5 May 1961, Volume 133, Number 3462 SCIENCE

CURRENT PROBLEMS IN RESEARCH

Information Theory

A brief survey is made of the recent developments and the current status of information theory.

H. H. Goldstine

It is with considerable diffidence that I attempt to discuss more or less authoritatively a subject that has been conceived and perfected by communication engineers and that is both an interesting and a novel application of mathematics. In the dozen years since its inception by Shannon, information theory has reached a state of perfection and completeness in which it is likely to remain for some time. At least this is the impression I have received from talking with some men who are specialists in this and related areas. This statement must, of course, be accepted with suitable caveats, since it sounds much like similar statements, made in the past by others, that later required extensive modification. But it does seem to me that an equilibrium has more or less been reached in the fundamentals of the subject, so that the mathematician can reasonably undertake a survey of the situation.

Entropy

Before discussing the newest developments, let me review briefly the elements of information theory (*1*). This subject was initiated in a tentative way by various authors (*2*) in the 1920's and ultimately developed as a comprehensive subject in 1948 by C. E. Shannon (*3*) in a now classic paper in the *Bell System Technical Journal*. Since then there has been considerable activity in the field. But before discussing the more recent activity, let me sketch some of Shannon's work. He started with the very plausible assumption that it is desirable to have a means of measuring the amount of information present in a given situation. He was able to construct in most cases of importance such a measure. Consider, for the sake of simplicity, an event chosen from a finite number of possible events. If all these events are a priori equiprobable, then the number, n, of them can be used as a measure of the amount of information present. There are, however, certain cogent reasons why it is better not to use n but to use, instead, the logarithm of n. To understand this reason, suppose one has two systems, one of which has m and the other n possible, equiprobable events. If one then considers the sum of the two systems, the number of possible events is now mn. It would be desirable that the measure of information content be not multiplicative but additive. Thus, one desires

$$f(mn) = f(m) + f(n)$$

and also if $m > n$, then in some sense the system with m events should contain more information than the one with n events; that is,

$$f(m) > f(n)$$

provided that $m > n$. It is an easy technical task to show that under these conditions

$$f(n) = C \ln n$$

for some positive constant C. Finally, it is customary to take the smallest nonzero amount of information as the unit of the system. But this smallest quantum is, of course, the binary digit, and thus

$$f(n) = \log_2 n$$

In the case that not all events are equally probable, it is still possible to define a measure for the information content for the system. To do this, suppose that the events have probabilities $p_1, p_2, \ldots, p_n$; then the "entropy" or measure of information content is by definition

$$H = -\Sigma^{n}_{i=1} p_i \log_2 p_i$$

Note that this measure is given in binary digits. By way of illustration, note that a decimal digit represents $\log_2 10 = 3.3$ binary digits, and a letter from our alphabet, $\log_2 26 = 4.7$ binary digits. If the $p_1, p_2, \ldots, p_n$ are all equal to $1/n$, then our last definition reduces to $\log_2 n$, as it should. If Shannon's axioms are slightly modified one obtains the concepts of information introduced much earlier by Fisher (*4*), in 1925; and if one makes the axioms slightly weaker, one obtains both their concepts, as well as other similar ones. This fact was shown in 1951 by Schutzenberger (*5*).

The introduction of the notion of an entropy into information theory establishes a close connection between thermodynamical systems and information-theoretical ones. This, according to von Neumann, is due to the fact that Boltzmann's definition of entropy is in terms of the number of a priori equiprobable states compatible with the macroscopic description of the state—that is, it corresponds to the amount of microscopic information missing in the macroscopic description.

Coding Theorem

Shannon's theory is principally concerned with those problems of communications engineering that naturally

The author is affiliated with the IBM Research Center, Yorktown Heights, N.Y. This article is adapted from a lecture delivered 28 December 1960 before the AAAS meeting in New York.

arise in the transmission of information over channels, and in particular over noisy ones—that is, ones where it is not certain that the information received is identical to that transmitted. It is clear that one can give an intuitively reasonable notion of rate of transmission of information and also of the capacity of a channel. To do these things, consider an arbitrary time T and let $N(T)$ be the number of distinct messages of duration T; then $[\log_2 N(T)]/T$ can be used to measure the *rate of transmission of information,* and the limit of this ratio, as T approaches infinity (it exists under a wide class of conditions), measures the capacity of the channel. Thus, in some sense the channel capacity is the maximum rate at which it is possible to communicate information over the channel.

One of Shannon's most remarkable results is the so-called coding theorem for noisy channels. This theorem, roughly speaking, guarantees that the channel capacity as just defined is precisely the maximum rate at which it is possible to receive information with arbitrarily high reliability (Below I state this remarkable theorem precisely). This is, in my opinion, one of the most surprising and important results in the entire theory. A priori, one might imagine that as the noise in a channel increases, the rate at which one can transmit information reliably down the channel would necessarily decrease to zero.

To see why this theorem is so noteworthy, let us consider how one might go about ensuring reliable transmission in the presence of noise. The most obvious thing to do is to repeat the message enough times to guarantee reliable reception. It can be shown that the price paid by this method for a very high reliability is a vanishingly low rate of transmission. For example, one can repeat each digit an odd number of times and decide at the receiver, by a majority rule, what was transmitted. It is clear how the rate of transmission goes down, and it is not too difficult to estimate how the reliability goes up as the number of iterations increases. This and other naive attacks all seem to lead to the same predicament, and this is the reason why Shannon's result is both so beautiful and so unexpected. Unfortunately, now, a dozen years after the discovery of the theorem, we still do not know how to achieve in general the result guaranteed by the theorem. This is not due to a desire for a result of great mathematical generality or a desire for elegance. Even in the case of the simplest nontrivial channel it is not known in complete detail how to proceed.

A little later we will return to this problem, which will lead us into that branch of information theory known as coding theory. Before proceeding, it is perhaps desirable to mention that Elias devised an iterated scheme of coding of considerable theoretical interest which achieves as high a degree of reliability as one desires and yet does not result in an arbitrarily low rate of transmission. It does not, it is true, achieve a rate of transmission equal to the channel capacity, but it has two interesting properties: for a fixed rate, the reliability can be increased at the cost of coding delay; for a fixed delay, the reliability can be increased at the cost of transmission rate. In a quite real sense Elias's scheme achieves its goal by introducing redundancy, much as natural language does (*6*).

Probabilistic Logics

Before going into more detail on modern developments and current status in information theory, I should like to discuss a different but closely related problem originally formulated in 1952 by von Neumann in a series of lectures at California Institute of Technology (*7*) and later reconsidered by Shannon and E. F. Moore (*8*), in 1956. Here the problem is not how to transmit data down noisy channels in the presence of noise but rather how to carry out logical functions in the presence of noise. That is, von Neumann inquired whether a reliable automaton could be designed in which logical elements could be used which were not perfect in the sense that each had a certain probability of malfunctioning. For his building blocks he used majority organs, whereas Moore and Shannon used relays.

The result is most remarkable and very elegant: It is possible to organize components which are individually unreliable into an automaton capable of arbitrarily high reliability. This is, to me, a result like Shannon's coding theorem for noisy channels in that it is, a priori, not obvious that the opposite is not true. Perhaps it is worth saying a few words to indicate how von Neumann achieved his result. He considered a majority organ, which is a three-input, one-output device where the output is the majority of the inputs. Suppose that we have such a majority organ with probability e of being in error and with probabilities f_1, f_2, f_3 that the input lines carry wrong signals. Then clearly $e + f_1 + f_2 + f_3$ is an upper bound for the probability that the output line is in error. From this we might conclude that the entire program is hopeless. However, we can manifestly obtain a much better upper bound under reasonable assumptions—namely, (i) the probabilities of error are independent, and (ii) under normal conditions all lines should be in the same state. Then the probability of at least two inputs being in error is clearly $P = f_1f_2\,(1 - f_3) + f_1f_3\,(1 - f_2) + f_2f_3\,(1 - f_1) + f_1f_2f_3 = f_1f_2 + f_1f_3 + f_2f_3 - 2f_1f_2f_3$. Thus, the probability Q of the output being in error is $(1 - e)\,P + e\,(1 - P) = e + (1 - 2e)\,P$. Now, if all $f_i \leqq f$, then $e + 3f$ is always an upper bound for Q and $e + (1 - 2e)\,(3f^2 - 2f^3) \leqq e + 3f^2$ is a better one, since $e + 3f^2 < f$ is possible. (In fact, for $f \sim e + 3e^2$, the two sides of the inequality are roughly equal.) If the conditions outlined above can always be maintained, then errors can indeed be kept under control. To do this without introducing a number of components that is increasing exponentially with the "logical depth" of the function—that is, the number of elements such as relays, tubes, and transistors, arranged in cascade fashion—von Neumann resorted to a very elegant trick. This consisted of transmitting all messages over N lines in parallel. A fiducial level $\Delta < \frac{1}{2}$ is chosen beforehand so that if ΔN lines, or less, all are stimulated, it is decreed that this is the one type message, say a 0, and if $(1 - \Delta)\,N$ or more are stimulated, then it is the other type, say a 1. Any other number is considered a malfunction of the system. Space does not allow me to go into details beyond noting the main result, which is roughly this: the probability of a malfunction is approximately

$$(6.4/\sqrt{N})\;10^{-8.6N/10,000}$$

where N is the number of lines in a bundle.

The drawback both in von Neumann's and in Moore-Shannon's work lies not in the concept or in any lack of beauty and depth of the analysis; in fact, on the contrary, these papers are both brilliant. The difficulty is that to achieve an arbitrarily low probability of a malfunction one must increase the size of the assemblage beyond bounds.

Coding Theory

Before going on to my main task of reporting on the current status of information theory, I think it would be desirable to say a few words on coding theory. I alluded most briefly to this topic in discussing Shannon's coding theorem for noisy channels. Recall that this theorem guarantees the existence of a coding of the input information and a decoding of the output information from the channel with an arbitrarily low probability of error, provided that the entropy of the information per unit of time is less than the channel capacity; for the entropy per unit of time in excess of the capacity, it is impossible. Now the essential task of coding theory is to discover these codings and decodings. This is one of the most important present-day areas of interest in the subject. Perhaps it would be well to say a few words on this topic, which received its initial impetus in a now classic paper by Hamming (*9*) in the 1950 *Bell System Technical Journal*. Since then a great many others have been instrumental in carrying forward the development of the theory (see references). It is not out of place, however, to mention Slepian (*10*), Bose and Ray-Chaudhuri (*11*), and Elias (*6*) as four of the leaders in the field.

To discuss coding one must realize that to achieve any sort of reliability one must introduce redundancy. In general, the problem faced by the coder is how to introduce redundancy into the system as efficiently as possible. In the next few paragraphs I describe briefly and geometrically some facts related to coding theory.

Let us consider all possible sequences of n binary digits. There are 2^n of these, and they form the vertices of the unit cube in an n-dimensional space. A code then consists of a designation of some subset of these vertices, and the received message is a transformation of this set into another set of these vertices. That is, it is a mapping of vertices onto vertices. If there is no error in reception, then the transformation is the identity, whereas if one error is introduced, the transformation carries a vertex into one an edge away; if r errors occur, the transformed point is r edges away.

It is not difficult to see that the transformed points are most probably near the input point and least probably far away. This suggests the means of decoding: one can associate with each vertex a set of vertices which are in an obvious sense quite close to the given vertex; the decoding consists of systematically classifying all these "closest" neighbors as being identical with the given vertex. These decoding schemes are called "maximum likelihood detectors." Two very important notions in this subject are group and parity check codes. Slepian showed them to be equivalent. The set of all binary sequences of length n form a group under addition modulo 2—the addition is element by element. (For this group the zero element is of course the sequence of all zeros.) Then a group code is any subset of this group that forms a subgroup.

These codes are of especial interest and permit easy definition of the "closest neighbors" of a vertex. They are known to be equivalent to parity check codes. Such codes are defined by assigning some digits in a sequence to be for information and keeping the remaining ones as check digits.

Another development in coding theory of a quite interesting nature is that concerned with correcting bursts of errors. The basic assumption made in this analysis is that in data transmission the noise on the channel is not randomly distributed but occurs in bursts. The problem of course is to construct codes which make it safe to transmit information in the presence of such noise bursts.

New Directions

Having said all these more or less preliminary things, I now turn more explicity to the modern developments and attempt to indicate, in a brief way, the course the subject has had. Once again it is worth remarking that such a survey as this necessarily ignores some work and also gives emphasis to other developments in a highly subjective fashion; it is to be hoped that time will show that the choices made are wise ones.

From 1948 until perhaps 1954 or 1955, men working in information theory were primarily interested in understanding and consolidating the pioneering results of Shannon. In some sense this period was more or less at an end by 1954–1955, when Feinstein (*12*) first established Shannon's conjectured theorem on transmission down noisy channels, and when Slepian introduced the notion of group codes. Not unnaturally, one of the main directions of post-1955 information theory has been to consolidate the subject more and more closely into the fold of "conventional mathematics."

One of the important directions of research in information theory has arisen from the esthetic desires of the mathematician to reformulate results from science and engineering into "polite" mathematical language. This was done in a notable way with the coding theorem of Shannon by the late A. Khinchin (*13*), and the result was a most remarkable and unexpected limit theorem in probability. In the simplest case the theorem is concerned with two sets of sequences of binary digits—the emitted and the received symbols—and with a non-negative function of two binary digits. (This function enables one to construct various conditional probabilities that arise.) Then the formulation of the coding theorem of Wolfowitz (*14*) ensures the existence of a positive channel capacity C such that for every a between 0 and 1, (i) there exists a positive constant k such that for every n there is a code with probability of error less than a and length at least

$$2^{nC - k\sqrt{n}}$$

and (ii) there exists a positive constant k' such that for every n there does not exist a code with probability of error less than a and length in excess of

$$2^{nC + k'\sqrt{n}}$$

The notion of the length of a code is not quite a self-evident one, and it might be well to explain in detail its meaning, particularly since this helps to illuminate the significance of the theorem. To do this, we will consider sequences $u = (x_1, x_2, \ldots, x_n)$, each of n binary digits. Now the channel down which each digit is transmitted carries out a chance experiment on each of these digits in turn. The experiment is such that its outcome, when performed on x_i, is 1 with probability p_i and 0 with probability $1 - p_i$. The sequence $v(u) = [y_1(u), y_2(u), \ldots, y_n(u)]$ of outcomes is again a sequence of binary digits. A sequence v of n binary digits is a possible received sequence in case, for some transmitted sequence u, the probability of v being $v(u)$ is not zero.

We can now follow Wolfowitz and define a code of length N corresponding to a positive number a as a set $[(u_1, A_1), (u_2, A_2), \ldots, (u_N, A_N)]$ where each u_i $(i = 1, 2, \ldots, N)$ is a transmitted sequence, where each A_i

($i = 1, 2, \ldots, N$) is itself a collection of sequences of n binary digits, where the probability that v (u_i) is in A_i must not be less than $1 - a$ for $i = 1, 2, \ldots, N$, and finally where the $A_1, A_2, \ldots, A_N$ are disjoint. For such codes the i-th word in a message is taken to be u_i; the recipient of a message now decodes in this manner: if he receives any sequence in A_j, he interprets this to mean that u_j was transmitted. In case the sequence is not in the union of $A_1, A_2, \ldots, A_N$, the result is completely ambiguous. Thus, the probability that a word can be correctly received is not less than $1 - a$. The ratio $(\log_2 N)/n$ is called the rate of transmission.

The study of this theorem of Shannon's has led researchers in two general directions. One group has worked on simplifying the proof of the theorem and casting the result itself into a more general probabilistic setting. The proofs depend upon combinatorial theorems and are now quite elementary in nature. The other group has been concerned with generalizing the theorem by relaxing the requirements of the hypotheses. The original theorem was concerned with discrete channels without memory—that is, discrete channels where no prior transmission can effect a subsequent one. Subsequent researches have been in the direction of relaxing these restrictions to include channels with finite memory; channels with a continuum of received symbols instead of the 0, 1 of a binary channel; channels with the channel probability function not completely known to the receiver, to the sender, or to either of them; and so on. Work in this direction has been done by Feinstein (*12*), Wolfowitz (*14*), Blackwell, Breiman, and Thomasian (*15*), McMillan (*16*), and others. Apparently, part 2 of the theorem has been the subject of much interest since the surprise occasioned by the possibility of coding at a rate $(\log_2 N)/n$ less than C perhaps made some people assume without proof the impossibility of coding at a higher rate. The net result of these two lines of work has been to produce a much more elegant theorem than that originally formulated, set in a more abstract and general milieu, with a considerably simpler proof than it originally had.

The second broad line of research in information theory has been in the field of coding itself. Suppose the total number of words in the sender's—and the receiver's—dictionary is N. These words may be enumerated as $1, 2, \ldots, N$, and then if there were no noise on the channel, they could be coded for transmission down a binary channel into sequences of length $n = \log_2 N$. If this were done, then the rate of transmission would be 1, which is the optimal value. However, when there is noise on the channel, redundancy must be introduced—that is, n must be chosen to be greater than $\log_2 N$, and the rate of transmission drops. How large n must be depends of course upon the channel capacity. In general it is clear that n must be large enough so that all words in coded form are quite different, to enable the receiver to distinguish them unambiguously.

As I mentioned above, instead of attempting optimal codings, workers in the field have devoted their efforts to constructing codes which are guaranteed to be safe against various types of disturbances. For example, Hamming in his original paper exhibited a code that can detect two errors and correct one of them. Recently Bose and a student, Ray-Chaudhuri, using theory of finite fields, gave a constructive method for finding a t-error correcting code for messages of length $n = 2^m - 1$, using at most mt parity symbols.

Bursts of errors are a frequent source of difficulty, and codes to combat this type of noise burst have been devised by Abramson (*17*), Fire (*18*), Meggitt (*19*), Muller (*20*), and others. These codes correct errors when they occur in adjacent positions.

They are all examples of Slepian group codes. Other particularly interesting types of such codes are the so-called cyclic ones—that is, ones where cyclic permutations of code words are again code words. Such codes have been studied by W. W. Peterson (*21*), Melas (*22*), Meggitt (*19*), Elspas (*23*), and Abramson. Most codes that have been discussed here are in fact of this kind. They permit relatively easy encoding and decoding. In these cases one does not need large dictionaries but can encode or decode by analytical means.

There was a period in which interest in codes dropped off, but with the advent of ever-larger computers there has been a substantial revival of interest in the subject. An interesting study would be that of contrasting the economics of establishing additional transatlantic or continental communication lines with the economics of installing very large computers to act as coders and decoders for messages. (Of course, the need for very large and very fast computers arises from the fact that the coding and decoding involve problems either of very large tables or of very rapid analytical calculations.) As far as I know, no such study has as yet been made.

A third line of research is that of Elias (*24*), to which I have alluded above, in which he has shown that any attempt to subject to the methods of information theory the von Neumann or Shannon-Moore approach to computers and automata either is impossible or will require at the least a very deep and penetrating analysis. His result, while a negative one, is nonetheless a landmark in the field. It indicates clearly that no coding theorem-like result can be expected except perhaps at a quite profound and as yet unknown level of understanding and sophistication.

Finally, a fourth line of research has been that of applying information theory to other disciplines. In general, most of these efforts have consisted of applying the concept of entropy to different situations. One of the most notable of these applications has been that of a U.S.S.R. school under the leadership of one of the great Russian probabilists, Kolmogoroff. He has enthusiastically promoted the role of entropy in probability theory. The most striking results have been achieved not in the field of information theory but, remarkably enough, in ergodic theory. This subject grew out of studies of analytical dynamics by G. D. Birkhoff (*25*) and von Neumann (*26*) over 30 years ago. Von Neumann had brought this subject to some state of completeness but had left a number of famous unsolved problems. Kolmogoroff and two of his students, Sinai and Rohlin (*27*), within the last several years have solved many of these problems quite simply by introducing the notion of entropy into the subject and showing that it is still another invariant. Work on seeing if still other invariants exist continues. Halmos and Kakutani are two of the key figures in this field in the United States today.

Last year Linnik (*28*) established the central limit theorem of probability under Lindberg's condition, using notions of information theory. Also Gelfand (*29*), Yaglom, and other Russian mathematicians have been applying entropy definitions to ever more general classes of stochastic processes and obtaining limit theorems on sums of independent random variables. This work really is much closer to probability theory than to information theory.

The concepts and results of information theory, together with those of Fourier series theory, have now become the basic themes of most work in communication theory today.

The only other applications I shall mention are two attributable to Mandelbrot (*30*). One lies in the field of statistical thermodynamics; in it he exploited a suggestion of Szilard's to apply various statistical concepts to thermodynamics. The other lies in the field of linguistics and concerns, in particular, stochastic properties of language. Mandelbrot considered discourse as a sequence of letters and words with various Markoff-type chain relationships between them. This leads to some quite unusual theorems in probability that have received empirical verification.

There have also been applications of information theory to biology, psychology, science, and statistics, but space does not permit me to go into these. Fortunately, however, books are available on these subjects, by Attneave (*31*), Brillouin (*32*), Kullback (*33*), and Quastler (*34*).

References and Notes

1. F. L. Stumpers has compiled a very complete bibliography on information theory, which has been published in various years in the *IRE Transactions on Information Theory* ["A bibliography of information theory," *IRE Trans. on Inform. Theory* **PGIT-2,** 1 (1953); **IT-1,** 31, 40 (1955); **IT-3,** 150 (1957)]. Forthcoming books by W. W. Peterson and J. Wolfowitz will have quite up-to-date bibliographies on the entire subject.
L. Szilard, *Z. Phys.* **32,** 753 (1925).
3. C. E. Shannon, *Bell System Tech. J.* **27,** 379, 623 (1948).
4. R. A. Fisher, *Proc. Cambridge Phil. Soc.* **22,** 700 (1925); *Contributions to Mathematical Statistics* (Wiley, New York, 1950).
5. M. P. Schutzenberger, *Compt. rend.* **232,** 925 (1951).
6. P. Elias, *IRE Trans. on Inform. Theory* **PGIT-4,** 29 (1954).
7. J. von Neumann, "Probabilistic Logics and the Synthesis of Reliable Organisms from Unreliable Components," *Automata Studies, Annals of Mathematics, Study No. 32* (Princeton Univ. Press, Princeton, N.J., 1956).
8. C. E. Shannon and E. F. Moore, *J. Franklin Inst.* **262,** 191, 281 (1956).
9. R. W. Hamming, *Bell System Tech. J.* **29,** 147 (1950).
10. D. Slepian, *ibid.* **35,** 203 (1956).
11. R. C. Bose and D. K. Ray-Chaudhuri, *Information and Control* **3,** 68 (1960).
12. A. Feinstein, *Foundations of Information Theory* (McGraw-Hill, New York, 1958).
13. I. A. Khinchin, *Uspekhi Math. Nauk* **8,** No. 3, 3 (1953) [translated in *Mathematical Foundations of Information Theory* (Dover, New York, 1958)].
14. J. Wolfowitz, *Illinois J. Math.* **1,** 591 (1957).
D. Blackwell, L. Breiman, A. J. Thomasian, *Ann. Math. Statistics* **29,** 1209 (1958).
B. McMillan, *ibid.* **24,** 196 (1953).
17. N. M. Abramson, *IRE Trans. on Inform. Theory* **IT-5,** 150 (1959).
18. P. Fire, "A Class of Multiple-Error-Correcting Binary Codes for Non-Independent Errors," *Sylvania Rept. No. RSL-E-2* (1959).
19. J. E. Meggitt, *IBM J. Research Develop.* **4,** 329 (1960).
20. D. E. Muller, unpublished work.
21. W. W. Peterson, *Error Correcting and Error Detecting Codes* (Wiley, New York, in press).
22. C. M. Melas, *IBM J. Research Develop.* **4,** 58 (1960).
23. P. Elspas, *IRE Trans. on Circuit Theory* **CT-6,** 45 (1960).
24. P. Elias, *IBM J. Research Develop.* **2,** 346 (1958).
G. D. Birkhoff, *Proc. Natl. Acad. Sci. U.S.* **17,** 656 (1931).
26. J. von Neumann, *ibid.* **18,** 70 (1932).
27. A. N. Kolmogorov, *Doklady Akad. Nauk S.S.S.R.* **124,** 754 (1959); see also V. A. Rohlin, *ibid.* **124,** 980 (1959); Y. Sinai, *ibid.* **124,** 768 (1959).
28. Y. V. Linnik, in *Theory of Probability and Its Applications* (1959).
29. I. M. Gelfand and A. M. Yaglom, *Uspekhi Math. Nauk* **12,** No. 1, 3 (1957).
30. L. Apostel and B. Mandelbrot, *Logique, Language et Théorie de l'Information* (Presses Universitaires de France, 1957); B. Mandelbrot, "Role of Sufficiency and of Estimation in the Statistical Structure of Thermodynamics," *IBM Research Rept. No. RC-235* (1960).
31. F. Attneave, *Applications of Information Theory to Psychology* (Holt, New York, 1959).
32. L. Brillouin, *Science and Information Theory* (Academic Press, New York, 1956).
33. S. Kullback, *Information Theory and Statistics* (Wiley, New York, 1959).
34. H. Quastler, *Information Theory and Biology* (Univ. of Illinois Press, Urbana, 1953); *Information Theory in Psychology* (Free Press, Glencoe, Ill., 1956).

Chromosome Cytology and Evolution in Primates

Study of chromosomes adds to our knowledge of evolutionary relationships among primates.

E. H. Y. Chu and M. A Bender

Since the chromosomes of higher organisms are characteristic of a species, they are of obvious value in taxonomic and evolutionary studies. The chromosomes of members of a given species are usually the same in number and form, but those of members of different species are frequently different. Thus the karyotype—the number of chromosomes, their lengths, their relative arm lengths, and other features—is a valuable morphological character, particularly because of its intimate association with the genetic makeup of the species.

Although the evolution of the primates has been a subject of great interest to biologists, it is only relatively recently that any attempt has been made to determine the interrelationships of these animals by comparative studies of their chromosomes. The main reason for this situation has probably been that technical difficulties are involved. Since most mammalian species possess a large number of small chromosomes, counting and observing individual chromosomes are very difficult. The recent development of techniques for the culture of diploid somatic cells, as well as the improvement of cytological procedures, has made it possible to determine not only the chromosome numbers but also the morphology of the chromosomes of a great variety of animals which have not been previously studied. These technical advances led to the discovery by Tjio and Levan (*1*) of the correct chromosome number of man, and stimulated a great many descriptive and experimental studies of mammalian, and especially human, cytogenetics.

The karyotypes of relatively few primates other than man have been deter-

The authors are members of the staff of the Biology Division of Oak Ridge National Laboratory, Oak Ridge, Tenn.