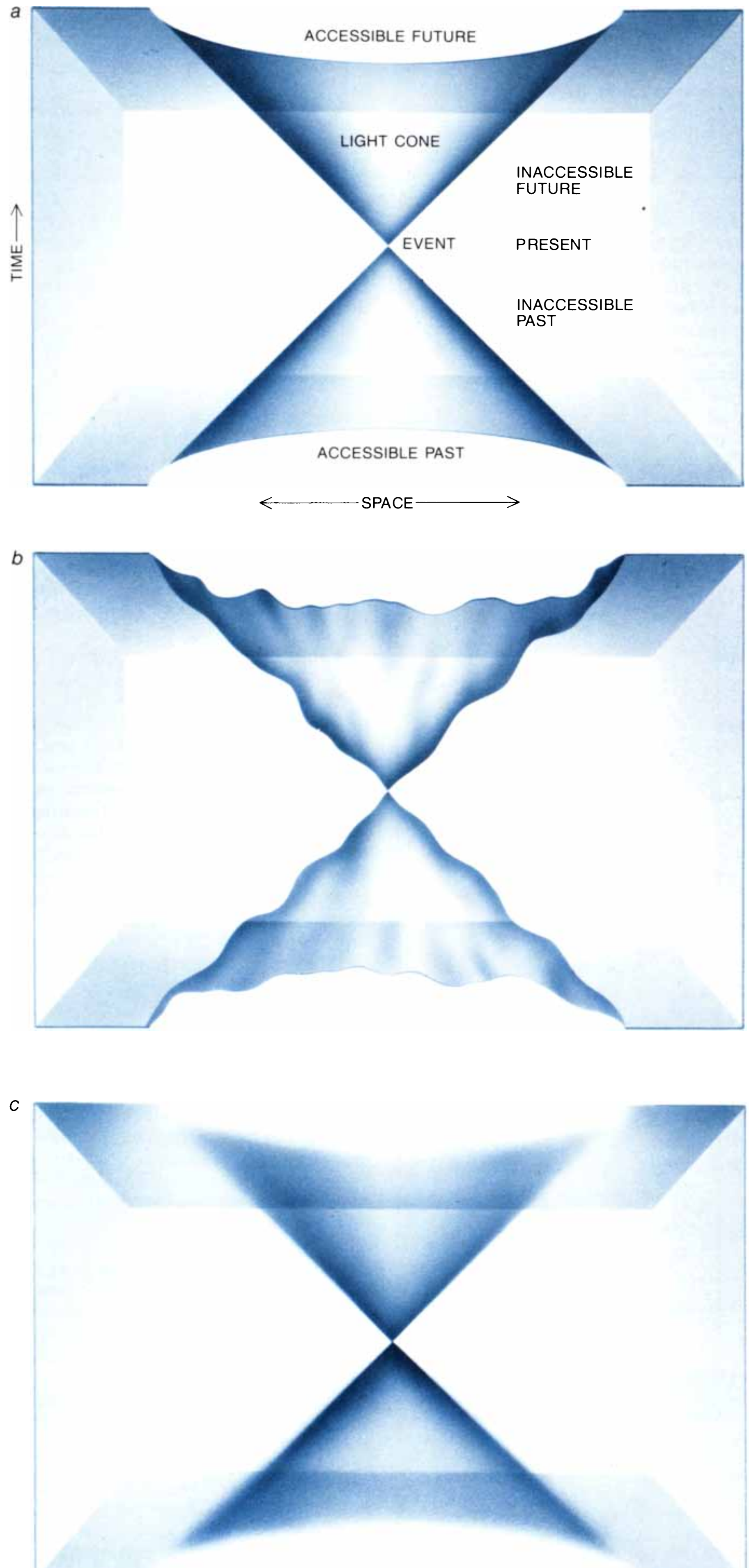
When space and time are measured in the same units, the world line of a photon (a quantum of light) is tilted at 45 degrees. The world line of any material object has a tilt from the vertical that is always less than 45 degrees, which is another way of saying that its velocity is always less than that of light. If the world line of any object or signal were tilted at more than 45 degrees from the vertical, the object or signal would seem to certain observers to be moving backward in time. By setting up a relay of faster-than-light signals a person could transmit information into his own past, thus violating the principle of causality. Such signals are outlawed by provisions of special relativity.

Consider two events on the world line of an unaccelerated observer. Suppose the events, in a particular frame of reference, are separated by four meters of space and five meters of time. Then in that frame the observer is moving at four-fifths the speed of light. In another frame his speed would be different, and so would the associated space and time intervals. There is one quantity, however, that would be the same in all frames of reference. This invariant quantity is called the "proper time" between the two events; it is the time interval measured by a clock the observer carries with him.

In the chosen frame of reference the world line between the two events is the hypotenuse of a right triangle having a base of four meters and a height of five meters. The "proper time" corresponds to the "length" of this hypotenuse, but it is calculated in an unusual way: by means of a "pseudo-Pythagorean theorem." As with the ordinary Pythagorean theorem, one first squares the sides of the triangle. In special relativity, however, the square of the hypotenuse is equal to the difference of the squares of the sides rather than to the sum.

In the present example the proper time is three meters. It remains three meters in the frame of reference of any unaccelerated observer. The invariance of the proper time is what unites space and time into the single entity spacetime. The geometry of spacetime, being based on a pseudo-Pythagorean theorem, is not Euclidean, but in many ways it is analogous. In Euclidean geometry, among all the lines joining two points a straight line can be defined as one whose length is an extremum. The same is true in the geometry of spacetime. In Euclidean geometry, however, the extremum is always a minimum, whereas in spacetime it is a maximum whenever the two points can be connected by a world line requiring no faster-than-light travel.

In 1854 the German mathematician G. F. B. Riemann invented a generalization of Euclidean geometry for curved spaces. Two-dimensional curved spaces had been studied since antiquity. They are called curved surfaces and are usually viewed from the perspective of ordinary three-dimensional Euclidean space. Riemann showed that a curved space can have any number of dimensions and that it can be studied intrinsically. It does not need to be imagined as

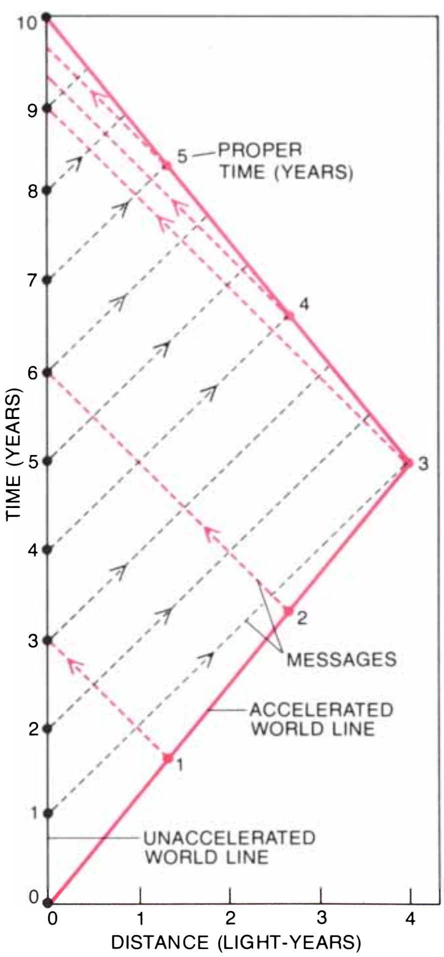


LIGHT CONE, which defines the regions of the universe accessible from a given point in space and moment in time, would become an ill-defined concept in a theory of quantum gravity. The cone (*a*) is a surface in four-dimensional spacetime, but it is drawn here with one spatial dimension suppressed. If gravitation is quantized, the shape of the cone can fluctuate wildly over short distances (*b*). Actually the fluctuations could not be perceived directly; instead the light cone would act as if it were fuzzy. As a result the question of whether two points in spacetime could communicate with each other (by means of signals moving no faster than the speed of light) could be given only a probabilistic answer (*c*).

embedded in a Euclidean space of higher dimension.

Riemann also pointed out that the physical space we live in may be curved. In his view the question could be decided only by experiment. How might one carry out such an experiment, at least in principle? Euclidean space is said to be flat. A flat space has the property that parallel straight lines can be drawn to create a uniform rectangular grid. What would happen if one attempted to draw such a grid on the surface of the earth in the belief that the earth is flat?

The result can be seen from an airplane on any clear day over the cultivated regions of the Great Plains. The land is subdivided by east-west and north-south roads into square-mile sections. The east-west roads often extend in unbroken lines for many miles, but not the north-south roads. Following a road northward, there are abrupt jogs to the east or west every few miles. The jogs are forced by the curvature of the earth. If the jogs were eliminated, the roads would crowd together, creating sections of less than a square mile.



WORLD LINE defines a path through space and time. Here two world lines are shown in a version of Einstein's twins paradox. The "bent" world line of the twin who undergoes acceleration at the turnaround point in his journey appears to be the longer one, but that twin records the shorter "proper time." Indeed, a straight line represents the longest interval between two points in a spacetime diagram. The emission and arrival times of the messages transmitted by the twins are shown.

In the three-dimensional case one can imagine building a giant scaffold in space out of straight rods of equal length joined at angles of precisely 90 degrees and 180 degrees. If space is flat, the construction of the scaffold would proceed without difficulty. If space is curved, one would eventually have to begin shortening the rods or stretching them to make them fit.

The same generalization Riemann introduced for Euclidean geometry can be applied to the geometry of special relativity. The generalization was carried out by Einstein between 1912 and 1915 with the help of the mathematician Marcel H. Grossmann. The result is a theory of curved spacetime. In Einstein's hands it was developed into a theory of gravitation. In special relativity gravitational fields are assumed to be absent and spacetime is flat. In a curved spacetime a gravitational field is present; indeed, "curvature" and "gravitational field" are synonymous.

Because Einstein's gravitational field theory is a generalization of special relativity, he called it general relativity. The name is a misnomer. General relativity is actually less relativistic than special relativity. The complete featurelessness of flat spacetime, its homogeneity and isotropy, are what ensure that positions and velocities are strictly relative. As soon as spacetime acquires "bumps," or local regions of curvature, it becomes absolute because position and velocity can be specified with respect to the bumps. Spacetime, instead of being merely a featureless arena for physics, is itself endowed with physical properties.

In Einstein's theory curvature is produced by matter. The relation between the amount of matter and the degree of curvature is simple in principle but complicated to calculate. Twenty functions of the coordinates of a point in spacetime are needed to describe the curvature at that point. Ten of the functions correspond to a part of the curvature that propagates freely in the form of gravitational waves, or "ripples of curvature." The other 10 functions are determined by the distribution of mass, energy, momentum, angular momentum and internal stresses in the matter, as well as by Newton's gravitational constant, G.

With reference to the mass densities encountered on the earth G is a very small constant. It takes a lot of mass to bend spacetime appreciably. The reciprocal quantity $1/G$ can be regarded as a measure of the "stiffness" of spacetime. In terms of everyday experience spacetime is very stiff. The entire mass of the earth induces a spacetime curvature that is only about a billionth the curvature of the earth's surface.

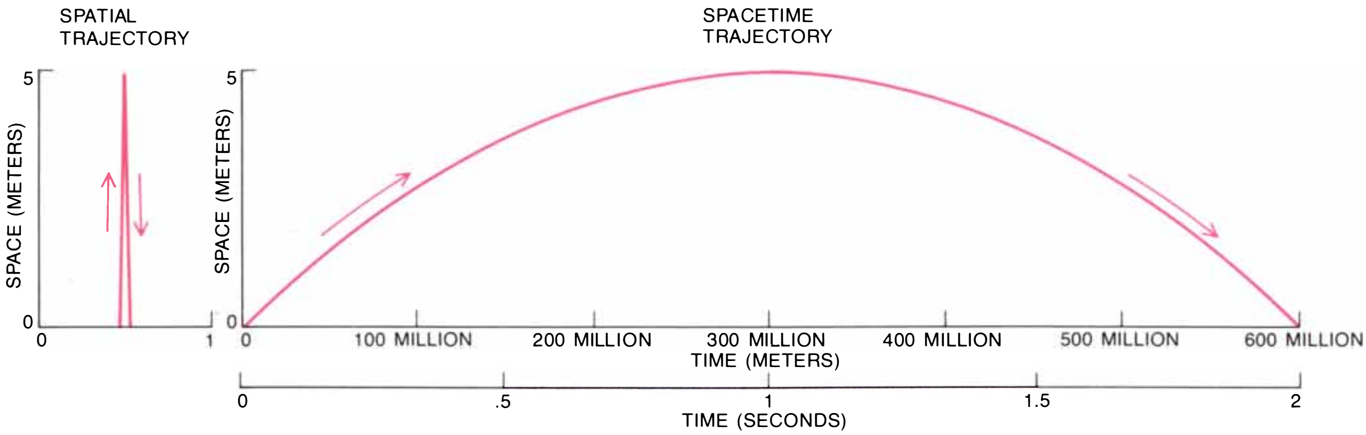
In Einstein's theory a freely falling or freely orbiting body follows a geodesic world line. A geodesic connecting two spacetime points is a world line of extremal length between them; it is a generalization of the concept of a straight line. If a curved spacetime is imagined as embedded in a flat space of higher dimension, a geodesic appears as a curved line.

The effect of curvature on a moving body has often been illustrated by a model in which a ball rolls on a distorted rubber sheet. The model is misleading in that it can represent only spatial curvature. In real life we are constrained to remain in the four-dimensional universe of space and time. Moreover, we cannot avoid motion in this universe, because we are hurtling forward in time. Time is the key element. It turns out that although space is curved in a gravitational field, curvature in time is much more important. The reason is the high value of the velocity of light, which is the magnitude relating the scale of space to that of time.

Near the earth the curvature of space is so slight that it cannot be detected by static measurements. Yet our headlong rush in time is so fast that in dynamical situations the curvature becomes noticeable, much as a slight bump on a speedway, which would pass unnoticed by a pedestrian, becomes a hazard to a fast-moving automobile. Although space near the earth appears flat to a high degree of precision, we can see the curvature of space*time* by merely throwing a ball into the air. If the ball is in the air for two seconds, it follows an arc with a height of five meters. Light travels 600,000 kilometers in two seconds. If one imagines the five-meter-high arc stretched horizontally to a length of 600,000 kilometers, the curvature of the arc is the curvature of spacetime.

Riemann's introduction of the idea of curved spaces initiated another rich area of mathematical study: topology. Boundaryless two-dimensional surfaces had been known to exist in an infinite variety of types that cannot be continuously deformed into one another; a sphere and a torus are two simple examples. Riemann pointed out that the same is true of higher-dimensional curved spaces, and he took some initial steps toward their classification.

Curved spacetimes (or, more precisely, models of curved spacetime) also exist in an infinite variety of topological types. As candidates for a description of the real universe some of the models can be rejected because they lead to paradoxes of causality or because in them known physical laws cannot be made to hold. There still remain a staggering number of possibilities.



CURVATURE OF SPACETIME in the presence of mass constitutes a gravitational field. When a ball is thrown five meters into the air (*left*), it stays aloft for two seconds. Its up-and-down motion reveals the curvature of spacetime near the surface of the earth. Although the curvature of the trajectory is readily visible, it is actually very small when space and time are measured in the same units. For example, seconds can be converted into meters by multiplying them by the speed of light, or 300 million meters per second. When that is done, the trajectory becomes an exceedingly shallow arc 600 million meters long and five meters high (*right*). The height is exaggerated.

One notable model of the universe was proposed by the Russian mathematician Alexander A. Friedmann in 1922. In special relativity spacetime is viewed as being not only flat but also infinite in extent in both space and time. In Friedmann's model every three-dimensional spatial cross section of spacetime is finite in volume and has the topology of a 3-sphere, a space that can be embedded in a four-dimensional Euclidean space in such a way that all its points are equidistant from a given point. The model has been a favorite of cosmologists ever since the expansion of the universe was discovered by Edwin P. Hubble in the 1920's. When Friedmann's model is combined with Einstein's theory of gravitation, it predicts a big bang at an initial moment of infinite compression, followed by an expansion that slows down over billions of years because of the mutual gravitational attraction of all the matter in the universe.

A Friedmann spacetime has the property that every closed curve drawn in it can be continuously shrunk to a point. Such a spacetime is said to be simply connected. The real universe may not have this property. The Friedmann model seems to describe very well the region of space within a few billion light-years of our own galaxy, but we cannot see the entire universe.

A simple example of a multiply connected universe is one whose structure is repeated ad infinitum, like a wallpaper pattern, in a given spatial direction. Every galaxy in such a universe is a member of an infinite series of identical galaxies separated by some fixed (and necessarily enormous) distance. If the members of a series are truly identical, it is questionable whether they should be considered distinct. It is more economical to view each series as representing just one galaxy. Hence a journey from one member of the series to the next returns a traveler to his starting point, and a line tracing such a journey is a closed curve that cannot be shrunk to a point. It is like a closed curve on the surface of a cylinder that goes around the cylinder once. The repeating universe is called a cylindrical universe.

Another example of a multiply connected structure, on a far smaller scale, is the "wormhole," introduced in 1957 by John Archibald Wheeler, now of the University of Texas at Austin. A two-dimensional wormhole can be constructed by cutting two circular holes in a two-dimensional surface and smoothly joining the cut edges [*see illustration on page 118*]. The procedure is the same in three dimensions; it is just harder to visualize.

Because the two holes can be at a great distance in the original space and yet close together through the "throat" connecting them, the wormhole has become a popular device in science fiction for getting from one place to another much faster than light can get there: just punch two holes in space, connect them and crawl through the throat. Unfortunately even if one could construct a hole puncher (which is doubtful), the scheme would not work. If the geometry of spacetime is governed by Einstein's equations, the wormhole is a dynamical object. It turns out that the two holes it joins are necessarily black holes and anything that gets into them can never get out again. What happens is that the throat "pinches off," and everything inside is crushed to infinite density before it can get to the other side.

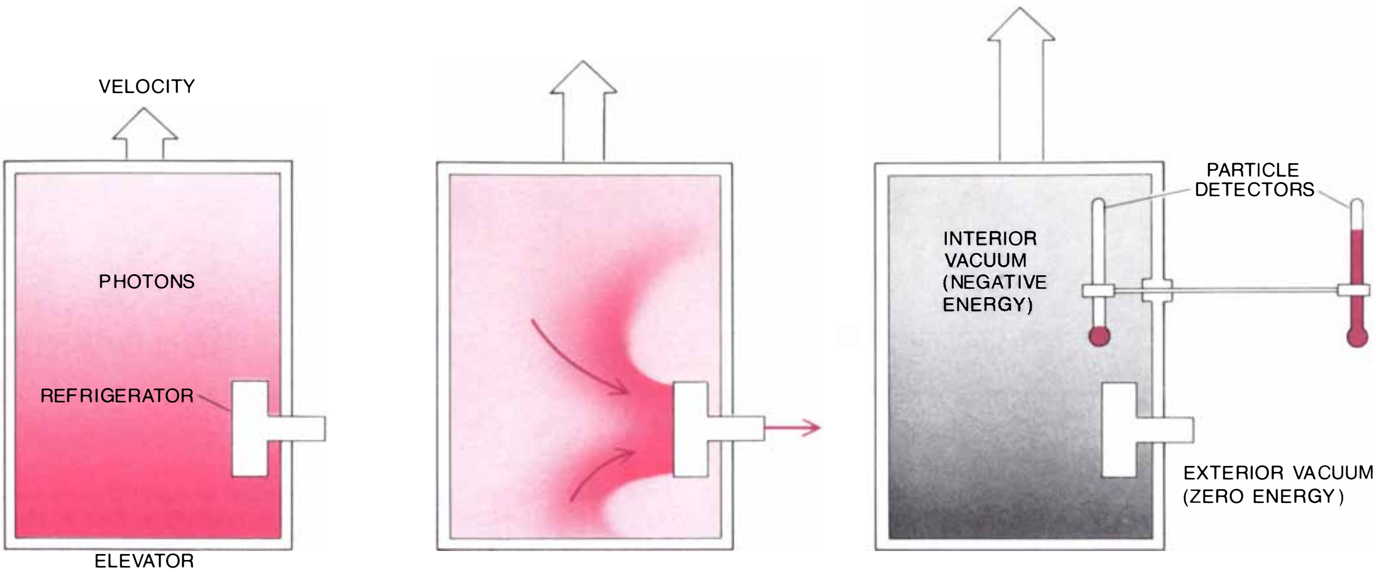
Quantum mechanics, the third component of quantum gravity, was invented in 1925 by Werner Heisenberg and Erwin Schrödinger, but their initial formulation took no account of the theory of relativity. Its success was nonetheless immediate and brilliant, for there lay waiting to be explained a host of experimental observations in which quantum effects dominate and relativity plays a minor or negligible role. It was known, however, that the electrons in some atoms reach speeds that are a sizable fraction of the speed of light, and so the search for a relativistic quantum theory was not long delayed.

By the mid-1930's it was fully understood that when the quantum theory is combined with relativity, a number of entirely new facts can be deduced. The two most fundamental facts are the following. First, every particle is associated with a type of field and every field is associated with a class of indistinguishable particles. No longer could the electromagnetic and gravitational fields be considered the only basic fields of nature. Second, there are two types of particles classified according to their (quantized) spin angular momentum. Those with spins $\frac{1}{2}\hbar$, $\frac{3}{2}\hbar$ and so on obey the exclusion principle (no two in the same quantum state); those with spins 0, $\hbar$, $2\hbar$ and so on are gregarious.

These astonishing consequences of uniting special relativity and quantum mechanics have been confirmed repeatedly in the past half century. Relativity and quantum theory together yield a theory that is greater than the sum of its parts. The synergistic effect is even more pronounced when gravity is included.

In classical physics flat, empty spacetime is called the vacuum. The classical vacuum is featureless. In quantum physics a much more complex entity, with a rich structure, is given the name vacuum. Its structure arises from the existence in it of nonvanishing free fields, that is, fields far from their sources.

A free electromagnetic field is mathematically equivalent to an infinite collection of harmonic oscillators, which can be represented as springs with attached masses. In the vacuum every oscillator is in its ground state, or state of



ACCELERATING ELEVATOR CAR is apparatus for a thought experiment that bears on the nature of the vacuum in quantum mechanics and on the effect either acceleration or gravitation has on the vacuum. The car is assumed to be empty and sealed, so that initially there is a perfect vacuum both inside the car and outside. When the acceleration begins, however, an electromagnetic wave is emitted by the floor, and the car becomes filled with a dilute gas of photons, or quanta of electromagnetic radiation (*left*). A refrigerator driven by some outside source of energy pumps out the photons (*middle*). When all the photons have been removed, photon detectors measure the energy of the vacuum both inside and outside (*right*). Because the instrument outside is accelerating through the vacuum, it responds to quantum-mechanical fluctuations of fields that permeate space even in the absence of particles. The detector inside is at rest with respect to the car and does not sense the fluctuations. It follows that the vacua inside and outside the car are not equivalent. If the "standard" vacuum outside the car is defined as having zero energy, the vacuum inside must have negative energy. The photons removed by the refrigerator would have to be restored in order to bring the energy up to zero. A gravitational field can also create a vacuum with negative energy.

lowest energy. When a classical (non-quantum-mechanical) oscillator is in its ground state, it is motionless at a well-defined point. This is not true of a quantum oscillator. If a quantum oscillator were at a definite point, its position would be known with infinite precision; the uncertainty principle then implies that it would have infinite momentum and energy, which is impossible. In the ground state of a quantum oscillator neither position nor momentum is precisely fixed. Both are subject to random fluctuations. In the quantum vacuum it is the electromagnetic field (and every other field) that fluctuates.

Although the field fluctuations in the quantum vacuum are random, they are of a special kind. They satisfy the principle of relativity in that they "look" the same to every unaccelerated observer, no matter what his velocity. This property can be shown to imply that the field is zero on the average and that the fluctuations increase in magnitude at shorter wavelengths. The net result is that an observer cannot exploit the fluctuations to determine his speed.

The fluctuations can, however, serve to determine acceleration. In 1976 it was shown by William G. Unruh of the University of British Columbia that a hypothetical particle detector undergoing constant acceleration would react to the vacuum fluctuations just as if it were at rest in a gas of particles (and hence not in a vacuum) with a temperature proportional to the acceleration. An unaccelerated detector would not react to the fluctuations at all.

The idea that temperature and acceleration can be related in this way has led to a rethinking of what is meant by "vacuum" and to a recognition that there are different kinds of vacua. One of the simplest nonstandard vacua can be created by repeating, in a quantum-mechanical context, a thought experiment first proposed by Einstein. A closed elevator car is imagined to be drifting freely in empty space. A "playful spirit" starts tugging on it, bringing it into a state of constant acceleration, top end forward. The walls of the car are assumed to be perfect conductors, impermeable to electromagnetic radiation, and the car itself is assumed to be completely evacuated, so that it contains no particles. Einstein introduced this imaginary scene as a way of illustrating the equivalence of gravitation and acceleration, but a reconsideration of it shows that several strictly quantum-mechanical effects can also be expected.

To begin with, at the moment the acceleration begins the floor of the car emits an electromagnetic wave that propagates to the ceiling and then bounces back and forth. (To show why the wave is emitted would require a detailed mathematical analysis of an accelerated electrical conductor, but the effect is analogous to the creation of the acoustic compression wave that would appear if the car were full of air.) If some dissipation in the walls of the car is temporarily allowed, the electromagnetic wave is converted into photons with a thermal energy spectrum, or in other words into black-body radiation characteristic of a certain temperature.

The car now contains a dilute gas of photons. A refrigerator with a radiator on the outside can be installed to get rid of the photons, at some cost in energy from an external power supply. The end result, when all the photons have been pumped out, is a new vacuum inside the car, one that is subtly different from the standard vacuum outside. The difference consists in the following. First, an Unruh detector that shares the acceleration of the elevator car, and that would react thermally to the field fluctuations if it were placed in the standard vacuum outside, displays no reaction inside. Second, the two vacua differ in their energy content.

Specifying the energy of a vacuum requires the resolution of some delicate issues in quantum field theory. I noted above that a free field is equivalent to a collection of harmonic oscillators. The ground-state fluctuations of the oscillators give the vacuum field a residual energy known as the zero-point energy. Because the number of field oscillators per unit volume is infinite, the energy density of the vacuum would also seem to be infinite.

An infinite energy density is an embarrassment. Theorists have introduced a number of technical devices to exorcise it. The devices are part of a general program, called renormalization theory, for handling various infinities that crop up in quantum field theory. Whatever device is adopted must be universal, in the sense that it is not tailored to a spe-

Today's Chevrolet

Braggin' Wagon

Cavalier

Room for the one that didn't get away Chevrolet Cavalier Wagon has more total room than Subaru DL, Nissan Sentra or Toyota Tercel wagons, so that means more comfort, more convenience and more cubic feet for you.

More than just more room Cavalier Wagon also gives you front-drive traction. And Cavalier's 2.0 Liter, high-compression engine with electronic fuel injection gives you more standard horsepower than the three leading import wagons.

Braggin' before it ever hits the road Before any Cavalier Wagon ever hits the road, it's already been through over 1,000 different inspections. Dedicated workers using computerized robots and lasers achieve a high level of precision fit and finish. There's even a computer to check the computer's work.

At Chevrolet we're working to bring you the cars and trucks you want and need—that's what Taking Charge is all about.

Some Chevrolets are equipped with engines produced by other GM divisions, subsidiaries, or affiliated companies worldwide. See your dealer for details.

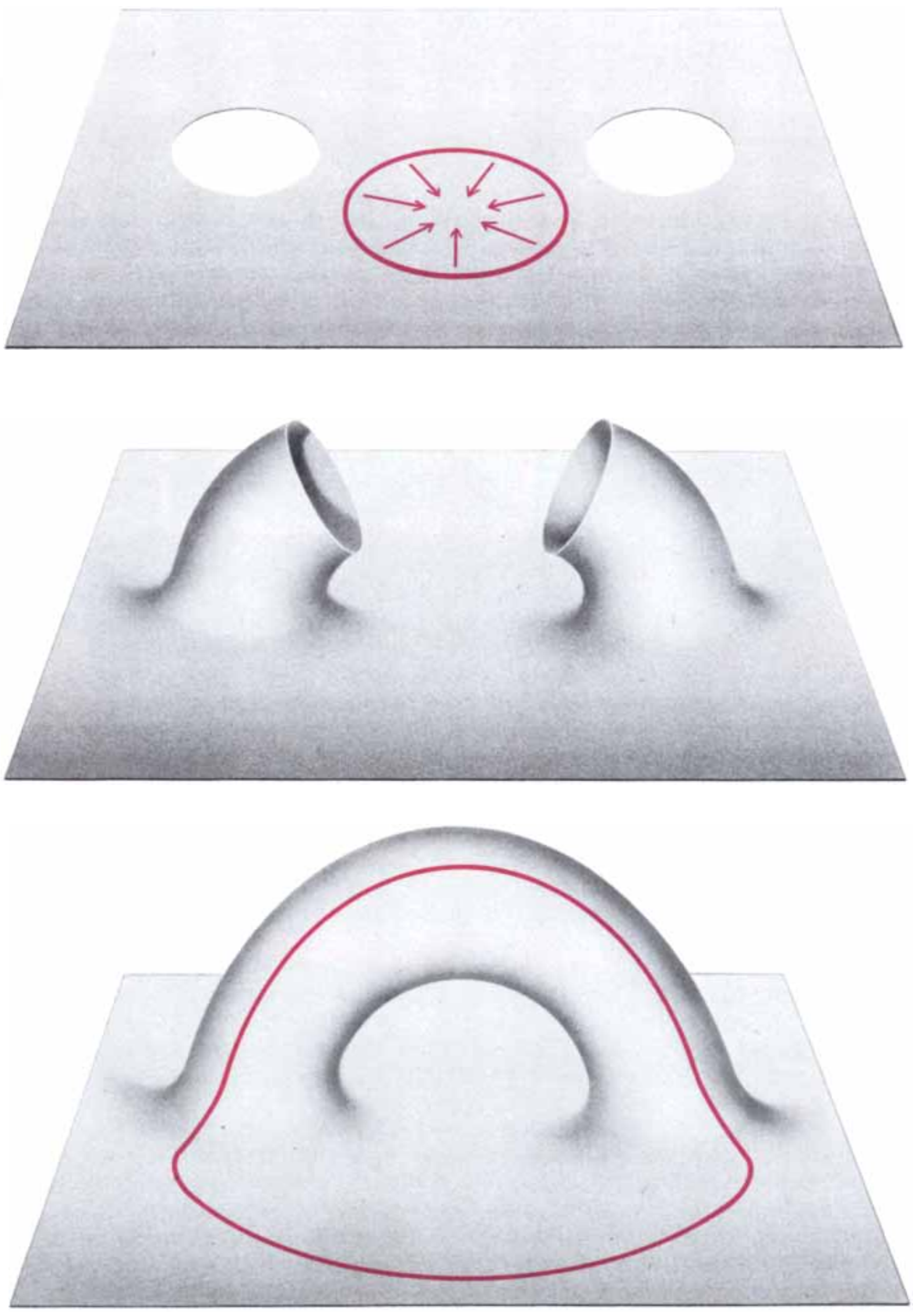
OFFICIAL U.S. CARS AND TRUCKS OF THE XIV OLYMPIC WINTER GAMES

Sarajevo '84

cific physical setting but can be applied uniformly in all settings. It must also yield a vanishing energy density in the standard vacuum. The latter requirement is essential for consistency with Einstein's theory because the standard vacuum is the quantum equivalent of flat, empty spacetime. If there were any energy in it, it would not be flat.

As a rule the various approaches to renormalization theory give identical results when they are applied to the same problem, which inspires confidence in their validity. When they are applied to the vacua inside and outside the elevator car, they yield zero energy density outside and a negative energy density inside. A negative vacuum energy is a surprise. What can be less than nothing? A moment's reflection, however, makes the reasonableness of the negative value apparent. Thermal photons must be added to the interior of the car to make an Unruh detector inside act as it would in the standard vacuum outside. When the photons are added, their energy brings the total interior energy up to zero, equal to that of the vacuum outside.

It must be stressed that such bizarre effects would be difficult to observe in practice. For accelerations common in daily life, even in high-speed machinery, the negative energy is far too small to be detected. There is one case, however, in which a negative vacuum energy has been observed, at least indirectly: in an effect predicted in 1948 by H. B. G. Casimir of the Philips Research Laboratories in the Netherlands. In the Casimir effect two clean, parallel, uncharged, microscopically flat metal plates are placed very close together in a vacuum. They are found to attract each other weakly with a force that can be attributed to a negative energy density in the vacuum between them.



"WORMHOLE" IN SPACETIME is a hypothetical structure that could alter the topology of the universe. On a flat plane a wormhole is formed by cutting two openings and stretching the cut edges into tubes that are then joined. On the original plane any closed curve can be shrunk to a point (*color*). A curve passing through the wormhole, however, cannot be shrunk. A wormhole in three-dimensional space or four-dimensional spacetime is not conceptually different.

The vacuum becomes even more complex when spacetime is curved. Curvature influences the spatial distribution of the quantum field fluctuations and, like acceleration, can induce a nonzero vacuum energy. Because curvature can vary from place to place, the vacuum energy can also vary, being positive in some places and negative in others.

In any consistent theory energy must be conserved. Assume for the moment that an increase in curvature causes an increase in the quantum vacuum energy. That increase must come from somewhere, and so the very existence of quantum field fluctuations implies that energy is needed to bend spacetime. It follows that spacetime resists bending. This is just like Einstein's theory.

In 1967 the Russian physicist Andrei Sakharov proposed that gravitation may be a purely quantum phenomenon, arising from vacuum energy, and suggested that Newton's constant *G*, or equivalently the stiffness of spacetime, may be computable from first principles. The proposal faces several difficulties. First, it requires that gravity be replaced, as a fundamental field, by some "grand unified gauge field" suggested by the known elementary particles. Here a fundamental mass must be introduced so that an absolute scale of units is still obtained; hence one fundamental constant is replaced by another.

Second, and perhaps more important, the calculated dependence of the vacuum energy on curvature turns out to yield a theory of gravity more complicated than Einstein's. Depending on the number and type of elementary fields chosen and on the method of renormalization, the vacuum energy, rather than increasing with increasing curvature, can even decrease. Such an inverse relation would mean that flat spacetime is unstable and that it would tend to wrinkle like a prune. Here I shall assume that the gravitational field is fundamental.

A true vacuum is defined as a state of thermal equilibrium at a temperature of absolute zero. In quantum gravity such a vacuum can exist only if the curvature is independent of time. When the curvature is time-dependent, particles can appear spontaneously in the vacuum (with

PROJECTION

PRESENTING THE END OF THE DARK AGES.
RCA INTRODUCES ITS BEST AND BRIGHTEST GENERATION OF BIG-SCREEN TV'S.

BIG, BRIGHT AND BEAUTIFUL. The picture you see on RCA's big 45-inch screen (measured diagonally) marks the beginning of a bright new era. Its remarkable 65% increase in peak brightness and 200% increase in contrast make it light-years ahead of earlier RCA projection sets. Watching projection television in the dark is a thing of the past.

Another improvement is a big, crowd-pleasing viewing angle that makes just about every seat in the house the best seat in the house.

VIDEO MONITOR CAPABILITY. Our high-performance giant is also a video monitor. Its convenient system of input/output jacks lets you hook up your video cassette recorder, videodisc player, and stereo for maximum enjoyment.

SIZE ISN'T EVERYTHING. RCA's best and brightest features our exclusive Digital Command Center. The total control remote control that enables you to switch from broadcast to VCR to videodisc viewing at the touch of a button.

All of this comes wrapped in the most compact cabinet in RCA big-screen history. To see our best generation of big-screen TV's ever, visit your RCA dealer.

It promises to be a most enlightening experience.

For more information and a free copy of "Living With Video" (a $2.50 retail value), write: RCA Consumer Electronics, Department 32-312P, P.O. Box 1976, Indianapolis, IN 46206.

WE'LL OPEN YOUR EYES. **RCA**

TELEVISION

the result, of course, that it is no longer a vacuum).

The mechanism of particle production can again be explained in terms of harmonic oscillators. When the curvature of spacetime changes, the physical properties of the field oscillators change too. Suppose an ordinary oscillator is initially in its ground state, subject to zero-point oscillations. If one of its properties, such as its mass or the stiffness of its spring, is changed, then its zero-point oscillations must adjust themselves to the change. After the adjustment there is a finite probability that the oscillator is no longer in its ground state but in an excited state. The phenomenon is analogous to the increased vibration induced in a vibrating piano wire when its tension is increased; the effect is known as parametric excitation. In the quantum field the analogue of parametric excitation is the production of particles.

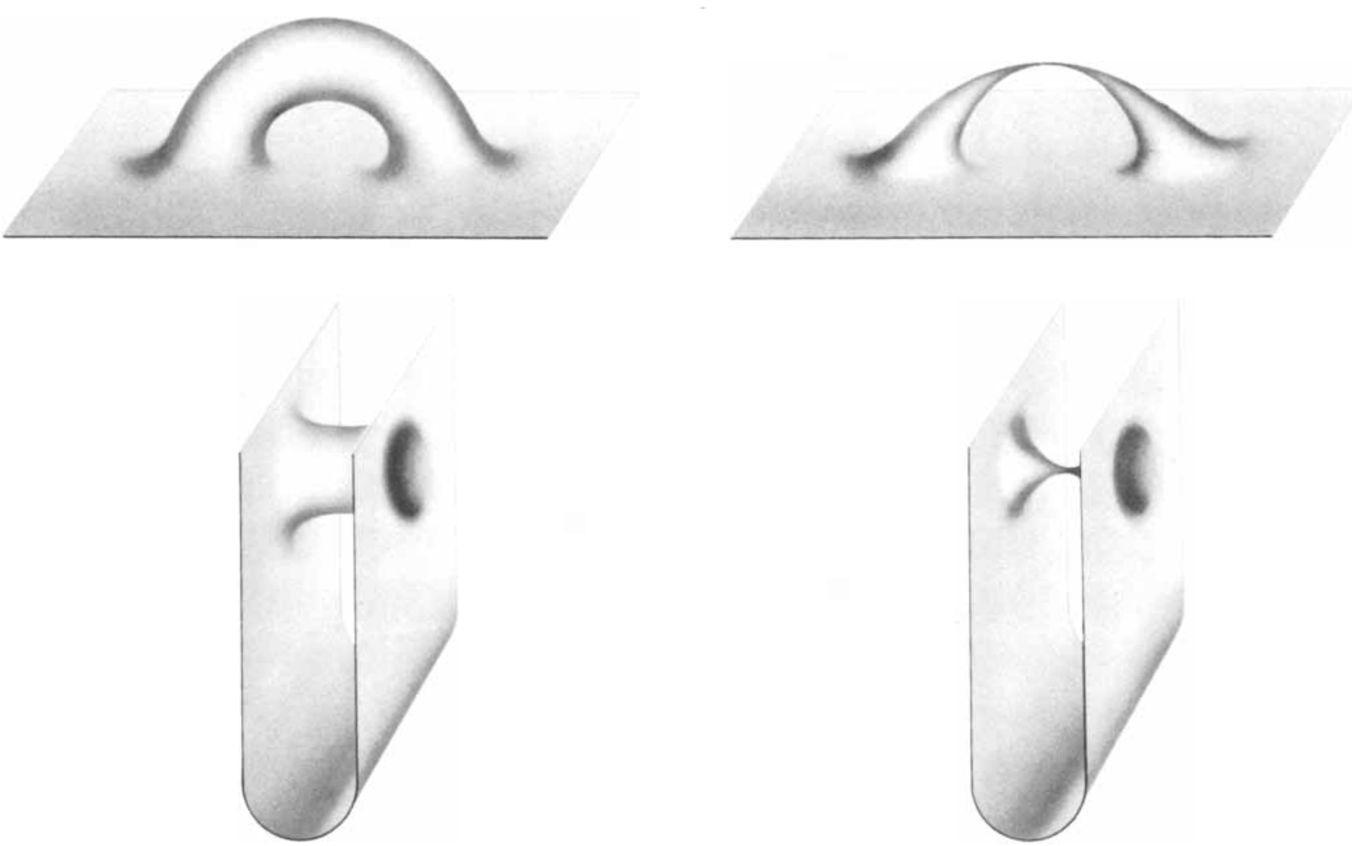
The particles generated by time-varying curvature appear randomly. It is not possible to predict in advance exactly where or when a given particle will be born. One can, however, calculate the statistical distribution of the particles' energy and momentum. Particle production is greatest where the curvature is greatest and changing most rapidly. It was probably very large in the big bang and could have had a major effect on the dynamics of the universe in its earliest moments. It is not implausible that particles created in this way could account for all the matter in the universe!

Attempts to calculate big-bang particle production were begun independently a decade ago by the Russian academician Yakov B. Zel'dovich and by Leonard E. Parker of the University of Wisconsin at Milwaukee. Many others have taken up the question since. Although several of the results are suggestive, none is definitive. Moreover, a major question hangs over the effort: What does one choose for the initial quantum state at the instant of the big bang? Here the physicist must play God. None of the proposals made so far seems uniquely compelling.

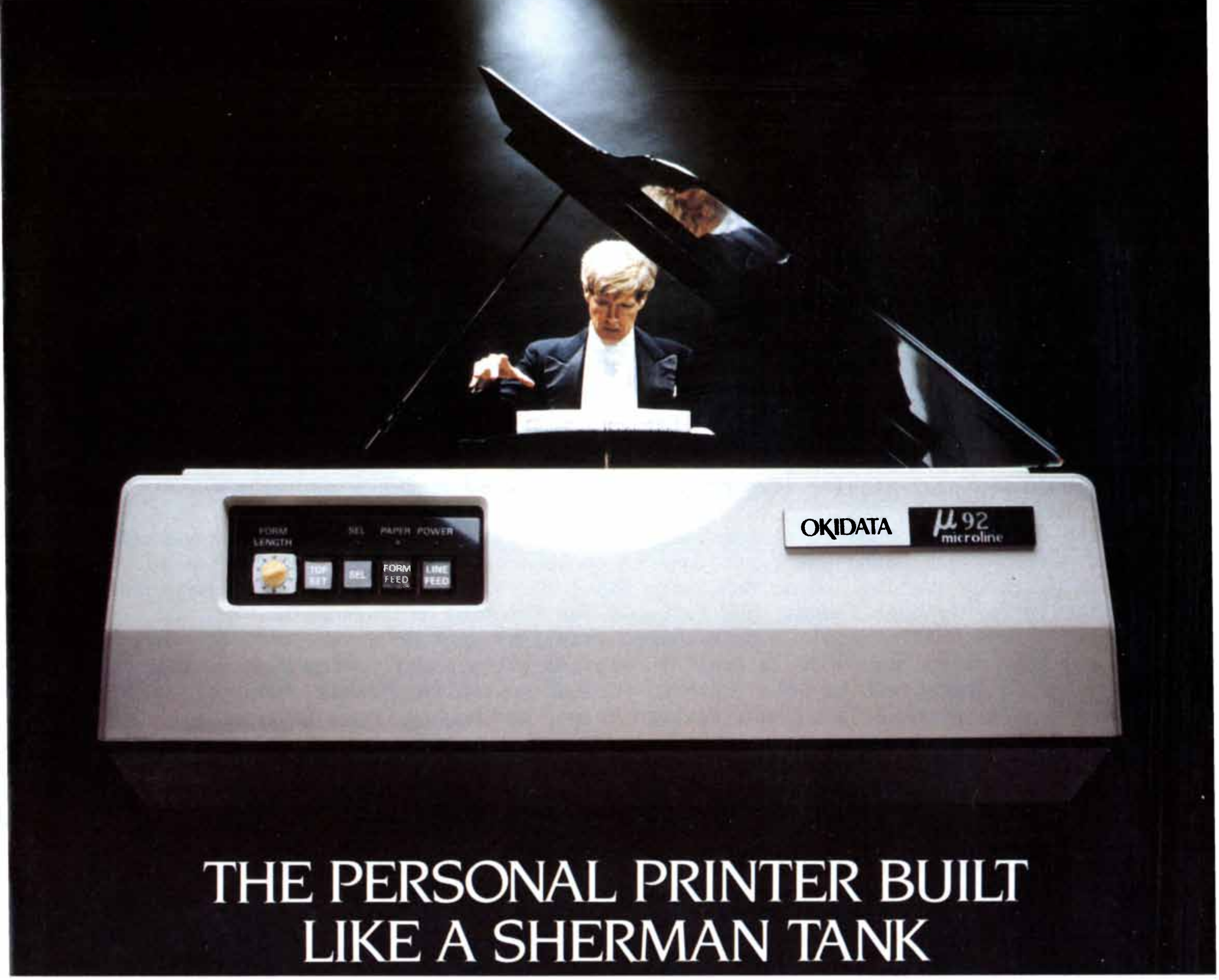
The one other event in the universe during which curvature should change rapidly is the collapse of a star to form a black hole. Here quantum-mechanical calculations have led to a real surprise, essentially independent of initial conditions. In 1974 Stephen W. Hawking of the University of Cambridge showed that the variation in curvature near a collapsing black hole creates a radiating stream of particles. The stream is steady and continues long after the black hole has become geometrically quiescent. It can continue because time seems to slow down in the enormous gravitational field near the "horizon" surface of a black hole; for an external observer all activity comes to a virtual standstill. Particles born closer to the horizon are delayed longer in their outward journey.

Although the delay in emission implies that there must be an enormous number of particles crowded near the horizon, each "waiting its turn" to escape, the total energy density in this region is actually negative and rather small. The positive energy carried by the particles is largely compensated for by an enormously negative vacuum energy that would be there if the particles were absent (for example, if the black hole had always existed and had never been created by gravitational collapse).

It can be shown that the particles emitted are statistically uncorrelated and that their energy spectrum is thermal. The black-body character of the radiation is perhaps its most important property. It allows both a temperature and an entropy to be assigned to a black hole. The entropy, which measures the



DISTANT REGIONS OF THE UNIVERSE could in principle be connected by a wormhole, suggesting that faster-than-light communications might be established between them; actually the scheme cannot work. In the drawing of a wormhole at the upper left the distance between the holes in the "outside world" is comparable to the distance through the "throat." In the wormhole at the lower left the outside distance is much greater. In the lower drawings the space represented by the plane appears to be bent, but that is only because it is viewed from the perspective of a higher-dimensional space; to an observer living in the plane it would appear approximately flat. Whether or not the throat is a short cut, it is not possible to pass through it. The reason is that a wormhole invariably connects two black holes. The throat "pinches off," as is shown at the right, and anything that enters it is crushed to infinite density before reaching the other side.



PERFORMS LIKE A CONCERT GRAND.

Why We Get Encores. Okidata takes center stage with a cast of printers that can't be outperformed. All eight dot matrix printers offer you more features for your money than you can find anywhere else. Pick your tempo: data processing at speeds from 80 to an exceptional 350 cps; to stress a point, enhanced and emphasized printing at up to 100 cps; outstanding letter quality printing at speeds three times faster than most daisywheels — up to 85 cps. Add a full range of graphics capabilities, downline loadable character sets for creating personalized typefaces and symbols, and your print repertoire is virtually unlimited.

We Play On and On. Our virtuosos feature rugged steel frames, laser-welded parts, and our long-life, non-ballistic print head warranted for up to one full year. With this tank-toughness you'd expect Okidata to have the lowest warranty claim rate in the industry. And we do: less than ½%.

In Tune with All Major Computers. We've designed each of these finely tuned instruments to be harmonious with all the major names in personal computers. And to give you more than you'd get from the major computer name printers. After all, we specialize in printers. The computer folks specialize in computers. (That's why MOST buy their printers from somebody else).

Larger Selection. Smaller Prices. Because we make more printers than anybody else, we can give you just the right one to fit your specific needs. Not to mention your budget. Suggested retail prices range from $449 to $2995. Call 1-800-OKIDATA (in NJ, 609-235-2600) for the dealer nearest you. Both you and your computer will enjoy the performance.

OKIDATA

Mt. Laurel, NJ 08054

A subsidiary of Oki Electric Industry Company Ltd.

SCIENCE/SCOPE

Very High Speed Integrated Circuit technology will be introduced in a system that lets U.S. Army and Marine Corps units automatically report navigation data and their locations to command centers. Hughes Aircraft Company VHSIC chips will be used in the Position Location Reporting System. The chips will significantly increase communications and encoding capability, and reduce vulnerability to jamming. The VHSIC program is being conducted by the Department of Defense to develop chips that will give military electronic systems a tenfold increase in signal processing capability. The high-speed, compact VHSIC chips will be more reliable and will require less power than integrated circuits now in use.

Seven of every 10 commercial communications satellites now in service throughout the world -- including the first four to be launched from the space shuttle -- were built by Hughes. Those four were SBS-3 for Satellite Business Systems, two Anik-C's for Telesat Canada, and Palapa-B for Perumtel of Indonesia. All are based on the HS-376 model, a drum-shaped satellite with an outer panel of solar cells that drops down when in orbit to provide extra power. Nine customers have ordered 27 spacecraft to date. Others include Australia, Brazil, Mexico, American Telephone & Telegraph, Western Union, and Hughes Communications.

A uniquely shaped waveguide antenna is one of 13 patentable innovations built into the Advanced Medium-Range Air-to-Air Missile. The antenna is configured to occupy a very small space and yet provide a low-frequency-band data link to launching aircraft. Its novel shape also minimizes interference and provides a moderate amount of cross-polarization, a feature that improves communications. Hughes designed and developed AMRAAM for the U.S. Air Force and Navy.

Computers are boosting productivity in the electronics business in many ways. At Hughes, for example, one system saved untold time and money in the design of a heat exchanger casting. The computer prepared engineering drawing dimensions and tolerance data in far less time than manual methods. Then, in simulating how numerical control machines would prepare the casting, it discovered the device would have been destroyed in one phase of manufacturing. Engineers thus were able to redesign the item even before a prototype was built.

"Diversification" characterizes Hughes Industrial Electronics Group, located in the Southern California communities of Carlsbad, Irvine, Newport Beach, Sylmar, and Torrance. Our facilities are like small companies with 500 to 2,500 employees, but each offers the resources and benefits of a multibillion-dollar company. Our diverse technologies include silicon and GaAs semiconductors, fiber optics, microwave and millimeter-wave communications, microprocessors, lasers, and solar cells. Send your resume to B.E. Price, Hughes Industrial Electronics Group, Dept. SE, P.O. Box 2999, Torrance, CA 90509. Equal opportunity employer.

Creating a new world with electronics

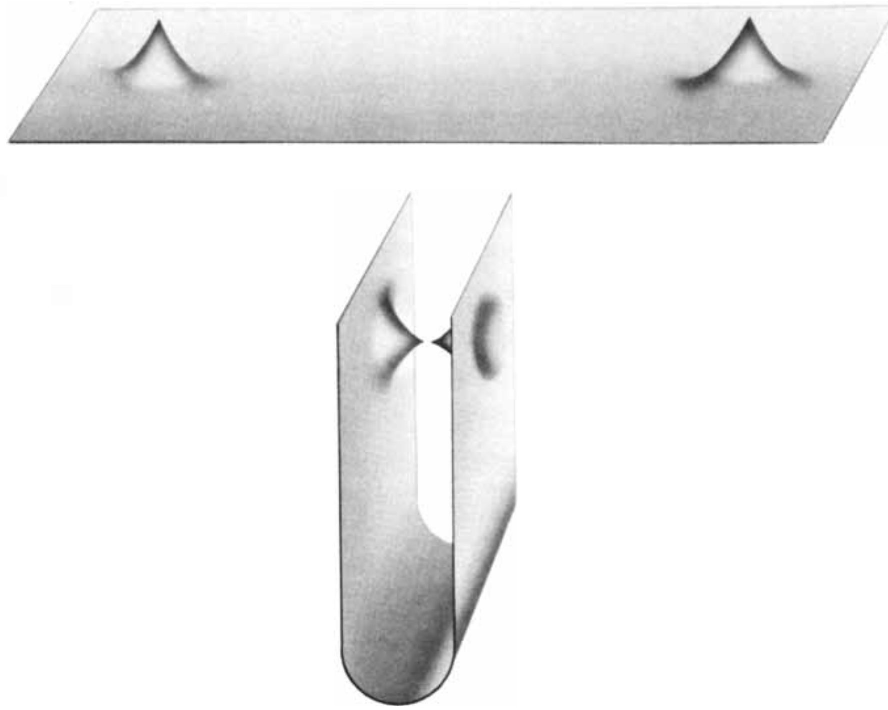
HUGHES

HUGHES AIRCRAFT COMPANY

For additional information please write to
P.O. Box 11205, Marina del Rey, CA 90295

thermodynamic disorder of the system, turns out to be proportional to the area of the horizon surface. For a black hole of stellar mass it is enormous: more than 19 powers of 10 larger than the entropy of the star that collapsed to form the black hole. The temperature, on the other hand, is inversely proportional to the mass and, if the mass is stellar, is more than 11 powers of 10 smaller than that of the parent star.

Because the quantity of radiation emitted by an object depends on its temperature, the Hawking radiation from an astrophysical black hole is utterly negligible. It becomes important only for "mini" black holes, those with a mass of less than about 10^{10} grams. The only conceivable way that mini black holes could have been formed is through compression during the big bang. It is possible they were copiously produced then, in which case they would have contributed significantly to the entropy of the universe.



FLUCTUATING TOPOLOGY, which is a feature of spacetime in some attempts to formulate a theory of quantum gravity, raises serious conceptual difficulties. Shown here are two representations of a wormhole that has just pinched off, leaving two "dimples." If such an event can take place, the reverse process should also be possible; in other words, the dimples should be able to coalesce into a new wormhole. The reversal seems plausible where the dimples appear to be close together but not where they appear to be far apart. The ideas of "close" and "far," however, depend on embedding the surface in a higher-dimensional space. For an observer within the surface the objects represented by the two drawings would be indistinguishable.

The energy of particles created by a time-varying curvature cannot be conjured up from nowhere. It is taken from spacetime itself. It follows that the particles act back on spacetime. Attempts have been made to calculate this "back reaction" in the case of the big bang, to determine its dynamical effect on the early universe. One goal has been to see whether the back reaction might suppress the infinite initial density of matter required in Einstein's classical theory. The infinite density is a barrier to all further inquiry. If it could be replaced by a merely enormous density, one might ask: What was the universe doing *before* the big bang?

In the 1960's Roger Penrose of the University of Oxford and Hawking showed that Einstein's classical theory is incomplete. It predicts the past or future occurrence of infinite densities and infinite curvatures under a variety of physically reasonable present conditions. A theory that predicts an infinite value for an observable quantity ceases to be able to predict beyond that point. Since physicists believe in the ultimate comprehensibility of nature, they expect that such a theory needs to be broadened, to include a wider variety of phenomena. The conservative view at present is that the inclusion of quantum effects is the only reasonable cure in sight for the incompleteness of Einstein's theory.

Calculations of the back reaction on the big bang are done by numerical simulation with a digital computer. So far they have given ambiguous results. One difficulty has been to determine, as input for the computer, a believable value for the combined energy density of the particles produced and the quantum vacuum on which they are superposed.

The effect of the back reaction is of special importance in the case of a black hole. The Hawking radiation steals both entropy and energy from a black hole. The mass of the hole consequently decreases. The rate of decrease is slow at first but picks up as the temperature rises. Ultimately the rate of change becomes so great that the approximations made in Hawking's calculations fail. It is not known what happens thereafter. Hawking thinks that his approximations continue to be qualitatively correct and that the black hole ends its life in a spectacular flash, momentarily leaving behind a "naked singularity" in the causal structure of spacetime.

Any singularity, whether naked or not, represents a failure of the theory. If Hawking is right, not only is Einstein's theory incomplete but also the quantum theory is incomplete. The reason is that for every particle born outside the horizon surface another is born inside. The two particles are correlated in the sense that an observer could detect "probability interference effects" if he could communicate simultaneously with both particles. Hawking assumes that the particles inside get crushed to infinite density and cease to exist. The moment they cease to exist the standard probability interpretation of quantum mechanics fails. Probability itself gets lost in the infinite crush.

An alternative and equally plausible assumption is that the very framework of quantum field theory one erects around Einstein's theory prevents both probability and information from being lost in the collapse. It is quite possible the back reaction becomes so extreme that the crush is prevented from becoming infinite. The horizon, which is a mathematical construct rather than a physical one, may not be there at all as a strict one-way barrier. The matter that collapsed to form the black hole may ultimately be accounted for, particle by particle. That there will be a final flash of Hawking radiation and enormous densities inside the hole no one doubts. The very pressures to which the nuclear particles are subjected, however, may convert them into photons and other massless particles, which may finally escape, carrying with them the small remaining energy and all the quantum correlations. These end products need carry none of the original entropy of the black hole. That has all been stolen by the Hawking radiation.

I come now to the deep and difficult part of quantum gravity. When a quantum effect, such as particle production or vacuum energy, reacts back on the curvature of spacetime, the curvature itself becomes a quantum object. A consistent theoretical framework demands that the gravitational field itself be quantized. For wavelengths that are long compared with the Planck length the quantum fluctuations of the quan-


THE · 1984 · CHRYSLER

tized gravitational field are small. They can be accurately represented if they are treated as a weak disturbance on a classical background. The disturbance can be analyzed the way an independent field would be. It contributes its share both to the vacuum energy and to particle production.

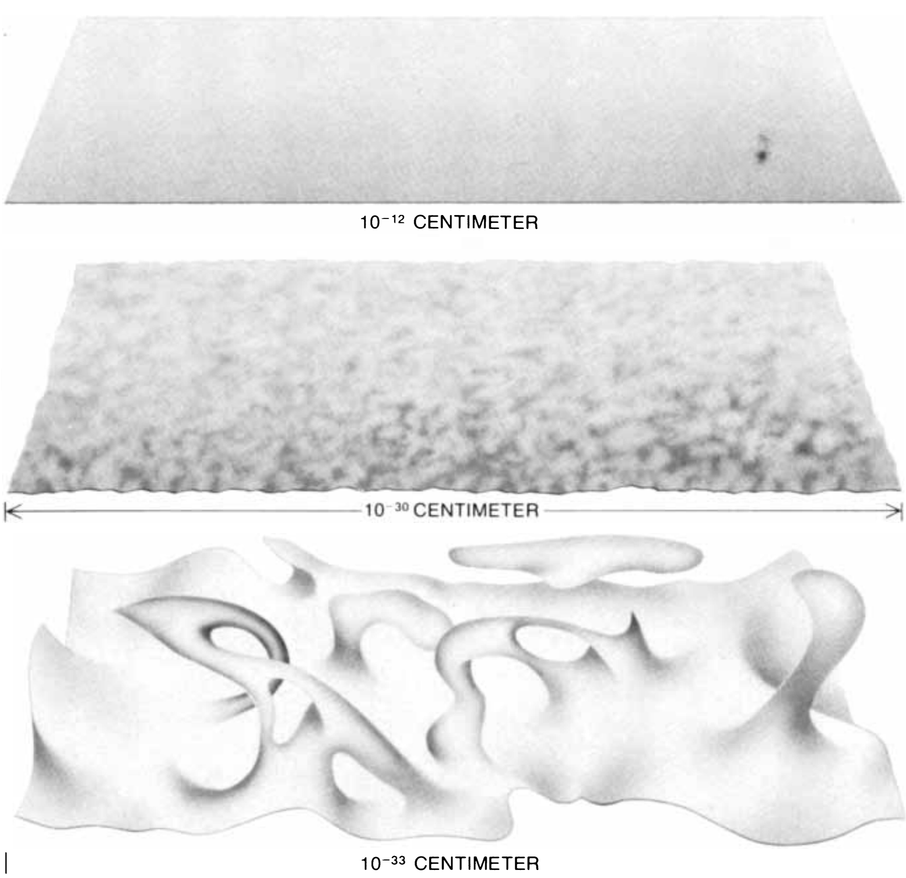
At Planckian wavelengths and energies the situation is profoundly more complicated. The particles associated with a weak gravitational field are called gravitons; they are massless and have a spin angular momentum of $2\hbar$. It is unlikely that individual gravitons will ever be observed directly. Ordinary matter, even an entire galaxy of ordinary matter, is almost totally transparent to them. Only when they reach Planckian energies do they interact appreciably with matter. At these energies, however, they are capable of inducing Planckian curvatures in the background geometry. The field with which they are associated is then no longer weak and the very concept of "particle" is ill-defined.

At long wavelengths the energy carried by a graviton distorts the background geometry. At shorter wavelengths it distorts the waves associated with the graviton itself. This is a consequence of the nonlinearity of Einstein's theory: when two gravitational fields are superposed, the resulting field is not equal to the sum of the two components. All nontrivial field theories are nonlinear. For some it is possible to deal with the nonlinearity by a method of successive approximations called perturbation theory, a name taken originally from celestial mechanics. The essence of the method is to refine an initial approximation by making a series of progressively smaller corrections. When perturbation theory is applied to quantized fields, it leads to infinities that must be eliminated by renormalization.

In the case of quantum gravity perturbation theory does not work. There are two reasons. First, at Planckian energies the successive terms of the perturbation series (that is, the successive corrections) are comparable in magnitude. Truncating the series at a finite number of terms does not yield a valid approximation; instead the entire infinite series must be summed. Second, the individual terms of the series cannot be consistently renormalized. In each order of approximation new classes of infinities appear, which have no counterparts in ordinary quantum field theory. They arise because in quantizing the gravitational field one quantizes spacetime itself. In ordinary quantum field theory spacetime is a fixed background. In quantum gravity the background not only reacts to the quantum fluctuations but also shares in them.

In a narrow technical response to these difficulties a few attempts have been made to sum infinite subsets of terms of the perturbation series. The results, notably the complete disappearance of infinities, are both suggestive and encouraging. The results must be viewed with caution, however, because severe approximations are made in obtaining them and the perturbation series is never summed in toto. They are nonetheless being used to compute improved estimates of the effect of the back reaction on the big bang.

From a broader point of view one must expect other problems to arise whose resolution cannot even be approached by summing series. A quantized spacetime is one whose causal structure is fluctuating and uncertain. At Planckian dimensions the very distinction between past and future becomes blurred. By analogy with the phenomenon of tunneling in atomic systems, which allows an electron to pass through an energy barrier it cannot climb over, one must expect processes that are not allowed in Einstein's classical theory, including faster-than-light travel over Planckian distances. How the probabilities of such processes should be calculated is largely unknown. In many cases it is not even known which are the right questions to ask. There are no experiments to guide us. It is therefore still possible to indulge in flights of fancy.



QUANTUM VACUUM, as envisioned by John Archibald Wheeler in 1957, becomes increasingly chaotic as one inspects smaller regions of space. At the scale of the atomic nucleus (*top*) space looks very smooth. At dimensions of 10^{-30} centimeter (*middle*) a certain roughness in the geometry begins to appear. At the scale of the Planck length, 1,000 times smaller still (*bottom*), the curvature and the topology of space are continually undergoing violent fluctuations.

One of the most persistent flights of fancy, referred to repeatedly in the literature on quantum gravity, is the idea of fluctuating topology. The basic notion, which was introduced by Wheeler in 1957, is the following. The vacuum fluctuations of the gravitational field, like those of all other fields, increase in magnitude at shorter wavelengths. If standard weak-field results are extrapolated into the Planck domain, the curvature fluctuations become so violent that they appear to be capable of tearing holes in spacetime and changing its topology. Wheeler imagines the vacuum as being in a state of perpetual turmoil, with wormholes (and more complicated structures) of Planckian size continually forming and disappearing. The turmoil is "visible" only at the Planck level. At a coarser level spacetime continues to appear smooth.

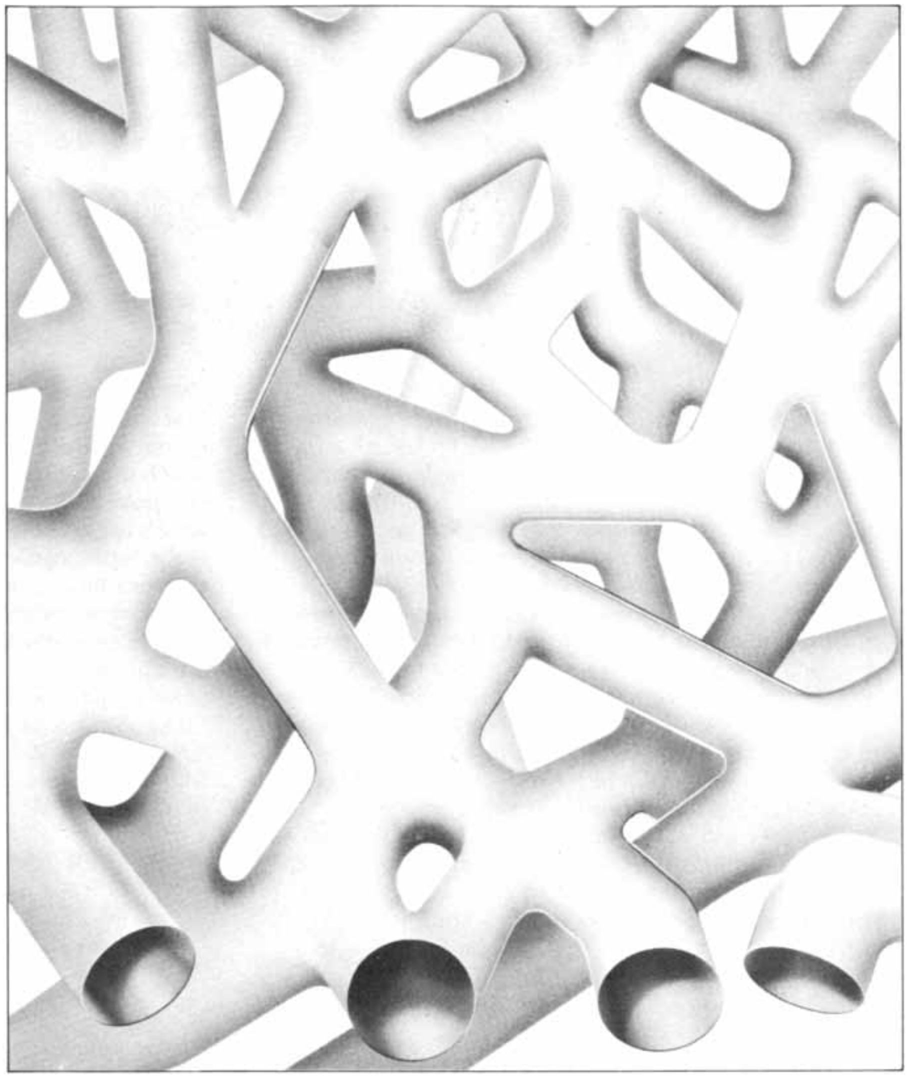
An immediate objection can be raised: Every topological change is necessarily accompanied by a singularity in the causal structure of spacetime, so that one is faced with the same difficulty that arises in Hawking's view of black-hole decay. Suppose, however, Wheeler's view is correct. One of the first questions that must be asked is: How much do the topological fluctuations contribute to the energy of the vacuum and how do they affect the resistance of spacetime (at the coarse level) to bending? To date no one has given a convincing answer, largely because no coherent picture has emerged of the topological transition process itself.

To appreciate just one of the obstacles in the way of constructing such a picture, consider the process shown in the illustration on page 123. The drawing gives two representations of the same event: a wormhole has just pinched off, leaving two residual "pseudopods" on a simply connected space. In one representation space is shown bent, in the other it is shown flat.

Now consider the reverse process: the formation of a wormhole. If there is a finite probability for a wormhole to disappear by pinching off, there is also a finite probability for one to form. Here a new difficulty arises. From the time-reversed point of view the illustration represents two pseudopods that have grown spontaneously in the quantum vacuum. In one representation the possibility that the two pseudopods may join to form a wormhole seems reasonable. In the other it does not. And yet the physical situation is the same in both drawings. The formation of the wormhole seems reasonable in the one case because the pseudopods appear to be close together. The "closeness," however, is not an intrinsic property of the spatial arrangement, as the other representation clearly shows. A notion of "closeness" requires the existence of a higher-dimensional space in which spacetime is embedded. Furthermore, the higher-dimensional space must be endowed with physical properties in order for the pseudopods to transmit a sense of nearness to each other. But then spacetime is no longer the universe. The universe is something more. If one remains faithful to the view that the properties of spacetime are intrinsic and are not the result of something outside, a coherent picture of topological transitions seems beyond reach.

Another difficulty with topological fluctuations is that they might undermine the macroscopic dimensionality of space. If wormholes can form spontaneously, the wormholes themselves can develop wormholes, ad infinitum. Space could evolve into a structure that, although three-dimensional at the Planck level, has four apparent dimensions or more at a larger scale. A familiar example of this process is the formation of a foam, which is made up entirely of two-dimensional surfaces but has a three-dimensional structure [*see illustration on this page*].

Because of difficulties such as these some physicists have suggested that the conventional description of spacetime as a smooth continuum fails at the Planck level and must be replaced by something else. What that something else consists of has never been made very clear. In view of the success of the continuum description over scales of length spanning more than 40 powers of 10 (60 if the possible failure is assumed to come only at the Planck level), it would seem equally reasonable to suppose that the continuum description is valid at all levels and that topological transitions simply do not exist.



DIMENSIONALITY OF SPACE is put into question by the possibility that spacetime has a complex topology. The surface shown is two-dimensional, but its topological connections give it the appearance of a three-dimensional object. It is conceivable that three-dimensional space perceived at a macroscopic scale actually has fewer dimensions but is topologically convoluted.

Even if the topology of space is unchanging, it is not necessarily simple, even at a microscopic level. It is conceivable that space could have a foam structure, built in from the beginning, in which case its apparent dimensionality could be greater than its real dimensionality. Its apparent dimensionality could also be less than its real dimensionality.

The latter possibility was proposed in a theory put forward by Theodor Kaluza in 1921 and by Oskar Klein in 1926. In the Kaluza-Klein theory space is four-dimensional and spacetime is five-dimensional. The reason space *appears* to be three-dimensional is that one of its dimensions is cylindrical, as in the universe discussed above, but with an important difference: the circumference of the universe in the cylindrical direction, instead of being billions of light-years, is only a few (perhaps 10 or 100) Planck units. As a result an observer who attempts to penetrate the fourth spatial dimension is almost instantly back where he started. Indeed, it is meaningless to speak of such an attempt, because the very atoms of which the observer is composed are vastly larger than the cylindrical circumference. The fourth dimension is simply unobservable as such.

It can nonetheless manifest itself in another way: as light! Kaluza and Klein showed that if their five-dimensional

Over a million people use Microsoft® software on their microcomputers. There's a reason.

Powerful simplicity. That's the concept behind every microcomputer software program we write. Powerful programs that allow you to do more with less effort. Programs that let you spend more time thinking about your problem...and less time thinking about the computer. It's a good reason why over a million people depend on Microsoft software in their everyday computing... and why the number is growing. Rapidly.

The value of experience. Microsoft made a giant step toward putting the power of computers in the hands of non-technical people when we wrote the first BASIC programming language for the very first microcomputer. It's an English-like programming language simple enough to teach in grade schools. Yet, it's so powerful that more application programs have been written in Microsoft BASIC than any other programming language.

Tools that help you think. Microsoft is an old-timer in one of the world's youngest industries. Microsoft BASIC was just the beginning. Today, we offer a broad range of proven tools for microcomputers. A complete line of programming languages. From the original BASIC to Pascal. Operating systems like XENIX™ and MS™-DOS that allow your computer to utilize a wide variety of programs. Products like the Microsoft SoftCard™ system which dramatically extends the capabilities of Apple® II computers. Management software such as Multiplan,™ which allows you to explore more alternatives to reach more intelligent business decisions. Better tools. Because they're powerfully simple.

Building a library. The great majority of microcomputers come equipped with Microsoft software. But you'll undoubtedly want more. Specialized programs that solve your particular problems. Programs for your industry, profession, or home. Programs that do more and ask less. Look to Microsoft, or to programs written in Microsoft languages. Software that's proven on millions of computers. Worldwide.

BETTER TOOLS FOR MICROCOMPUTERS

MICROSOFT®

MICROSOFT CORPORATION
10700 NORTHUP WAY
BELLEVUE, WASHINGTON 98004

spacetime is treated mathematically in exactly the same way that four-dimensional spacetime is treated by Einstein, their theory is equivalent to Maxwell's theory of electromagnetism combined with Einstein's theory of gravitation. The components of the electromagnetic field are implicit in the equation satisfied by the curvature of spacetime. Kaluza and Klein thus invented the first successful unified field theory, a theory that gave a geometric explanation of electromagnetic radiation.

In a sense the Kaluza-Klein theory was too successful. Although it united Maxwell's and Einstein's theories, it predicted nothing new and so could not be tested against other theories. The reason was that Kaluza and Klein put constraints on the way spacetime is allowed to curve in the extra dimension. If the constraints had been removed, the theory would have predicted new effects, but the effects did not seem to correspond to reality. Hence the theory was viewed for many years as a beautiful curiosity and relegated to a closet.

The Kaluza-Klein theory was brought out of the closet in the 1960's when it was realized that the new gauge theories that were attracting increasing attention could be reformulated as Kaluza-Klein theories in which space is endowed with not merely one but several extra microscopic dimensions. It began to appear that all of physics might be explainable in geometric terms. It then became important to ask what happens if the constraints on the curvature in the compact dimensions are removed.

One thing that happens is that curvature fluctuations are predicted in the extra dimensions; the fluctuations are manifested as massive particles. If the circumference of the extra dimensions is 10 Planck units, the mass of the associated particles is roughly a tenth of the Planck mass, or about a microgram. Because the energy needed to create such particles is enormous, they almost never get produced. It therefore makes little practical difference whether the constraints on curvature fluctuations are imposed or not. Problems do remain. The chief one is that the extreme curvature of the extra dimensions gives rise to a very large energy density in the classical vacuum. The large vacuum energy is ruled out by observation.

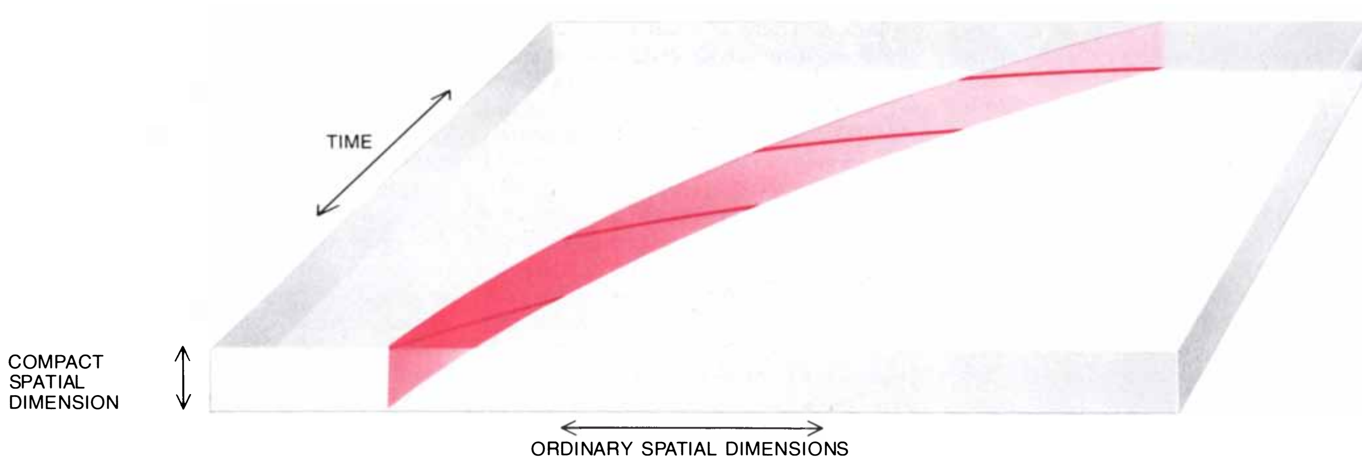
Kaluza-Klein models have never been pursued with undivided attention, and their role in physics is still uncertain. In the past two or three years, however, they have again been much scrutinized, this time in connection with the remarkable generalization of Einstein's theory known as supergravity, invented in 1976 by Daniel Z. Freedman, Peter van Nieuwenhuizen and Sergio Ferrara and (in an improved version) by Stanley Deser and Bruno Zumino.

One of the inadequacies of standard Kaluza-Klein models is that they predict the existence only of particles with spin angular momenta of 0, $\hbar$ and $2\hbar$, and even these particles are either massless or supermassive. Nowhere is there room for the particles of ordinary matter, most of which have spin angular momentum $\frac{1}{2}\hbar$. It turns out that if Einstein's theory is replaced by supergravity and if spacetime is given the Kaluza-Klein treatment, a true union of all the spin varieties is achieved.

In the "super" Kaluza-Klein model that is currently most popular seven extra dimensions are added to spacetime. These dimensions have the topology of a 7-sphere, a space that has some fascinating properties of its own. The resulting theory is extraordinarily rich and complicated, specifying huge supermultiplets of particles. The masses of the particles are still either zero or extremely large, but it is possible that a "breaking" of the symmetry of the 7-sphere will give some of the particles a more realistic mass. The large energy of the classical vacuum also persists, but it might be canceled by a negative quantum vacuum energy. It is not yet known whether these strategies for rectifying the theory will succeed. Indeed, much work will be needed to determine exactly what the theory actually states.

If Einstein could come back in spirit to witness what has become of his theory, he would certainly be astonished, and I think pleased. He would be pleased that physicists at last, after years of hesitation, have come to accept his view that theories that are mathematically elegant deserve to be studied even if they do not seem to correspond immediately to reality. He would also be pleased that physicists now dare to hope a unified field theory may be attainable. He would be particularly pleased to find his old dream that all of physics may be explainable in geometric terms seems to be coming true.

Above all he would be astonished. Astonished that the quantum theory still stands pristine and unmoved in the midst of it all, enriching field theory and itself being enriched by it. Einstein never believed the quantum theory expresses ultimate truth. He never reconciled himself to the indeterminism it implies and thought it would someday be replaced by a nonlinear field theory. Exactly the opposite has happened. The quantum theory has invaded Einstein's theory and transmuted it.



ADDITIONAL SPATIAL DIMENSIONS, beyond the three known ones, could exist if they had a "compact" form. For example, a fourth spatial dimension could be rolled up into a cylinder with a circumference of perhaps 10^{-32} centimeter. Here the hypothetical compact dimension has been "unrolled" and is represented as the vertical axis in a spacetime diagram. The path of a particle can therefore have a cyclical component; each time it reaches the maximum extent of the compact dimension it finds itself back where it started. The observed path is a projection of the actual path onto the macroscopic spacetime dimensions. If the path is a geodesic one, it can have the appearance of an electrically charged particle moving in an electric field. A theory of this kind was introduced in the 1920's by Theodor Kaluza and Oskar Klein, who showed it could explain both gravitation and electromagnetism. There has been a resurgence of interest in such theories.

Prey Switching in a Simple Ecosystem

On the island of Newfoundland, which has few species of mammals, lynxes preyed on snowshoe hares until the hare population "crashed." Then the lynxes switched to caribou calves, and the cycle continues

by Arthur T. Bergerud

In most parts of the world the ecological chain along which matter and energy are transferred from simple photosynthetic organisms to carnivorous mammals by way of the intervening herbivores is an intricate one. In any small area there are generally many species of prey and predators. For each predator there are several types of prey and for each prey several types of predators. In some regions, however, geographic conditions have led to the formation of a relatively simple ecological system including only a few species. Such ecosystems can provide a laboratory for the study of the principles that govern the size of animal populations. One of the ecosystems is on the large island of Newfoundland, where there are only 14 indigenous mammals.

The array of animals in Newfoundland is not only small but also unbalanced: nine of the 14 species are carnivorous. In a region where there are so many carnivores and so few herbivores for them to eat, any shift in the prey species can have strong reverberations; the damping of the effects that would take place in a more complex ecosystem is absent. Such a dramatic effect was observed in Newfoundland early in this century when the caribou and the arctic hare, two of the chief prey animals on the island, began to decrease in number. By means of a long and thorough investigation it was determined that in both instances the cause of the decrease was a rapid increase in the number of lynxes.

For many centuries the lynx was rare in Newfoundland. In the 1860's, however, the snowshoe hare was introduced to the island, and with this species as a source of food the lynx population quickly expanded. By the early part of this century the snowshoe hares had reached their ecological limits. The snowshoe hare population "crashed," that is, diminished rapidly, leaving the lynx without its main food item. As a result the lynx, a resourceful predator, began to prey on caribou calves and the other species of hare: the arctic hare. When the snowshoe hare increased again, the lynx returned to preying on it.

The cycle has been repeated many times. Thus in recent years the lynx, the snowshoe hare and the caribou have been connected in a triangle in which the populations of the two prey species oscillate in alternating cycles of increase and decrease. The lynx also limits the number of arctic hares, once numerous in Newfoundland but now rare. The introduction of a single new species into the simple ecosystem of the island has had a long, tangled skein of consequences for the equilibrium of the system. The unraveling of this skein could illuminate the dynamics of animal populations in finer-spun ecological webs.

The ecological system of Newfoundland is quite distinct from that of the nearby parts of the Canadian mainland. Several factors have contributed to this isolation. There is evidence that when the Wisconsin ice sheet advanced some 18,000 years ago, the ice did not cover all Newfoundland. Part of the island could have been what is called a refugium: an area where local mammals could survive. As a result some native species were preserved and continued to evolve independently of the mammals on the mainland. Nine of the 14 indigenous mammals in Newfoundland are distinct enough from those on the mainland to be classified as subspecies, which is evidence of thousands of years of genetic isolation.

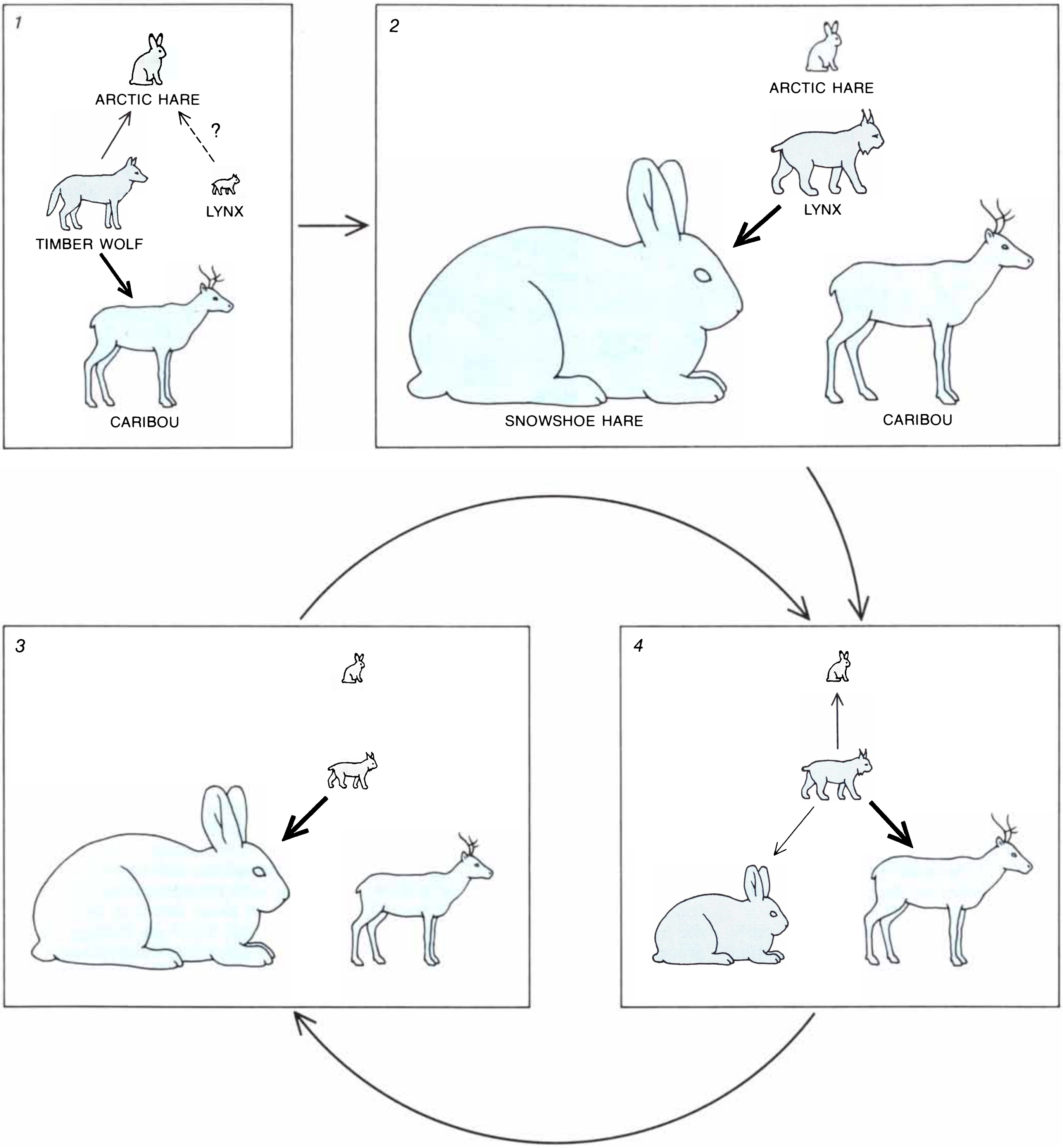
Some species, particularly predators, could have reached the island on ice floes or ice bridges in the postglacial period. The sea, however, has continued to act as a barrier to a full biological exchange between the island and the coastal areas of mainland Canada. In Labrador, which is only 17.6 kilometers from Newfoundland across the Strait of Belle Isle, there are 34 mammalian species. On Cape Breton Island, which is 112 kilometers southwest of Newfoundland and is connected to the mainland, there are 38 species.

The sea around Newfoundland has acted not only as a barrier, reducing the overall number of species, but also as a filter, selecting certain types of mammals and preventing others from reaching the island. The adventurous predators were more likely than their prey to colonize the island. There were originally seven carnivorous quadrupeds on the island: the timber wolf (*Canis lupus*), the red fox (*Vulpes vulpes*), the lynx (*Lynx lynx*), the otter (*Lutra canadensis*), the black bear (*Ursus americanus*), the weasel (*Mustela erminea*) and the pine marten (*Martes americana*). The five herbivores on which these predators fed before the arrival of the snowshoe hare are the Canadian beaver (*Castor canadensis*), the North American meadow mouse (*Microtus pennsylvanicus*), the caribou (*Rangifer tarandus*), the arctic hare (*Lepus arcticus*) and the muskrat (*Ondatra zibethica*). There are also two species of bats.

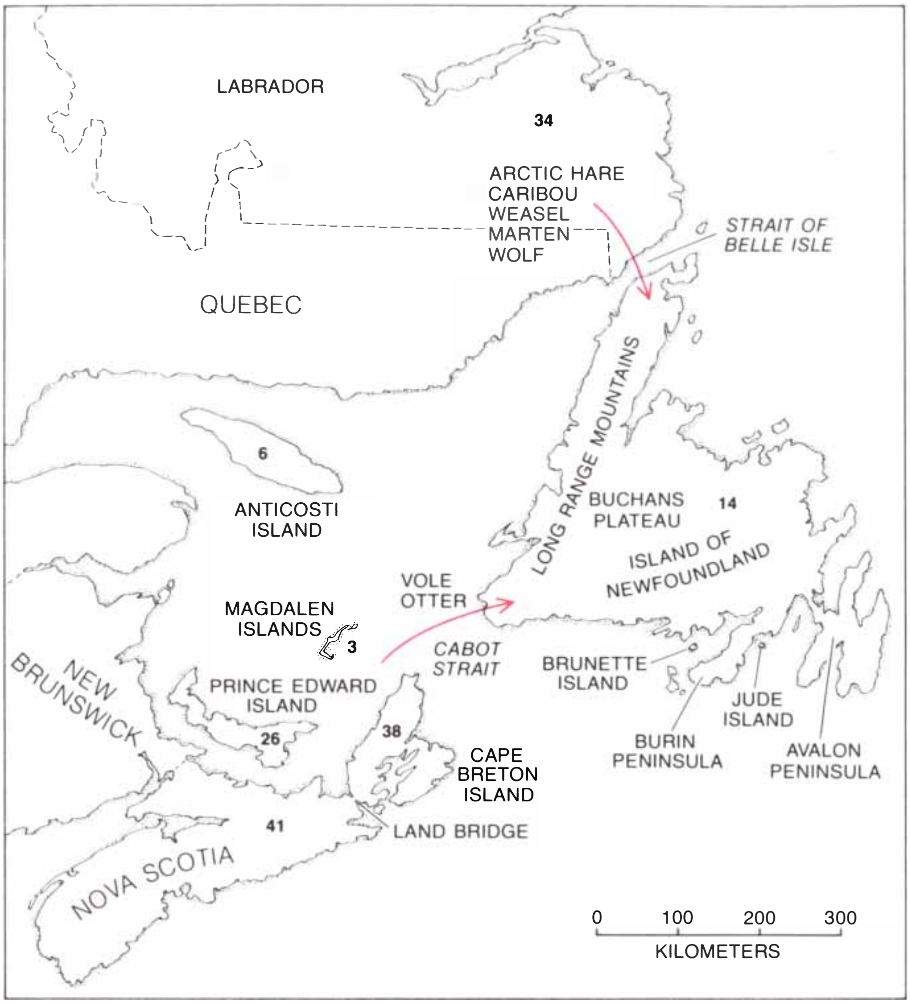
In this array the usual pyramid of species is inverted. Since biological systems, like all other systems of energy transfer, have a limited efficiency, the total mass of predators is much less than the total mass of prey. By the same token the number of predator species is generally smaller than the number of prey species. Simple inverted faunas such as the one in Newfoundland are fragile, partly because they constitute an exception to the rules that normally govern food chains. My own experience in Newfoundland shows just how vulnerable such a system can be. From 1956 to 1967 I served as provincial biologist for the Newfoundland Division of Wildlife. My main task was to find the cause of a sharp decrease in the number of caribou. What I found was also relevant to other species.

In 1900, when the interior of Newfoundland was uninhabited except for the small tribe of Micmac Indians, the caribou herds were quite large. A. A. Radclyffe Dugmore and J. G. Millais, two early naturalists, estimated that at the end of the 19th century the caribou on the island numbered between 150,000 and 200,000. These estimates, however, could be no more than educated guesses; Dugmore and Millais had made no attempt to count or sample the caribou herds systematically. On the basis of interviews with older hunters, records of the caribou harvest and the size of the available range, I have estimated that the population of caribou in Newfoundland in 1900 was 40,000, still a considerable number of animals for an island with an area of only 100,000 square kilometers.

Only 25 years later the caribou was



PREY-SWITCHING CYCLE in Newfoundland is shown schematically. For millenniums up to the middle of the 19th century the main predator on the caribou was the timber wolf. There were few native lynxes (*1*). Snowshoe hares were brought to the island as a source of food in 1864. They multiplied rapidly, reaching a peak in about 1900 (*2*). The lynxes, feeding on snowshoe hares, also multiplied rapidly. The wolf became extinct in about 1911. In about 1915 the snowshoe hare population crashed and the lynxes switched to caribou calves and arctic hares, considerably reducing the number of both prey species (*3*). Meanwhile the snowshoe hares had begun an intrinsic 10-year cycle of population growth and decline. When the hare population peaked, the lynx switched back to them, allowing the caribou population to grow again (*4*). For decades the lynx has been switching between the caribou and the snowshoe hare in a 10-year cycle. Arctic hares have become quite rare; they now live mostly in upland areas and their population fluctuates less than the other populations.



ORIGINAL FAUNA of Newfoundland includes only 14 species of mammals. Nearby regions of mainland Canada have many more species. Five species in Newfoundland may have come from Labrador on ice floes or ice bridges; two may have come from the Maritime Provinces. The sea has been a barrier to species and a filter of the types of animals that reached the island. As a result the food-chain pyramid of species in Newfoundland is inverted: nine of the 14 species are carnivorous and only five are herbivorous. Such an inverted fauna is fragile.

almost extinct. In about 1925 Dugmore estimated that there were only 200 animals left. This remarkable decrease took place even though the timber wolf, which for millenniums had been the chief predator on the caribou, became extinct in about 1911.

In the late 19th century and the early 20th the native arctic hare population also decreased from its original size, although the original number of arctic hares and the rate of decrease are not as well documented as those of the caribou. The number of arctic hares that lived in Newfoundland before the island was heavily settled can only be guessed at, but most observers agree that the territory in which arctic hares lived once extended down the Northern Peninsula to Buchans Plateau and the southern Long Range Mountains. From there the range stretched east along the south coast to the Burin and Avalon peninsulas. This domain includes much of the area of the island.

The clearest sign of a major drop in the population of arctic hares is the reduction in the hare's range since 1900. Outram Bang, a naturalist who described seven of Newfoundland's 14 native species, wrote in 1913: "The arctic hare is now very rare and local in Newfoundland, occurring only on the tops of the high mountains." By the 1950's the arctic hares were limited to the highlands of the northern and southern Long Range Mountains and Buchans Plateau and extended along the south coast to the Bay d'Espoir. Even in this area the animals were rare. The total arctic hare population of Newfoundland in recent decades has probably been less than 1,000.

When I arrived on the island, there was much more evidence on the decrease of the caribou than there was on that of the arctic hares. The investigation of the caribou's decline concentrated initially on the factors that influence the mortality rate among caribou calves. In many species of mammal the young age classes are the weakest link in the survival of the population. Indeed, the reduction in the caribou herds appeared to be due in large part to an increased death rate among the calves.

The adult males in the caribou herd spend much of the year foraging apart from the females. In mid-October the sexes meet for a mating period that lasts for about 10 days. In this time some 80 percent of the cows are impregnated. Among the females that are not impregnated some become pregnant during a shorter period of sexual receptivity about 10 days later.

The gestation period of the caribou is approximately 229 days. Because of the synchronization of mating, almost all the calves are born in a 14-day period that begins around May 24. The females calve in large groups at specific sites, and they generally return to the same calving ground every year. Most of the calving grounds in Newfoundland are open, marshy areas surrounded by scrub forest. When the calving season ends, some herds scatter into the woods to escape flying insects.

Each spring from 1957 through 1964 biologists found many dead calves on the calving grounds. The remains were easily found because the mother stayed near the calf for several days after it had died. Many of the dead calves had abscesses on their neck resulting from infection by a common pathogenic bacterium: *Pasteurella multocida* mucoid type A. The immediate cause of death was septicemia, or blood poisoning, from the infection.

Since the female caribou calve on a well-synchronized annual cycle and the fraction of females that give birth is fairly constant from year to year, it is not hard to calculate the number of new members that would be added to a caribou herd each year in the absence of perturbing influences. On the basis of such calculations it appeared that the mortality rate among caribou calves was quite high and that the calves whose remains were found represented only a minority of those that had died. From 1958 through 1967 an average of 27 percent of the calves disappeared within two weeks of their birth. By counting females with calves and those without calves it was estimated that by October about 70 percent of the females that had given birth were missing their offspring. After October, however, the death rate dropped sharply; most of the calves that made it to October survived the winter. It was the mortality in the first four months of life that was limiting the size of the caribou herds.

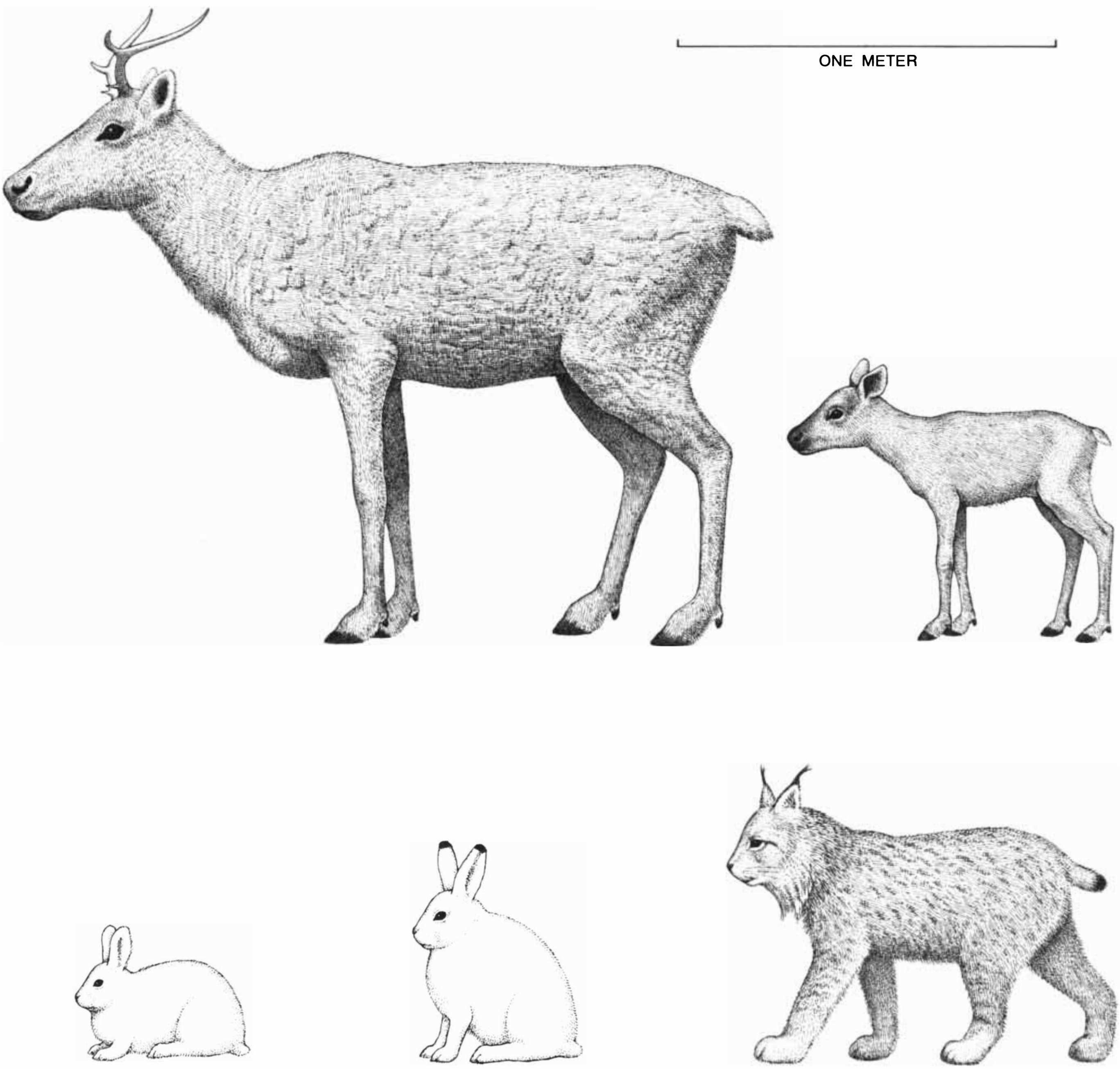
In addition to the missing calves and the concentration of the deaths in the spring there were other puzzles. First, at birth the ratio of male calves to female calves was 52 to 48. By the fall the ratio was reversed and there were 62 females to 38 males. Second, the overall mortality was greatest in the calving grounds at low elevations. This was an unexpected finding because in such areas the wind-chill factor was at its lowest. In the alpine areas, where the climate was considerably harsher, the death rate was lower. Most intriguing of all, as the herds were counted a cyclical pattern of survival began to emerge. Annual counts showed that the fraction of calves surviving increased in 1958 and 1959, then declined steadily from 1960 through 1963.

The pieces of the puzzles began to fall into place in 1964 when the cause of the abscesses was discovered. In the spring of that year a dead calf was found with four puncture wounds in its neck but no abscesses. It seemed probable that the other dead calves had also had puncture wounds but that the wounds had been concealed by the abscesses.

The limited fauna of Newfoundland made the search for the cause of the punctures relatively simple. Among the local predators only the canine teeth of the lynx were of the right size and arrangement to cause the four evenly spaced wounds. When the skull of a lynx was obtained for verification, the lynx canines fit neatly into the punctures.

Saliva from the mouth of a lynx was cultured and was found to contain *Pasteurella multocida*. Next three caribou calves were put in a pen with a lynx. The lynx seized each calf in turn by the neck. Observers kept the predator from killing its prey, and the calves were removed from the pen. Abscesses with *Pasteurella* infections developed quickly, and the calves died respectively four, five and 15 days after being bitten.

This piece of detective work suggest-



FOUR SPECIES IN THE PREY-SWITCHING CYCLE of Newfoundland are shown. The caribou (*Rangifer tarandus*) has a synchronized cycle of births. Calves are born on long-established calving grounds during a two-week period in May and June. In the 1950's and 1960's many calves were found dead on the calving grounds with abscesses on their neck. The primary defenses of the artic hare *Lepus arcticus* (*bottom middle*) and the snowshoe hare *Lepus americanus* (*bottom left*) are alertness and speed. The snowshoe hare inhabits the boreal forest, which in winter has loose-packed snow. The arctic hare lives on the windswept tundra, which has hard-packed snow. Both hares are shown in their winter coloration. The lynx (*Lynx lynx*) stalks its prey from the underbrush and seizes the prey by the neck.

CAUSE OF ABSCESSES on the dead caribou calves was discovered in 1964. The body of a calf with four puncture wounds in the neck but no abscesses was found. When a lynx skull was obtained, the four large canine teeth fit neatly into the wounds. Culturing saliva from a lynx yielded the pathogenic bacteria that had previously been found in the abscesses. It was evident that predation by lynxes was responsible for the mortality among the caribou calves.

ed that predation by lynxes was what was causing the attrition of the caribou herds. The dead calves found in the field were the ones that had escaped after being bitten, only to die of septicemia. Biologists explored the wooded areas on the edges of the calving grounds and discovered the remains of other calves.

Like other feline predators, lynxes often grasp their prey by the neck and drag it under overhanging vegetation. The lynx apparently waits in the underbrush at the edge of the calving ground for an unwary calf to approach. The lynx dashes out, seizes the calf by the neck and drags it back into the brush. Male calves are more vulnerable to attack than females because the males tend to be more exploratory, often leaving their mother far behind to explore the forest edge. Female calves tend to stay closer to their mother. The adult female caribou in Newfoundland stands about 1.2 meters at the shoulder and weighs about 100 kilograms. This size is more than sufficient for her to drive off an attacking lynx weighing about 10 kilograms.

To test the hypothesis that the lynx was responsible for the decline of the caribou herds lynxes were removed by trapping from the area around two calving grounds. When this was done, there was a statistically significant increase in the fraction of calves that survived to join the herd. Caribou were then introduced onto several small islands along the coast of Newfoundland where there were no lynxes. As an experimental control caribou were also introduced into mainland habitats where the local caribou had become extinct but where lynxes were still present. On the small islands the caribou increased at a rate near the maximum calculated for the species. In the habitats where lynxes were present the rate of increase was half the maximum. In this same period the size of the main Newfoundland herd changed very little. It consisted of 6,100 animals in 1961 and 6,200 in 1966.

One of the places where the caribou had increased rapidly after their introduction was Jude Island. All the animals were introduced by the Newfoundland Wildlife Service under the direction of Eugene Mercer. Mercer was particularly closely involved in the work on Jude Island. Radio transmitters were attached to two lynxes and the animals were released on the island to observe their effect on the herd in an environment that had been free of predators. In the first season 65 percent of the calves were killed by lynxes, whereas in the preceding year all the calves had survived.

The work on Jude Island also revealed one reason the caribou dam stays near the body of her calf for several days. The movements of the lynxes, tracked by the radio transmissions, made it clear that after being driven away from its prey by the dam the lynx often remained hidden in the brush nearby waiting for her to leave. When the transmitter-bearing lynxes were removed from the island, the rate of calf survival immediately returned to its former high level.

After the experiment on Jude Island even the most skeptical observers were convinced that the lynx, which had never been reported to be a significant predator on caribou on the Canadian mainland, was the major factor in limiting the caribou population of Newfoundland. The timber wolf, once the chief constraint on the caribou population, was gone and the lynx had taken its place in the island's simple ecosystem.

If the lynx could regulate the caribou population, it seemed possible that it could also regulate the number of arctic hares. The two highland calving grounds where the rate of calf survival was highest were also the areas where arctic hares were still occasionally reported. On the other hand, it was possible that the snowshoe hares had reduced the population of native hares by competing with them for food. It was also possible that a change in the flora of the island had eliminated the indigenous hare's food supply.

To determine whether the increase of the arctic hares was limited by the amount of available food, two male and two female arctic hares were released on Brunette Island, about 16 kilometers south of Newfoundland in Fortune Bay. There were no mammalian predators or snowshoe hares on the island. Moreover, like much of Newfoundland, Brunette Island was a suitable habitat for arctic hares. Much of the island consists of subalpine and alpine zones, and the tundra hill summits are covered by mosses and ericaceous shrubs. Six reproductive seasons later there were 1,000 arctic hares on the island. Like the caribou, in the absence of lynxes the hares had multiplied at a rate close to the maximum for the species.

The release on Brunette Island proved that food was not a barrier to the growth of the arctic hare population. Arctic hares from Brunette were subsequently employed to test the other hypotheses; they were introduced to islands where there were snowshoe hares but no lynxes. As a control for the experiment arctic hares were introduced into areas of Newfoundland where there were both snowshoe hares and lynxes.

On the outlying islands where there were snowshoe hares but no lynxes the arctic hares survived and their population increased, although not as rapidly as it did on Brunette Island. In the Newfoundland habitats, where there were lynxes, none of the introductions succeeded. It was clear that the indigenous hare can withstand competition from the snowshoe hare but cannot withstand predation by the lynx, at least in the lowland areas of Newfoundland.

The field tests confirmed that the lynx was limiting the abundance of both the caribou and the native hare on Newfoundland. From the release experiments and a study of the historical records of the island's fauna the outlines of an unusual dynamic system including the lynx, the caribou and the two species of hare began to emerge. For thousands of years before Newfoundland was settled by white men arctic hares, wolves and caribou coexisted on the windswept barrens and massifs of the island. In the early period the lynx was uncommon. Indeed, it was so rare that expert observers disagree on whether it should be considered native to the island at all.

The populations of wolves, caribou, lynxes and arctic hares undoubtedly fluctuated, but stabilizing mechanisms kept the small fluctuations from leading to the extinction of any one species. Then in 1864 snowshoe hares were brought to Newfoundland to help feed isolated fishermen on the west coast. The introduction of the new type of hare disrupted the delicate population dynamics that had prevailed among the indigenous species.

When an animal such as the snowshoe hare arrives in a new environment, it often tends to increase rapidly. After a few decades the introduced population reaches its highest level. By 1890 the entire island of Newfoundland was experiencing such a "high" in the snowshoe hare population. The density of the animals was so great that some inhabitants were able to trap 1,000 hares in a single season.

The initial high that follows the introduction of a species is generally followed by a decline as the inherent limits of the ecological niche occupied by the species begin to exert their effects. By about 1915 the snowshoe hare population had begun to decline rapidly.

In much of Canada the lynx is the chief predator on the snowshoe hare. As the snowshoe hare population was increasing rapidly in Newfoundland toward the end of the 19th century the lynx population was also increasing, although the curve of the lynx increase lagged behind that of the hares. This is a typical pattern in areas where the lynx and the snowshoe hare form an obligatory prey-predator pair. In such areas the curves representing the two populations take a regular sinusoidal form. The snowshoe hare population increases until it is limited by the food supply or by an intrinsic biological mechanism. The hare population then declines because its rate of reproduction has decreased. With fewer hares present the reproductive rate of the lynx also decreases and the lynx population decreases. In the absence of external perturbations the complete cycle takes from nine to 11 years.

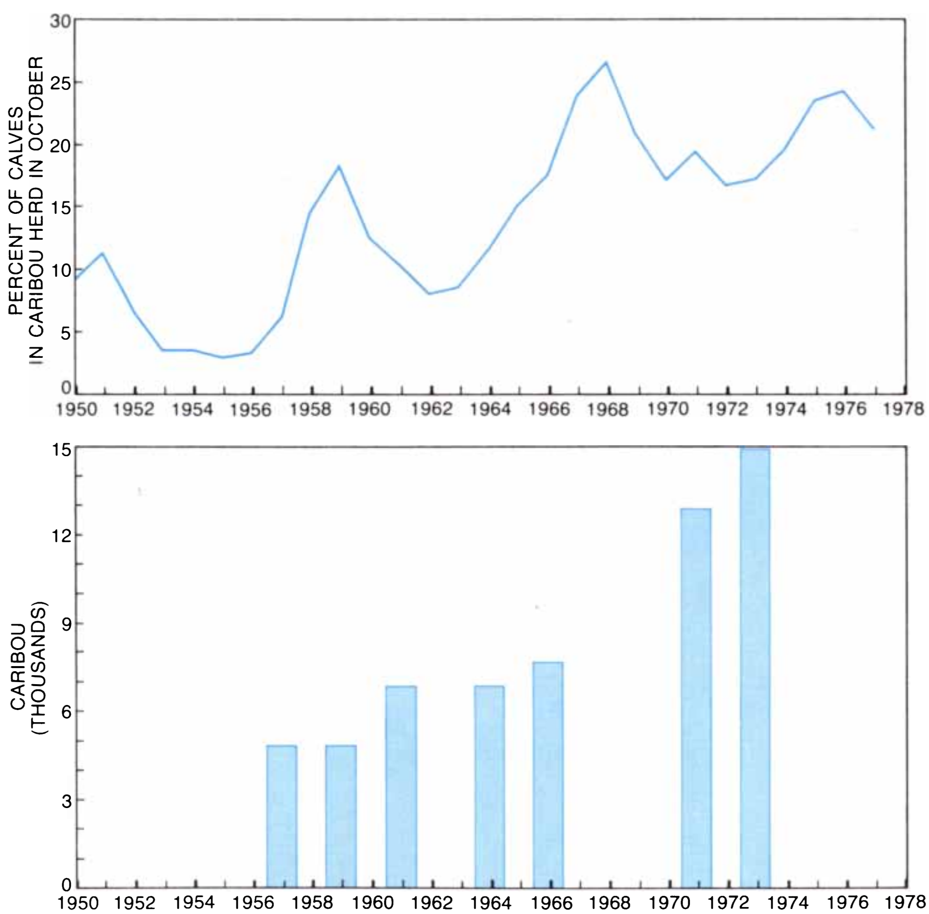
By the early years of this century the lynx had spread across Newfoundland from the west coast. Lynxes were reported within the city limits of St. John's, the capital on the Avalon Peninsula, which extends from the southeast coast. A bill to exterminate the lynx was proposed in the Newfoundland Assembly in 1904, which suggests how rapidly the lynx had increased. The lynx population was ultimately regulated, however, not by human factors but by ecological ones.

The snowshoe hare population remained at a peak from 1896 to 1915. (The oscillations of the hare and lynx populations had not yet become established.) When the hares crashed, the abundant lynxes were left without their main food item. Lynxes are resourceful scavengers and flexible predators. When food is particularly short, adult lynxes stop breeding but rarely starve. Instead they switch to alternative wild prey species, to domestic animals or even to ocean refuse deposited on the shore.

In Newfoundland the only alternative wild prey were mice, arctic hares and caribou calves. The lynxes presumably began to stalk caribou calves and the few available arctic hares. The new prey were not abundant enough to allow the lynxes to continue to increase or maintain their numbers, but they did mitigate the loss of the chief source of food.

At about the time lynxes began to take large numbers of caribou calves the death rate among the adult caribou rose for a quite different reason. The Newfoundland railroad, completed in about 1900, intersects the path along which the island's main caribou herd once migrated south in the fall. From 1911 to 1925 hunters, exploiting the railroad as a convenient means of pursuing the animals, began harvesting the herd in excess. In each of the years 1911, 1912, 1914 and 1915 hunters killed more than 5,000 caribou.

Because the high attrition among adult caribou came when the lynxes were beginning to prey on the calves, there were few new recruits to replenish the losses caused by the hunters. In a few years the 40,000 caribou of 1900 were all but wiped out. In the two dec-



SURVIVAL RATE of caribou calves determined the size of the main caribou herd in the 1950's and 1960's. The curve in the upper panel shows the fraction of the main herd consisting of calves in October. The fraction corresponds to the number of calves that survived the spring and summer. The peaks in calf survival in 1951, 1959, 1968 and 1976 coincided with snowshoe hare peaks on the island. When the snowshoe hare population crashed, calf survival declined. The bars in the lower panel show the total size of the herd in the interior of Newfoundland in the years for which data are available. When the prey-switching cycle was most intense, the size of the herd was determined by calf survival. Because of a decrease in the number of lynxes and the introduction of new prey species, the intensity of the cycle has recently diminished.

ades required for the catastrophic reduction of the caribou herd the snowshoe hare population had begun to fluctuate in the cyclical pattern observed in other parts of Canada. Peaks in the number of snowshoe hares were noted in 1920, 1931, 1940–1943, 1951–1952, 1959–1960, 1969 and 1976.

At the highs in the snowshoe hare cycle the lynxes switched back to their main prey and the rates of survival among caribou calves increased. Historical records show the caribou herds grew rapidly in about 1940, 1950 and 1960. During the 1960 snowshoe hare peak the chief caribou herds that were counted increased from 4,800 animals in 1959 to 6,100 in 1961. Thus the predatory adaptation of the lynx meant that the populations of both caribou and arctic hares were tightly coupled to the snowshoe hare population.

In recent decades the cycles of caribou-calf survival have become less pronounced; even at the low points in the cycle a larger fraction of calves survive than was the case when the cycle was at its most intense. This damping has occurred partly because the lynx population on the island has declined as a result of another cyclical phenomenon: women's fashions.

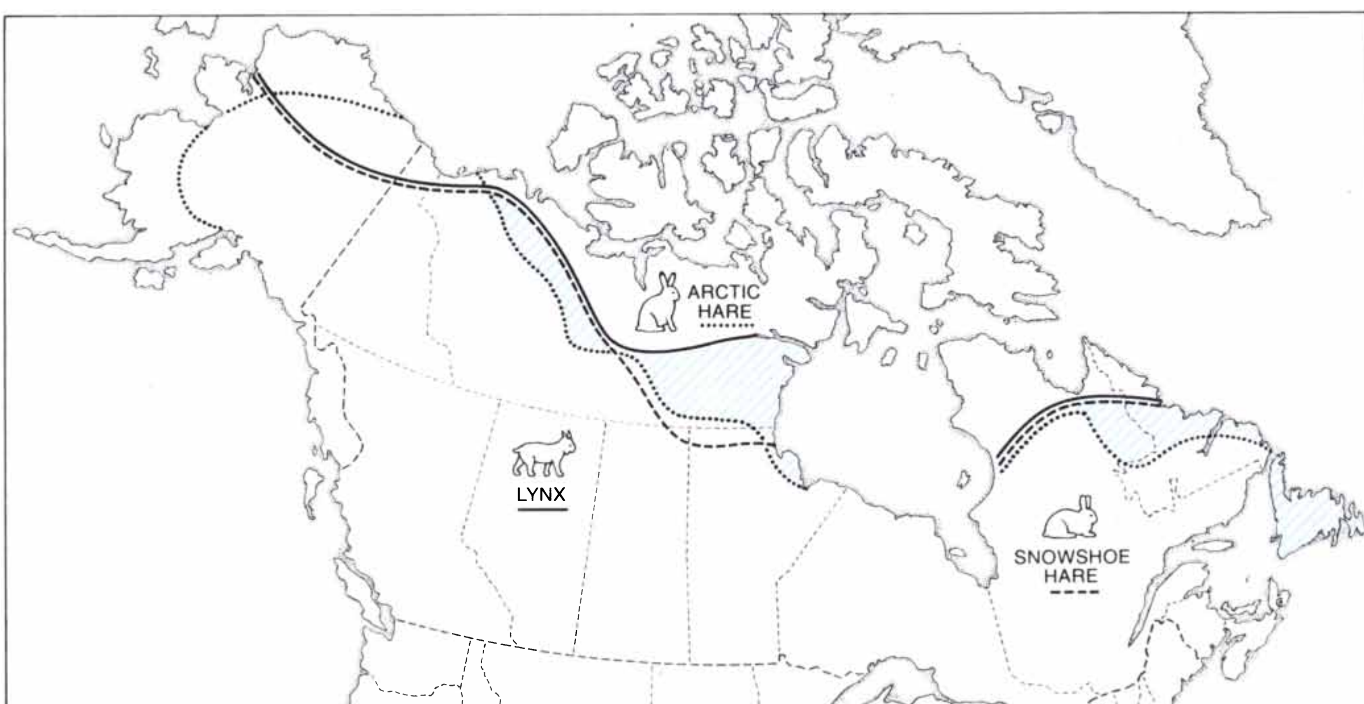
Long-haired animal pelts have recently become fashionable again after a hiatus. Trapping the lynx has therefore become economically rewarding, and the density of the lynx population in Newfoundland has dropped. Moreover, several new types of prey have been imported to Newfoundland, including the ruffed grouse and the spruce grouse. The island's food chain has become more complex, contributing an additional damping effect to the oscillations of predator populations.

Much of the effect the lynx has had on the caribou in Newfoundland stems from the fact that the caribou has adapted over a long period to predation by the wolf. The strategies that worked well against the wolf, however, were actually a handicap against a new predator with different hunting habits. For example, the large caribou aggregations that form after calving in the spring give the caribou a substantial advantage in protecting the calves from wolves. Wolves often hunt by day and stalk their prey in the open. When the caribou herd sights a wolf, its main defensive strategy is high-speed flight. The size of the herd at calving time enhances vigilance: it increases the probability that one female will see the wolf coming from afar. Moreover, the eye of the caribou is well adapted to detecting motion by day, a valuable asset against a diurnal predator moving in the open. By the same token, the unobstructed terrain of the calving ground makes it easier to detect a stalking wolf.

This combination of physiological and behavioral characteristics, however, is of little value when the predator is a lynx. The lynx hunts mostly at night and ambushes its prey rather than bringing it down in full-speed chase. Lynxes often crouch and freeze during their approach; this neutralizes the defenses of the caribou, whose eyes are better adapted to detecting motion than to detecting form. Caribou generally do not react with alarm to a man standing still, and they may be similarly unreactive to a crouched lynx.

When a predator is stalking a herd of herbivores, it often selects a particular animal to attack because that animal differs in some way from the others in the herd. Wolves choose young, sick, old or crippled caribou; almost any kind of conspicuous difference can make a caribou attractive to a wolf. With lynxes, however, this kind of selection is focused exclusively on the young. For a lynx the most conspicuous target is an active, curious calf at the edge of the calving ground near the fringe of vegetation that serves the predator as stalking cover.

It may be that predation by lynxes in the absence of wolves has pointed a new direction for the evolution of the Newfoundland caribou. Natural selection may currently favor the dispersal of calves rather than their concentration on the open calving ground, which makes them an obvious target. Moreover, the bond between the dam and the calf may be strengthened. A calf that



RANGE OF THE ARCTIC HARE in Canada (*color*) could be limited in part by predator-prey relations. The southern boundary of the range of the arctic hare coincides with the northern boundary of the range of the lynx: the range of the arctic hare overlaps only slightly that of the lynx (*hatched area*). The range of the lynx and the range of the snowshoe hare, however, largely overlap. The lynx and the snowshoe hare have a low foot loading: the pressure exerted on the ground by the feet when the animal runs. Thus they are well equipped for running on the soft snow of the boreal forest. The arctic hare has a higher foot loading for running on the hard snow of the tundra. In the boreal forest the arctic hare flounders and is vulnerable to the lynx. In the absence of lynxes the arctic hare might range farther south.

Relational Data Base

IBM

To: Dick
From: Bill
Subject: IBM Technology

I've been reviewing some of our past and present technological achievements, and it occurred to me that the scientific, engineering, and academic communities might like to know more about them. Will you select a topic from the following list?
Thanks.

- Vacuum tube digital multiplier
- IBM 603/604 calculators
- Selective Sequence Electronic Calculator (SSEC)
- Tape drive vacuum column
- Naval Ordnance Research Calculator (NORC)
- Input/output channel
- IBM 608 transistor calculator
- FORTRAN
- RAMAC and disks
- First automated transistor production
- Chain and train printers
- Input/Output Control System (IOCS)
- STRETCH computer
- "Selectric" typewriter
- SABRE airline reservation system
- Removable disk pack
- Virtual machine concept
- Hypertape
- System/360 compatible family
- Operating System/360
- Solid Logic Technology
- System/360 Model 67/Time-Sharing System
- One-transistor memory cell
- Cache memory
- Relational data base
- First all-monolithic main memory
- Thin-film recording head
- Floppy disk
- Tape group code recording
- Systems Network Architecture
- Federal cryptographic standard
- Laser/electrophotographic printer
- First 64K-bit chip mass production
- First E-beam direct-write chip production
- Thermal Conduction Module
- 288K-bit memory chip
- Robotic control language

Bill - It's a tough choice but I'd go with relational data base. It's a software development that's making a big impact. Dick

PARTS

PARTNO	PNAME
P107	BOLT
P113	NUT
P125	SCREW
P132	GEAR

SUPPLIERS

SUPPNO	SNAME
S51	ABC Co.
S57	XYZ Co.
S63	LMN Co.

PRICES

PART#	SUPP#	PRICE
P107	S51	0.59
P107	S57	0.65
P113	S51	0.25
P113	S63	0.21
P125	S63	0.15
P132	S57	5.25
P132	S63	7.50

Figure 1. *Relational data base consisting of three tables.*

WHICH SUPPLIERS HAVE PARTS FOR LESS THAN $0.50?

SQL QUERY

USER INPUT:

```
SELECT PART#, PRICE, SNAME
FROM   PRICES, SUPPLIERS
WHERE  SUPP# = SUPPNO
       AND PRICE < 0.50
```

QBE QUERY

USER INPUT: DRAW PARTS, SUPPLIERS, SKELETON

COMPUTER-GENERATED TABLES AND USER INPUT:

SUPPLIERS	SUPPNO	SNAME
	_Sn	_Na

PRICES	PART#	SUPP#	PRICE
	_Pt	_Sn	_Pr <0.50

P.	_Pt	_Pr	_Na

RESULTS

PART#	PRICE	SNAME
P113	0.25	ABC Co.
P113	0.21	LMN Co.
P125	0.15	LMN Co.

Figure 2. *An example of using IBM's very-high-level data base languages, SQL and QBE, to satisfy a request involving two tables from Figure 1. The SQL commands are expressed in a standardized block format; an example of the most common form for extracting data is:*

SELECT some data (column names)
FROM some file (table names)
WHERE certain conditions, if any, are to be met (rows)

QBE is initiated simply by typing the table name on the display screen, and the screen returns a skeleton table with column names in it. In this example, the user builds a new table in the blank skeleton by typing "example elements" (e.g., __Pt) under existing tables and in the blank skeleton. The example elements are formed by typing an underline followed by any mnemonic the user desires. Note that "P." simply means to present the results.

With business information growing at the rate of two file drawers per office worker per year, and with increasing amounts of it stored in electronic data bases, new techniques are required to allow easy, yet controlled, access by workers who lack computer expertise.

Starting in 1970, IBM researchers formulated, implemented, and tested prototype relational data base systems. This new approach in data base processing virtually eliminates the need for computer experience among users.

The relational model opened the way to more flexible, easy-to-use data base systems. The two relational data base management systems marketed by IBM for intermediate and large computer systems—Structured Query Language/Data System, introduced in 1981, and IBM DATABASE 2, introduced in 1983—allow users to update, retrieve, insert, delete, and otherwise manipulate data merely by specifying *what* they want to do, without having to tell the computer *how* to do it.

These relational systems are especially "friendly" because of the familiar, easy-to-interpret manner in which users see the data—as two-dimensional, rectangular tables ("relations"), with all information arranged in columns and rows.

IBM developed two very-high-level languages, Structured Query Language (SQL) and Query-By-Example (QBE), to access the relational data bases. Both are easy to learn, easy to apply, and immensely powerful. The innovative concept of QBE, which had a significant influence on display-screen interfaces, uses a two-dimensional programming approach. All queries are made directly onto a blank "skeleton" table appearing on a display screen. The user extracts data by a fill-in-the-blanks mode. SQL is a linear language that comes very close to "speaking English." It may be used both by ad hoc users at terminals and by programmers to embed SQL statements in application programs.

The non-navigational nature of the relational data base model endows QBE and SQL with extreme flexibility. Since there are no predefined information pathways to negotiate, the user is free to make all manner of ad hoc queries—an essential feature for applications where information needs change rapidly.

One of the most important IBM innovations in relational data base technology was a compiler approach to execute SQL statements. Replacing an interpretive approach, this compilation technique reduces the overhead cost of implementing the SQL language by using a precompiler to generate a tailored data access routine before execution time. The access routine, because it is tailored to one specific program, and is reusable, runs much more efficiently than a generalized interpreter.

In addition, the compilation technique uses a very sophisticated optimizer, which chooses economical access paths to the data. The compiler approach allows data base query in a high-level, easy-to-use language, yet also provides efficient program execution.

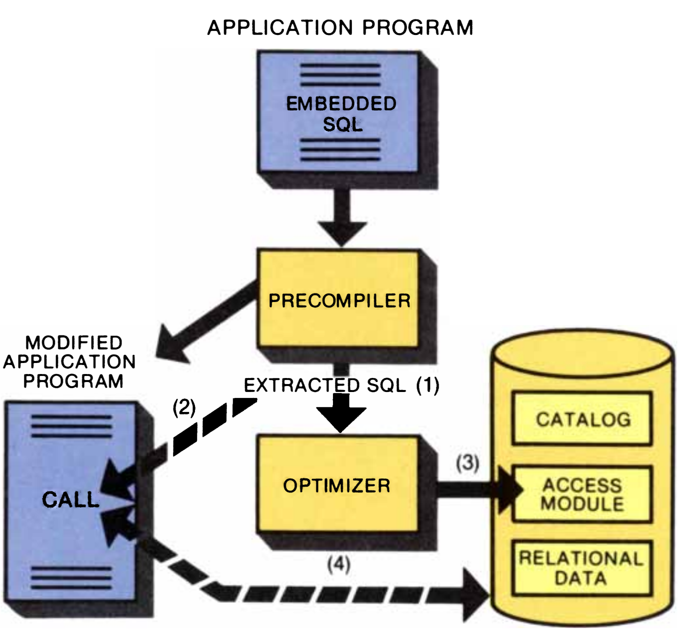


Figure 3. *The compiler approach is the key to IBM's efficient execution of SQL (very-high-level relational data base language) statements. This diagram illustrates the execution of application programs with embedded SQL statements. Programs are first processed by a precompiler, which extracts SQL statements from the application program (1). The precompiler also replaces the SQL statement in the host program with a CALL to the access routine (2). By very sophisticated analysis of available paths to the data, the optimizer chooses an economical path for the specific SQL statement, which is implemented as an access module (3). When the programs are executed, all the access modules for that program are loaded to provide targets for the modified CALLS (4).*

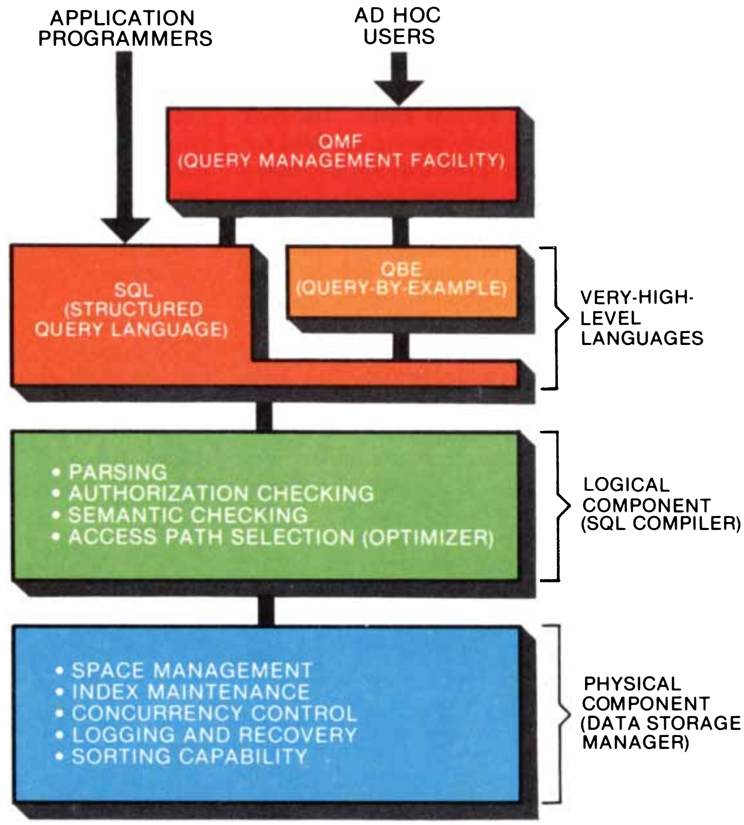


Figure 4. *This generalized architecture is the basis for IBM's relational data base products. It enables different types of users to access data easily, and yet is designed to handle complex programming tasks efficiently while providing the full function of a data base management system.*

Many scientists and programmers throughout IBM contributed to the development of relational data base technology, and researchers continue to explore future applications for the office environment and network users. These contributions are only part of IBM's continuing commitment to research, development, and engineering. IBM®

For free additional information on relational data base, please write: IBM Corporation, Dept. 813E/3N35 Old Orchard Rd., Armonk, N.Y. 10504 stays close to a protective mother is likelier to survive predation by a lynx than a calf that can run fast, even though the calf that can run fast is better able to escape a wolf.

If the caribou adapts, its rate of survival could increase rapidly, because the reciprocal adaptation of the lynx must operate within narrow limits. The lynx cannot depend only on the caribou for food; it depends primarily on the snowshoe hare. Any change in its hunting behavior cannot alter the pattern that makes it an effective hunter of hares. A new evolutionary relation between the lynx and the caribou should therefore favor the caribou.

Like the predation of the lynx on the caribou, the relation between the lynx and the arctic hare has not been established over a long period of close coevolution. The two species have mostly separate domains. On a map of Canada the southern boundary of the arctic hare's range coincides closely with the northern boundary of the range of the lynx. The territories of the lynx and the snowshoe hare, on the other hand, are mostly overlapping: the predator ranges about as far north as the prey.

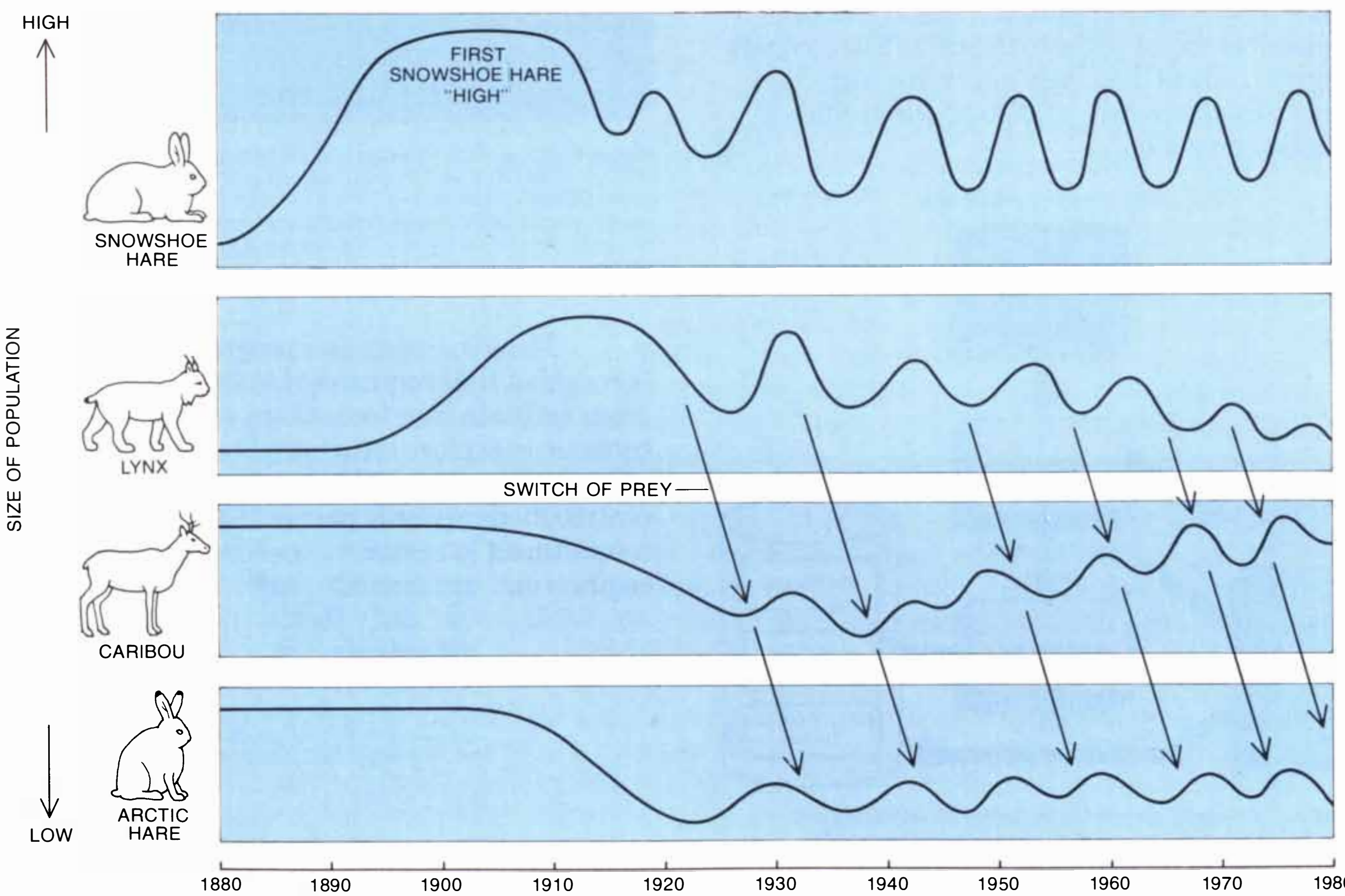
Most ecologists would explain the distribution of the three species according to the concept of biomes, or life zones: the arctic hare lives in the tundra biome, whereas the lynx and the snowshoe hare live in the boreal-forest biome. Predator-prey relations, however, could also influence the range of the arctic hare. It could be that in the absence of lynxes the arctic hare in Canada would range somewhat to the south of its current domain.

The success of the lynx in stalking the two species of hare depends on, among other things, the vegetation and the snow conditions. In forest habitats the lynx stalks and ambushes its prey, and the snowshoe hare sits motionless on the alert. The outcome of an encounter between them depends mainly on the distance between them when the prey sights the predator and runs.

Mercer has pointed out that as the lynx and the snowshoe hare sprint through the loose snow of the boreal forest they are evenly matched in one critical respect. Both animals have a low foot loading: the pressure exerted by the feet on the ground when the animal runs. As a result they can run quickly through the forest without sinking into the soft snow.

The arctic hare, on the other hand, is adapted to the hard-packed snow of the northern tundra. It has a foot loading more than twice that of the snowshoe hare. In soft snow the arctic hare flounders, greatly reducing its chances of escaping a lynx. This does not imply that snow conditions alone determine the range of the arctic hare. It is the presence of soft snow and predators with a low foot loading that constitutes the fatal combination.

The switching of predators from one prey species to another is currently a particularly promising area of investigation in ecology. The switches help to explain how stability can be maintained in small and fragile ecosystems. When the primary prey declines, a resourceful predator can shift to a secondary prey. Through this mechanism the predator holds its place in a food web that otherwise would break down. Switching also removes predation pressure from a de-



POPULATION DYNAMICS of the snowshoe hare, the lynx, the caribou and the arctic hare suggest that the driving force of the prey-switching system is the intrinsic 10-year cycle of the snowshoe hare. The first peak in the snowshoe hare population in Newfoundland was the highest; subsequent peaks have been somewhat lower. Before the snowshoe hares came to the island lynxes were so rare that it is disputed whether the animal is a native species. When the snowshoe hare population grew, the number of lynxes increased rapidly. Although there were some 40,000 caribou in 1900, as a result of predation by lynxes and human hunters there may have been 1,000 or fewer by 1925. The arctic hare population is in a "predator pit": it continues to exist because it is so scarce. The species cannot get out of the pit. If the number of arctic hares increases, the lynxes quickly reduce it when they switch after the next cyclical decline of snowshoe hares.

creasing prey population and enables it to survive.

Even so, switching can result in the decreasing prey population's surviving only at a very low density. This is true of the arctic hare in Newfoundland. The high density of lynxes, which is maintained by the snowshoe hare, ensures that the arctic hare density on the island will remain low.

In such a system the arctic hare is said to be in the "predator pit." Another instance of a prey species in such a situation is the caribou in Ontario and British Columbia. In those provinces the caribou forms one leg of a triangular system that includes the wolf and the moose. In both areas the caribou coevolved with the wolf in the absence of the moose. In the 19th century the moose spread into the areas from the south and east. The addition of the moose to the food base caused the wolf population to increase. Caribou, however, are easier for wolves to hunt than moose are, and the wolf takes caribou whenever it can.

In parts of Ontario and British Columbia the caribou are in the predator pit. They persist because they are rare. If the caribou increase, wolves switch back to them as a source of food and the caribou numbers fall. They cannot escape from the pit as long as the moose population continues to support a high density of wolves.

In some instances the prey species in such a system might survive not only because its numbers are low but also because the population of the species is spread so thin that hunting the animals is inefficient in terms of energy. For example, the present carrying capacity of Newfoundland for the arctic hare is defined by the minimum area per hare that is needed to make meetings between lynxes and their prey quite infrequent.

Indeed, such meetings must be rare enough for the survival of young arctic hares to compensate for the mortality among the adults. If the number, and hence the density, of arctic hares increases, this condition will be violated. Predation by lynxes will quickly restore the equilibrium. If this is so, spatial relations could turn out to be a highly significant feature of the dynamics of predator-prey systems.

The most significant conclusion to emerge from the study of the fauna of Newfoundland, however, is the considerable and unpredictable effect of introducing a new species into a simple ecosystem. What inhabitant of the island in 1864 could have imagined that the importation of a few snowshoe hares as a source of food for hungry fishermen would 100 years later be the cause of cycles in the survival of caribou calves and limit the population of arctic hares? In interfering with insular and fragile natural systems man must always tread softly.

The Aerodynamics of Human-powered Land Vehicles

A bicycle and its rider are strongly impeded by their resistance to the flow of air. Aerodynamic stratagems have brought vehicles that can go 60 miles per hour on a level road without assistance

by Albert C. Gross, Chester R. Kyle and Douglas J. Malewicki

For decades the principles of aerodynamics have been applied with great success to improving the speed and efficiency of aircraft, automobiles, motorcycles and even competitive skiers and skaters. Vehicles powered by human energy, however, were virtually ignored until quite recently, which is strange in view of the fact that air resistance is by far the major retarding force affecting them. With a bicycle, for example, it accounts for more than 80 percent of the total force acting to slow the vehicle at speeds higher than 18 miles per hour. Here we undertake to explain this neglect and to show what attention to aerodynamics is beginning to do for the performance of human-powered land vehicles.

Looking first at the bicycle, one sees that it has remained almost the same in form for nearly a century. The Rover Safety Cycle, which was introduced in England in 1884, could easily pass for a modern bicycle; it lacks only a seat brace, which would have formed the modern diamond frame, and a few components such as brakes and multiple gears. Almost from the beginning the designers and users of bicycles recognized the importance of aerodynamics, but artificial constraints on design largely prevented the application of the necessary technology. It was as obvious then as it is now that wind forces at the bicycle-racing speed of from 20 to 30 m.p.h. are enormous.

Before 1900 the crouched posture of the bicycle racer had become common as a means of reducing air resistance. Another practice adopted before 1900 was to put a multiple-rider bicycle ahead of a single racer to shield him from the wind. In 1895 the Welsh wheelman Jimmy Michael rode 28.6 miles in one hour behind a four-man lead bicycle. In 1899 Charles ("Mile-a-Minute") Murphy of the U.S. gained international fame by pedaling one mile at 63.24 m.p.h. on a bicycle traveling behind a train of the Long Island Rail Road on a board path built for the occasion.

In 1912 Étienne Bunau-Varilla of France patented a streamlined enclosure for a bicycle and its rider that was inspired by the shape of the first dirigible balloons. Versions of this bicycle and its descendants set speed records in Europe from 1912 to 1933. In 1933 Marcel Berthet of France covered 31.06 miles in one hour riding a streamlined rig named the Vélodyne; his pace was more than 3 m.p.h. faster than anyone riding a standard bicycle had gone for one hour.

In the same year the French inventor Charles Mochet built a supine recumbent bicycle (with the rider pedaling while lying on his back) that he later streamlined. With a professional racer, François Faure, this "Vélocar" set a number of speed records between 1933 and 1938. Mochet and Faure hoped the records would be recognized by the Union Cycliste Internationale, the world governing body for bicycle racing. They were not.

Indeed, in 1938 the Union banned the use of aerodynamic devices and recumbent bicycles in racing; the rule is still in force. The ban has been a serious deterrent to the development of high-speed bicycles and is one of two major reasons the bicycle has remained nearly unchanged for so long. (The other reason is that in the developed countries the shift to the automobile has made the bicycle less important for transportation than it once was.)

By its ruling the Union essentially classified improvements in the aerodynamics of bicycles and other technological changes as "cheating." (It is perhaps fortunate that the Union was not active when a Scotch-Irish veterinary surgeon, John Boyd Dunlop, developed the pneumatic tire for bicycles in 1887, otherwise people might now be riding bicycles and possibly automobiles with solid steel wheels.) To its credit, however, the Union has gradually begun to relax its restrictions on changes in aerodynamics, although recumbents are still forbidden. Since 1976 skintight one-piece suits have become common in international bicycle racing. Streamlined helmets, teardrop cross sections for frame tubing, streamlined brake levers and other aerodynamically improved components have been allowed. In fact, technological change in all forms of human-powered vehicle is flourishing at a rate unmatched since the heyday of the bicycle in the 19th century.

This rapid change can be partly attributed to a series of events in California. In 1973 one of us (Kyle) and Jack H. Lambie, a consultant in aerodynamics who was working independently, built and tested the first two streamlined bicycles in the U.S. Unlike their predecessors, Kyle and Lambie actually measured the reduction in drag achieved by streamlining. They did so by conducting numerous coast-down tests, in which an unpowered vehicle is allowed to decelerate on a level surface. In this condition the deceleration of the vehicle is proportional to the total retarding forces acting on it; instruments measure either the speed or the deceleration. Kyle and Lambie, publishing their results independently, both concluded that the total drag forces on a bicycle could be reduced by more than 60 percent with a vertical, wing-shaped fairing that completely encloses the bicycle and the rider. (It was not until some two years later that either Kyle or Lambie learned that similar vehicles had been built earlier in Europe.)

In 1974 Ronald P. Skarin, an Olympic cyclist for the U.S., set five world speed records riding the Kyle streamlined bicycle at the Los Alamitos Naval Air Station. Because of this success, Kyle and

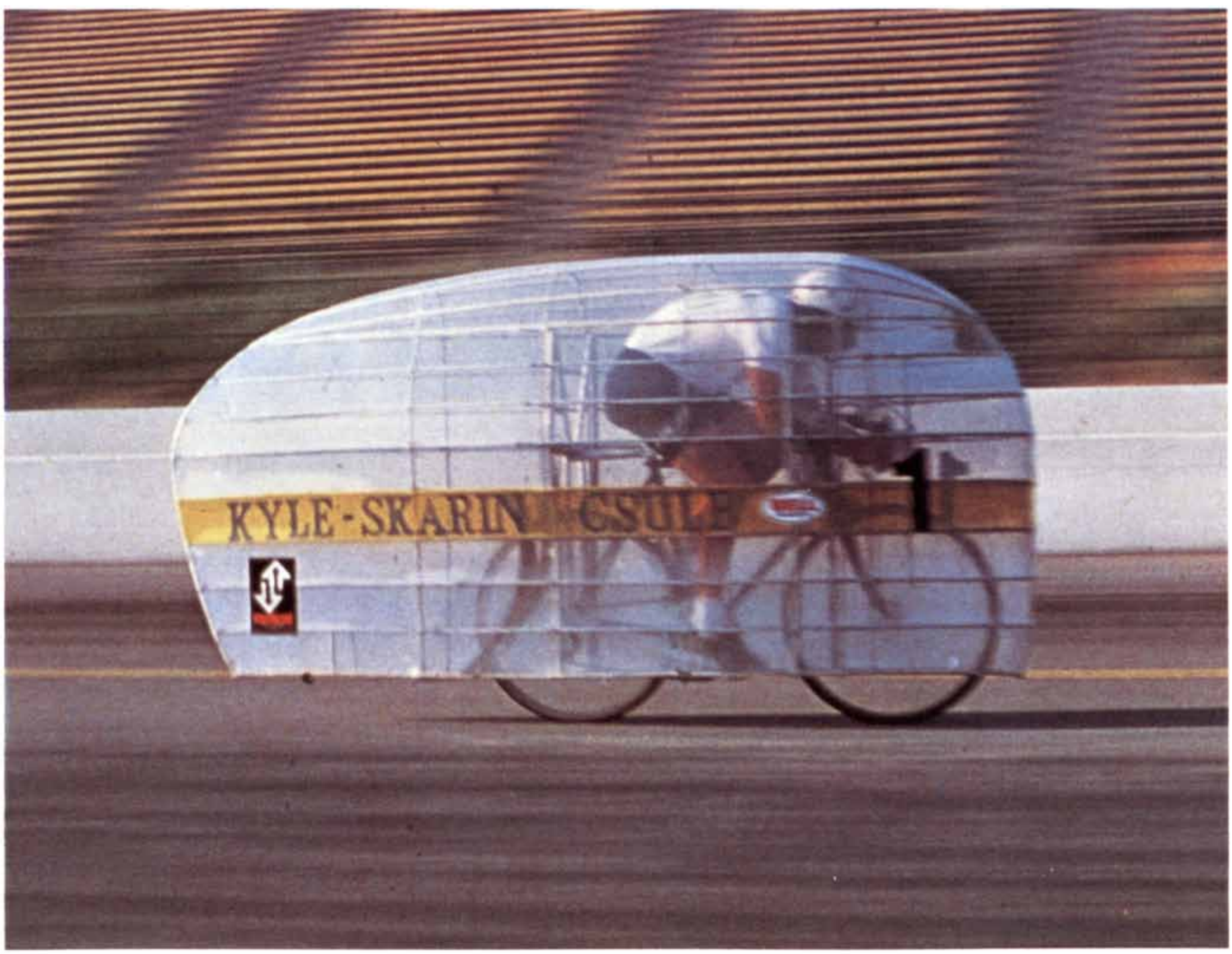
Lambie decided to organize a race for unrestricted human-powered vehicles. On April 5, 1975, at Irwindale, Calif., 14 distinctive vehicles competed in this historic first race. Many of them were recumbents, some with the rider pedaling supine (face up) and some with the rider prone (face down). Some were propelled by both hand and foot power. The winner at 44.87 m.p.h. was a streamlined tandem bicycle designed by Philip Norton, a high school teacher in Edgewood, Calif. The pedalers were Norton and Christopher Deaton, who is a skilled racing cyclist but not a world-class competitor. (The fastest an unaided standard racing bicycle has been ridden is 43.45 m.p.h., a record set in 1982 by Sergei Kopylov of the U.S.S.R., a cyclist of world class.)

Faced with the policy of the Union Cycliste Internationale against streamlining, the competitors in this race founded the International Human Powered Vehicle Association in 1976. Its purpose was to sanction competitions in which human-powered vehicles would be under no restrictions of design. Since then in dozens of races held in many countries the machines have become much more sophisticated and speeds have risen steadily. Four vehicles have broken the U.S. automobile speed limit of 55 m.p.h. (Each one received an honorary speeding ticket from the California Highway Patrol.) Among them is a third-generation streamlined quadricycle designed by Norton.

At present the world's fastest human-powered vehicle is the Vector Tandem, a gracefully streamlined two-person recumbent. It was built by a team headed by Allan A. Voigt, an engineer who as president of Versatron Research, Inc., primarily designs aerospace servomotors. (The pedalers ride supine and facing in opposite directions.) In 1980, with a flying start of about one mile of acceleration, it covered 200 meters along the track of the Ontario Motor Speedway in California at 62.92 m.p.h. Later that year the Vector Tandem averaged 50.5 m.p.h. for 40 miles on Interstate Route 5 between Stockton and Sacramento.

These extraordinary speeds are almost entirely the result of attention to aerodynamics. A cyclist traveling at 20 m.p.h. typically displaces approximately 1,000 pounds of air per minute. When the machine and the rider are not streamlined, they leave a substantial wake and exact a high cost in human energy.

Two types of aerodynamic drag affect the performance of a bicycle: pressure (or form) drag and skin-friction drag. Pressure drag results when the flow of air fails to follow the contours of the moving body. The separation changes the distribution of the air pressure on the body. If the separation takes place toward the rear of the body, the air pressure there becomes lower than it is on the forward surface, causing drag.

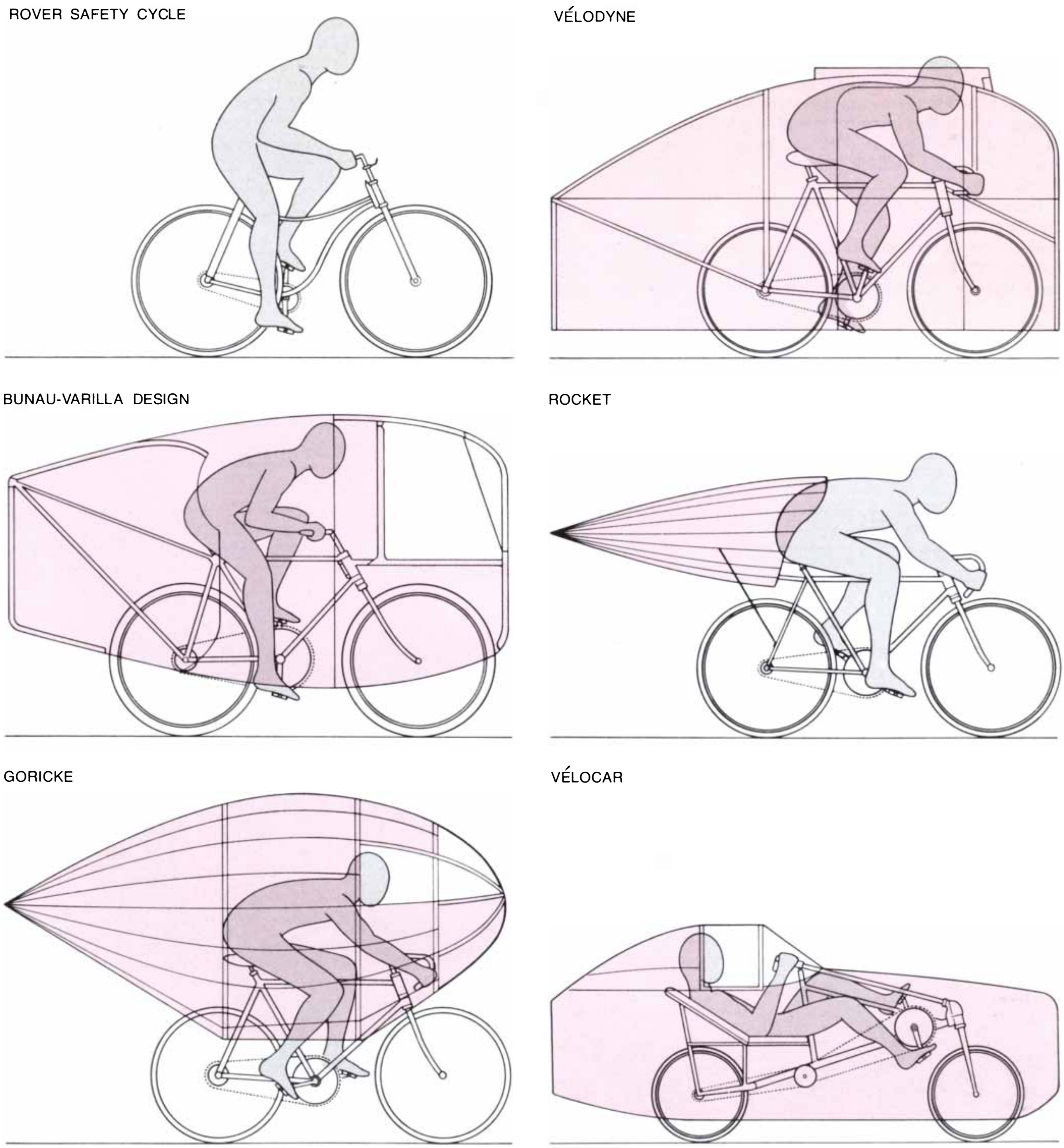


STREAMLINED RACING BICYCLE designed by one of the authors (Kyle) and ridden by Ronald P. Skarin, an Olympic cyclist for the U.S., is shown setting the world record of 31.88 miles per hour for one hour of pedaling from a standing start. The key to the performance was the streamlined fairing that reduced the aerodynamic resistance of the rider and the bicycle. Skarin established the new speed record in 1979 at the Ontario Motor Speedway in Ontario, Calif. Except for the fairing the vehicle was basically a standard racing bicycle.

Skin-friction drag results from the viscosity of the air. It is caused by the shearing forces generated in the boundary layer: the layer of air immediately next to the surface of the body.

Blunt configurations such as the cylinders, spheres and other shapes found on a bicycle are aerodynamically inefficient because the airflow separates from the surfaces. Low-pressure regions form behind the objects, resulting in pressure drag hundreds of times greater than skin-friction drag. In contrast, air flows smoothly around a streamlined shape. The air closes in behind as the body passes. Pressure drag is greatly reduced and skin-friction drag becomes more important.

For the highest efficiency a vehicle should be designed to minimize the transfer of unrecoverable energy to the air by the two types of drag. At the present level of technology aerodynamic drag absorbs from 40 to 50 percent of the fuel energy consumed by an automobile or a truck at 55 m.p.h. Since the bicycle has lower power, weight and



EARLY IMPROVEMENTS in human-powered land vehicles resulted in the introduction of the Rover Safety Cycle in England in 1884. In 1912 and 1913 Étienne Bunau-Varilla of France obtained patents for a streamlined design; similar bicycles set many speed records. The Goricke was developed in Germany in 1914. The Vélodyne was ridden 31.06 miles in one hour (a new record) by Marcel Berthet of France in 1933. From the same year is the Rocket, designed by Oscar Egg. Another French vehicle, the Vélocar, set several speed records between 1933 and 1938. Most of the drawings are based on data from the Wolfgang Gronen Archive at Binningen in Switzerland.

rolling resistance and poor streamlining, aerodynamic drag accounts for an even higher percentage of the energy consumed at speeds above 10 m.p.h.

A term employed to describe the aerodynamic efficiency of a shape is the drag coefficient. An inefficient shape such as a sphere will have a drag coefficient of, say, 1.3, whereas a streamlined shape such as a teardrop will have one of less than .1. Hence an object of teardrop shape can move with less than a tenth of the loss of energy incurred by an object of cylindrical shape.

For land-transportation vehicles the aerodynamic resistance is almost directly proportional to the product of the frontal area and the drag coefficient. For convenience we call the product the effective frontal area. In discussing which of two vehicles has less aerodynamic drag it is not sufficient to compare drag coefficients; the size of the vehicle must also be taken into account. That is done in the concept of the effective frontal area. An ordinary bicycle and its rider will have an effective frontal area of from 3.4 to six square feet, whereas a streamlined human-powered vehicle can have one of less than .5 square foot.

The force of aerodynamic drag increases as the square of the velocity. Power is proportional to the product of drag force and velocity, so that the power necessary to drive an object through the air increases as the cube of the velocity. Hence a modest increase in speed requires an enormous increase in power. A cyclist who suddenly doubles his output of power when he is traveling at 20 m.p.h. will increase his speed to only about 26 m.p.h.

Conversely, reductions in aerodynamic drag affect speed less than one might think. If the air drag is cut in half at 20 m.p.h., a cyclist who does not change his power output will speed up to about 24.4 m.p.h. The reason is that the rolling resistance remains constant. If that resistance could be ignored, doubling the horsepower or reducing the effective frontal area by half would again get the speed up to about 26 m.p.h.

In sum, high speeds require extremely high aerodynamic efficiency. The Vector Tandem, receiving an input of slightly more than one horsepower from each of its two riders, attained a speed of 62.92 m.p.h. For a standard bicycle to achieve that speed would require more than 6 h.p. That level of power from a human rider is clearly impossible.

Designers and riders can reduce the aerodynamic drag on human-powered vehicles in three major ways. First, they can cut down the amount of energy wasted by the vehicle's interaction with the air. They do it by streamlining (reshaping the front and rear of blunt objects to minimize the pressure drag) and

Embark on a wondrous journey across the universe of science

SCIENTIFIC AMERICAN LIBRARY

An incomparable collection of enduring value

Out of *Scientific American's* collaboration with scientists in every field of endeavor, and after nearly a decade of planning, the SCIENTIFIC AMERICAN LIBRARY now offers a limitless resource for those intrigued by the wonder of science.

The SCIENTIFIC AMERICAN LIBRARY acquaints you with a world of understanding and shares with you the major ongoing advances at the frontiers of scientific investigation. Written by scientists who have made significant contributions, each volume rewards you not merely with information but with true understanding, as well as a command of the new concepts with which research proceeds.

As you receive your volumes you will see how they combine to make up a reference library of enduring value, one that embraces the unity of scientific understanding.

Share in this exciting enterprise with us. Simply fill in the attached card or coupon and we will send your splendid premier volume, POWERS OF TEN, to examine *free* for 15 days. You will be billed just $19.95 plus postage and handling for each volume, to be paid only *after* you have decided to add the volume to your personal library.

While you may cancel at any time with absolutely no notice, membership will bring you a cumulative library of scope and focus unparalleled in the literature of science.

YOUR PREMIER VOLUME

POWERS OF TEN

Philip and Phylis Morrison and the Office of Charles and Ray Eames

The essential survey of the universe from its largest dimension to its smallest

"If I were marooned on a desert island with just this one felicitous, densely informative book on all of science, I would count myself a lucky man. Having read it through, nothing on earth could prevent me from swimming off, any distance, to get another one like it."

LEWIS THOMAS, M. D.,
author of *The Lives of a Cell*

An atlas of the universe and a guide to its exploration, POWERS OF TEN takes you on an extraordinary adventure in magnitudes. This spectacular journey, with its sweeping changes in perspective, proceeds in ordered steps from the largest known dimension (10^{25} meters) to the smallest (10^{-16} meters).

You will find yourself abroad in a cosmos that in surprise and splender beggars the past creations of man's innocent imagination. In vivid color illustrations, POWERS OF TEN shows you the universe scaled from the human dimension. You will know better not only where you are but who you are.

The SCIENTIFIC AMERICAN LIBRARY offers POWERS OF TEN as its premier volume because it sets out the agenda for the books that follow.

MUSICAL SOUND

John R. Pierce

John Pierce shares with you his joy in the discovery that understanding from physics and mathematics enlarges and enhances one's appreciation of music. He traces the connection between Pythagoras and Galileo to the current musical innovators — Pierre Boulez, John Cage, and Edgard Varese — now working at the frontiers of sound and music.

John Pierce imbues readers with his confidence that rational inquiry into the highly subjective experience of music will open up new realms to the enjoyment of that experience.

Two 33⅓-rpm sound sheets demonstrate acoustic illusions.

SUBATOMIC PARTICLES

Steven Weinberg

As everyone knows, the atom, once considered irreducible, is composed of more fundamental particles: electrons, protons, and neutrons. But how do we know this?

In THE DISCOVERY OF SUBATOMIC PARTICLES we come to know the answer. With the author we reenact the experiments that disassembled the atom. We are there as J. J. Thomson, Ernest Rutherford, and James Chadwick discover, in turn, the electron, the proton, and the neutron. We gain a confident understanding of how the laws of physics govern each event we encounter.

HUMAN DIVERSITY

Richard C. Lewontin

The relevance of science to human values finds compelling demonstration in our understanding of the genetics of our species. In HUMAN DIVERSITY, Richard Lewontin shows that each of us differs from all others because of the interaction of genetic and environmental differences coupled with chance events that occur during our development.

Lewontin makes clear that the human genetic endowment confers a plasticity of psychic and physical development; each of us has the potential to develop any of a great diversity of identities within the human ambit.

FOSSILS

George Gaylord Simpson

In FOSSILS, noted paleontologist George Gaylord Simpson sets out the splendor of evidence upon which the theory and facts of evolution rest. He explains how "the primary record of the history of life is written in the successive strata of rocks as in the pages of a book. Fossils may be called the writing on those pages."

In his lifework and in this book, Simpson presents a picture of evolution very different from that of "nature red in tooth and claw" depicted in the first crude Darwinian popularizations and persistent still in the popular imagination.

THE SOLAR SYSTEM

Roman Smoluchowski

Standing on the moon, men have seen the Earth "small and blue and beautiful." Beyond doubt, the space expeditions of the U. S. and the U. S. S. R. must be reckoned as the most successful, as well as the most extravagant, ventures in the popularization of science.

In THE SOLAR SYSTEM, Roman Smoluchowski assembles a wealth of images from satellites and earthbound telescopes to give us a magnificent picture of the Earth and its cosmic neighborhood.

ON SIZE AND LIFE

Thomas A. McMahon and John Tyler Bonner

A biologist and an engineer bring their disciplines to bear in showing how nature obeys its own physical laws in fashioning living things. They demonstrate how size relates to shape and how this in turn controls the upper limit to the sizes of animals and plants.

Beyond the diversity of life, the authors show orderly ratios and relations that surprise and delight: the time for 50 percent of a mammal's growth is about three percent of a lifetime, and every mammal, no matter how long it can live, will take about 330 million breaths, sustaining 1.5 billion heartbeats.

PERCEPTION

Irvin Rock

The eye, we say, is a camera. Consider, however, that an object appears to us unchanging in size, shape, or orientation despite changes in our position with respect to it. How can we explain this constancy of perception when we know that the size, shape, and orientation of an image projected by the lens of the eye upon the retina must vary as we move?

In PERCEPTION, Irvin Rock explores the problems posed by both correct and illusory perception and examines the traditions of thought which inform contemporary investigations of the subject.

THE LIVING CELL

Christian de Duve

A trillion cells compose the human body, each cell about four orders of magnitude smaller than the world in reach of our unaided senses. In THE LIVING CELL, Christian de Duve brings the new understanding in the form of a wonderfully illustrated dynamic anatomy.

De Duve invites us to accompany him on an engrossing expedition through a world far removed from everyday experience yet vital to our very being. A bold but reassuring guide, de Duve puts the wisdom and knowledge gained from a lifetime of research at our service with his narrative, nontechnical, and often anecdotal style.

In addition to these volumes, leading scientists are at work on the exciting future books in the SCIENTIFIC AMERICAN LIBRARY. James Watson is writing on DNA, David Hubel on vision and the brain, Julian Schwinger on space and time. In the pages of the SCIENTIFIC AMERICAN LIBRARY, these Nobel Laureates and other distinguished scientists will explore the world as we know it.

Special Offer to Readers of *Scientific American*
Membership Reservation Form

To SCIENTIFIC AMERICAN LIBRARY
P. O. Box 646
Holmes, Pennsylvania 19043

Please enroll me as a member of the SCIENTIFIC AMERICAN LIBRARY. Send me the premier volume — POWERS OF TEN — by return mail. Thereafter, I am entitled to receive for inspection a new volume every two months, as it is published, for 15 day's *free examination*.

I need send no payment now. For each volume I keep I will be billed $19.95 plus postage and handling. This price is guaranteed for at least one year, regardless of any price increases asked of new members.

I understand that I may cancel my membership at any time.

NAME ______________________
(Please print clearly)

ADDRESS ______________________

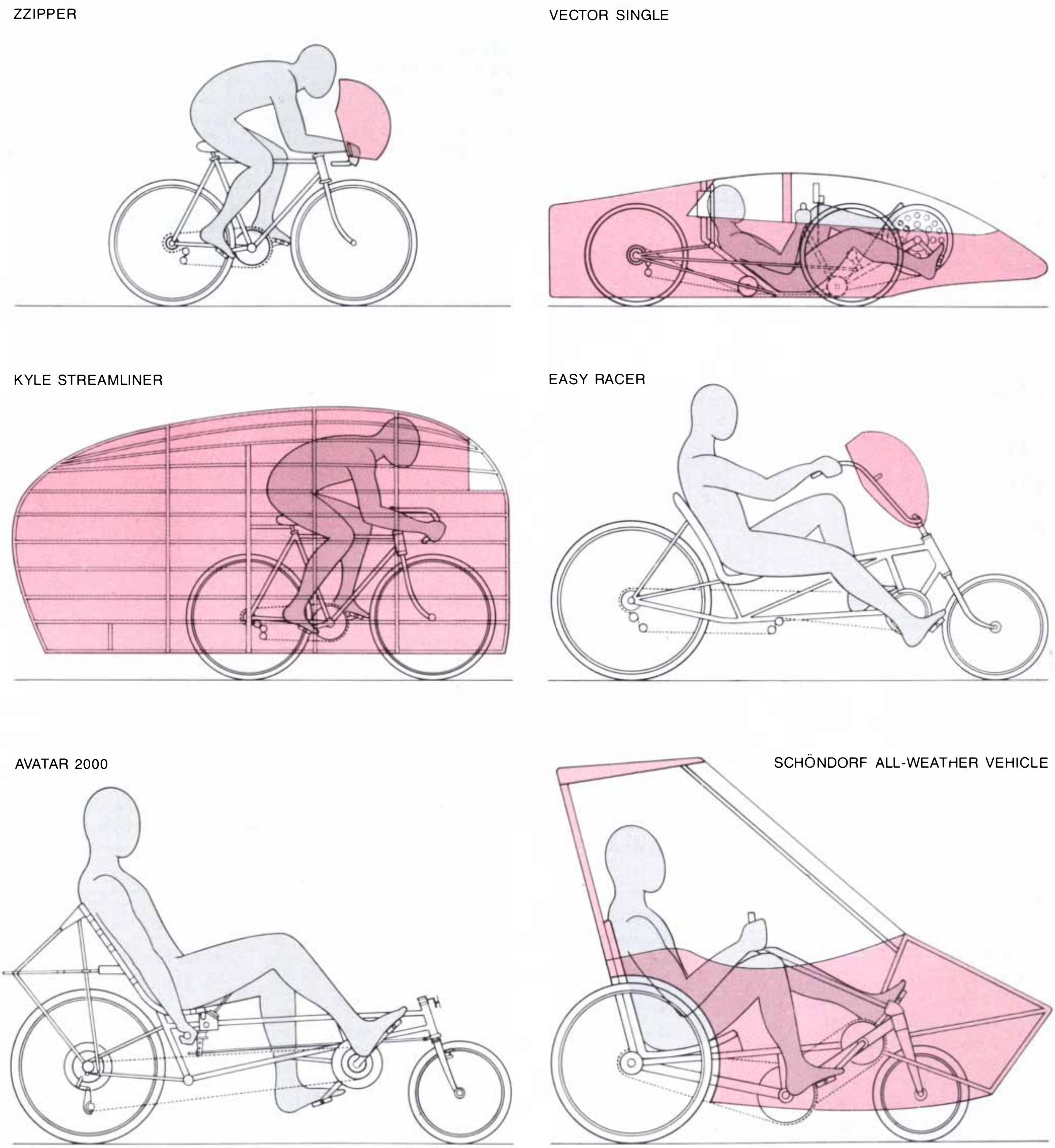
CITY ______________ STATE ________ ZIP ________

Canadian residents: U. S. dollars please

SSA123

by smoothing rough surfaces to minimize skin-friction drag. Second, the amount of air encountered during each second of forward travel can be reduced. This is done by lowering the effective frontal area of the vehicle-rider combination. The same effect can be achieved by riding at higher altitudes. Third, the rider can find air moving in such a way that it provides a tail wind. Here the most effective approach is drafting, that is, riding closely in the wake of another vehicle.

At high altitudes the atmosphere is less dense and bicyclists encounter less air. In Mexico City (elevation 7,414 feet, where the air is only 80 percent as dense as it is at sea level) cycling records are from 3 to 5 percent faster than records made at lower altitudes. At La Paz, Bolivia (elevation 12,000 feet), sea-level



MODERN HUMAN-POWERED VEHICLES make intensive use of streamlining to reduce the aerodynamic drag of the vehicle-rider combination. The simplest is the Zzipper, which is a partial fairing mounted in front of the rider. The Kyle Streamliner dates from 1973. A design that is meant for touring and commuting rather than for racing is the Avatar 2000; it utilizes the advantages of a recumbent position for the rider. The Vector Single, which has a full fairing, is theoretically capable of reaching almost 62 m.p.h. with an input of one horsepower from the rider. The Easy Racer is a recumbent designed mainly for touring or commuting, but it has also been raced. The last vehicle is one of the all-weather recumbents designed by Paul Schöndorf in Germany for elderly and handicapped people.

records could theoretically be improved by 14 percent. On the moon, where there is no atmosphere and only one-sixth the gravitational attraction, a suitably equipped bicyclist could theoretically ride at 238 m.p.h. with a very modest input of .1 h.p.

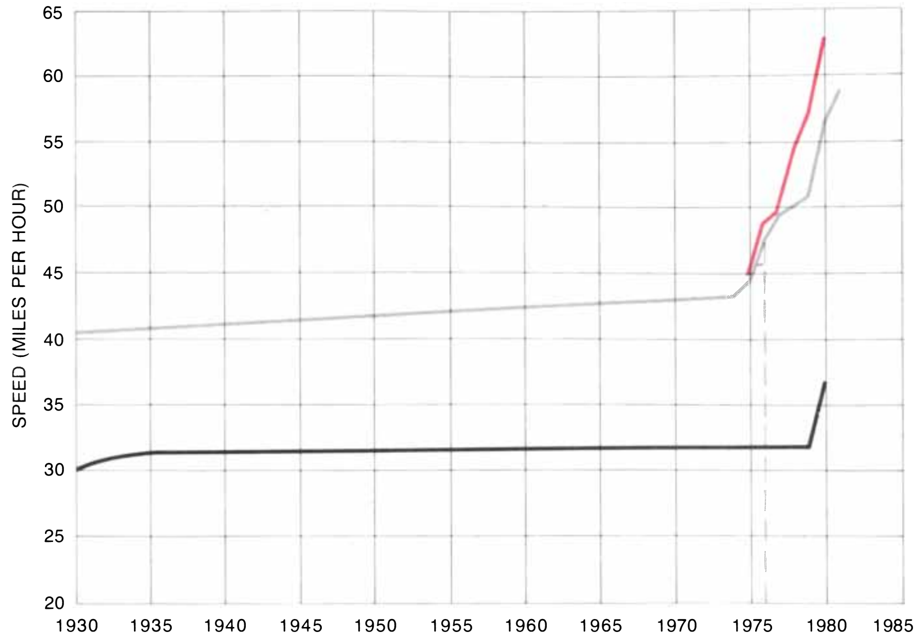
Analyzing the relation in which 80 percent of the power generated by a cyclist traveling on level ground at 18 m.p.h. goes to overcome air resistance, one finds that about 70 percent of the power consumption is due to the air's resistance to the rider and 30 percent to the air's resistance to the bicycle. This finding leads to the conclusion that to improve the performance of the standard bicycle one must first improve the aerodynamics of the rider.

For riders who race, the restrictions of the Union Cycliste Internationale leave little room for improvements beyond what has already been done in adopting the crouched position, the streamlined helmet, the skintight suit and the streamlining of components of the bicycle. As Voigt has calculated, even with a "perfect" bicycle (no aerodynamic drag on the machine at any speed and tires with no rolling resistance) the aerodynamic drag on the rider alone would severely hamper improvements in performance. According to Voigt, a crouched rider on a conventional racing bicycle could reach a maximum velocity of about 34 m.p.h. with a power input of 1 h.p. On a perfect bicycle the same rider making the same effort could achieve 38 m.p.h.

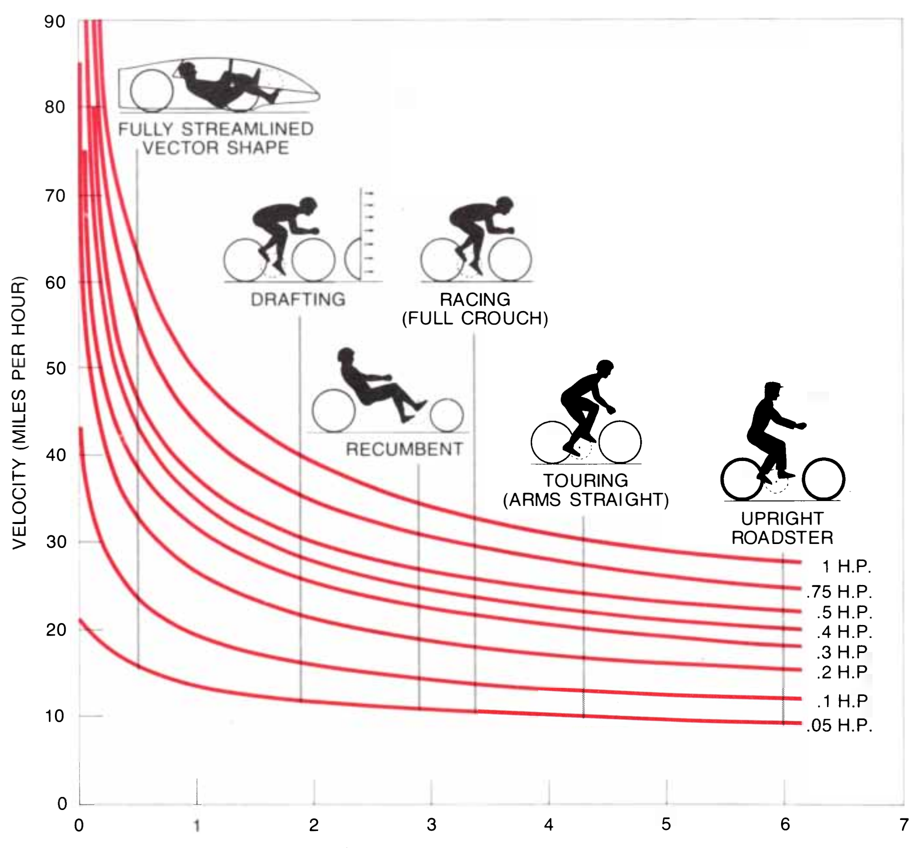
For the millions of noncompetitive cyclists who simply want a more efficient ride, several aerodynamic improvements are possible. They can be ranked in order of cost, beginning with the cheapest: a partial fairing such as the Zzipper, developed and manufactured by Glen Brown of Santa Cruz, Calif. It is a small, transparent, streamlined shield mounted in front of the rider. For about $60 a rider can lower the aerodynamic drag by about 20 percent, achieving a speed increase of some 2.5 m.p.h. for a 1-h.p. input.

Another effective way of reducing aerodynamic drag is to ride a recumbent bicycle. (The machine would cost several hundred dollars more than a basic touring bicycle.) The pioneers in this field are Gardner Martin of Freedom, Calif., designer of the Easy Racer, and David Gordon Wilson of the Massachusetts Institute of Technology, designer of the Avatar 2000. Because of the smaller frontal area presented by the recumbent rider, wind resistance decreases by 15 to 20 percent, resulting in about the same speed increase as is achieved by the Zzipper fairing.

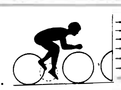
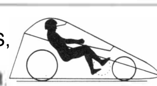
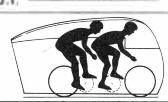
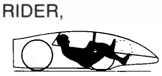
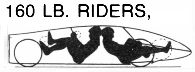
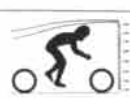
The recumbent bicycle offers other advantages. It is more comfortable to ride than a standard bicycle. In acci-



SPEED RECORDS of human-powered land vehicles have risen rapidly since the formation in 1976 of the International Human Powered Vehicle Association, which puts no restrictions on design. The time of the founding is indicated by the broken line. For many years before then the rules of the Union Cycliste Internationale, which banned streamlined vehicles from sanctioned bicycle competition, had kept speed records virtually unchanged. The curves represent records for multiple riders for 200 meters with a flying start (*color*), single riders under the same conditions (*gray*) and riders who pedaled for one hour at the maximum effort (*black*).



EFFECT OF STREAMLINING is to improve the performance of human-powered vehicles at all levels of power input. The upright roadster is the least streamlined vehicle, the Vector shape the most streamlined. Drafting means to follow closely behind another vehicle, here a bicycle. A good athlete can produce 1 h.p. for about 30 seconds, a healthy nonathlete for about 12 seconds. They can sustain an output of .4 and .1 h.p. respectively for about eight hours. The effective frontal area is the product of the drag coefficient and the projected frontal area.

	DESCRIPTION			FORCES AT 20 M.P.H. (POUNDS)	AERODYNAMIC DATA			ROLLING-RESISTANCE COEFFICIENT	LEVEL-
					DRAG COEFFICIENT	FRONTAL AREA (SQUARE FEET)	EFFECTIVE FRONTAL AREA (SQUARE FEET)		HORSE-POWER REQUIRED
STANDARD BICYCLES	BMX (YOUTH OFF-ROAD RACER)	30-LB. BIKE, 120-LB. RIDER, KNOBBY TIRES, 20-IN. DIA., 40 P.S.I.		5.52 2.10	1.1	4.9	5.4	.014	146
	EUROPEAN UPRIGHT COMMUTER	40-LB. BIKE, 160-LB. RIDER, TIRES 27-IN. DIA., 40 P.S.I.		6.14 1.20	1.1	5.5	6	.006	140
	TOURING (ARMS STRAIGHT)	25-LB. BIKE, 160-LB. RIDER, CLINCHER TIRES, 27-IN. DIA., 90 P.S.I.		4.40 .83	1	4.3	4.3	.0045	100
	RACING (FULLY CROUCHED)	20-LB. BIKE, 160-LB. RIDER, SEWUP TIRES, 27-IN., DIA., 105 P.S.I.		3.48 .54	.88	3.9	3.4	.003	77
IMPROVED MODELS	AERODYNAMIC COMPONENTS (FULLY CROUCHED)	20-LB. BIKE, 160-LB. RIDER, SEWUP TIRES, 27-IN. DIA., 105 P.S.I.		3.27 .54	.83	3.9	3.2	.003	73
	PARTIAL FAIRING (ZZIPPER, CROUCHED)	21-LB. BIKE, 160-LB. RIDER, SEWUP TIRES, 27-IN. DIA., 105 P.S.I.		2.97 .54	.70	4.1	2.9	.003	67
	RECUMBENT (EASY RACER)	27-LB. BIKE, 160-LB. RIDER, CLINCHER TIRES, 20-IN. FRONT, 27-IN. REAR, 90 P.S.I.		2.97 .94	.77	3.8	2.9	.005	75
	TANDEM	42-LB. BIKE, TWO 160-LB. RIDERS, CLINCHER TIRES, 27-IN. DIA., 90 P.S.I.		5.32 2.66 1.62 .81	1	5.2	5.2	.0045	66
	DRAFTING (CLOSELY FOLLOWING ANOTHER BICYCLE)	20-LB. BIKE, 160-LB. RIDER, SEWUP TIRES, 27-IN. DIA., 105 P.S.I.		1.94 .54	.50	3.9	1.9	.003	47
RECORD HOLDERS	BLUE BELL (TWO WHEELS, ONE RIDER)	40-LB. BIKE, 160-LB. RIDER, SEWUP TIRES, 20-IN. FRONT, 27-IN. REAR, 105 P.S.I.		.61 .8	.12	5	.6	.004	27
	KYLE (TWO WHEELS, TWO RIDERS)	52-LB. BIKE, TWO 160-LB. RIDERS, SEWUP TIRES 105 P.S.I.		1.44 .72 1.12 .56	.2	7	1.4	.003	24
	VECTOR SINGLE (THREE WHEELS)	68-LB. BIKE, 160-LB. RIDER, SEWUP TIRES, 24-IN. FRONT, 27-IN. REAR		.51 1.02	.11	4.56	.5	.0045	29
	VECTOR TANDEM (THREE WHEELS)	75-LB. BIKE, TWO 160 LB. RIDERS, SEWUP TIRES, 24 IN. DIA.		.62 .31 1.78 .89	.13	4.7	.6	0045	23
THEORETICAL LIMITS	PERFECT BIKE	NO ROLLING RESISTANCE, NO DRAG ON BIKE		3.07 0	.8	3.8	3	0	59
	DRAGLESS RIDER	ROLLING RESISTANCE INCLUDES RIDER'S WEIGHT		1.33 .81	1.1	1.2	1.3	.0045	41
	PERFECT RECUMBENT	DRAG ON RIDER ONLY		.72 0	.6	1.2	.7	0	14
	PERFECT PRONE BIKE	DRAG ON SMALL BUT STRONG RIDER		.51 0	.6	.8	.5	0	10
	PERFECT PRONE STREAMLINER			.07 0	.05	1.4	.07	0	1
	MOTOR PACING	42-LB., BIKE, 160-LB. RIDER, MOTORCYCLE ROAD-RACING TIRES, 70 P.S.I.		0 1.21			VARIES WITH SPEED	.006	23
	MOON BIKE	25-LB. BIKE, 160-LB. RIDER, 15-LB. SPACE SUIT		0 .15			0	.0045	3

GROUND, NO WIND		EFFECT OF HILLS	
ALL-DAY TOURING	MAXIMUM SPEED	STEADY SPEED UP	STEADY-SPEED COASTING
10.1	27.8	12.2	19.8
11.3	27.6	10.9	24
13.1	31.1	12.2	27.7
14.7	33.9	13	31.2
15	34.6	13	32.2
15.4	35.7	13.1	33.9
14.4	35.2	12.5	33.7
15.2	36.6	13	35.2
17.5	41	13.6	41.7
22.5	58.6	12.9	77.4
23.3	56.6	14	69.9
21.8	61.2	11.3	90.1
25.6	72.5	13	108.4
16.7	35.9	13.4	34.7
18.4	45.8	13.3	50.3
27.1	58.3	16.8	66.9
30.4	65.3	23.2	65.3
58.3	125.9	25.6	174.5
29.4	294	12.6	?
237.5	2,375	78.4	?

dents that do not involve an encounter with an automobile it is much safer, since the rider is closer to the ground (making falls less serious) and the feet are forward (making a head injury less likely in a fall). A problem is that a recumbent is hard to see on a road and so is perhaps more vulnerable to automobiles; the problem can be relieved somewhat by mounting on the vehicle a long, thin pole with a flag.

At the top of the expense ladder is a bicycle with a full fairing. The Vector Single, a one-rider version of the Vector Tandem, is the best example of a fully faired, enclosed, pedal-powered vehicle. (It is the machine portrayed on the cover of this issue of *Scientific American.*) According to Voigt, the vehicle is theoretically capable of reaching 61.7 m.p.h. with a 1-h.p. input, an increase of 28.2 m.p.h. over what has been done with a standard racing bicycle. A Vector Single costs about as much as a first-class racing bicycle.

In going up or down a hill a fully streamlined vehicle retains its advantage over a conventional bicycle. Although the Vector Single weighs about 80 pounds, compared with about 25 pounds for a standard bicycle, it can climb moderate hills as fast as or faster than the bicycle. With an input of .4 h.p. a bicycle can climb a 2.5 percent grade at about 16 m.p.h. and a 6 percent grade at about 11 m.p.h. With the same input the Vector can climb the two grades at 20.5 and 11 m.p.h. respectively.

Downhill the difference between the two machines is remarkable. The bicycle can descend a 2.5 percent grade at 29.5 m.p.h., the Vector at 54. On a 6 percent grade the bicycle can reach a speed of 39 m.p.h. and the Vector can exceed 100. Such potential speeds mean that if streamlined human-powered vehicles become common, careful attention must be given to the design of brakes and suspension and to the stability of the vehicle.

Since the aerodynamic drag force is proportional to the square of the relative velocity, head winds, tail winds and even crosswinds can drastically change both aerodynamic drag and the power requirements. For example, a bicyclist going at 18 m.p.h. in still air must increase his power output by 100 percent to maintain that speed against a head wind of 10 m.p.h. Usually a bicyclist confronting a head wind slows down and tries to maintain his customary leg force and pedaling cadence by shifting gears. This is one reason bicycles with multiple gears are desirable even for level country.

A tail wind makes the bicyclist go faster with his customary input of power. In general moving air will speed up or slow down a bicycle by about half the wind speed. When one bicyclist rides in the wake of another, the power requirements of the drafting rider are reduced by about 30 percent. The forward bicyclist creates an artificial tail wind.

The closer the rear bicycle follows the leader, the more pronounced the drafting effect is. One can think of the rear rider on a tandem bicycle as drafting extremely closely. Tandem riders use 20 percent less power per rider than two separate cyclists.

When the riders in a line of drafting bicyclists take turns in the lead position, the entire group can travel much faster than a single rider. In a pursuit race of 4,000 meters (almost 2.5 miles) a four-rider team can go about 4 m.p.h. faster than a single bicyclist. Typically a group of bicycle tourists of equal ability can travel from 1 m.p.h. to 3 m.p.h. faster than any rider alone. The larger the group is, the faster it should be able to travel (up to, say, a dozen riders).

Artificial winds created by passing automotive traffic can increase a bicyclist's speed from 1 m.p.h. to 3 m.p.h. for periods of about seven seconds. The larger the passing vehicle, the more substantial the effect. A steady stream of traffic can enable a bicyclist to sustain a speed from 3 m.p.h. to 6 m.p.h. higher than would otherwise be possible for a given energy input.

When a bicyclist rides directly in the wake of a motor vehicle, quite remarkable speeds can be attained. The practice is called motor pacing. On August 25, 1973, Allan V. Abbott, a physician in California, achieved a record of 138.674 m.p.h. motor pacing along a measured mile at the Bonneville Salt Flats in Utah. John Howard, a U.S. Olympic cyclist, is attempting to break Abbott's record and to achieve a motor-pacing speed in excess of 150 m.p.h.

Although the findings we have described are significant in their own right, one wonders if they will have any practical application beyond their effect on speed records. For a large fraction of the world's bicycle riders it seems unlikely that the work will have much immediate utility. For example, in the

PERFORMANCE OF HUMAN-POWERED VEHICLES is summarized. The numbers listed under forces for each vehicle represent air resistance and rolling resistance respectively. The five columns at the far right represent respectively the horsepower required at 20 m.p.h. as a percentage of the touring rider's performance; the all-day touring speed in miles per hour at an output of .1 h.p.; the maximum speed at an output of 1 h.p.; the steady speed in miles per hour up a 5 percent grade at an output of .4 h.p., and the coasting speed down the same grade.

many developing countries where the bicycle is the chief means of transportation most riders travel at about 7 m.p.h., often with a substantial load; aerodynamic drag becomes more important than other impediments to bicycle motion only at speeds above 10 m.p.h. Even here the work on aerodynamics makes a contribution. Without it designers would not know why they should largely ignore aerodynamics for slow-moving human-powered vehicles.

For bicycles intended for slow but sure progress it makes sense to decrease rolling resistance by improving tires and by paving roads. Designers should also reduce the bicycle's weight to facilitate climbing hills. The recent introduction of "mountain bikes" in the U.S. is a step in the direction of making lightweight bicycles durable enough for rugged or unpaved roads.

In several ways the knowledge gained by the recent research on the aerodynamics of human-powered vehicles can be directly useful. Although the standard bicycle is likely to be the predominant representative of the class for many more years because of its public acceptance, low cost, simplicity and mechanical reliability, it offers plenty of scope for innovation. For example, a light, simple and inexpensive front fairing will substantially improve the performance of the standard bicycle. The recumbent bicycle may come into greater use by commuters and tourists because of its efficiency and comfort.

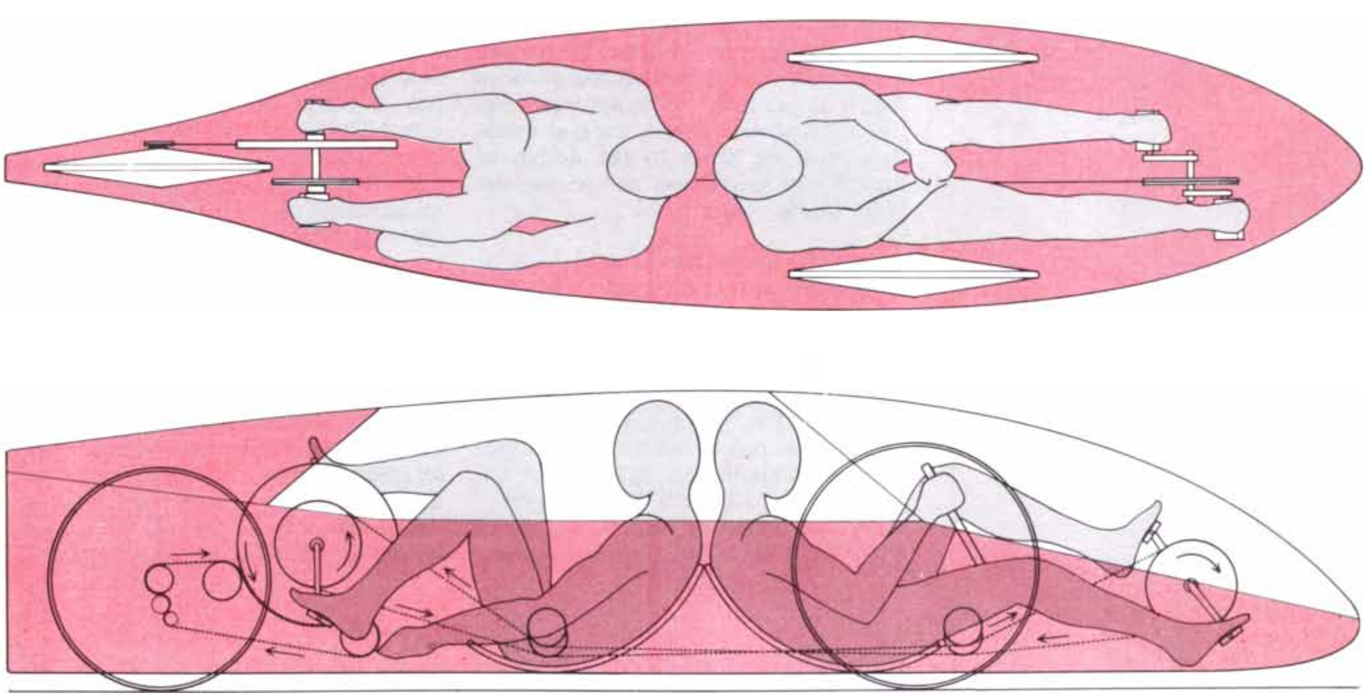
A further application of the technology would be to fit a recumbent with a small, lightweight motor of low horsepower. The motor would serve mainly as an aid in accelerating from a stop and in climbing hills. Fitted also with as much streamlining as would be consistent with the need for ventilation and stability, the machine would be a true moped. (The machines now sold under that name are not really motor-pedal vehicles but merely underpowered motorcycles.)

The recent research has inspired inventors to develop several special-purpose human-powered vehicles. Paul Schöndorf, a professor of engineering at the Fachhochschule in Cologne, has built a series of easily pedaled, all-weather recumbent tricycles for the elderly and the handicapped. Similar vehicles would serve well in retirement communities. Douglas Schwandt of the Veterans Administration's Rehabilitation Engineering Research and Development Center at Palo Alto, Calif., has built hand-cranked tricycles and bicycles for paraplegics. William Warner, a paraplegic who once held the record for hand-powered vehicles in the races sponsored by the International Human Powered Vehicle Association, says a disabled person can propel such a vehicle much faster than a standard wheelchair and thereby can gain a new sense of freedom and mobility. (The present record of 25.09 m.p.h. was set in 1981 by Ascher Williams of the Palo Alto rehabilitation center.)

In principle a fully enclosed, streamlined human-powered vehicle could be quite useful in transportation. A rider could travel at speeds of from 20 to 30 m.p.h. in all kinds of weather. As such vehicles are now designed, however, they would not serve on the open road. They lack adequate ventilation, visibility, maneuverability and such safety features as lights and windshield wipers. Most of them are not easy to get into or out of.

To produce a practical vehicle of this kind would require an investment and an engineering effort comparable to that made in producing a new automobile. Even then the pedaled vehicle would not be safe in traffic that included a large number of motor vehicles. One must conclude that a fully enclosed human-powered vehicle will not be a practical form of transportation until fuel shortages remove most motorized vehicles from the roads or until special roadways are built for pedaled machines.

Far likelier is the development of lighter and more fuel-efficient automobiles employing much of the technology we have described. One of us (Malewicki) has already built such a vehicle, a single-passenger machine weighing 230 pounds. It holds records for fuel economy at the freeway speed of 55 m.p.h. with a gasoline engine (157.2 miles per gallon) and with a diesel engine (156.3 m.p.g.). The diesel record was set on a trip from Los Angeles to Las Vegas, during which the average speed was 56.3 m.p.h. A trend toward such vehicles could help to extend fuel resources and ironically might postpone the time when the human-powered vehicle will have come fully into its own.



VECTOR TANDEM is shown in plan and elevation. It is a companion vehicle to the Vector Single portrayed on the cover of this issue of Scientific American. The Tandem, receiving an input of a bit more than 1 h.p. from each of its two riders, who are positioned back to back, set the speed record of 62.92 m.p.h. for 200 meters in 1980. (The riders had a flying start of more than one mile.) Traveling on an interstate highway in California later that year the Vector Tandem managed an average speed of 50.5 m.p.h. on a trip of 40 miles.

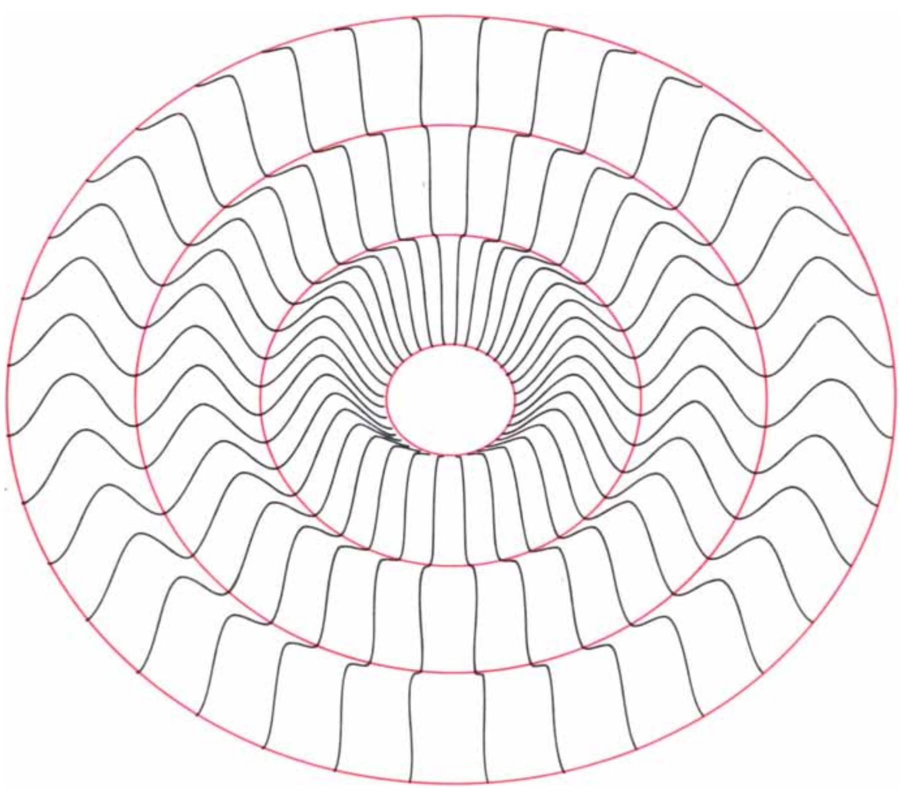
The Interpretation of Visual Illusions

The visual system apparently organizes ambiguous retinal images according to rules of inference that exploit certain regularities in the external world

by Donald D. Hoffman

Vision is a process of inference. What you see when you look around depends not only on what is there to be seen but also on how your visual system organizes and interprets the images that fall on your retinas. An intriguing demonstration of this aspect of perception is presented by the apparent surface that is formed by rotating a cosine wave around a vertical axis viewed obliquely [*see illustration on this page*]. When you first look at the figure, it appears to be organized into a set of raised concentric rings, with the boundaries between the rings delineated approximately by the colored circular contours. If you turn the page upside down, however, the organization changes: now each colored contour, instead of lying in a trough between two rings, appears to trace the crest of a ring. (Try it.) Evidently the visual system does more than passively transmit signals to the brain. It actively takes part in organizing and interpreting them.

This finding raises three questions. First, why does the visual system need to organize and interpret the images formed on the retinas? Second, how does it remain true to the real world in the process? Third, what rules of inference does it follow? The answers to these questions call for a closer examination of such figures.



AMBIGUOUS SURFACE is made by rotating a cosine wave about a vertical axis. The surface initially appears to be organized into raised concentric rings, with the colored circular contours lying in the troughs between the rings. When the page is turned upside down, however, the organization appears to change: each colored contour is now seen to trace the crest of a ring.

One reason the visual system organizes and interprets retinal images is simply that many possible configurations in the real world are consistent with any given retinal image. In other words, retinal images need organization and interpretation because they are fundamentally ambiguous. Their ambiguity is due in part to the fact that the world is three-dimensional and each retina is essentially two-dimensional. To describe the world in its full three-dimensional glory necessarily involves some rather sophisticated inferences by the visual system, inferences that for the most part proceed without any conscious awareness. For example, the cosine surface at the right, like your retinal image of it, is two-dimensional. Yet it appears, quite compellingly, as three-dimensional. The appearance of depth is entirely inferred, or, to put it another way, hallucinated. This conclusion should be cause for some concern. If, as I suggest, such hallucinations are not an exception but the rule, and if they are in fact a necessary concomitant of visual perception, how can one justify one's faith in perception? How is it still possible that in general seeing is believing?

What is needed for an understanding of vision, therefore, is an explanation of why such visual inferences usually bear a nonarbitrary relation to the real world. A promising line of investigation begins with the observation that the visible world, far from being completely chaotic, obeys certain laws and exhibits numerous regularities. If the visual system is adapted to exploit these laws and regularities in its organization and interpretation of retinal images, and if it is constrained somehow to prefer the interpretation that is most credible, given both the image and a knowledge of these laws and regularities, then it might be possible to understand how it is that one's visual hallucinations bear a nonarbitrary and even useful relation to the external world.

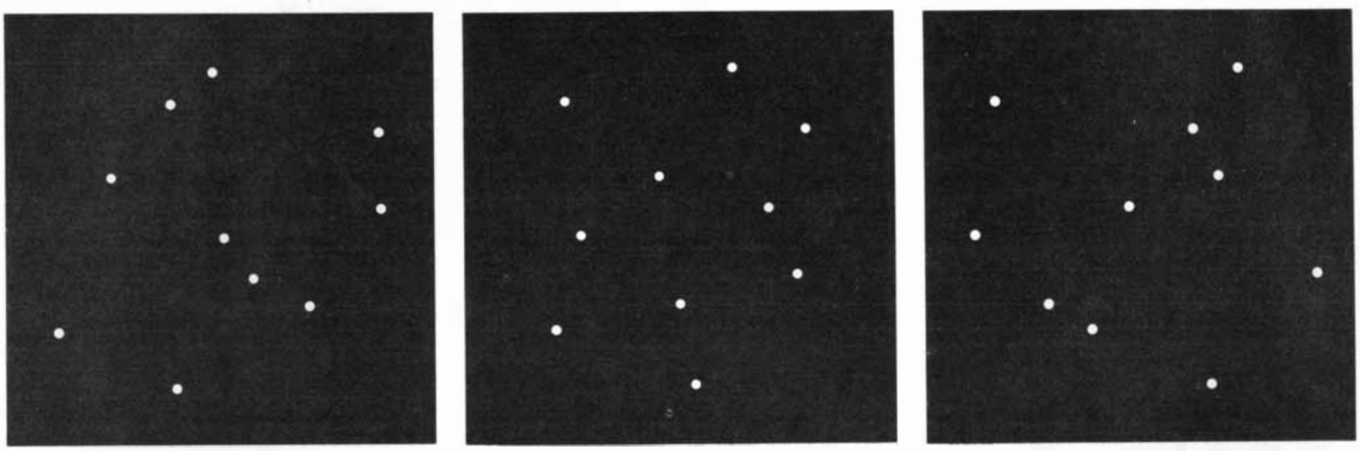
A particularly clear example of this approach is the research into visual mo-


100% GRAIN NEUTRAL SPIRITS. 94.6 PROOF. IMPORTED BY SOMERSET IMPORTERS, LTD. N.Y. © 1983
Suddenly, the idea of importing a Christmas tree from England begins to make sense.
Tanqueray Gin. A singular experience.
Send a gift of Tanqueray Gin any 28-6148. Void where prohibited.

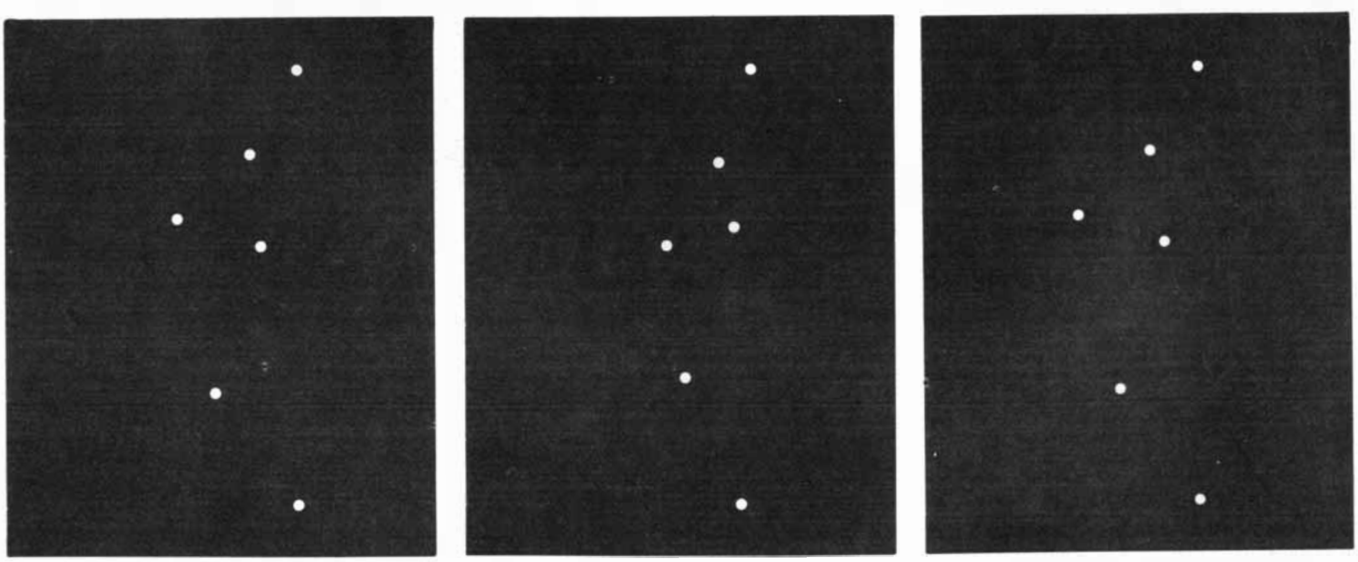
tion perception done by Shimon Ullman of the Massachusetts Institute of Technology. Ullman has explored the remarkable ability of the human visual system to perceive the correct three-dimensional structure and motion of an object solely from its moving two-dimensional projection, an ability Hans Wallach and Donald N. O'Connell of Swarthmore College call the kinetic-depth effect. For instance, if a transparent beach ball with tiny light bulbs mounted randomly on its surface is set spinning in a dark room, one immediately perceives the correct spherical layout of the lights [*see upper illustration below*]. When the spinning stops, so does the perception of the spherical array. How does one see the correct three-dimensional structure when infinitely many three-dimensional structures are consistent with the moving two-dimensional retinal projection? Ullman showed mathematically that if the visual system exploits the laws of projection, and if it exploits the fact that the world contains rigid objects, then in principle a unique and correct interpretation can be obtained. In particular he showed that three views of four noncoplanar light bulbs are enough to solve the problem. The key point is that an inference rule, based on a law (the law of projection) and a regularity (namely the fact that the world includes rigid objects), enables the visual system to make a correct interpretation.

At this stage, however, a puzzle arises. The same mathematical precision that shows the rigidity regularity is sufficient in principle to interpret the rotating beach ball also shows the rigidity regularity by itself is insufficient to interpret a similar display. This display was first devised by Gunnar Johansson of the University of Uppsala as an example of what he calls biological motion [see "Visual Motion Perception," by Gunnar Johansson; SCIENTIFIC AMERICAN, June, 1975]. Johansson put small light bulbs on the major joints of a person and took motion pictures as the person moved about in a dark room. A single frame of such a film looks like a random collection of white dots on a black background. When the film is set in motion, however, one immediately sees the correct three-dimensional structure of the dots and recognizes that there is an invisible person walking about [*see lower illustration below*].

When my colleague Bruce E. Flinchbaugh, who is now at Bell Laboratories, and I considered this problem, what puzzled us was that it is possible to see the correct three-dimensional structure even though, according to Ullman's results, one lacks the appropriate information to do so. To infer a correct three-dimensional structure on the basis of the rigidity regularity it is necessary to have three snapshots of at least four nonco-



ROTATING SPHERE is seen when the three dot patterns represented here are shown in rapid succession. The visual system seems to be adopting the most rigid three-dimensional interpretation for the moving dots that is consistent with the two-dimensional projections.



WALKING PERSON is seen when these dot patterns are shown in rapid succession. In this case the visual system seems to adopt the most rigid and planar three-dimensional interpretation that is consistent with the two-dimensional motions of the dots. The display is based on an experiment conducted by Gunnar Johansson of the University of Uppsala in which small light bulbs were put on a person's major joints (shoulder, elbow, wrist, hip, knee and ankle) and a motion picture was made as the person moved about in a dark room.

planar points in a rigid configuration. In biological motion displays, on the other hand, at best only pairs of points are rigidly connected, such as the ankle and the knee or the knee and the hip. Rigid quadruplets of points just do not exist.

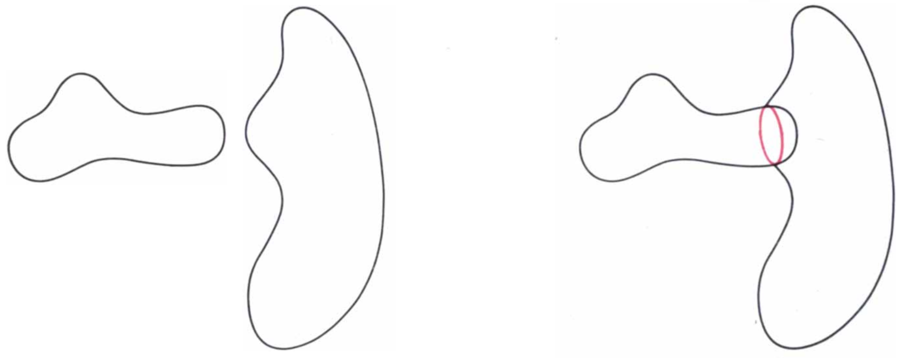
The rigidity regularity, then, is insufficient by itself, leading us to ask: What further regularity might the visual system be exploiting? After several false starts it occurred to us there is an anatomical regularity that might do the trick. Each weight-bearing limb of most animals is constrained, because of the construction of its joints, to swing in a single plane in a normal gait. We call this the planarity regularity.

In fact, the planarity regularity is sufficient to correctly interpret biological-motion displays of gait. The correct three-dimensional structure can be inferred either from three snapshots of two points swinging rigidly in a plane or from two snapshots of three points (such as an ankle, a knee and a hip) forming rigid pairs and swinging in one plane. These results comport nicely with Johansson's observation that only two or three frames of his films need be seen before subjects correctly perceive the biological motion. In addition it turns out not only that all three-dimensional motions governed by the planarity regularity can be given a correct interpretation but also that whenever an interpretation is found for image motion based on the planarity regularity or the rigidity regularity the interpretation is correct.

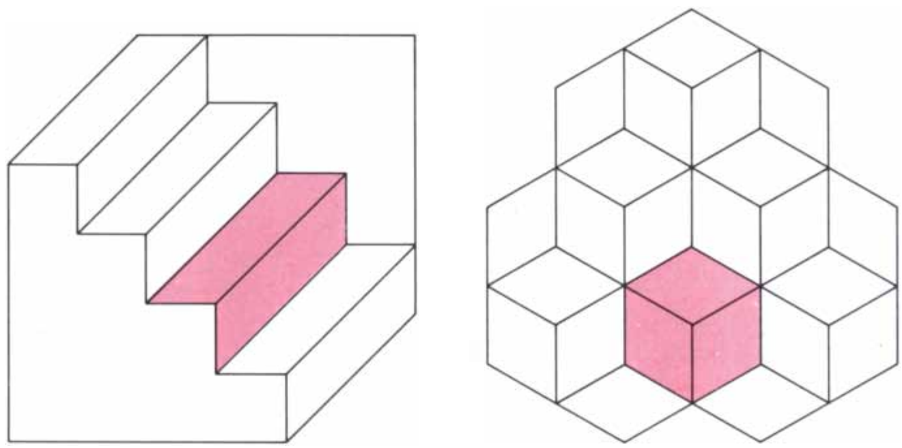
In short, the probability that the interpretation is wrong is zero, assuming infinite resolution in the image, or slightly greater than zero given less than perfect resolution. Hence nonrigid structures cannot masquerade as rigid ones, and nonplanar motions cannot be misconstrued as planar ones. Once again laws and regularities prove to be central in explaining how the visual system achieves a unique and correct interpretation of a retinal image.

Let us now return to the cosine surface. Its main interest is that it reveals the visual system organizing shapes into parts, an organization that is quite useful for the task of recognizing an object from its shape. The cosine surface also reveals that turning a shape upside down can alter this organization. Is the visual system, then, capricious in its organization? That is unlikely. If it is not governed by whim, however, it must be governed by rules for defining parts. And if the rules are not to be arbitrary, they must be grounded in some law or regularity in the external world.

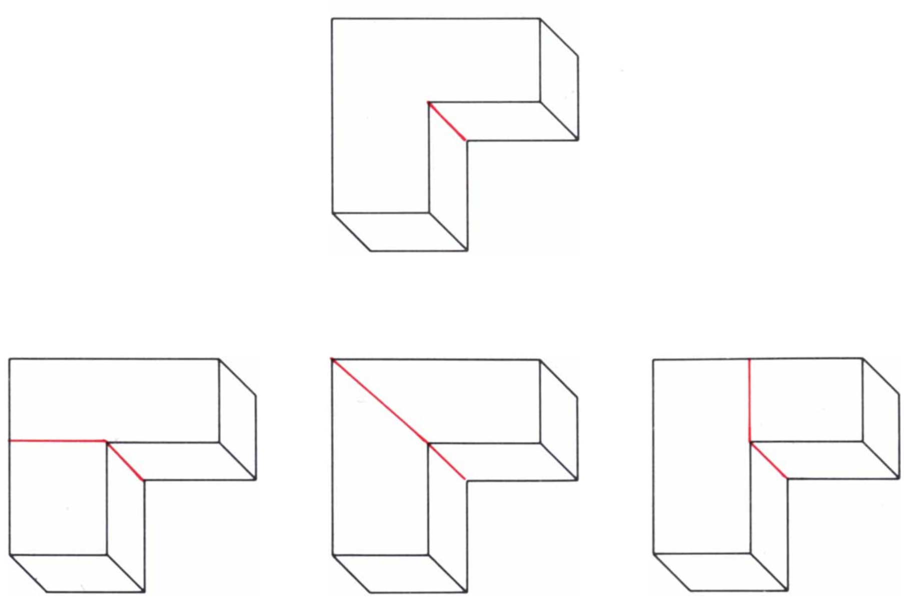
This line of reasoning led Whitman A. Richards of M.I.T. and me to seek a law or regularity that could motivate a set of rules for partitioning surfaces. The regularity we found to be relevant is the following transversality regularity:



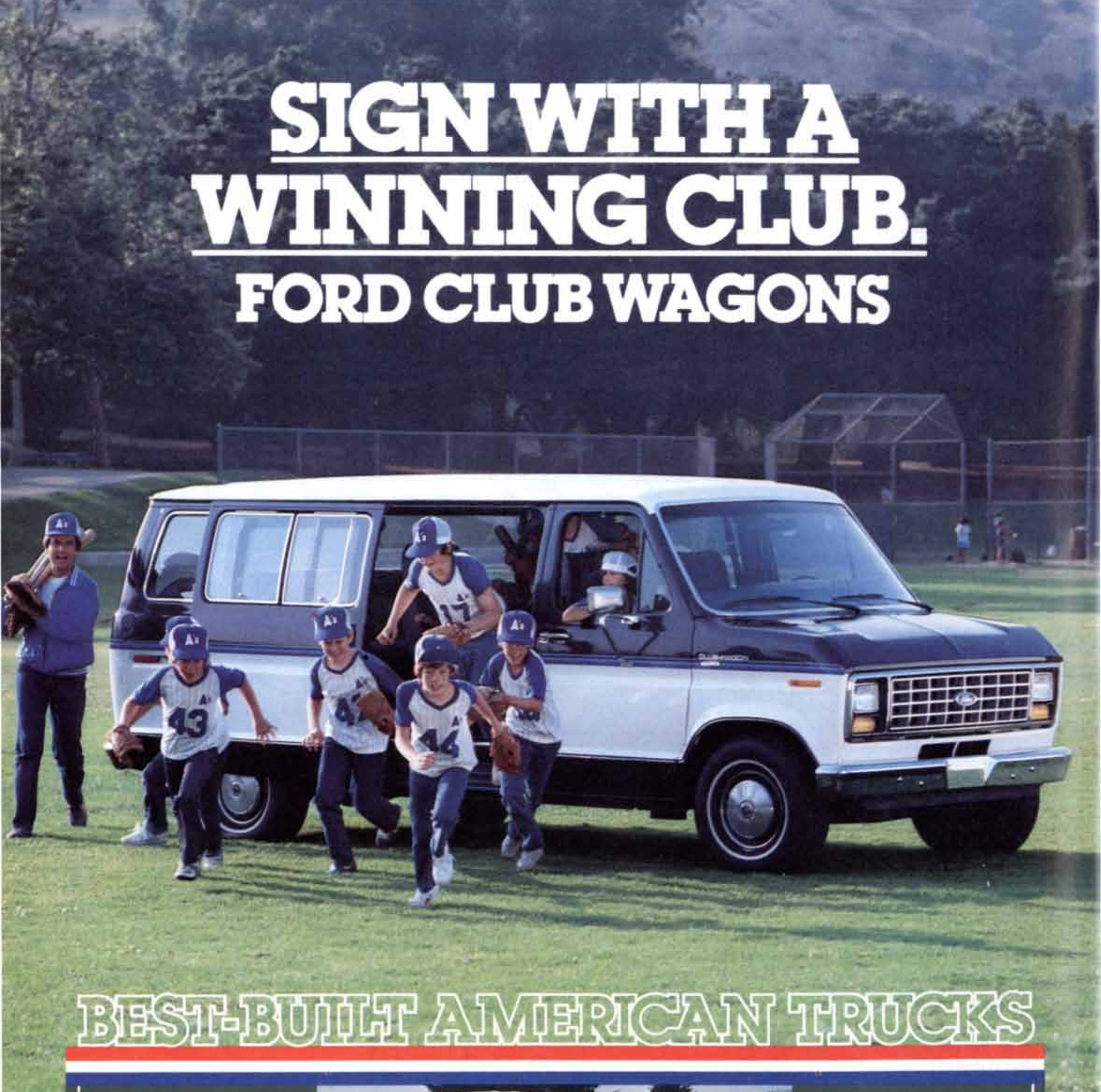
TRANSVERSALITY, a kind of regularity commonly observed in the external world, underlies a unified account of several visual illusions. According to the rule of transversality (as defined by Whitman A. Richards and the author), when any two surfaces penetrate each other at random, they always meet at a concave discontinuity, indicated here by the colored contour.



PARTITIONING RULE based on the transversality regularity is demonstrated with the aid of these two figures that reverse when they are looked at steadily. In both cases the apparent boundaries of the different parts of the perceived shape change when the "figure" becomes the "ground" and vice versa. For example, in the case of the reversing-stairway illusion (*left*), first published by H. Schröder in 1858, the two colored panels, which in one view appear to be parts of one step, suddenly seem to be parts of two adjacent steps when the stairway reverses. Similarly, in the stacked-cube illusion (*right*) the three diamond-shaped colored panels can be seen either as the faces of one cube or, when the figure reverses, as the faces of three different cubes.



ELBOW-SHAPED BLOCKS show that the rule partitioning shapes at concave discontinuities is appropriately conservative. The rule does not give a closed contour on the top block because three different perceived partitions seem possible, as illustrated by the bottom three blocks.



Ford leads off with the best-built van/passenger van. With out-front design, Ford Club Wagons can out-tow and out-carry any station wagon. There's passenger seating for up to 12 (15 in the roomy Super Wagon)! It's built tough, with body-on-frame construction. No wonder Ford is a winning club!

Get it together–Buckle up

TOWS UP TO 3,000 LB. MORE THAN CHEVY OR DODGE! Ford scores with power, rated up to 10,000 lb. towing capacity when properly equipped. Plus the most powerful gas or diesel engines.

Quality is Job 1.

Ford's commitment to quality results in the best-built American trucks, based on a survey of owner problems during the first three months of ownership of 1983 trucks. And the commitment continues in 1984.

BUILT FORD TOUGH

Ford

When two arbitrarily shaped surfaces are made to penetrate each other at random, they always meet at a contour of concave discontinuity of their tangent planes [*see top illustration on page 157*]. Although the transversality regularity may sound esoteric, it is actually a familiar part of everyday experience. A straw in a soft drink, for instance, forms a circular concave discontinuity where it meets the surface of the drink. A candle in a birthday cake, the tines of a fork in a piece of steak, a cigarette in a mouth—all are examples of this ubiquitous regularity.

On the basis of the transversality regularity one can propose a first rule for partitioning a surface: Divide a surface into parts along all contours of concave discontinuity. This rule cannot help with the cosine surface because it is entirely smooth. The rule must first be generalized somewhat, as will be done below. In its nongeneralized form, however, it can elucidate several well-known perceptual demonstrations.

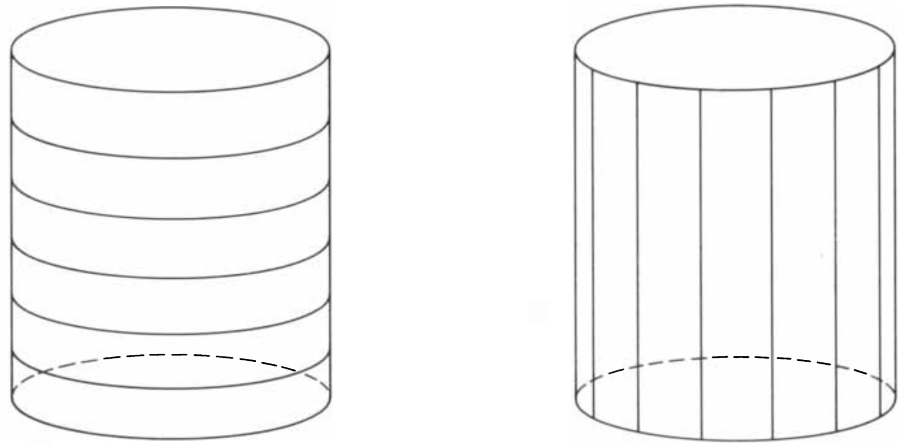
For example, the rule makes the obvious prediction that the parts of the staircase shown in the middle illustration on page 157 are its steps, each step lying between two successive lines of concave discontinuity in the staircase. The rule also makes a less obvious prediction. If the staircase undergoes a perceptual reversal, such that the "figure" side becomes "ground" and vice versa, then the step boundaries must change. This conclusion follows because only concave discontinuities define the step boundaries, and what looks like a concavity from one side of a surface must look like a convexity from the other. Thus when the staircase reverses, convex and concave discontinuities must reverse roles, leading to new step boundaries. You can test this prediction yourself by looking at the step that has color on each of its two faces. When the staircase appears to reverse, note that the colored panels are no longer on a single step but rather on adjacent steps.

This prediction can be confirmed with a more complicated demonstration such as the stacked-cubes test seen in the same illustration. The three colored faces, which at first appear to be on one cube, are seen to be on three cubes when the figure reverses.

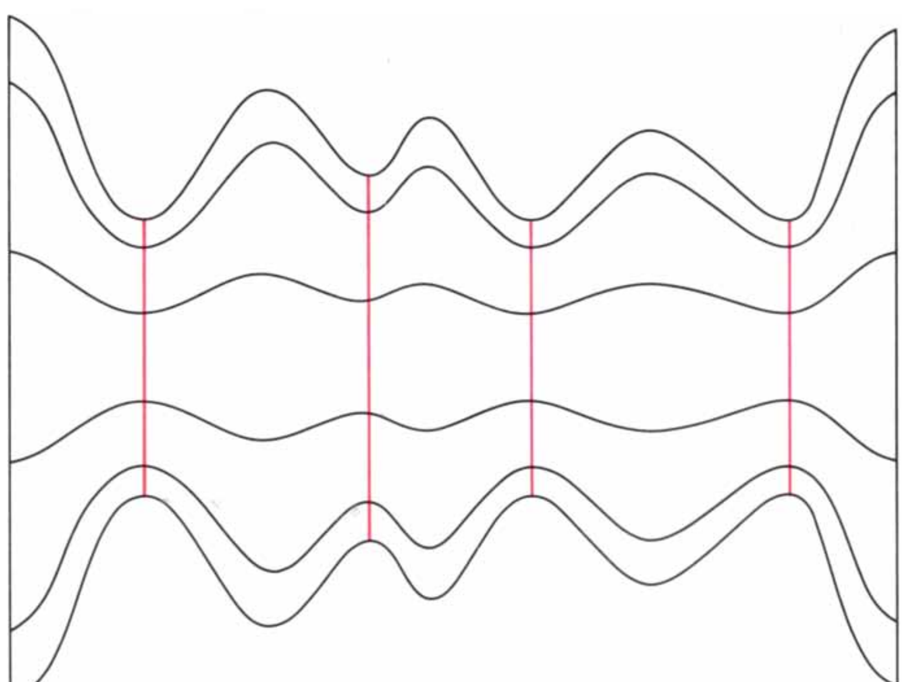
A further prediction follows from this simple partitioning rule. If the rule does not define a unique partition of some surface, then the appropriate way to divide the surface into parts should be perceptually ambiguous (unless there are additional rules that can eliminate the ambiguity). A clear confirmation of this prediction can be seen with reference to the elbow-shaped block in the bottom illustration on page 157. The only concave discontinuity is the vertical line in the crook of the elbow. As a consequence the rule does not define a unique partition of the block. Perceptually there are three plausible ways to cut the block into parts. All three ways rely on the contour defined by the partitioning rule, but they complete it along different paths.

Even this simple partitioning rule leads to interesting insights into the perception of shape. To explore the cosine surface and other smooth surfaces, however, the rule must be generalized. This requires a brief digression into the differential geometry of surfaces in order to understand three important concepts: surface normal, principal curvature and line of curvature. Fortunately, although these concepts are quite technical, they can readily be given an intuitive characterization.

The surface normal at a point on a surface can be thought of as a needle of unit length sticking straight out of the surface at that point, much like the spines on a sea urchin. All the surface normals at all points on a surface are collectively called a field of surface normals. Usually there are two possible fields of surface normals on a surface; they can be either outward-pointing or inward-pointing. For example, a sphere can have the surface normals all pointing radially out like spines or all pointing in toward its center. Let us adopt the convention that the field of surface normals is always chosen to point into the figure. Thus a baseball has inward normals whereas a bubble under water has outward normals. Reversing the choice of figure and ground on a surface im-



LINES OF CURVATURE are easily pictured on an idealized cylindrical drinking glass. The lines of greatest curvature (*left*) are circles; the lines of least curvature (*right*) are straight lines.



PART BOUNDARIES, as defined by the generalized, smooth-surface partitioning rule, are represented by the colored contours on this arbitrarily shaped surface. The black lines are the lines of greatest curvature whose minimums give rise to the colored partitioning contours.

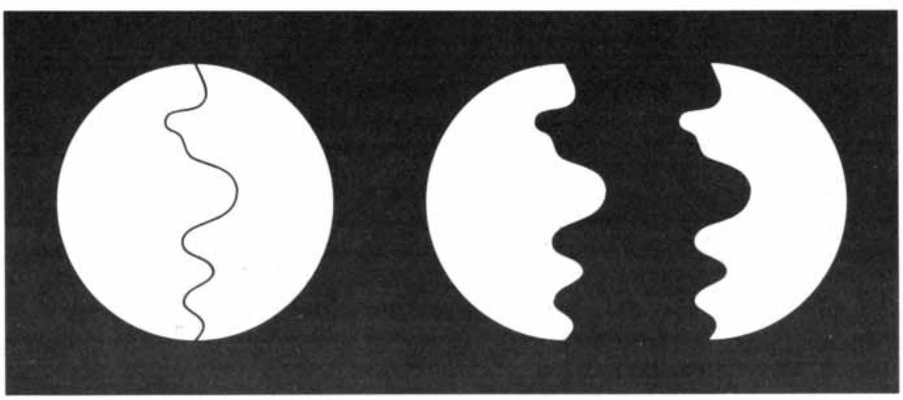
plies a concomitant change in the surface normals. A reversal of the field of surface normals induces a change in sign of each of the principal curvatures at every point on the surface.

It is often important to know not only the surface normal at a point but also how the surface is curving at the point. The 18th-century Swiss mathematician Leonhard Euler discovered that at any point on any surface there is always a direction in which the surface curves the least and a second direction, always at right angles to the first, in which the surface curves the most. (In the case of a plane or a sphere the surface curvature is identical in all directions at every point.) These two directions are called principal directions, and the corresponding surface curvatures are called principal curvatures. By starting at some point and always moving in the direction of the greatest principal curvature one traces out a line of greatest curvature. By moving instead in the direction of the least principal curvature one traces out a line of least curvature. On a drinking glass the family of lines of greatest curvature is a set of circles around the glass. The lines of least curvature are straight lines running the length of the glass [*see top illustration on preceding page*].

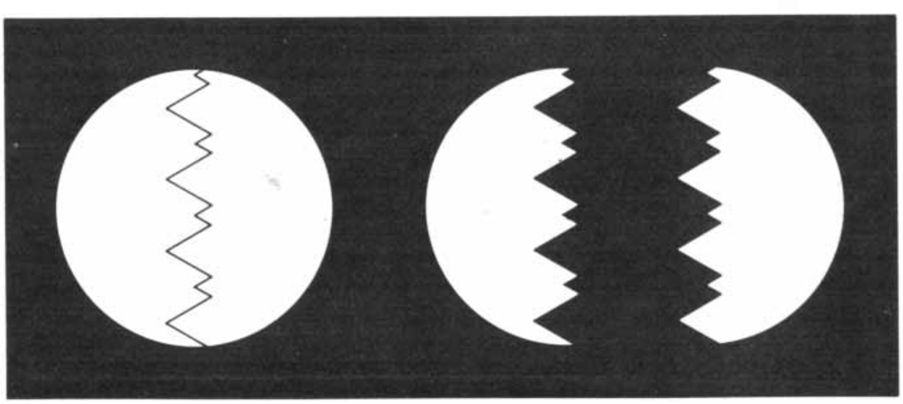
With these concepts in mind the transversality regularity extends easily to smooth surfaces. Suppose wherever a surface has a concave discontinuity one smoothes the discontinuity somewhat, perhaps by stretching a taut skin over it. Then a concave discontinuity becomes, roughly speaking, a contour where the surface has locally the greatest negative curvature. More precisely, the generalized version of transversality suggests the following generalized partitioning rule for surfaces: Divide a surface into parts at negative minimums of each principal curvature along its associated family of lines of curvature [*see bottom illustration on preceding page*].

This rule partitions the cosine surface along the colored circular contours. It also explains why the parts are different when the page is turned upside down: the visual system then reverses its assignment of figure and ground on the surface (perhaps owing to a preference for an interpretation that places the object below the observer's viewpoint rather than above it). When figure and ground reverse, so does the field of surface normals, in accordance with the convention mentioned above. Simple calculations show, however, that when the normals reverse, so does the sign of the principal curvatures. As a result minimums of the principal curvatures must become maximums and vice versa. Since minimums of the principal curvatures are used for part boundaries, it follows that these part boundaries must also move. In sum, parts appear to change because the partitioning rule, motivated by the transversality regularity, uses minimums of the principal curvatures, and because these minimums relocate on the surface when figure and ground reverse.

The transversality regularity, in short, provides an underlying unity for explanatory accounts of the perception of parts in both smooth and rough surfaces. It also underlies an explanation of another well-known class of visual illusions: reversing plane curves. A good example of this phenomenon is the reversing figure devised by Fred Attneave of the University of Oregon [*see upper illustration on this page*]. He found that by simply scribbling a line through a circle and separating the two halves one can create two very different-looking contours. Evidently, as Attneave points out, the appearance of the contour depends on which side is taken to be part of the figure, not on any prior familiarity with the contour [see "Multistability in Perception," by Fred Attneave; SCIENTIFIC AMERICAN, December, 1971].



REVERSING PLANE CURVE, constructed by Fred Attneave of the University of Oregon by scribbling a line through a circle and separating the two halves, shows that the apparent shape of the resulting contour depends on which side of the line is perceived as the figure.

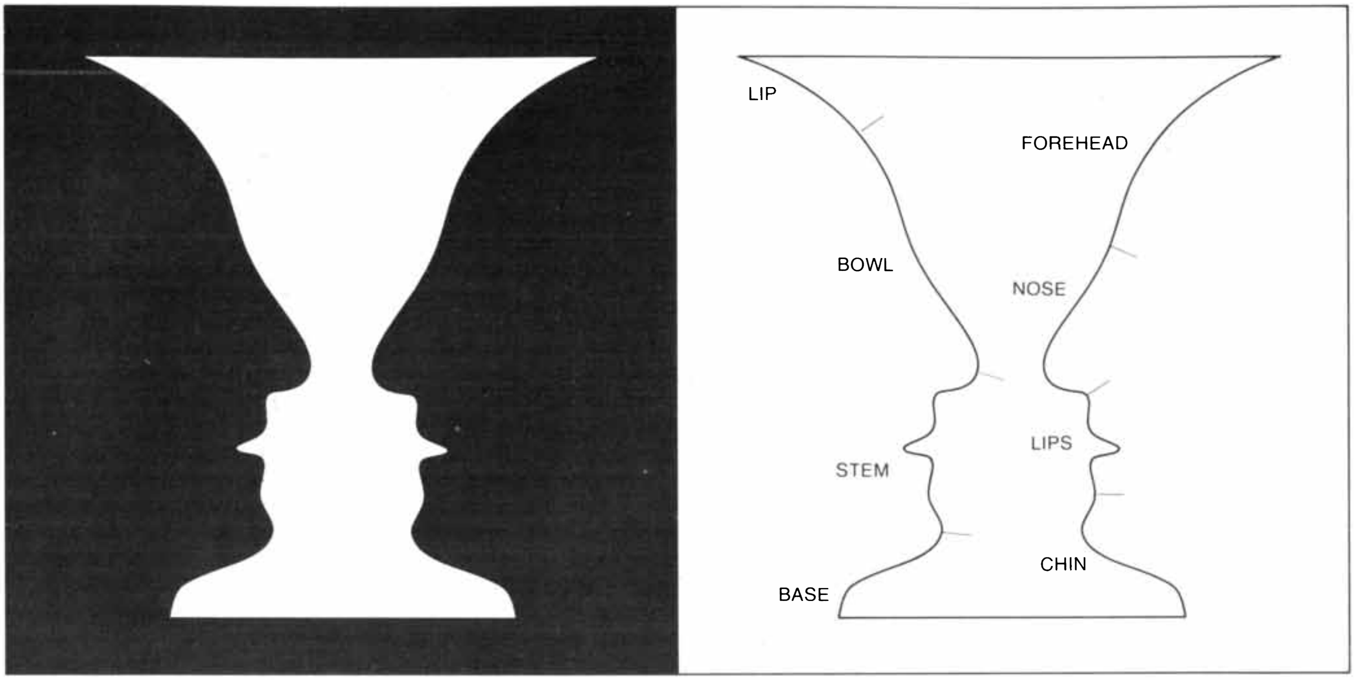


SIMILAR REVERSING FIGURE can be made with a plane curve that is not smooth. One can see the resulting jagged contour either as an alternating chain of tall and short mountains or, in the reversed figure-ground assignment, as a chain of tall mountains with twin peaks.

How does the transversality regularity explain this phenomenon? The answer involves three steps: (1) a projection of the transversality regularity from three dimensions onto two dimensions, (2) a brief digression on the differential geometry of plane curves and (3) the formulation of a partitioning rule for plane curves.

The two-dimensional version of the transversality regularity is similar to the three-dimensional version. If two arbitrarily shaped surfaces are made to penetrate each other at random, then in any two-dimensional projection of their composite surface they will always meet in concave cusps. To paraphrase it loosely, concave cusps are always formed in a silhouette at points where one part stops and another begins. This suggests the following partitioning rule for plane curves: Divide a plane curve into parts at concave cusps. This rule cannot apply to Attneave's demonstration because his demonstration relies on a contour that is everywhere smooth. The rule must again be generalized. Nevertheless, in its nongeneralized form it can account for a version of Attneave's demonstration that is not everywhere smooth.

In the lower illustration at the left the same jagged contour can look either like an alternating chain of tall and short mountains or, for the reversed



FACE-GOBLET ILLUSION, devised by Edgar Rubin in about 1915, can be seen either as a pair of facial profiles or as a goblet (*left*). If a face is taken to be the figure, partitioning the figure by reference to minimums of curvature divides the contour into chunks corresponding to a forehead, a nose, a pair of lips and a chin; if the goblet is taken to be the figure, defining the part boundaries by minimums of curvature divides the contour into a lip, a bowl, a stem and a base (*right*). Principal-normal lines in both cases point into the figure.

figure-ground assignment, like a chain of tall mountains with twin peaks. The contour is carved into parts differently when figure and ground reverse because the partitioning rule uses only concave cusps for part boundaries. What is a concave cusp if one side of the contour is figure must become a convex cusp when the other side is figure, and vice versa. There is a parallel between this example and the reversible staircase discussed above.

Before generalizing the rule to smooth contours let us briefly review two concepts from the differential geometry of plane curves: principal normal and curvature. The principal normal at a point on a curve can be thought of as a unit-length needle sticking straight out of the curve at that point. All the principal normals at all points on a curve together form a field of principal normals. Usually there are two possible fields of principal normals, one field on each side of the curve. Let us adopt the convention that the field of principal normals is always chosen to point into the figure side of the curve. Reversing the choice of figure and ground on a curve implies a concomitant change in the field of principal normals. What is important to note is that because of the convention forcing the principal normals to point into the figure, concave parts of a smooth curve have negative curvature and convex parts have positive curvature.

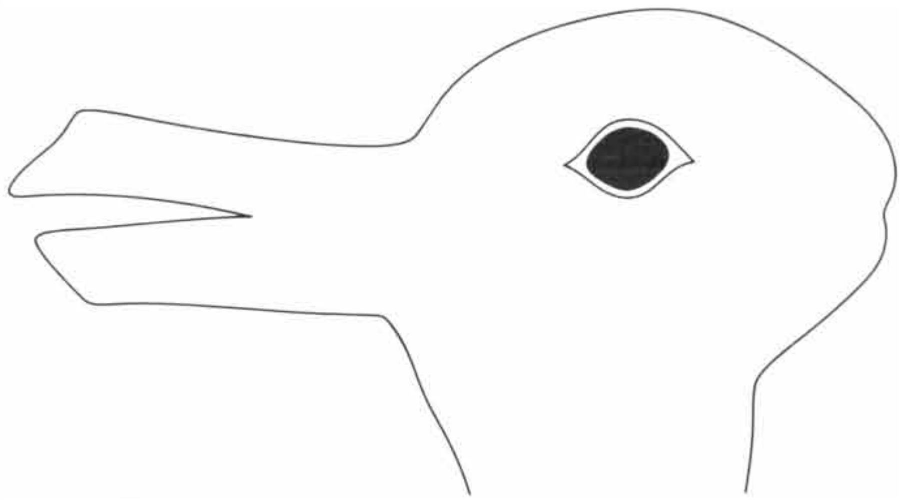
It is an easy matter now to generalize the partitioning rule for plane curves. Suppose wherever a curve has a concave cusp one smoothes the curve a bit. Then a concave cusp becomes a point of negative curvature having, locally, the greatest absolute value of curvature. This observation leads to the following generalized partitioning rule: Divide a plane curve into parts at negative minimums of curvature.

Now it is possible to explain why the two halves of Attneave's circle look so different. When figure and ground reverse, the field of principal normals also reverses in accordance with the convention, and when the principal normals reverse, the curvature at every point on the curve must change sign. In particular, minimums of curvature must become maximums and vice versa. This repositioning of the minimums of curvature leads to a new partitioning of the curve by the partitioning rule. In short, the curve looks different because it is organized into fundamentally different chunks, or units. Note that if one chooses to define part boundaries by inflections, or by both maximums and minimums of curvature, then the chunks would not change when figure and ground reverse.

A clear example of two very different partitions for one curve can be seen in the famous face-goblet illusion devised by Edgar Rubin in about 1915 [*see illustration above*]. If a face is taken to be figure, the minimums of curvature divide the curve into chunks corresponding to a forehead, nose, upper lip, lower lip and chin. If instead the goblet is taken to be figure, the minimums are repositioned, dividing the curve into new chunks corresponding to a base, a couple of parts of the stem, a bowl and a lip on the bowl. It is probably no accident that the parts defined by minimums are often easily assigned verbal labels.

Demonstrations have been devised that, like the face-goblet illusion, allow more than one interpretation of a single contour but that do not involve a figure-ground reversal. A popular example is the rabbit-duck illusion [*see top illustration on next page*]. Because such illusions do not involve a figure-ground reversal, and because as a result the minimums of curvature never change position, the partitioning rule must predict that the part boundaries are identical for both interpretations of each of the contours. This prediction is easily confirmed. What is an ear on the rabbit, say, becomes part of a bill on the duck.

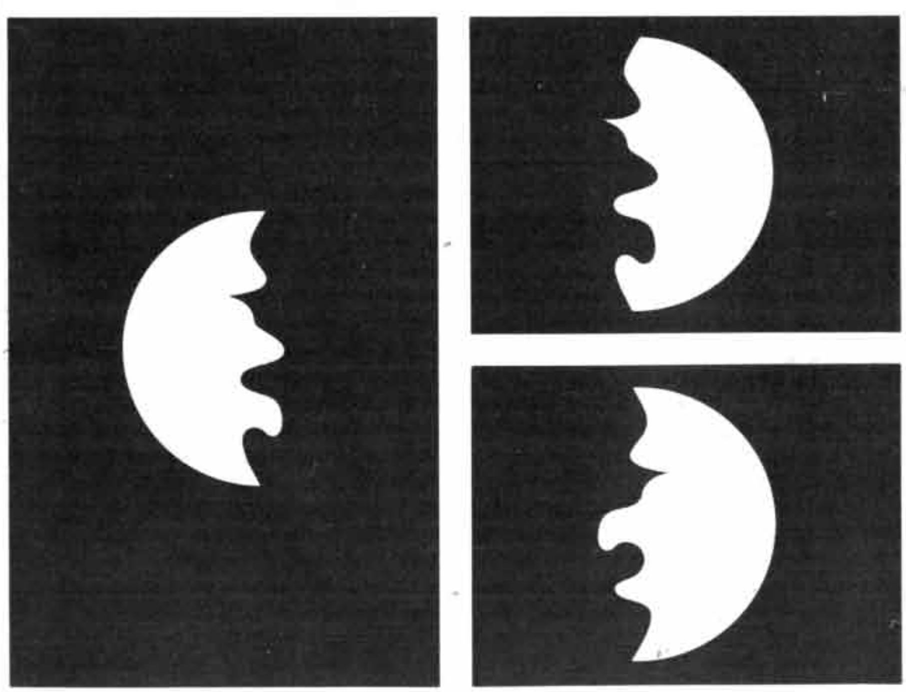
If the minimums rule for partitioning curves is really obeyed by the human visual system, one would expect it to predict some judgments of shape similarity. One case in which its prediction is counterintuitive can be seen in the bottom illustration on the next page. Look briefly at the single half-moon on the left side of the illustration. Then look quickly at the two half-moons at the right and decide which one seems more like the first one. In an experiment done on several similar figures, Aaron F. Bobick of M.I.T. and I found that almost all subjects chose the half-moon at the lower right as the more similar one. Yet if you look again, you will find that the bounding contour for the half-moon at



REVERSING-ANIMAL ILLUSION does not involve a reversal of figure and ground. Accordingly the part boundaries defined by the minimums of curvature do not change position when the interpretation changes. The rabbit's ears turn into the duck's bill without moving.

the upper right is identical with that of the left half-moon, only it is figure-ground reversed. The bounding contour of the lower half-moon has been mirror reversed, and two parts defined by minimums of curvature have been swapped. Why does the lower one still look more similar? The minimums rule gives a simple answer. The lower contour, which is not figure-ground reversed from the original contour, has the same part boundaries. The upper contour, which is figure-ground reversed from the original, has different part boundaries.

In summary, vision is an active process whose function is to infer useful descriptions of the world from changing patterns of light falling on the retinas. These descriptions are reliable only to the extent that the inferential processes building them exploit regularities in the visual world, such as rigidity, planarity and transversality. The discovery of relevant regularities and the mathematical investigation of their power in guiding visual inferences are promising directions for the investigator seeking an understanding of human vision.



HALF-MOON TEST demonstrates that judgments of the similarity of shapes can be correctly predicted by the minimums-of-curvature partitioning rule. At first glance the half-moon at the lower right seems to resemble the single half-moon at the left more than the one at the upper right does. Closer inspection, however, reveals that the bounding contour of the upper-right half-moon is identical with that of the half-moon at the left, whereas the bounding contour of the lower-right half-moon has been mirror reversed and has also had two parts interchanged.

A Smart Investment. Give your family a colorful Christmas this year with our TRS-80 Color Computer 2. It's the perfect way to join the computer revolution. And, at $80 off, it's the affordable way, too!

Easy to Use. Just connect your TV, plug in a ready-to-run Program Pak™ and you're all set. Prepare a household budget, monitor your investments, teach your kids or play thrilling outer space and sports games.

Deluxe Features. Both models feature a typewriter-quality keyboard in an attractive white case. Use the built-in Color BASIC language and our entertaining 308-page manual to start writing your own programs with color and sound. Want to create stunning high-resolution graphics and write advanced programs?

Then select our sale-priced 16K Extended BASIC Color Computer 2. Create incredible drawings and designs, business and engineering charts, even animation with simple one-line commands.

Expandable. Best of all, Color Computer 2 can expand easily as your computing skills grow. Add a cassette recorder to store and load your own programs, a telephone modem to access tomorrow's news and stock reports, a printer, joysticks or a color mouse for more exciting game action, disk drives and more.

The Perfect Gift to Give . . . Or Get! Come into a Radio Shack Computer Center, participating store or dealer near you and get a demonstration of our spectacular Color Computer 2.

Radio Shack®

The biggest name in little computers®

A DIVISION OF TANDY CORPORATION

Send for your free 1984 TRS-80 Computer Catalog.

Mail To: Radio Shack, Dept. 84-A-702
300 One Tandy Center, Fort Worth, Texas 76102

NAME ____________________

COMPANY ____________________

ADDRESS ____________________

CITY ____________ STATE ______ ZIP ________

TELEPHONE ____________________

Sale ends Dec. 31, 1983. TV not included. Prices apply at participating Radio Shack stores and dealers.

THE AMATEUR SCIENTIST

Funny things happen when drops of oil or other substances are placed on water

by Jearl Walker

What happens when a drop of oil or some other liquid is placed on the surface of water? If it is soluble in water, it disappears quickly into the bulk liquid. If it is insoluble, as paraffin oil is, it remains where it is placed. If it has an intermediate solubility, as olive oil has, it may spread over the surface of the water, eventually becoming invisibly thin; it will pool or bead on the surface if more drops are added. What determines the behavior of the drop?

The answer to the question remains somewhat uncertain. Part of it is that the behavior of a drop placed on a water surface is determined by the complex interactions (mostly electrical) of the molecules at the interface. Recently Frode Wissing of the Royal Dental College in Denmark sent me a manuscript about his experiments on the solubility of oleic acid (a major component of olive oil) and paraffin oil in water. His work, which was designed as a classroom demonstration, reveals several curious pieces of the solubility puzzle.

The solubility of a chemical compound in water depends on the extent of the bonding between the molecules of the compound and those of the water. The degree of solubility results from a competition between the bonds that hold each molecule together and the alternative bonding opportunities offered by the other substance.

Organic compounds range widely in their solubility in water. Life on the earth would not exist without this variability. The insoluble organic compounds have component groups of atoms that form few bonds (and in some cases none) with water molecules. Such groups are said to be hydrophobic, as is a molecule that is insoluble because of such groups. The term is misleading in that it implies a repulsion between the molecule (or a group) and the water. The effect arises not from repulsion but from the fact the bonding is so weak that the cohesion of the water keeps the hydrophobic compound out.

Paraffin oil is an example of a hydrophobic compound. It consists entirely of hydrophobic groups of carbon and hydrogen atoms. Since hydrocarbon groups form few bonds or none with water molecules, a drop of paraffin oil stays intact when it is placed on water. The drop pushes down on the water surface because of its weight, but it cannot dissolve in the water or even spread out over the surface.

Many other organic molecules are partially soluble in water because at least some of their atomic groups bond with water. The more such groups a compound has, the more soluble it is. For example, the sugar glucose dissolves readily in water because it has six hydroxyl (OH) groups that bond well to water molecules.

Oleic acid is intermediate to paraffin and glucose. One end of the molecule has a COOH group that bonds well to water. Such a group is said to be hydrophilic. The rest of the molecule, which is a chain of hydrocarbon groups, is hydrophobic. This dual nature of the oleic acid molecule is responsible for its ability to spread over the water in a layer only one molecule thick.

Wissing set out to illustrate the differences in solubility of paraffin oil (strictly hydrophobic) and oleic acid (partly hydrophilic). He did his experiments in petri dishes that were nine centimeters in diameter and absolutely clean. (Any contamination, even from a fingerprint, deposits a monolayer on the water surface and may alter the results.) A dish was placed on an overhead projector and surrounded by dark paper to make the light pass only through the dish. The arrangement yielded a magnification of about 10 for the events in the dish, more than enough for a class to follow as Wissing added drops of organic fluids to water in the dish. I repeated his experiments in my kitchen satisfactorily even though I substituted pure olive oil for the oleic acid.

Wissing first demonstrated the insolubility of paraffin oil by placing a drop of it on distilled water. Since the oil is insoluble, it merely rested on the water in a lenslike shape. As more drops were put on the water some coalesced to form a larger lens, but none spread over the surface or dissolved.

Wissing then did three similar experiments with dishes containing other solutions: hydrochloric acid, sodium hydroxide and ammonium hydroxide, each in the dilute concentration of .1 molar. In each dish the oil formed a lens, neither dissolving in the solution nor spreading over its surface.

Repeating the demonstrations with a drop of oleic acid, Wissing got more dramatic results. He had dyed the oleic acid with a small amount of Sudan III Red so that he could see it better. When he put a drop of oleic acid on pure water, the water surface developed a shock wave, as if someone had thumped the dish. The drop spread quickly over the surface, leaving a film so thin that even the red dye did not show up. Some of the acid also formed into a large lenslike drop and several smaller ones.

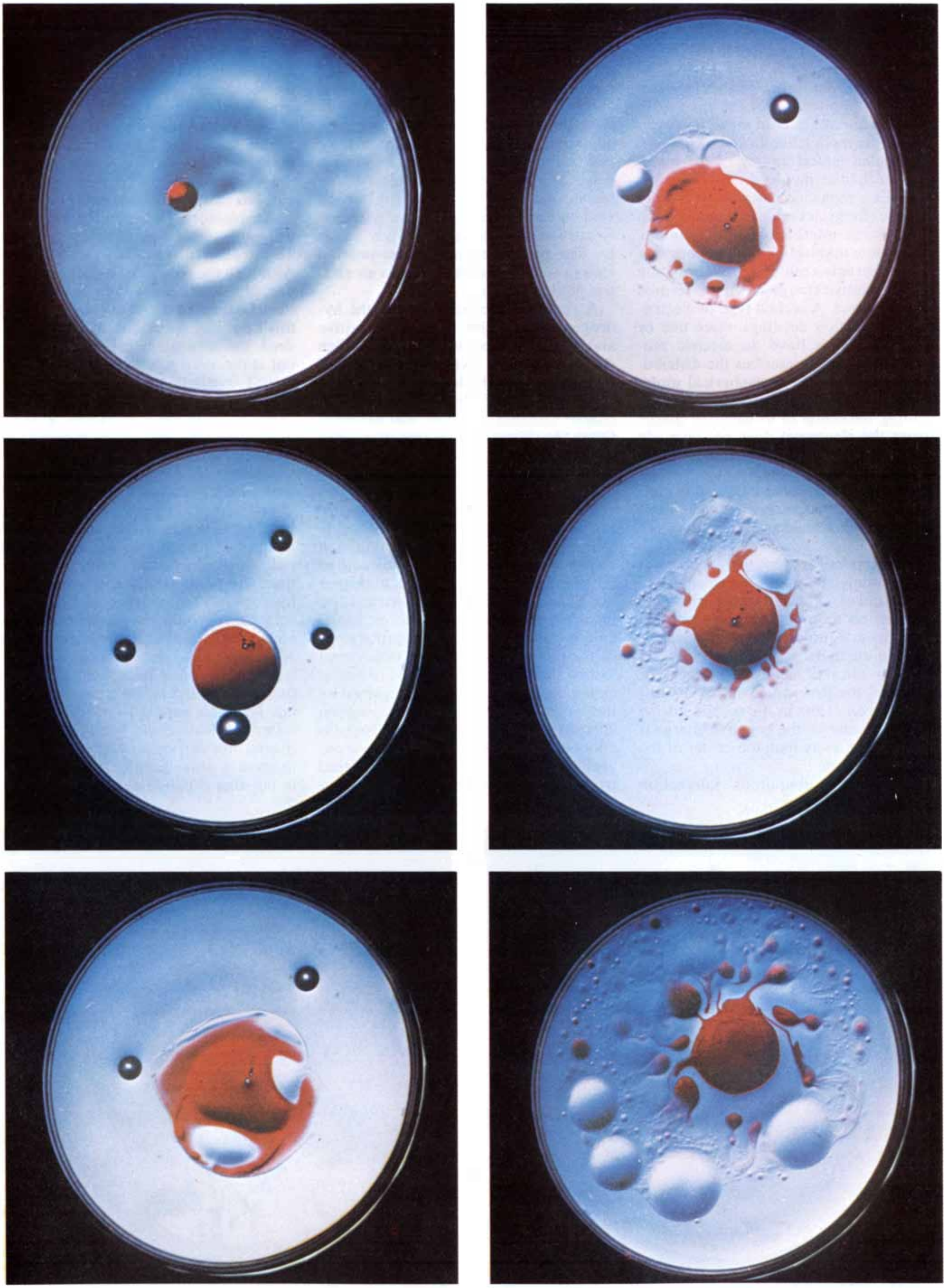
In my experiments I deployed a tiny drop of olive oil by flicking it from a straightened paper clip. (I had previously cleaned the clip with detergent and then heated one end to red heat.) The drop immediately disappeared as it spread over the surface, forming a layer much too thin to be observed directly or even to generate the kind of optical interference seen when an oil slick lies on water. The surface must have been almost completely covered with the olive oil, because a second drop held together for several minutes before spreading gradually into a layer that was quite visible in some places and at least thick enough in others for optical interference to create colored bands on it.

When Wissing added a drop of oleic acid to a dish of dilute hydrochloric acid, it again spread over the entire surface and formed a large drop. This time the lens was flatter, nearly filling the dish. Oleic acid added to a solution of sodium hydroxide formed an irregular lens. The acid flowed slowly outward from the drop in thin streams. After 20 minutes the lens had disappeared, leaving a turbid solution in the dish.

A drop added to the ammonium hydroxide zipped over the surface in small streams, each one bordered with clear zones, and then disappeared. The surface became calm after the acid stopped reacting with the ammonium hydroxide. Reaction products lay on the surface, but the bulk liquid showed no turbidity.

Wissing set himself the task of explaining those variations in terms of the forces acting between the oleic acid and the different solutions. The simplest force is gravity, which pulls a drop of oleic acid down into the solution, tending to spread it over the surface. Another force, arising from the displacement of the solution, is buoyancy.

More complex (and interesting) are



Frode Wissing's experiments in which a drop of oleic acid consumes paraffin drops on water the many electrical interactions that can take place between the molecules in the drop and those in the solution. I shall discuss the simple quantum-mechanical models of these forces and then follow Wissing's application of the models to his experiments. Bear in mind, however, that all the models are flawed by the lack of detailed understanding of the quantum mechanics involved.

When a molecule has sites with a net electric charge, it is able to bond through electrostatic interactions. For example, a molecule that has a site of net positive charge attracts a molecule that has a site of net negative charge in what is termed an ionic bond. A second type of electrostatic attraction develops when one or both molecules have an electric moment. (This term describes the distribution of charge in a nonspherical molecule.) Although the molecule may be neutral, the center of its negative charge (from the electrons) does not coincide with the center of its positive charge (from the protons). Such a state sets up an electric field surrounding the molecule. When two such molecules sample each other's electric field, they are attracted.

A third possible attractive force arises when a molecule that has a strong electric moment is near another molecule that has no moment. The electric field from the first molecule shifts the charge distribution in the second molecule, creating an attraction between them. For example, the first molecule might shift the electron orbits in the second one so that the center of the negative charge is then farther away than the center of the positive charge.

A virtually ubiquitous interaction between molecules with permanent electric moments is the hydrogen bond. Here attention is directed to a hydrogen atom lying between the molecules and serving to hold them together. Water is cohesive because of the hydrogen bond. A water molecule consists of one oxygen atom and two hydrogen atoms, together forming a wide *V* with the oxygen at the apex. Although the molecule has no net charge, the electron associated with each hydrogen atom is strongly attracted to the large oxygen nucleus. The molecule thus has an internal charge distribution that creates an electric field around it.

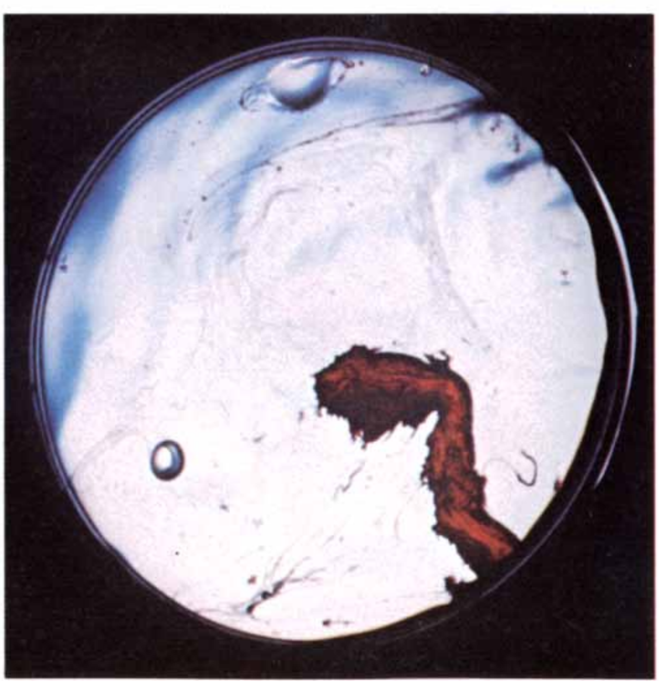
A simpler description is that the hydrogen ends of the *V* are left positive and the oxygen apex is left negative. One should resist the oversimplification of stating that the hydrogen atoms have been stripped of their electrons. The charge distribution is not that extreme. Often the charge distribution of a water molecule is said to be polar to indicate the presence of an electric moment.

Imagine two adjacent water molecules aligned as is shown in the top illustration on page 170. The molecule at the left presents one of its hydrogen atoms to the oxygen in the molecule at the right, thereby forming a hydrogen bond. The attraction is an electrical one in which the positive end of one molecule is pulled toward the negative apex of the other. The hydrogen atom is not lost by the molecule at the left (which is called the hydrogen donor) or gained by the molecule at the right (the hydrogen acceptor). The intermediate hydrogen is about a third of the way between one oxygen and the next and is said to be shared by the two molecules, even though it is still more strongly held by the donor.

Although this picture of a hydrogen bond serves well, I should point out that no mathematical model yet devised has proved to be completely successful in explaining the bonding. The interaction of two water molecules is truly not the simple static one I have described. In bulk water the stablest arrangement of the molecules arises when each molecule has hydrogen bonds with four adjacent molecules. To two of them the central molecule acts as a hydrogen donor. To the other two it acts as an acceptor. Although the arrangement is the stablest one, the hydrogen bonds can still be stretched, rotated and broken. Without this flexibility water could not flow. Indeed, the arrangement of molecules is not static even when water lies undisturbed in a container. Bonds are constantly being broken and re-formed.

Water frequently dissolves a compound because the water molecule is small enough to move close to the compound's molecules. If the compound has sites with a net charge, the water forms ionic bonds through an electrostatic attraction as it presents either its oxygen atom or a hydrogen atom to the charged site. At places on the molecule where there are polar groups the water forms hydrogen bonds with the compound.

In general only oxygen and nitrogen atoms participate in hydrogen bonding. A carbon atom cannot do so because it tends to surround itself with hydrogen (to form a hydrocarbon group) and does not leave an exposed end to which a water molecule could bond. Moreover, it does not set up an electric moment to form a polar group. This difference in bonding capabilities between carbon



The effect of ammonium hydroxide on paraffin and oleic acid

and oxygen is the main reason the hydrocarbon groups in oleic acid are hydrophobic whereas the COOH group at the end of the molecule is hydrophilic.

Additional bonding can be provided (although rarely) by the natural dissociation of water into H+ and OH− ions. They are attracted to sites on the compound that have an electric moment or a net charge. Finally, the compound might be able to dissociate to provide a hydrogen atom for a water molecule. Then the negative site left on the compound attracts the positive water ion.

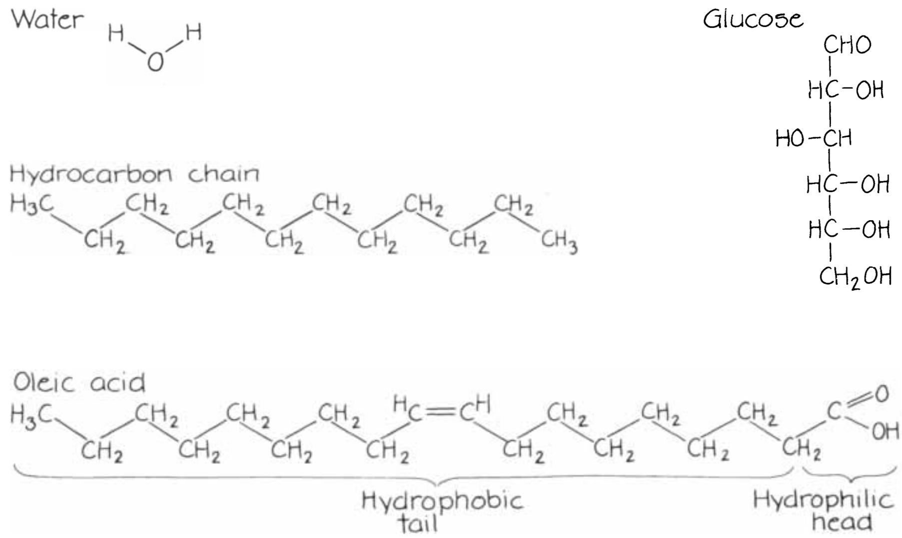
When two spherical and neutral molecules are close, there should be no electrical interaction. Because they are spheres, the center of each molecule's negative charge coincides with the center of its positive charge and so they lack electric moments. Ionic attraction is absent. Still, they can attract each other through a curious circumstance.

In the simple quantum-mechanical picture of the molecules the orbits of the electrons become synchronized, creating instantaneous electric moments that set up fleeting electric fields. The fields give rise to an attraction between the molecules. The attraction is usually termed the van der Waals force.

In a simple picture of the van der Waals force one first imagines the electrons in orbit in adjacent molecules. Each molecule should "see" the other as electrically neutral because the positive and negative charges within the molecules are equal. When the molecules are close enough, however, the two groups of electrons interact and develop synchronized orbiting.

One can then imagine that at a given instant a molecule has a momentary separation of charge, giving it an electric moment. The same is true of the other molecule. For that instant the two molecules attract each other. The attraction may fluctuate as time passes, but over time the average of the attraction is not zero; indeed, it is strong enough to hold the molecules together. Paraffin oil is held together by van der Waals forces.

Armed with this handful of models for intermolecular forces, Wissing was able to explain his experiments and to propose new ones. When he put a drop of paraffin oil on water, there was essentially no bonding between the hydrocarbons in the molecule and the water molecules. Hydrogen bonding is impossible between the two types of molecules. Other types of bonding are too uncommon to compete against the hydrogen bonding maintaining the cohesion of the water or the van der Waals force maintaining the cohesion of the oil drop. Hence the oil drop is merely pulled downward into the water surface slightly by gravity. When the drop is placed on the other solutions, the same thing happens because there is still little chance of bonding at the interface.



Some key molecular structures

The story behind the behavior of the drops of oleic acid now becomes clearer. When a drop of it is deposited on pure water, bonding immediately begins at the interface. The hydrophilic end of the oleic acid molecule is attracted to the water, but the opposite end and much of the length of the molecule are not. The hydrophilic end is commonly referred to as the head of the molecule and the opposite end is called the tail. The molecules of oleic acid at the interface rotate so that the heads line up toward the underlying water surface, leaving the tails upward.

Two types of bonding develop between the head and the water. Since the head and a water molecule are both polar, their electric moments end up in a configuration that leads to attraction. (More properly the attraction should be described as hydrogen bonding. A water molecule presents one of its hydrogen atoms to the exposed oxygen atom in the head in order to form the bond.)

The other kind of bonding involves the dissociation of the head group, which loses a hydrogen ion (a proton) to a water molecule. The loss leaves the head negatively charged (COO−) and the newly formed water ion positively charged (H_3O+). The two attract each other with ionic bonds because of their opposite electric charges.

Further bonding then begins along the interface. When the oleic acid molecules become oriented with their heads toward the water, they bond to each other in an orderly way. Adjacent heads form a hydrogen bond: the OH in one head aligns with the oxygen in a neighboring head. More bonds form between the hydrocarbons in the chain that makes up the rest of the molecule. At some sites the adjacent lengths of hydrocarbon attract each other through the van der Waals force.

The oleic acid at the edge of the drop bonds to the water as the acid molecules rotate into formation. Through the van der Waals force their motion pulls fresh oleic acid molecules onto the water, where they bond and then pull still more molecules from the bulk of the drop. The drop thins as more of its liquid moves to the edge and into molecular alignment. Soon it becomes an invisible sheet one molecule thick. The sheet cannot be stretched beyond this stage because of the bonding between its adjacent, aligned oleic acid molecules.

When a small amount of oleic acid is deposited on water, only a portion of the water surface is covered by the monolayer. With excess oleic acid the monolayer forms and the excess is left in a lenslike drop (or in several drops) on the surface. When the surface of the water has become covered with oleic acid molecules, no more of them form bonds with the water. They are left with only their van der Waals forces and the other electrical interactions that make them cohere. These surplus molecules bond themselves into the leftover drops.

When a drop of oleic acid is put on dilute hydrochloric acid, the process is much the same except for one thing. This time the abundance of hydrogen and chlorine ions in the bulk liquid precludes any systematic dissociation at the head of the oleic acid molecules. The molecules still rotate into alignment with heads down and tails up, but now their bonds with the water are almost all hydrogen bonds. Again the drop of oleic acid spreads over the surface.

It is a different story when a drop of oleic acid is placed on a dilute solution of sodium hydroxide. The solution neutralizes the acid, yielding sodium oleate. At the concentration in Wissing's experiment the oleate molecules form into tiny drops called micelles. The molecules at the interface between a micelle and the surrounding water have their polar heads outward and their hydrocarbon tails inward. The heads bond with the surrounding water through hydrogen bonds and sometimes also through an electrostatic attraction created when a head group loses a hydrogen ion to a water molecule.

The formation of micelles can be seen clearly in Wissing's experiment. Soon after a drop of oleic acid falls onto the surface of the sodium hydroxide the drop sends out surface streams that gradually disappear. At the same time micelles form just below the streams and the drop. The bulk solution gets turbid as the micelles obstruct the passage of light through the solution.

Ammonium hydroxide also neutralizes oleic acid, but no micelles form. As the oleic acid is neutralized it is vigorously expelled from the drop in streams. The drop and the streams move rapidly over the surface. I do not understand why the acid is expelled.

The next set of experiments done by Wissing fascinates me. He began with clean water to which three or four drops of oleic acid were added. The monolayer again formed and the excess oleic acid was left in a large drop. In the photograph at the top left on page 165 the drop rests on the water, which is still shimmering after the sudden formation of the monolayer.

Away from the drop of oleic acid Wissing deposited five drops of paraffin oil. When the first one was added, the drop of oleic acid developed a broad extension. After the fifth drop of paraffin oil was put into the dish the oleic acid drop sent a stream over to the nearest paraffin drop and began to consume it. For several minutes the oleic acid slid over to the drop of oil, climbing it and eventually dissolving into it, turning the oil pink. The oleic acid drop was then fatter and more spread out.

The photograph at the middle left on page 165 shows a second drop of paraffin oil just before it was consumed by the oleic acid drop. Traces of the first drop (already consumed) can still be seen on one side of the oleic acid drop. The photograph at the bottom left was made after a second and third drop had been eaten by the oleic acid drop. The rest of the experiment continues in the photographs running from top to bottom at the right.

Neither Wissing nor I can fully explain this demonstration. He suggests that the presence of the paraffin oil offers a means whereby the hydrocarbon groups of the excess molecules of oleic acid can bond with the hydrocarbon groups in the oil. My guess is that initially the molecules in the oil drop tug on the tails of the oleic acid molecules in the surrounding monolayer. Since the molecules of the monolayer are held together, the tug is transmitted to the oleic acid drop, with the result that part of the drop is pulled toward the oil drop.

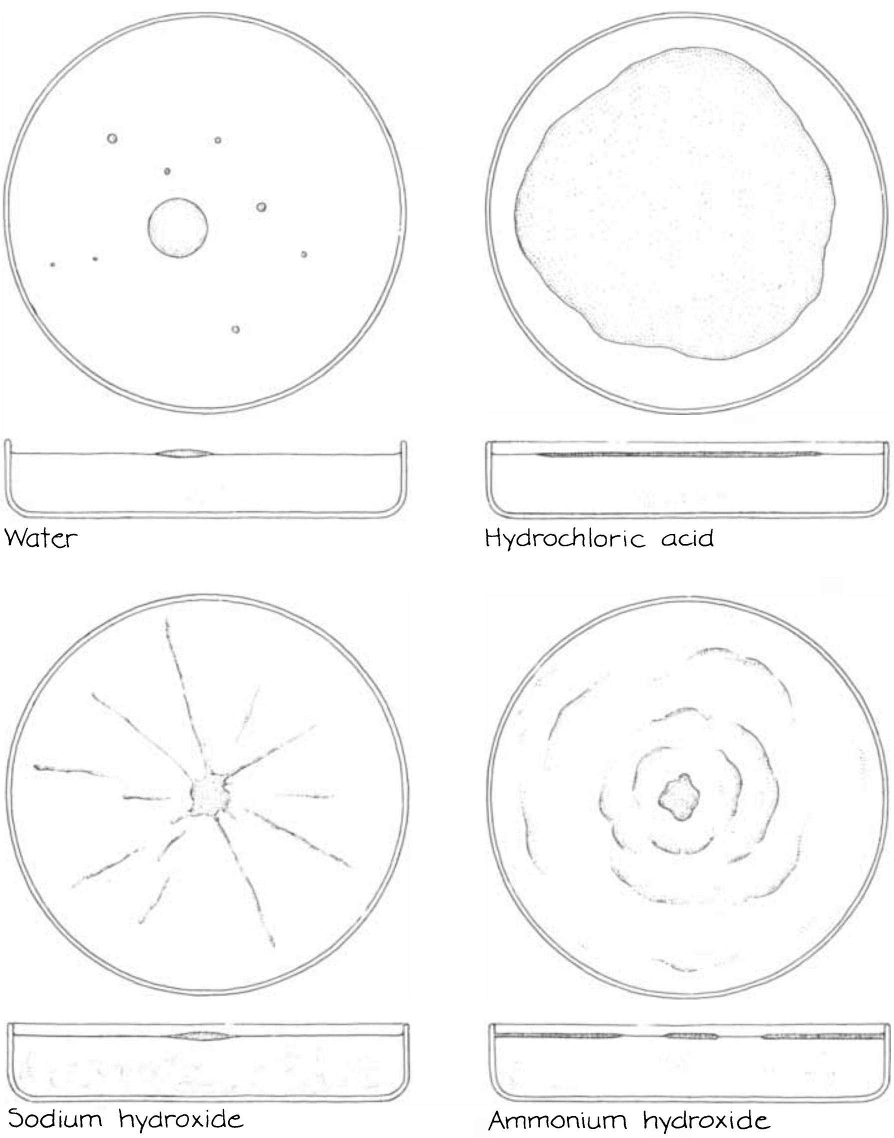
I am greatly puzzled by this experiment. Why does one drop of oil cause the drop of oleic acid to extend whereas five oil drops make the oleic acid start consuming the paraffin? I think the reason has to do with the horizontal force on the monolayer from the drops of paraffin. Suppose that before the first drop of paraffin is put in place the dish is covered with only a monolayer and the drop of excess oleic acid. Because the liquid in the drop of acid tends to slide off to the sides, it exerts a horizontal force on the monolayer.

When the first oil drop is added, it creates more force on the monolayer. Four additional drops of oil create even more force. By now an oil drop has begun to bond with the nearby oleic acid molecules surrounding it.

From which direction will a fresh batch of oleic acid come? Preferentially from the direction of the oleic acid drop because of the force with which it pushes on the monolayer. The drop of oleic acid droops onto the monolayer. This extension is then pulled by the monolayer (because of the bonding at the oil drop) and pushed by the hydrostatic pressure of the fluid in the rest of the oleic acid drop. Eventually the extension reaches and then devours the drop of oil.

Wissing sometimes changes the experiment while the oil drops are being gobbled up by the oleic acid. He pours one or two milliliters of concentrated ammonium hydroxide near one side of the oleic acid drop. The drop begins to dance wildly as the new compound neutralizes it. Micelles begin to form. The scene on the classroom screen is vivid with red drops and streams in crazy motion until the abundance of micelles blocks the passage of light through the petri dish.

Wissing's final experiments are just as entertaining. In one of them he deposits enough oleic acid on water so that a



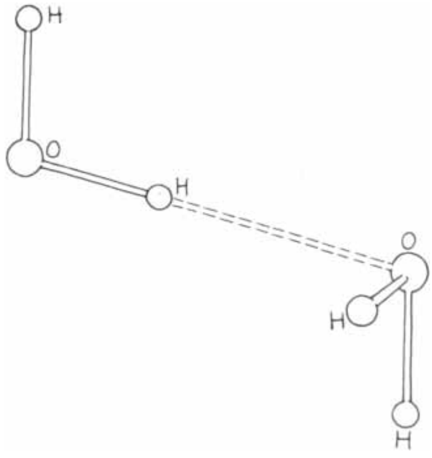
The action of oleic acid drops on four solutions drop of the acid remains. He puts on the middle of it a drop of Triton X-100, a commercial detergent. When the detergent has had a chance to penetrate the acid drop and reach the water, the surface of the solution is suddenly cleared of the oleic acid.

In another experiment Wissing deposits a tiny drop of oleic acid on water and adds enough paraffin oil to make an oil drop about three centimeters in diameter. Then additional oleic acid placed near the paraffin oil encircles it, forming a complicated pattern of drops that continue to move for a long time. In a separate demonstration Wissing shows how a drop of oleic acid on water can be maneuvered when a cotton swab wet with concentrated ammonium hydroxide is brought nearby. With a sudden jab of the swab he makes the drop of oleic acid jerk across the water surface. When he corrals it near the side of the dish, the drop oscillates in the vapor given off by the swab.

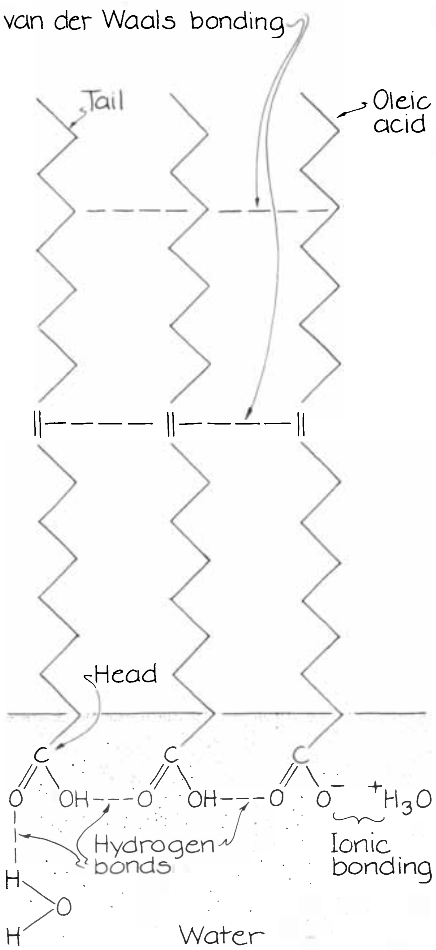
The formation of a monolayer by partially soluble organic compounds played a major role in the early work on determining the size of molecules. In 1890 Lord Rayleigh employed a minute amount of olive oil to estimate the size of its molecules. A fine platinum wire was dipped in the oil and then weighed on a sensitive balance. Some of the oil was released onto a circular water surface 84 centimeters in diameter and the wire was reweighed. The difference in the two measurements was the weight of the oil then covering the water surface.

To monitor the extent of spreading by the monolayer Rayleigh sprinkled the water with fine grains of camphor. As camphor dissolves in water on one side of a grain, it decreases the surface tension of the water. The larger surface tension on the other side jerks the grain. When Rayleigh added olive oil, the surface tension was too low for the camphor dance. He could therefore monitor the extent of the oil layer by watching the camphor grains. Adding just enough oil to stop all the grains from dancing, he then had the water surface covered evenly with a monolayer. From the surface area and the weight of the drop he computed the thickness of the layer as being about 17 angstrom units. Rayleigh thought this number was also the length of the molecules forming the monolayer.

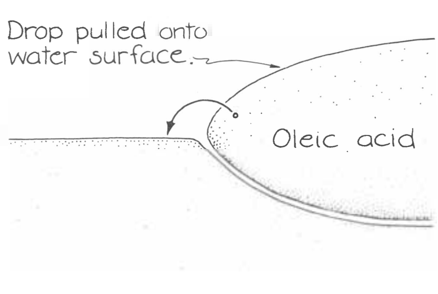
Richard E. Crandall and Jean F. Delord of Reed College have developed a modified form of Rayleigh's experiment for their students in introductory physics. In the exercise a student calculates the length of the oleic acid molecule, the average bond length along the molecule and the mass of the carbon atom. The first step is to sprinkle a water surface with lycopodium powder. When a drop of oleic acid is added, the monolayer pushes the powder outward into a circle



The hydrogen bond of water



A monolayer of oleic acid



How a monolayer is formed

and then contracts into an irregular perimeter. Before the circle contracts the student sketches its boundary on graph paper. Later the sketch is measured to ascertain the area of the circle.

The student is given several facts. The solution of oleic acid is prepared with 200 parts per million of acid in alcohol, measured in terms of volume. The size of a drop of acid is approximately .05 cubic centimeter. When the molecules of acid are aligned with heads down and tails up, a head occupies an area that can be approximated as a square with a side .1 the length of the molecule. The masses of O, C and H are in the ratio of 16 : 12 : 1. The density of oleic acid is about .895 gram per cubic centimeter.

By multiplying the concentration of the oleic acid and the volume of the drop the student calculates the volume of oleic acid that spreads over the water. Dividing the volume of the acid by the surface area of the monolayer yields the height of the layer. That height is also the length of the oleic acid molecule.

The next step is to estimate the bond length, that is, the distance between two adjacent carbon atoms along the length of the molecule. For simplicity you can consider the double bond at the center of the molecule as a single bond; the rest of the bonds along the length of the hydrocarbon chain are true single bonds. You also might ignore the nonlinear structure at the head of the molecule. In this way you are dealing with 17 bonds between carbon atoms along the molecule, or 18 if the hydrogen at the tail is included. Assume that the bonds lie along a straight line. When their number is divided into the length of the molecule, you find that the bond length is on the order of one angstrom.

Multiplying the volume of oleic acid by the density of the acid gives the mass of the monolayer. Find the mass of each molecule by means of the estimate that the head of the molecule is a square with a side .1 the length of the molecule. Since the molecule's length is known by now, the area of the square of a head is easy to compute.

The number of oleic acid molecules is computed by dividing the area of the monolayer by the area of each head. Dividing the mass of the layer by the number of molecules yields the mass of each molecule. Finally, the mass of a single carbon atom is determined by counting the number of each type of atom in the molecule and employing the known ratios of their masses.

Molecules are notoriously difficult to picture because of their smallness and complexity. Yet from Rayleigh's experiments one can estimate their size. From Crandall and Delord's laboratory exercise one can determine bond length. With Wissing's colorful demonstrations one can even imagine the ordering of the molecules along the interface between water and an organic compound and can make sense of why some organic compounds spread over water and others do not.

STATEMENT OF OWNERSHIP, MANAGEMENT AND CIRCULATION (required by 39 U.S.C. 3685). *1A.* Title of publication: Scientific American. *1B.* Publication number: 00368733. *2.* Date of filing: September 27, 1983. *3.* Frequency of issue: monthly. *3A.* Number of issues published annually: 12. *3B.* Annual subscription price: U.S. and its possessions, 1 year $24; all other countries, 1 year $33. *4.* Complete mailing address of known office of publication: 415 Madison Avenue, New York, N.Y. 10017. *5.* Complete mailing address of the headquarters or general business offices of the publisher: 415 Madison Avenue, New York, N.Y. 10017. *6.* Full names and complete mailing address of publisher, editor, and managing editor: Publisher, Gerard Piel, 415 Madison Avenue, New York, N.Y. 10017. Editor, Dennis Flanagan, 415 Madison Avenue, New York, N.Y. 10017. Managing Editor, none. *7.* Owner: Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017; Atwell & Co. (for Estate of John Hay Whitney), P.O. Box 456, Wall Street Station, New York, N.Y. 10005; Buckley & Co. (for Elinor G. Mannucci), c/o Fiduciary Trust Co. of New York, P.O. Box 3199, Church Street Station, New York, N.Y. 10008; George S. Conn, 20 Bittersweet Lane, Mount Kisco, N.Y. 10549; Cudd & Co. for (a) Helen Rosenwald Snellenburg Trust, (b) Robert L. Rosenwald Sr. Trust, (c) Trust U/W Lessing J. Rosenwald for benefit of Edith G. Rosenwald and children, (d) I. M. Scott, The Chase Manhattan Bank, 1 Chase Manhattan Plaza, New York, N.Y. 10015; Martin M. Davidson, 11 River Lane, Westport, Conn. 06880; Dengel & Co. (for Nathan W. Levin), c/o Fiduciary Trust Co. of New York, P.O. Box 3199, Church Street Station, New York, N.Y. 10015; Richard E. Deutsch, c/o The Chase Manhattan Bank, P.O. Box 1508, Church Street Station, New York, N.Y. 10008; Egger & Co. (for Bayard Ewing), c/o The Chase Manhattan Bank, P.O. Box 1508, Church Street Station, New York, N.Y. 10008; Dorothy D. Eweson, c/o Brunt & Co., P.O. Box 6626, Journal Square, Jersey City, N.J. 07306; Excelsior Fund, 545 Madison Avenue, New York, N.Y. 10022; Dennis Flanagan, c/o Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017; William T. Golden, Room 4201, 40 Wall Street, New York, N.Y. 10005; C. John Kirby, c/o Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017; Frasier W. McCann, Suite 1706, 680 Fifth Avenue, New York, N.Y. 10019; Donald H. Miller, Jr., c/o Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017; Eleanor Jackson Piel, Jr., 320 Central Park West, New York, N.Y. 10025; Gerard Piel, c/o Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017; Jonathan Piel, c/o Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017; The President & Fellows of Harvard College Trust, 70 Federal Street, Boston, Mass. 02110; Barbara K. Rosenwald, P.O. Box 3625, Norfolk, Va. 23514; Julius Rosenwald II, c/o The Chase Manhattan Bank, P.O. Box 1508, Church Street Station, New York, N.Y. 10008; Rowe & Co. (for Rhode Island Charities Trust), c/o Industrial National Bank of Rhode Island, Trust Department, 100 Westminster Street, Providence, R.I. 02903; Stanley Schaefer, 418 Yerba Santa Avenue, Los Altos, Calif. 94022; Joan R. Scott, c/o The Chase Manhattan Bank, P.O. Box 1508, Church Street Station, New York, N.Y. 10008. *8.* Known bondholders, mortgagees, and other security holders owning or holding 1 percent or more of total amount of bonds, mortgages or other securities: none. *9.* For completion by nonprofit organizations authorized to mail at special rates (Section 423.12, DMM only). The purpose, function, and nonprofit status of this organization and the exempt status for Federal income tax purposes: Not applicable to this organization. *10.* Extent and nature of circulation: *A.* Total number of copies (net press run): average number of copies each issue during preceding 12 months, 797,619; actual number of copies of single issue published nearest to filing date, 841,615. *B.* Paid circulation: *1.* Sales through dealers and carriers, street vendors and counter sales: average number of copies each issue during preceding 12 months, 149,573; actual number of copies of single issue published nearest to filing date, 201,626. *2.* Mail subscription: average number of copies each issue during preceding 12 months, 524,073; actual number of copies of single issue published nearest to filing date, 503,800. *C.* Total paid circulation (sum of *10B1* and *10B2*): average number of copies each issue during preceding 12 months, 673,646; actual number of copies of single issue published nearest to filing date, 705,426. *D.* Free distribution by mail, carrier or other means, samples, complimentary, and other free copies: average number of copies each issue during preceding 12 months, 6,784; actual number of copies of single issue published nearest to filing date, 7,128. *E.* Total distribution (sum of *C* and *D*): average number of copies each issue during preceding 12 months, 680,430; actual number of copies of single issue published nearest to filing date, 712,554. *F.* Copies not distributed: *1.* Office use, leftover, unaccounted, spoiled after printing: average number of copies each issue during preceding 12 months, 6,614; actual number of copies of single issue published nearest to filing date, 6,394. *2.* Return from news agents: average number of copies each issue during preceding 12 months, 110,575; actual number of copies of single issue published nearest to filing date, 122,667. *G.* Total (sum of *E*, *F1* and *2*—should equal net press run shown in *A*): average number of copies each issue during preceding 12 months, 797,619; actual number of copies of single issue published nearest to filing date, 841,615. *11.* I certify that the statements made by me above are correct and complete. (Signed) George S. Conn, Vice-President and Gen. Manager.

Test Drive the Leica R4.

We invite you to experience the perfection of the Leica® R4. We invite you to hold, inspect, and then test the Leica that has been acclaimed "...a miracle of modern technology." A limited production SLR of astonishing versatility, precision and beauty. With flawless Leitz® optics, two exposure measurement systems, five operation modes, and a series of advances unequalled in the history of photography.

The Leica R4 is Leica's proudest achievement. A photographic instrument designed for perfection, not price. Like every Leica since the beginning, it is lavished with attention to detail. Hand finishing. Hand assembly. Hand testing. No shortcuts or compromises. Because the Leica R4 is designed and built never to fail. And to have no limit to its useful life.

We invite you to experience the Leica R4. And through it enter Leica's world of superior photography. To see and appreciate for yourself the perfection of the photographic image that is uniquely Leica's.

You can test drive the Leica R4 now, at your local Leica dealer. And to show his appreciation for your visit and your interest, your Leica dealer will enter your name in a special drawing for the ultimate sports coupe—the BMW 633CSi.* We think you'll find this high performance German import, the perfect accessory to your new Leica.

For the name of your nearest Leica dealer write E. Leitz, Inc., 24 Link Drive, Rockleigh, N.J. 07647 or call (800) 223-0507, ext. 801. In New Jersey call (800) 223-0509, ext. 801.

And you could win a BMW 633 CSi to drive it home in.

*No purchase necessary. Entries must be made through your Leica dealer before January 31, 1984. The winner will be notified by March 30, 1984. Void where prohibited by law.

ANNUAL INDEX

The following index lists all the authors and articles that appeared in SCIENTIFIC AMERICAN *during 1983. Also indexed are "Metamagical Themas," "Mathematical Games," "Computer Recreations" and "The Amateur Scientist."*

AUTHORS

Fast and restless.

The new Delta-15 printer kicks out AnyCalc, easy as 1, 2, 3!

In a 9 to 5 world full of changing spreadsheet data, you need a business printer that moves fast. A printer that constantly fires out printed information. A printer called Delta-15.

Delta has the ability to print multi-copy spreadsheets at an intense 160 cps. Its throughput never rests. In fact , it never even wavers in speed. That means that Delta constantly fits the most work into every single second.

Delta comes in a 10″ or 15½″ carriage size. It offers you the flexibility of standard 8K parallel and serial interface. And has the ability to underline, accept macro instruction, and print characters that range from full graphics to everyday printing to scientific notations.

Plus, as always, you get our unique 180 day warranty (90 days on print head).

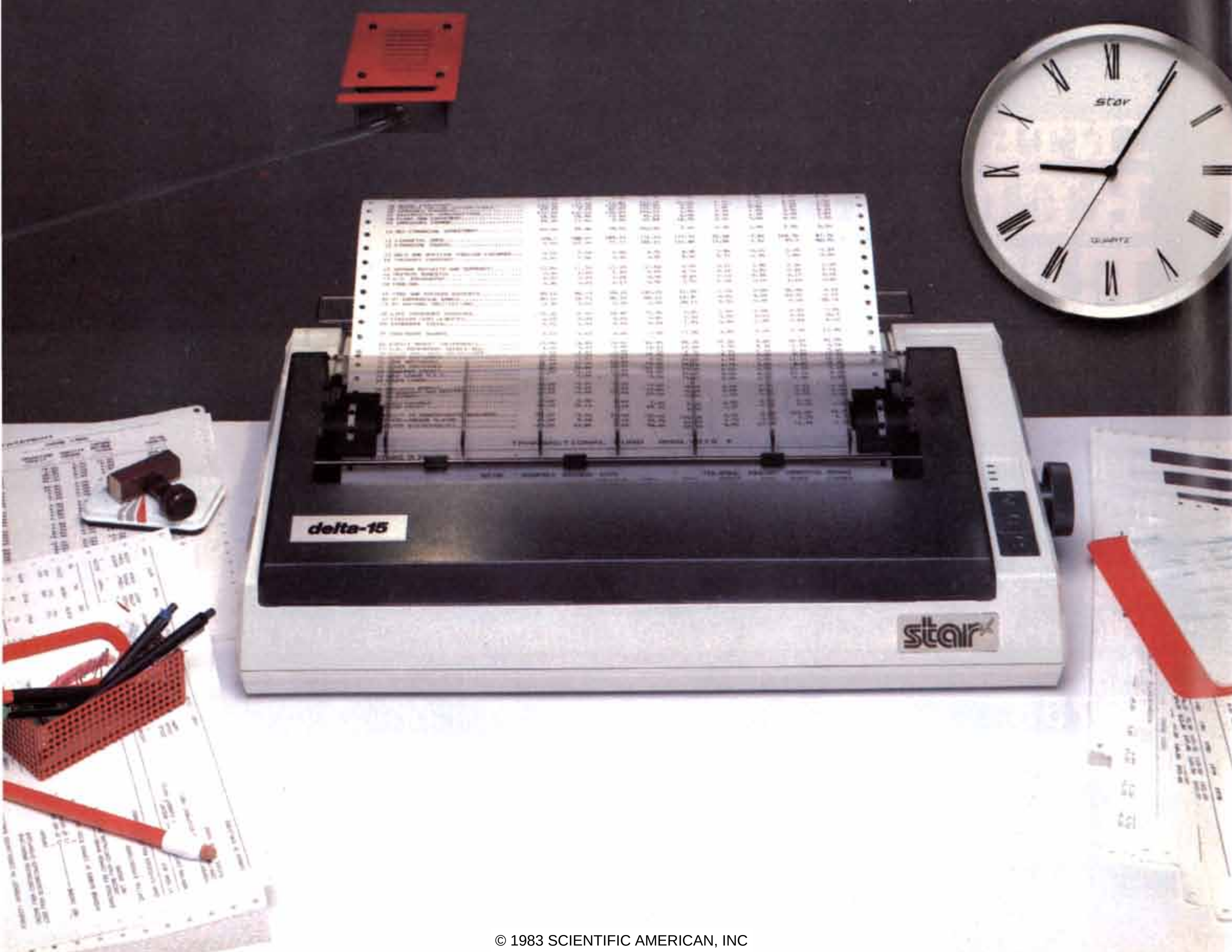
So for everyone who needs their spreadsheet data "yesterday," this is as close as you can come! The new fast and restless Delta-15 business printer from Star.

star™ micronics • inc

THE POWER BEHIND THE PRINTED WORD.

Computer Peripherals Division

P.O. Box 612186, Dallas/Ft. Worth Airport, TX 75261 (214) 456-0052



ARTICLES

METAMAGICAL THEMAS

MATHEMATICAL GAMES

COMPUTER RECREATIONS

THE AMATEUR SCIENTIST

BIBLIOGRAPHY

Readers interested in further explanation of the subjects covered by the articles in this issue may find the following lists of publications helpful.

COMPUTER RECREATIONS

COMPUTATION: FINITE AND INFINITE MACHINES. Marvin L. Minsky. Prentice-Hall, Inc., 1967.

MATHEMATICAL STRUCTURES FOR COMPUTER SCIENCE. Judith L. Gersting. W. H. Freeman and Company, 1967.

ALGORITHMS + DATA STRUCTURES = PROGRAMS. Niklaus Wirth. Prentice-Hall, Inc., 1976.

INTRODUCTION TO AUTOMATA THEORY, LANGUAGES, AND COMPUTATION. John E. Hopcroft and Jeffrey D. Ullman. Addison-Wesley Publishing Company, 1979.

MODERN ICEBREAKERS

THE NORTHEAST PASSAGE AS A COMMERCIAL WATERWAY. Terence Armstrong in *Vega 1880–Ymer 1980: Årsbok 1980 Svenska Sällskapet för Antropologi och Geografi,* pages 86–129; 1980.

SHIPPING ROUTES, ICE COVER AND YEAR-ROUND NAVIGATION IN THE CANADIAN ARCTIC. Balaram B. Dey in *Polar Record,* Vol. 20, No. 129, pages 549–559; September, 1981.

THE SOVIET NORTH IN THE ELEVENTH FIVE-YEAR PLAN, 1981–85. Terence Armstrong in *Polar Record,* Vol. 21, No. 130, pages 66–69; January, 1982.

IO

MELTING OF IO BY TIDAL DISSIPATION. S. J. Peale, P. Cassen and R. T. Reynolds in *Science,* Vol. 203, No. 4383, pages 892–894; March 2, 1979.

THE ROLE OF SO_2 IN VOLCANISM ON IO. B. A. Smith, E. M. Shoemaker, S. W. Kieffer and A. F. Cook II in *Nature,* Vol. 280, No. 5725, pages 738–743; August 30, 1979.

THE GALILEAN SATELLITES AND JUPITER: VOYAGER 2 IMAGING SCIENCE RESULTS. Bradford A. Smith, Laurence A. Soderblom et al. in *Science,* Vol. 206, No. 4421, pages 927–950; November 23, 1979.

THE GALILEAN MOONS OF JUPITER. Laurence A. Soderblom in *Scientific American,* Vol. 242, No. 1, pages 88–100; January, 1980.

HEAT FLOW FROM IO (JI). Dennis L. Matson, Gary A. Ransford and Torrence V. Johnson in *Journal of Geophysical Research,* Vol. 86, No. B3, pages 1664–1672; March 10, 1981.

SATELLITES OF JUPITER. Edited by D. Morrison. University of Arizona Press, 1982.

TWO CLASSES OF VOLCANIC PLUMES ON IO. Alfred S. McEwen and Laurence A. Soderblom in *Icarus,* Vol. 55, No. 2, pages 191–217; August, 1983.

AN EARLY IRON AGE FARM COMMUNITY IN CENTRAL EUROPE

EUROPEAN PREHISTORY. Sarunas Milisauskas. Academic Press, 1978.

FROM BRONZE AGE TO IRON AGE: ECONOMIC, HISTORICAL AND SOCIAL CONSIDERATIONS. R. Peroni in *Italy before the Romans: The Iron Age, Orientalizing, and Etruscan Periods,* edited by David and Francesca R. Ridgway. Academic Press, 1979.

RURAL ECONOMY IN THE EARLY IRON AGE: EXCAVATIONS AT HASCHERKELLER 1978–1981. Peter S. Wells. Peabody Museum of Harvard University, 1983.

THE DNA HELIX AND HOW IT IS READ

DNA: RIGHT-HANDED AND LEFT-HANDED HELICAL CONFORMATIONS. Alexander Rich, Gary J. Quigley and Andrew H.-J. Wang in *Biomolecular Stereodynamics: Vol. 1,* edited by R. H. Sarma. Adenine Press, 1981.

THE ANATOMY OF A-, B-, AND Z-DNA. Richard E. Dickerson, Horace R. Drew, Benjamin N. Conner, Richard M. Wing, Albert V. Fratini and Mary L. Kopka in *Science,* Vol. 216, No. 4545, pages 475–485; April 30, 1982.

MECHANICS OF SEQUENCE-DEPENDENT STACKING OF BASES IN *B*-DNA. C. R. Calladine in *Journal of Molecular Biology,* Vol. 161, No. 2, pages 343–352; October 25, 1982.

ORDERED WATER STRUCTURE AROUND A *B*-DNA DODECAMER: A QUANTITATIVE STUDY. Mary L. Kopka, Albert V. Fratini, Horace R. Drew and Richard E. Dickerson in *Journal of Molecular Biology,* Vol. 163, No. 1, pages 129–146; January 5, 1983.

SEQUENCE-DEPENDENT CONFORMATION OF AN *A*-DNA DOUBLE HELIX. Z. Shakked, D. Rabinovich, O. Kennard, W. B. T. Cruse, S. A. Salisbury and M. A. Viswamitra in *Journal of Molecular Biology,* Vol. 166, No. 2, pages 183–201; May 15, 1983.

BASE SEQUENCE AND HELIX STRUCTURE VARIATION IN *B* AND *A* DNA. Richard E. Dickerson in *Journal of Molecular Biology,* Vol. 166, No. 3, pages 419–441; May 25, 1983.

QUANTUM GRAVITY

GENERAL RELATIVITY: AN EINSTEIN CENTENARY SURVEY. Edited by S. W. Hawking and W. Israel. Cambridge University Press, 1979.

QUANTUM GRAVITY II. Edited by C. J. Isham, R. Penrose and D. W. Sciama. Oxford University Press, 1981.

PREY SWITCHING IN A SIMPLE ECOSYSTEM

THE DISTRIBUTION AND ABUNDANCE OF ANIMALS. H. G. Andrewartha and L. C. Birch. University of Chicago Press, 1954.

ECOLOGICAL GENETICS. E. B. Ford. John Wiley & Sons, Inc., 1965.

OF PREDATION AND LIFE. Paul L. Errington. Iowa State University Press, 1967.

THE AERODYNAMICS OF HUMAN-POWERED LAND VEHICLES

MAN-POWERED VEHICLE DESIGN CRITERIA. Chester R. Kyle and Walter Edelman in *Proceedings of the Third International Conference on Vehicle System Dynamics held at Virginia Polytechnic Institute and State University, August 12–15, 1974, Blacksburg, Virginia,* edited by H. K. Sachs. Swets & Zeitlinger, Amsterdam, 1975.

PROCEEDINGS OF THE FIRST HUMAN POWERED VEHICLE SCIENTIFIC SYMPOSIUM. Edited by Alan Abbott. IHPVA, P.O. Box 2068, Seal Beach, Calif. 90740, 1981.

BICYCLES & TRICYCLES: AN ELEMENTARY TREATISE ON THEIR DESIGN AND CONSTRUCTION. Archibald Sharp. Reprint of 1896 edition by the MIT Press, 1982.

BICYCLING SCIENCE. Frank R. Whitt and David G. Wilson. The MIT Press, 1982.

THE INTERPRETATION OF VISUAL ILLUSIONS

DIFFERENTIAL GEOMETRY OF CURVES AND SURFACES. Manfredo P. do Carmo. Prentice-Hall, Inc., 1976.

VISION: A COMPUTATIONAL INVESTIGATION INTO THE HUMAN REPRESENTATION AND PROCESSING OF VISUAL INFORMATION. David Marr. W. H. Freeman and Company, 1982.

THE MODULARITY OF MIND. Jerry Fodor. The MIT Press, 1983.

THE AMATEUR SCIENTIST

THE HYDROPHOBIC EFFECT: FORMATION OF MICELLES AND BIOLOGICAL MEMBRANES. Charles Tanford. John Wiley & Sons, Inc., 1980.

INTERATOMIC AND INTERMOLECULAR FORCES. John P. McTague in *Encyclopedia of Physics,* edited by Rita G. Lerner and George L. Trigg. Addison-Wesley Publishing Company, Inc., 1981.


WINE & FOOD
Taste of the Holidays
During the holidays, the kitchen is the heart of our Sonoma Valley home. Here, the marriage of wine and food fills the air with wonderful aromas and memories of vintages past.
WINE
It's a time for California sparkling wine, and we express this festive tradition with Sebastiani Brut Three-Star, a premium sparkling wine made entirely from Pinot Noir grapes grown in Sonoma County. Though it requires extra-time and special care, we produce Brut Three-Star by the classic methode Champenoise, a method that dates back to seventeenth century France. In each individual bottle, bubbles were formed naturally during a secondary fermentation, and fifteen months of aging on the yeast has developed a depth and complexity that make Brut Three-Star a sparkling wine of extra-special character.
FOOD
Once reserved for special occasions only, sparkling wine is now enjoyed in combination with many fine foods.
We enjoy Brut Three-Star with different hors d'oeuvres and also with various shellfish, particularly fresh dungeness crab. In a classic match between sparkling wine and food, my wife Vicki has created a Caviar Pastiera that marries the rich flavors of caviar, egg, cheese and onion with the crisp, toasty flavors of Brut Three-Star.
If you would like to sample some of our family holiday recipes and learn more about our wine, please write to the address below for your free brochure, "Wine and Food: Taste of the Holidays."
From my family to your family . . . Buon Natale!
Sam J. Sebastiani
Sebastiani Vineyards, Dept. G, P.O. Box AA, Sonoma, CA 95476
© 1982 SEBASTIANI VINEYARDS
Sebastiani VINEYARDS
ESTABLISHED 1825
SONOMA BRUT
Sebastiani
SONOMA COUNTY SPARKLING WINE
BRUT
METHODE CHAMPENOISE
PRODUCED AND BOTTLED BY
SEBASTIANI CELLARS, BONDED WINERY 4831, GRATON, CALIFORNIA
ALC. 12.3% BY VOL. 750 ml (25.4 fl. oz.)

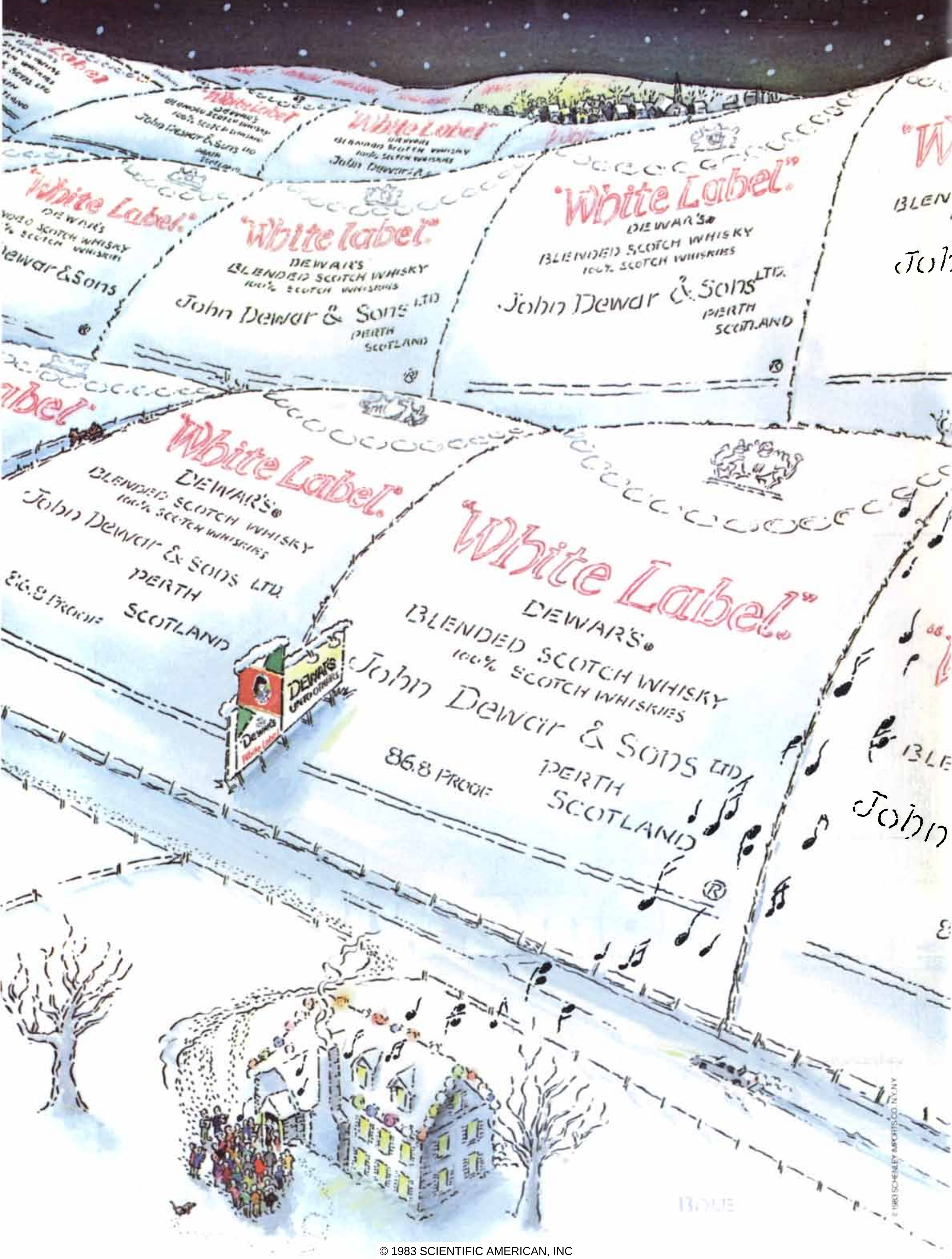
We're giving people who've been trained to build in quality on the line a chance to judge the finished product on the road.

And just as important as the feedback we get from the employees is the pride they gain from being involved. Because building pride is an important part of building better cars.

We believe in taking the extra time, giving the extra effort and paying attention to every detail. That's what it takes to provide the quality that leads more people to buy GM cars and trucks than any other kind. And why GM owners are the most loyal on the road.

That's the GM commitment to excellence.

Nobody sweats the details like GM.


"White Label."
DEWAR'S®
BLENDED SCOTCH WHISKY
100% SCOTCH WHISKIES
John Dewar & Sons LTD
86.8 PROOF
PERTH
SCOTLAND
DEWAR'S
© 1983 SCHENLEY IMPORTS CO. N.Y., N.Y.