621.391(091)

Paper No. 1177
RADIO SECTION

A HISTORY OF THE THEORY OF INFORMATION

By E. COLIN CHERRY, M.Sc., Associate Member.

(The paper was first received 7th February, and in revised form 28th May, 1951.)

SUMMARY

The paper mentions first some essential points about the early development of languages, codes and symbolism, picking out those fundamental points in human communication which have recently been summarized by precise mathematical theory. A survey of telegraphy and telephony development leads to the need for "economy," which has given rise to various systems of signal compression. Hartley's early theory of communication is summarized, and Gabor's theory of signal structure is described.

Modern statistical theory of Wiener and Shannon, by which "information" may be expressed quantitatively, is shown to be a logical extension of Hartley's work. A Section on calculating machines and brains attempts to clear up popular misunderstandings and to separate definite accomplishments in mechanization of thought processes from mere myths.

Finally, a generalization of the work over the whole field of scientific observation is shown, and evidence which supports the view that "information plus entropy is an important invariant of a physical system" is included.

(1) INTRODUCTION

In recent years much has been heard of the "theory of communication" or, in a wider scientific circle, the "theory of information": in scientific books, articles and broadcast talks, many concepts, words and expressions are used which are part of the communication engineer's stock-in-trade. Although electrical communication systems have been in use for over a century it is only comparatively recently that any attempt has been made to define what is the commodity that is being communicated and how it may be measured. Modern communication theory has evolved from the attempts of a few mathematically-minded communication engineers to establish such a quantitive basis to this most fundamental aspect of the subject. Since the idea of "communication of information" has a scope which extends far beyond electrical techniques of transmission and reception, it is not surprising that the mathematical work is proving to be of value in many, widely diverse, fields of science.

It may therefore be of interest to communication engineers to consider historically the interrelation between various activities and fields of scientific investigation on which their subject is beginning to impinge. The scope of *information theory* is extensive, and no attempt will be made here to present an exposition of the theory; it is in any case still in process of development. This paper should be regarded purely as a history; the references appended may be found useful by those wishing to study the subject in detail.

(1.1) Languages and Codes

Man's development and the growth of civilizations have depended in the main on progress in a few activities, one of the most important of which has been his ability to receive, to communicate and to record his knowledge. Communication essentially involves a language, a symbolism, whether this is a spoken dialect, a stone inscription, a cryptogram, a Morse-code signal or a chain of numbers in binary digital form in a modern computing machine. It is interesting to observe that as technical applications have increased in complexity with the passage of time, languages have increased in simplicity, until to-day we are considering the ultimate compression of information in the simplest possible forms. It is important to emphasize, at the start, that we are not concerned with the meaning or the truth of messages; semantics lies outside the scope of mathematical information theory. The material which is to be communicated between a transmitting mind and a receiving mind must, however, be prearranged into an agreed language. We are not concerned with the "higher" mental processes which have enabled man to find a word for a concept, even less with the formation of a concept; we start only from the point when he already has a dictionary.

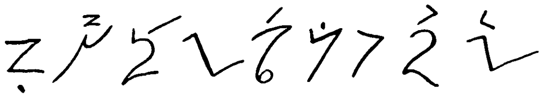
A detailed history of spoken and written languages would be irrelevant to our present subject, but nevertheless there are certain matters of interest which may be taken as a starting-point. The early writings of Mediterranean civilizations were in picture, or "logographic" script; simple pictures were used to represent objects and also, by association, ideas, actions, names, and so on. Also, what is much more important, phonetic writing was developed, in which sounds were given symbols. With the passage of time, the pictures were reduced to more formal symbols, as determined by the difficulty of using a chisel, or a reed brush, while the phonetic writing simplified into a set of two or three dozen alphabetic letters, divided into consonants and vowels.[1, 2]

In Egyptian hieroglyphics we have a supreme example of what is now called *redundancy* in languages and code; one of the difficulties in deciphering the Rosetta stone lay in the fact that a polysyllabic word might give each syllable not one symbol but a number of different ones in common use, in order that the word should be thoroughly understood.[3] (The effect when literally transcribed into English is one of stuttering.) On the other hand the Semitic languages show an early recognition of redundancy. Ancient Hebrew script had no vowels; modern Hebrew has none, too, except in children's books. Many other ancient scripts show no vowels. Slavonic Russian went a step further in condensation: in religious texts, commonly used words were abbreviated to a few letters, in a manner similar to our present-day use of the ampersand, abbreviations such as lb and the increasing use of initials, e.g. U.S.A., Unesco, O.K.

But the Romans were responsible for a very big step in speeding up the recording of information. The freed slave Tyro invented shorthand in about 63 B.C. in order to record, verbatim, the speeches of Cicero. This is not unlike modern shorthand in appearance (Fig. 1), and it is known to have been used in Europe until the Middle Ages.[4]

Related to the structure of language is the theory of cryptograms or ciphers, certain aspects of which are of interest in our present study, in connection with the problem of coding.[24]

Written contributions on papers published without being read at meetings are invited for consideration with a view to publication.
This is an "integrating" paper. Members are invited to submit papers in this category, giving the full perspective of the developments leading to the present practice in a particular part of one of the branches of electrical science.
The paper is a modified version of one presented by the author to the Symposium on Information Theory held under the auspices of the Royal Society in September, 1950. The original paper was given only restricted publication.
Mr. Cherry is Reader in Telecommunications, University of London.



Nemo fideliter diligit quem fastidit nam et calamitas querula.

Fig. 1.—Roman shorthand (orthographic).

Ciphering, of vital importance for military and diplomatic secrecy, is as old as the Scriptures. The simple displaced alphabet code, known to every schoolboy, was most probably used by Julius Caesar.[3] There are many historic uses of cipher; e.g. Samuel Pepys's diary[4] was entirely ciphered "to secrete it from his servants and the World"; also certain of Roger Bacon's scripts have as yet resisted all attempts at deciphering. A particularly important cipher is one known as "Francis Bacon's Biliteral Code": Bacon suggested the possibility of printing seemingly innocent lines of verse or prose, using two slightly different founts (called, say, 1 and 2). The order of the 1's and 2's was used for coding a secret message: each letter of the alphabet was coded into five units, founts 1 and 2 being used as these units.[3] Now, this code illustrates an important principle which seems to have been understood since the early civilizations —*information may be coded into a two-symbol code.** There are numerous examples: bush-telegraph signals of many Congo tribes used drum beats with high and low pitch;† long and short smoke signals are another case. Nowadays we have the Morse code, in dots and dashes, and many similar codes.

Such two-symbol codes are the precursors of what is now called "binary coding," as used in pulse-coded telephony and high-speed digital computing machines. The transmitted information is coded into a series of electrical pulses and blank spaces, often referred to as a "yes-no" code. The ancient Celts invented a script, of interest in this connection, which is known as the Ogam[4] script and is found in Ireland and Scotland. Most scripts have developed into structures of complex letters, with curves and angles, difficult to chip in stone, but the Celts seem to have consciously invented this script, using the simplest symbol of all—a single chisel stroke—discovering that this was all that is necessary (Fig. 2). This script could nowadays be directly written in electrical pulses: thus we could use positive pulses, negative pulses, double pulses and blank spaces. Our modern need is similar to that of the Celts, for it is of great advantage economically to be able to write using only one symbol—a pulse, or a chisel stroke.

With the introduction of the famous dot-dash code by S. F. B. Morse in 1832, a new factor was consciously brought in. This was the statistical aspect. Morse purposely designed his code so that the most commonly used letters were allocated the shortest symbols (Fig. 3). It is of course true that in every language the most commonly used words have unconsciously become the shortest. The need for such a statistical view has been appreciated for a hundred years, and as time has progressed this has assumed increasing importance. Certain types of message are of very common occurrence in telegraphy; e.g. some commercial expressions and birthday greetings, and these should be coded into very short symbols. By the year 1825 a number of such codes were in use. The modern view is that messages having a high probability of occurrence contain little information and that any mathematical definition we adopt for the expression "information" must conform with this idea, that the information conveyed by a symbol, a message or an observation, in a set of such events, must decrease as their frequency of occurrence in the set increases. Shannon has emphasized the importance of the statistics of language in his recent publications,[24] and has referred to the relative frequencies of letters in a language, and of the digrams (or letter-pairs) such as *ed*, *st*, *er*, and of the trigrams *ing*, *ter*, and so on, which occur in the English language. His estimate is that English has a redundancy of at least 50%, meaning that if instead of writing every letter of an English sentence (or its coded symbol) we were suitably to symbolize the digrams, trigrams, etc., the possible ultimate compression would be at least two to one. If such a perfect code were used, any mistake which might occur in transmission of a message could not be corrected by guessing. Language statistics[3] have been essential for centuries, for the purpose of deciphering secret codes and cryptograms. The first table of letter frequencies to be published was probably that of Sicco

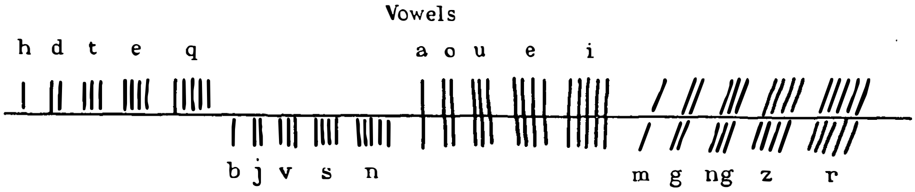


Fig. 2.—The Ogam (Celtic) script.

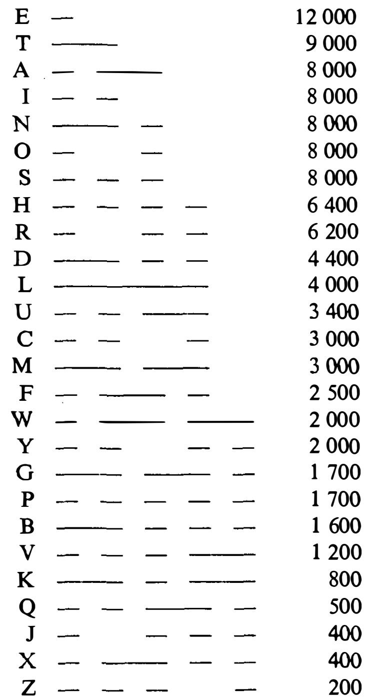


Fig. 3.—Morse's original code, showing relation to quantities of type found by him in a printer's office.

* Strictly, only "logically communicable" information.
† See CARRINGTON, J. F.: "The Drum Language of the Lokele Tribe," *African Studies* (Witwatersrand University Press, 1944).

Simonetta of Milan in the year 1380; another used by Porta in 1658 included digrams also.

Modern mathematical symbolism illustrates a language possessing a high degree of compression of information. The Greeks had been limited largely to geometry, algebra eluding them because of the lack of a symbolism. Descartes's application of formulae to geometry and, even more important, Leibnitz's great emphasis of symbolism are two outstanding developments in the compression of mathematical information. The importance of symbolism is indeed prominent throughout the modern evolution of mathematics, as its generalizations have increased; Russell and Whitehead's treatment of the bases of mathematics (1910) as a generalization of ordinary logic was written almost entirely without words.[6] During the last century, the idea has emerged of mathematics as the "syntax of all possible language" and as the language of logic. In particular Peano invented the symbolism used for symbolic logic.

Leibnitz not only considered mathematical symbolism as a universal "language of logic," but he was a great advocate of language reform. Already, in 1629, Descartes had considered the possibility of creating an artificial, universal, language, realizing that the various languages of the world were, by virtue of their complex evolution, utterly illogical, difficult and not universal. However, his dream did not materialize until 1661, when the Scotsman George Dalgarno published his "Ars Signorum." All knowledge was first to be grouped into 17 sections, under a heading represented by a consonant; each section was then to be headed with titles such as "politics," "natural objects," etc., and further divided into subsections, represented by a vowel, then into sub-subsections and so on, consonants and vowels alternating.[2] Every word, always pronounceable, thus denoted an object or an idea by a sequence of letters representing selections from the prearranged sections, subsections, etc.; this notion of selection is very relevant to the modern theory of information and will be discussed more fully later.

The work of Leibnitz, on the construction of a rational grammar, has no real connection with information theory, since it is essentially concerned with the meaning of a sentence; however, the ideas of Descartes and Dalgarno, to a limited extent, are connected with the theory. Information theory is applicable when an idea has already been expressed as a sentence or a succession of letters or symbols, where each successive symbol represents a selection out of a prearranged language or code, possessing a certain statistical structure. In his recent work, Shannon[24] has illustrated the idea of the building-up of a "message" as a stochastic process, that is, as a series of words, each one being chosen entirely on a statistical basis, depending on the one, two, three . . . words immediately preceding. That such sequences of words can bear some resemblance to an English text merely illustrates the accuracy of the statistical tables used, although no meaning is conveyed by the resulting "message."

In this connection there is a certain historical interest in the following quotation from Jonathan Swift's "Gulliver's Travels." Gulliver has paid a visit to the Academy of Lagado, and amongst a series of abusive descriptions of imaginary research programmes, describes that of the Professor of Speculative Learning.

> The professor observed me looking upon a frame which took up a great part of the room; he said that by this contrivance the most ignorant person may write in philosophy, poetry and politics. This frame carried many pieces of wood, linked by wires. On these were written all the words of their language, without any order. The pupils took each of them hold of a handle, whereof there were forty fixed round the frame and, giving them a sudden turn, the disposition of the words was entirely changed. Six hours a day the young students were engaged on this labour; the professor showed me several volumes already collected, to give to the world a complete body of all the arts and sciences. He assured me that he had made the strictest computation of the general proportion between the numbers of particles, nouns and verbs. . . .*

(2) COMMUNICATION THEORY

Perhaps the most important technical development which has assisted in the birth of information theory is that of telegraphy. With its introduction, the idea of speed of transmission arose, and when its economic value was fully realized, the problems of compressing signals exercised many minds, leading eventually to the concept of "quantity of information" and to theories on times and speed of signalling. In the year 1267 Roger Bacon[4] suggested that what he called "a certain sympathetic needle" (i.e. lodestone) might be used for distance communication. Porta and Gilbert, in the sixteenth century, wrote about "the sympathetic telegraph," and in 1746 Watson, in England, sent electric signals over nearly 2 miles of wire. In 1753 an anonymous worker used one wire for each letter of the alphabet, but in 1787 Lomond used one wire-pair and some code. Gauss and Weber invented a 5-unit code in 1833, later to be named after Baudot, in honour of his life's work in automatic telegraphy. The introduction of carrier waves, during the First World War, was made practicable by G. A. Campbell's invention of the wave filter, and the frequency-division-multiplex method of transmission rapidly developed. This principle of allocating simultaneous signals into "frequency-bands" has been the mainstay of electrical communication and remained unchallenged until the Second World War.

Related techniques which have greatly urged the development of general communication theory are those of telephony and television. Alexander Graham Bell's invention of the telephone in 1876 has particular significance in relation to the physiologist's interest in communication theory, to which more reference will be made in Section 3; otherwise it is, from our present point of view, purely a technological development, setting up problems similar to those of telegraphy. However, early in the history of television, 1925–27, the very great bandwidth required for detailed "instantaneous" picture transmission was appreciated, and this was brought to a head with the introduction of the techniques of cathode-ray tubes, mosaic cameras and other electronic equipment rendering high-definition practicable. Great masses of information had to be read off at high speed at the camera end, transmitted and reassembled at the receiver. Now that the human eye had been brought into the communication link, with its ability to recognize shape, the problem of phase-distortion became all important, although Sallie Pero Meade had already been concerned with phase troubles in long-distance telegraphy,[8] in 1928. Furthermore, other major theoretical studies were forced by the great bandwidths required for television; in particular the noise problem, the transient response problem, and the problem of amplifying widebands of energy. Noise is of particular interest to our subject, as we are all aware; it is the ultimate limiter of the communication of information. The subject of noise is itself a vast one and cannot be treated in this short history, except only to mention the names of the chief pioneers, Einstein (1905),[49] Schottky (1918),[50] Johnson (1928)[52] and Nyquist (1928).[53]

But to return for a moment to the First World War: wireless had been developed from the laboratory stage to a practical proposition largely owing to the field work of Marconi and the early encouragement of the British Post Office, at the turn of the century.[4] Sidebands were soon discovered by many people, but fruitless controversies arose about whether they did or did not exist. The advantages of reducing the bandwidth required

* SWIFT, JONATHAN: "The Voyage to Laputa" (1726), Chap. 5. (The extract given has been condensed.)

for the transmission of a telephony signal were appreciated, but the early theories of modulation were very vague and lacked mathematical support. In 1922 John Carson[9] clarified the situation with a paper showing that the use of frequency modulation, as opposed to amplitude modulation, did not compress a signal into a narrower band. He also made the important suggestion that all such schemes "are believed to involve a fundamental fallacy," a fact we now know well. At that time it was well known that only one sideband needs to be used, since both contain the same information.* Curiously enough, although it was recognized, very early on, that carrier telephony requires the same bandwidth as the spoken word, nobody seems to have stated clearly that carrier telegraphy requires a finite bandwidth until long after such systems had been in use. In those days, the inventors of wireless telegraphy naïvely imagined that the "frequency" was "on" for a time and "off" for a time. The priority of recognition of this necessity of bandwidth again probably belongs to Carson.

After Carson, the next major contribution to the theory of frequency modulation was due to Balth. van der Pol in 1930, who used the concept of "instantaneous frequency" as originally defined by Helmholtz† as rate of change of a phase angle. In 1936, Armstrong published an account of the first practical frequency-modulation system[14] which appeared to refute Carson's results. The new principle was used of limiting a carrier to constant amplitude, by which means the strongest of a number of carrier waves, received simultaneously, might be separated out (the "capture effect"). The advantage was claimed for the system that a number of stations might be sited close together, and the strongest received at a given point without interference from the others. Also the system offered advantages with regard to signal/noise ratio, a point not considered by Carson in 1922.

In 1924, Nyquist[10] in the United States and Küpfmüller[11] in Germany simultaneously stated the law that, in order to transmit telegraph signals at a certain rate, a definite bandwidth is required, a law which was expressed more generally by Hartley[12] in 1928. This work of Hartley's has a very modern ring about it; he defined information as the successive selection of symbols or words, rejecting all "meaning" as a mere subjective factor, and showed that a message of N symbols chosen from an alphabet or code of S symbols has S^N possibilities and that the "quantity of information" H is most reasonably defined as the logarithm, that is $H = N \log S$. Hartley also showed that in order to transmit a given "quantity of information" a definite product, bandwidth × time, is required. We shall later be considering the more modern aspects of this theory of Hartley's, which may be regarded as the genesis of the modern theory of the communication of information.

All the early modulation theories took as a basic signal the continuous sine wave or, at the best, a continued periodic waveform. Such applications of Fourier analysis give "frequency descriptions" of signals and are essentially timeless. The reverse description of a signal, as a function of time, falls into the opposite extreme, as if the values of the signal at two consecutive instants were independent. Practical signals, whether speech or coded symbols, are of finite duration and at the same time must, to be reasonable, be considered to occupy a certain bandwidth. The longer the signal time element Δt, the narrower the frequency band Δf, or, as we may say, the more certain is its frequency.

Gabor[15] took up this concept of uncertainty in 1946, and associated the uncertainty of signal time and frequency in the form $\Delta t \, . \, \Delta f \simeq 1$, by analogy with the Heisenberg uncertainty relation of wave mechanics. In this, he is most careful to point out that he is not attempting to explain communication in terms of quantum theory, but is merely using some of the mathematical apparatus. Gabor points out that our physical perception of sound is simultaneously one of time *and* frequency, and that a method of representation may be used which corresponds more nearly to our acoustical sensations than do either the pure frequency-description or time-description. The basic signals on which such a representation is based must be finite in both time and frequency bandwidth. Using reasoning closely related to that of Hartley,[12] Gabor shows that there must be only two independent data expressed by these basic signals, per unit of the product of time and frequency (i.e. $2ft$ data). These data (which may be, for example, the Fourier coefficients) are all we know about the signal; from this information the signal may be reconstructed. There are many choices for the basic signals, but one of particular interest uses sine waves modulated by a Gaussian probability function. These are distinguished by their property that their effective bandwidth-time product is the smallest out of all possible signals and hence they overlap as little as possible. Such a signal is regarded as a "unit of information" and is called by Gabor a "logon."* Further, their Fourier transforms have the same mathematical law; hence the representation (which is made as a kind of matrix) is symmetrical in frequency and time. Also they have the advantage that the notions of amplitude and of phase can be applied to these signals as well as, and with more *physical* justification than, to continous sine-waves. In this work, Gabor operates with complex signals, which have no negative frequencies; more recently the theory of such complex signals has been extended by J. A. Ville, in France.[16]

By this time in the history of communication it had been realized for several years that in order to obtain more economical transmission of speech signals, in view of this bandwidth-time law, something drastic must be done to the speech signals themselves, without seriously impairing their intelligibility. These considerations led to what is known as the "vocoder," an instrument for analysing, and subsequently synthesizing, speech. It is fair to say that this arose out of a study of the human voice and the composition of speech, which is itself of early origin; for example, Graham Bell and his father had studied speech production and the operation of the ear, while more recently there has been the work of Sir Richard Paget[17] and of Harvey Fletcher.[18] In the year 1939 Homer Dudley[19, 20] demonstrated the "voder" at the New York World's Fair. This instrument produced artificial voice sounds, controlled by the pressing of keys and could be made to "speak" when manually controlled by a trained operator. In 1936 Dudley had demonstrated the more important vocoder: this apparatus is essentially a means for *automatically* analysing speech and reconstituting or synthesizing it. The British Post Office also started, at about this date, on an independent programme of development, largely due to Halsey and Swaffield.[21] In simple terms it may be said that the human voice employs two basic tones: those produced by the larynx operation, called "voiced sounds," as in *ar*, *mm*, *oo*, and a hissing or breath sound, for which the larynx is inoperative, as in *ss*, *h*, *p*, etc. Speech contains much which is redundant to information or intelligence and which is therefore wasteful of bandwidth; thus it is unnecessary to transmit the actual voice tones, but only their fluctuations. At the transmitter these fluctuations are analysed and sent over a narrow bandwidth. At the same time another signal is sent to indicate the fundamental larynx pitch, or if absent, the hissing, breath, tone. At the

* Carson, J.: Patents, 1915. Espenscheid: In demonstrations, 1922.
† Helmholtz: "Die Lehre von den Tonemfindungen" 1862).

* This has nowadays come to be known as a unit of "structural information"; it relates only to a specified channel and not to the particular signal being transmitted, or to the code used. The "logons" are the units out of which any message, which the channel is capable of transmitting, may be considered to be constructed.

receiver, these signals are made to modulate and control artificial locally produced tones from a relaxation oscillator or a hiss generator, thus reconstituting the spoken words.

The minimum overall bandwidth required is (at least on paper) about 275 c/s, a compression of about 10/1; such a method may be regarded as a voice-operated coder, which automatically codes and decodes the voice. Two decades earlier Küpfmüller had suggested that a compression ratio of 40/1 in the magnitude of *ft* could be achieved, for the transmission of a single letter, if a telegraph code were used rather than the human voice. Another method to reduce the bandwidth of the signal, called "frequency compression," has been described by Gabor.[22] In the transmitter a record of the speech is scanned repeatedly by pick-ups themselves running, but with a speed different from that of the record. Thus a kind of Doppler effect is produced, reducing the bandwidth of the transmitted signal, which in the receiver is expanded to its original width by a similar process. It is of course impossible to reduce *all* frequencies in the same ratio, since this would imply stretching the time scale; what the apparatus does is rather to reduce the acoustical frequencies, leaving the syllabic periods unchanged. It has been demonstrated that fourfold compression is possible with hardly any, and sixfold compression with slight, loss of intelligibility, but there is some loss of quality. This can be avoided, according to Gaber's theory, by matching the scanning frequency continuously to the voice pitch, but this has yet to be demonstrated experimentally.

We have so far considered, in the main, the frequency aspect of the transmission of signals, i.e. questions of bandwidth; of frequency spectra of signals; of amplitude, phase or frequency modulation and so on. This frequency aspect absorbed most attention during the rapid developments of the nineteen-twenties and early nineteen-thirties. In the last few years, however, it is the time aspect of which we hear so much: we hear of pulse modulation, of pulse-code modulation and of time-division multiplex. Referring back into history, the earliest suggestion for the simultaneous transmission of two messages, over one line without frequency separation, seems to have come from Edison, who introduced the "duplex" and "quadruplex" systems in 1874 (Fig. 4).* With this system one message, sent

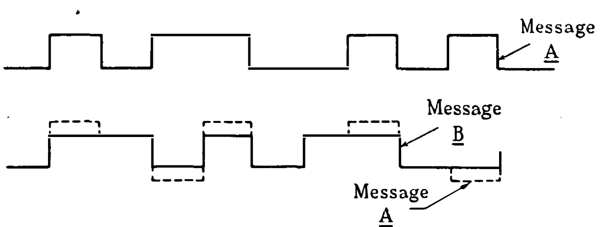


Fig. 4.—Edison's system of duplex telegraphy.

in Morse code, was read at the receiving end by a polarized relay; the second message was transmitted as an amplitude modulation of the first signal and was read by an unpolarized relay, the first message merely acting as a carrier wave and so being ignored by this unpolarized relay. The important principle was employed here that two messages can be sent simultaneously, over the same bandwidth that is required for one, if the power is increased. Although not explicitly stated in this form in his paper, Hartley[12] has implied that the quantity of information which can be transmitted in a frequency band of width B and time t is proportional to the product: $2Bt \log S$, where S is the number of "distinguishable amplitude levels." Hartley has considered messages consisting of discrete symbols, e.g. letters or Morse code, and also messages consisting of continuous

* Heaviside made similar suggestions for duplex telegraphy in 1873.

waveforms, such as speech and music. He observes that the latter signals do not contain infinite information since "the sender is unable to control the waveform with complete accuracy." He approximates the waveform by a series of steps, each one representing a *selection* of an amplitude level. Such a representation is nowadays referred to as *amplitude quantization* of the waveform. For example, consider a waveform to be traced out on a rectangular grid (Fig. 5), the horizontal mesh-width

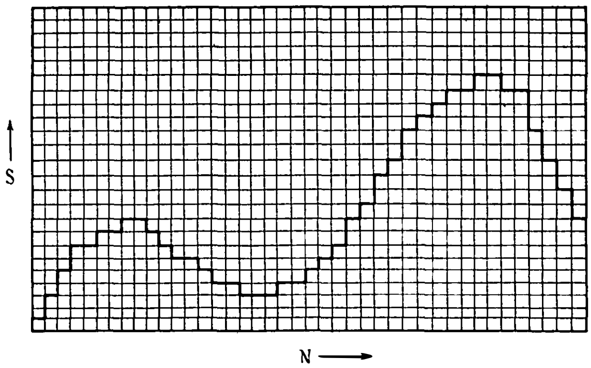


Fig. 5.—"Quantizing" of a continuous wave.

representing units of time (equal to $1/2B$ in order to give the necessary $2Bt$ data in a time t), and the vertical the "smallest distinguishable" amplitude change; in practice this smallest step may be taken to equal the *noise level*, n. Then the quantity of information transmitted may be shown to be proportional to

$$Bt \log \left(1 + \frac{a}{n}\right)$$

where a is the maximum signal amplitude, an expression given by Tuller,[23] being based upon Hartley's definition of information.

The total transmitted quantity of information may be held constant, but the magnitudes B, t and a changed; thus bandwidth or time may be traded for signal power. This principle has been given practical embodiment in various systems of pulse modulation[25] and time-division multiplex. Although a number of such systems had appeared in patent form in the 1920's, their application was mainly to improvement of transmitter efficiency. Their development was delayed until a few years before the 1939–45 War, partly owing to the lack of suitable pulse and microwave techniques. Particular mention should be made of the work of Reeves and Deloraine,* who patented time-division-multiplex systems[26] in 1936. In such systems the waveform is not transmitted in its entirety but is "sampled" at suitable intervals (about $1/2B$), and this information is transmitted in the form of pulses, suitably modulated in amplitude, width, number or time position. Reeves† also proposed another system, which uses directly the idea of amplitude quantization, as envisaged by Hartley; the waveform is automatically restricted, at any instant, to one of a number of fixed levels (as illustrated in Fig. 5) before being sampled and transmitted as a pulse signal. This assimilation of telephony into telegraphy becomes even more complete if the quantized pulses are coded.[26] Thus the information which must be transmitted is given by the number of quanta in each sample, and this number, having one of S possible values, may be coded into a train of pulses in, for example, binary (yes-no) code. Binary code is the simplest

* Reeves, A. H.: British Patents Nos. 509820 and 521139 (U.S. Patent No, 2262838). British Patents No. 511222 (U.S. Patent No. 2266401; French Patent Nos. 833929 and 49159).

† Reeves, A. H.: British Patent No. 535860 (U.S. Patent No. 2272070. French Patent No. 852183).

code; the number of "distinguishable levels" is reduced to two, and information cannot be communicated with less.

A waveform quantized in amplitude and time, as in Fig. 5, can have S^N possible "states"; by regarding these states as an alphabet of all the possible waveforms, Hartley's law gives the information carried by one waveform as $H = KN \log S$, which is finite. Since Hartley's time this definition of information as *selection* of symbols has been generally accepted, variously interpreted and gradually crystallized into an exact mathematical definition.

Any quantitative description of the information in a message must be given in statistical terms; the information conveyed by a symbol must decrease as its probability of occurrence increases. With probabilities attached to the various symbols $P_1 P_2 \ldots P_i \ldots$ in a message, or to the various "states" of a waveform, Hartley's law may be reinterpreted so as to define the average information in a long sequence of n symbols as

$$H_n = - \Sigma P_i \log P_i$$

an expression which has been evolved in various ways by several different authors, in particular Shannon[24] and Wiener,[37] during the last few years.*

This expression for the information is similar to that for the entropy† of a system with states of probabilities $P_1, P_2 \ldots P_i \ldots P_n$, using the term in the Boltzmann statistical sense. Probably the first detailed discussion of the identity between information and entropy was made by Szilard[48] as early as 1929, who, in a discussion on the problem of "Maxwell's demon," pointed out that the entropy lost by the gas, due to the separation of the high- and low-energy particles, was equal to the information gained by the "demon" and passed on to the observer of this "experiment." In his recent publications, Shannon has forged this theory of communication of information[24] into a coherent theory, using the Boltzmann statistical definition of entropy as a basis.

One of Shannon's principle contributions to communication theory is his expression for the maximum capacity of a channel. This gives the greatest quantity of information which may be communicated in a time t over bandwidth B, in the presence of white Gaussian noise, as

$$Bt \log \left(1 + \frac{P}{N}\right)$$

where P and N are the signal and noise powers.

This formula superficially resembles that, given above, due to Tuller, which was derived by somewhat loose argument from Hartley's definition of information. Shannon's formula, however, gives a true maximum; it emerged from his recognition of communication theory as a branch of statistical mechanics.

For this maximum rate of communication to be achieved the message itself, being limited as regards its power, must be so encoded as to have the statistical structure of Gaussian noise. In this case the coding achieves maximum entropy in the channel; the information in the messages is then perfectly compressed. Nevertheless, the practical methods whereby such ideal systems of coding may be achieved still remain to be discovered; certain error-correcting codes have, however, been developed.[33]

Some reference has already been made to Shannon's treatment of a message as a stochastic process and to the idea of redundancy. An encoding system exists which, in the presence of noise, introduces just sufficient redundancy to overcome this noise, giving as low a frequency of errors as desired (we are all familiar with the idea of having to repeat words or sentences on a noisy telephone). This theorem and others which have been expounded by Shannon, in precise mathematical form, are of the most general application, not only to communication engineering, but to many other fields of science.

* Also, in Great Britain, by W. S. Percival, in unpublished work.

† The relation to the entropy concept of statistical mechanics is a little indefinite; for example, there is a subtle distinction between the ordinates of spoken language sounds, which possess energy, and the letters of a written language, which do not. The term "selective entropy" has been suggested for this application to communication theory.

The related problem of designing the apparatus comprising the communication channel, on a statistical basis, was considered earlier by Wiener,* who dealt with the design of a filtering system to be optimum in combating noise, showing that an ideal response characteristic exists which minimizes the mean-square error between waveforms of the input signals and of the output signals plus noise. A time delay is the price paid; the longer the time delay, the less is the chance of error. A related problem considered by Wiener is that of prediction. He shows that a predictor circuit may be designed which reproduces the incoming waveforms with minimum mean-square error, in advance of their arrival by a short time; in this case the longer this time advance, the greater is the chance of error. In both problems a certain delay period is required as the price of a solution; it is the price of the increased information. However, the choice of a mean-square error is an arbitrary one, and this criterion may not be appropriate in every problem.

(3) BRAINS—REAL AND ARTIFICIAL

The operation of a computing machine is of the same nature as that of any electrical communication channel; information is supplied from a "source," suitably coded, transmitted, operated on in various ways, and passed to the output. From the information theory point-of-view there are, however, certain differences. First, a computing machine is usually "noiseless" in that it cannot be allowed to make a single mistake,† since this mistake would render all subsequent calculations invalid; it may, however, possess a limiting accuracy, set by the limited digital capacity. Secondly, the machine comprises many individual communication channels.‡ Thirdly, the questions of the language statistics and coding such as arise in electrical communication are replaced by problems of "programming"; the development of the most recent digital computing machines, such as the Eniac and the Edsac, and the Ace[29] in Great Britain, the so-called "electronic brains," primarily for application to problems in mathematical physics and in pure mathematics, has raised complex problems in "programming," i.e. the breaking down of mathematical operations into the most elementary steps and the logical feeding of these steps into the machine together with *a priori* data referring to the particular calculation.§ The surprising thing is that, once the mathematical processes have so been broken down, both these fundamental steps and the actions required of the machine are few in number and elementary in principle; such simple processes as adding, subtracting, moving-up one, etc. It is the automatic feeding-in of the sequence of instructions which distinguishes these modern machines from the manually operated desk types, and especially the facility of changing the sequence according to criteria evaluated during the course of calculation.

Descartes, and more particularly Leibnitz, had visions of computing machines—"reasoning machines" as they were con-

* WIENER, N.: "The Interpolation, Extrapolation and Smoothing of Stationary Time Series" (John Wiley, 1949).

† Self-checking procedures are now a recognized essential part of high-speed digital computer operation; see Reference 33 (also for a further bibliography on this subject). Such a facility involves redundancy and, in accordance with Shannon's theory, requires extra channel capacity or a lower speed of operation.

‡ Either the serial system may be used, in which all the digits of a binary number are transmitted in time sequence, or the parallel system, in which they are sent simultaneously over different channels; this is analogous to the time-bandwidth alternative, as in electrical communication channels.

§ The punched card system, for storing numbers and subsequently entering them into a machine, was first conceived by Herman Hollerith, in 1889; the scheme is similar to that used for weaving of patterns in cloth, by the Jacquard loom, which is itself essentially a problem in coding.

sidered.* But the lack of technique prevented practical construction until Charles Babbage, while Professor of Mathematics at Cambridge between 1829 and 1838, commenced construction of two different kinds of automatic digital computing machine (unfortunately never completed), one of which bears considerable resemblance, in its basic structure, to our modern machines.[29,30,31] This "analytical engine" possesses three component parts: a store for data or for intermediate results of a calculation, which could be read off as desired, a "mill" for performing arithmetical operations and an unnamed unit (actually a "controller") for selecting the correct data from the store, for selecting the correct operation for the "mill," and for returning the result to the store.

Over a 100 years later, the first successful mechanical digital computer was operated, the Harvard Mark I calculator by Professor Aiken,[28] using the fundamental principles envisaged by Babbage.[32] The first electronic machine was the Eniac (designed by Eckert and Mauchly), with its inherent advantages of speed. Although the choice of a suitable scale or radix is quite arbitrary for the representation of numbers in a machine, there are great advantages offered by a binary scale, using only the numbers 0 and 1, since the physical elements used for representing numbers in a machine usually have two mutually exclusive states. For example, a switch, a relay or a valve can either be *on* or *off*. Such method of representing numerical information is another example of the age-old coding principle: information can be represented by a two-symbol code.

A suitable diagrammatic notation, for representing schematically the functional operation of these digital machines, was suggested by von Neumann, and later modified by Turing, being adapted from the notation proposed by McCulloch and Pitts for expressing the relations between parts of the nervous systems. This pair of workers[34] had applied the methods of mathematical logic to the study of the union of nerve fibres, by synapses, into networks, while Shannon had applied Boolean algebra to electric-circuit switching problems. Now the nervous system may be thought of, crudely speaking, as a highly complex network carrying pulse signals, working on an on/off basis; a neuron itself is thought to be a unit, like a switch or valve which is either on or off. The rise of a common notation for expressing the actions of the nervous system and of binary computing machines at least recognizes a crude analogy between them.

It is this analogy which has been greatly extended and has to-day become of considerable scientific interest. One is naturally led to extend the analogy to thought processes and to the possible design of "reasoning machines," thus fulfilling the dream of Leibnitz. Just as arithmetic has led to the design of computing machines, so we may perhaps infer that symbolic logic may lead to the evolution of "reasoning machines" and the mechanization of thought processes.

Such possibilities, together with the success already achieved by the automatic computing machine, have caught the popular imagination during the last few years. On the whole, the approach of the scientist, to this field, has been wise and cautious, but its news value has somewhat naturally led to exaggerations in the lay Press; thus, phrases such as "electronic brains" and "machines for writing sonnets" are now commonly heard, and may have led to some scepticism in certain scientific circles. However, practical accomplishments are now forcing one to the conclusion either that "mechanized thinking" is possible, *or* that we must restrict our concept of "thinking." Much analysis of "mechanized thinking" has been promoted by the theory of (intellectual) games.[35] Machines which play "noughts and crosses" and, incidentally, will always win or draw are comparatively simple;* the existence of such machines does not often cause surprise, because this game is considered determinate —presumably owing to the very limited numbers of moves— whereas chess or card games give one the feeling that "judgment" is involved. However, Shannon has recently dealt with the problem of programming a computor for playing chess,[36] concluding that a machine is constructible, in principle, which can play perfect chess, but that owing to the great number of possible moves this would be impracticable. Nevertheless one could be made which would give its opponents a very good game. Turing had also previously referred to such possibilities in unpublished work. There are two important points which are emphasized by most writers concerned with "mechanized thinking." First, the machine acts on instructions given to it by the designer; as an illustration, Shannon observes that at any stage in a game of chess, played against his machine, the next move which the machine will make is calculable by its designer, or by one who understands its programming. Again, every single step in a calculation, carried out by a digital computor, could well be done by a human—the machine is merely far quicker.† The second point of emphasis is the importance of the *programming* rather than the machine in the metal. As Norbert Wiener has been most careful to stress,[37] it is not the machine which is mechanistically analogous to the brain, but rather the *operation* of the machine plus the instructions fed to it. If a machine can ever be said to learn by its own mistakes, and improve its operation,[33] it can do this only, as Shannon emphasizes in connection with chess, by improving its programming.

A calculating machine cannot create information, it can only operate on, or transform, information supplied to it. Similarly when we solve a set of simultaneous equations we do not really obtain new information; the various steps and the final solution represent transformations (or "codings") of the information which was contained implicitly in the original equations.

To digress for a moment, it was apparent during the years immediately preceding the Second World War that the ideas, basic concepts and methods of communication engineering were of wide applicability to other specialized branches of science. For 200 years no single man has been able to compass the whole of science; the intensity of specialization has steadily increased and has necessarily been accompanied by much duplication of work. Nowadays, to venture out of one's own recognized domain of research is to invite an accusation of dilettantism. The lead was taken by Norbert Wiener,[37] who, with Rosenbleuth, named this field of applicability *cybernetics*.‡ The needs of the war brought matters to a head, with the urgency of developing not only high-speed computing machines, but automatic predictors, automatic gun-laying mechanisms and other automatic following or "self-controlling" systems, and to these two scientists should be given the credit for calling attention to the need for a general study to include, not only these automatic mechanisms, but certain aspects of physiology, the central nervous system and the operation of the brain, and even certain problems in economics concerning the theory of booms and slumps. The common thread here, linking these topics, whether mechanical, biological or mathematical, is the idea of the communication of information and the production of a self-stabilizing control action. Apart from a study of the mechanical governor by Maxwell, in 1868, the first mathematical treatment of the stabilization of a dynamic system by feeding information back from the output or "receiver"

* Pascal constructed an adding machine (using numbered wheels) in 1642; Leibnitz built a digital multiplying machine in 1694. The modern desk computing machines have originated from these. The "reasoning machines," to which later reference will be made, were visualized as dealing with problems of logic and not merely for arithmetic computations.

* For example, work has been carried out at the National Physical Laboratory. See Davies, D. W.: "Science News No. 16" (Penguin Books, 1950), p. 40.

† However, if a machine contains an element which permits it to make a *random* choice at any stage (e.g. a table of random numbers), this deterministic view becomes somewhat modified.

‡ From the Greek κυβερνητης, meaning "steersman." The word *cybernétique* was originally coined by Ampère (1834) in "Essai sur la philosophie des sciences," to mean the "science of government."

end to the input or "transmitter" end was made by H. S. Black, in a study of electrical feedback amplifiers[38] in 1934, and later developed, largely due to the efforts of Nyquist[39] and of Bode,[40] into an exact mathematical method and a system of design. The extension of the principles to electromechanical or to purely mechanical systems was a logical and natural one, and the design of automatic following systems, such as those for anti-aircraft guns, for automatic pilots in aircraft, etc., need no longer proceed entirely on a trial-and-error basis.

For these automatic control systems, the term "servo mechanism" has been coined. The existence of numerous controls in the body accounts partly for a common interest with physiology. For example, there is homeostasis, or the involuntary control of body temperature, of heart rate, blood pressure and other essentials for life, while voluntary control is involved in muscular actions, such as those required for walking along a narrow plank; the simplest movement of a limb may involve multiple feedback actions. If a stabilized servo mechanism has its feedback path open-circuited, so that the magnitude of its error cannot be measured at the input end and so automatically corrected, it is liable to violent oscillation; an analogous state of affairs in the human body has been mentioned by Wiener,[37] called ataxia, corresponding to a nervous disorder which affects the control of muscular actions. The analogies in physiology are countless; Wiener goes even so far, in developing the analogy between the operations of a digital computing machine and of the brain and central nervous system, as to compare certain mental, functional, disorders (the layman's "nervous breakdowns") to the breakdown of the machine when overloaded with an excess of input instructions as, for example, when the storage or "memory circuits" cannot store enough instructions to be able to tackle the situation. Note again, the emphasis is on the operation of the machine together with its instructions; no material damage may have occurred.

One is led instinctively to ask whether such analogies are not modern examples of a kind of animism,* though these analogies do not imply any attempt to "explain" life on a mechanistic basis or to explain the body as a machine in the sense of Descartes, who observed that the action of the body, apart from the guidance of the will, "does not appear at all strange to those who are acquainted with the variety of movements performed by the different automata, or moving machines fabricated by human industry. . . . Such persons will look upon this body as a machine made by the hand of God."

Early invention was greatly hampered by an inability to dissociate mechanical structure from animal form. The invention of the wheel was one outstanding early effort of such dissociation. The great spurt in invention which began in the sixteenth century rested on the gradual dissociation of the machine from the animal form, with a consequential improvement in method and performance, until machines eventually became completely functional. The development of machines had a converse effect, and the body came to be regarded as nothing but a complex mechanism: the eyes as lenses, the arms and legs as levers, the lungs as bellows, etc. Julien de la Mettrie, in about 1740 wrote, "[he thought] that the faculty of thinking was only a necessary result of the organization of the human machine," a materialist view which so greatly disturbed the "vitalists" of the time.[6]† Since "animistic thinking" has been recognized as such by inventors and scientists, its dangers are largely removed and turned to advantage, though amongst laymen it subconsciously exists to-day (as in the use of the expression "electronic brain"). Physics and biology have gone hand in hand; e.g. Harvey's discovery[6] of the circulation of the blood (1616) owes much to the work being carried out on air pumps in which Harvey was interested. In more modern times, Helmholtz attempted to unify physics, physiology and aesthetics, in his studies of music and hearing. Electrical communication owes a debt to physiology; thus Graham Bell,[4] the son of A. M. Bell, who was an authority on phonetics and defective speech and who invented a system of "visible speech," became a Professor of Physiology at Boston in 1873 and invented the first telephone after constructing a rubber model of the tongue and soft parts of the throat. The telephone receiver also was modelled on the structure of the human ear. At the present day, comparative studies of the central nervous system and the operation of automatic computing machines will undoubtedly be to mutual advantage.

* Using the word in the sense of LEWIS MUMFORD: "Technics and Civilization."

† The mention of organization here is significant. Nowadays there is speculation as to whether the fundamental distinction between a living and a dead organism is that the former constantly reduces its entropy (increases organization) at the expense of that of its environment; here loss of entropy is identified with information which the living object is constantly taking in. For example, see SCHRÖDINGER, E.: "What is Life?"

Although reflex response[6] had been observed in the 16th century and the essential function of the spinal cord had been discovered in 1751, the relation between function and structure remained elusive until 1800. In 1861 Broca fixed on the area of the cortex concerned with speech, while Thomas Young, in 1792, had settled that part associated with the eye.[6] The "on-or-off" action of nerve cells was first discovered by Bowditch in 1871, but the neuron theory, that the entire nervous system consists of cells and their outgrowths, has been developed only during the present century.[6] That the intensity of nerve signals depends on the frequency of nervous impulses was observed by Keith Lucas, in 1909—work which was subsequently carried to an extreme elegance with the assistance of modern amplifier and oscillograph technique by Professor Adrian[6] in the late nineteen-twenties.

It is most certain that further studies in physiology will lead to new developments in electrical techniques which in turn will reflect back; new theories and generalities may emerge, leading to greater understanding of machine capabilities. The study of the behaviour of these machines and methods of their control or programming may cast new light on logic, as Turing has suggested. Already there are signs of far-reaching developments.

The philosopher John Locke considered the content of the mind to be made up of ideas, not stored statically like books on a shelf, but by some inner dynamical process becoming associated in groups according to principles of "similarity," "contiguity," or "cause and effect." The word "idea" meant "anything which occupied the mind" or "any object of the understanding."* The first major experimental work, inherently implying a physiological basis for operation of the mind, was not carried out until Pavlov, starting about 1898, studied "patterns of behaviour" of animals. He produced salivation in a dog by showing it certain objects which had already been associated in the dog's mind, by previous repetition, with food—the conditioned reflex. It seems likely that it must have been understood, in a descriptive way, that an action was taking place here similar to that which we now call feedback. Wiener[37] has taken the view that conditioned reflexes enter into the field of cybernetics; that, to give the extremes, the encouragement of actions which lead to pleasure in the body and the inhibition of those which lead to pain may possibly be regarded as feedback actions, suggesting interconnection between different parts of the nervous system. Further, he observed that a conditioned reflex is a "learning mechanism" and that "there is nothing in the nature of the computing machine which forbids it to show conditioned reflexes." Again, the word "machine" here includes its instructions.

Experimental work on what may loosely be termed the

* LOCKE, J.: "Essay Concerning Human Understanding" (1690).

"behaviourism" of machines is at present being conducted, in Britain, by Grey Walter.[41] The inaccessibility and complexity of the central nervous system and of the brain render direct analysis overwhelmingly difficult; the brain may contain 10^{10} nerve cells, whereas the most complex computing machine has only some 10 000 relay units, so how can they possibly be compared? Grey Walter, in his experiments, has started on the simplest scale, building a moving machine having only two "effector" units (linear and rotary motion) and two "receptor" units (by light and touch), and has observed that "the behaviour is quite complex and unpredictable." The use of this toy is justified on the grounds "not that it does anything particularly well, but that it does anything at all with so little."

One principal method of direct analysis of the workings of the brain is the electro-encephalographic method; the wave-like rise and fall of potential on the surface of the brain (the "alpha rhythm"), first observed by Berger[6] in 1928, have been found to possess a complex structure, which varies with the mental state of the subject—asleep or awake—relaxed or concentrating, etc., particularly with relation to the visual sense.[41] Study of these waveforms, in both normal and epileptic cases, is slowly leading to a detection of "pattern," crudely analogous to the decoding of a cypher by search for its structure, as by counting letter frequencies, etc., although the "cypher" here is overwhelmingly complex. In contrast, the relative simplicity of a computing machine calls for infallibility in its various elements, though as the complexity is increased, some redundancy may perhaps be afforded. In a recent paper,[33] Hamming has dealt with the problem of programming a digital machine "which is always liable to make a mistake" using error-detecting and error-correcting codes. This facility requires redundancy in the coding, so that the price to be paid is a slightly lower speed of operation. The possibility of such "self-correction" was first pointed out by Shannon,[24] who, as we have already mentioned, observed that a coding method may always be found which, by introduction of some redundancy, is the optimum for combating noise.

One of the most important and humane applications of the theory of information, which concerns both the biological and the mechanistic fields, is the substitution of one sense for another which has been lost. Early work[4] in this field* includes that of A. M. Bell, who invented a system of "visible speech" for the education of deaf mutes; braille, invented in 1829, involves the learning of a raised-dot code which employs permutations of six positions.† The first machine to convert print directly into sound was the "Optophone," invented by the Frenchman Fournier d'Albe in 1914, while Naumberg in the United States designed a machine (the "Visagraph") for translating a printed book into embossed characters, which unfortunately was slow in operation and costly.

The possibility of a machine which can directly read printed type and convert this into intelligible sounds or "feels" is restricted by the fact that the print may be in different sizes and types; this therefore raises the difficult question of *Gestalt* (perception of form). How do we recognize an object by its shape, irrespective of size and orientation? Or recognize a friend's voice? Or recognize the shape of a ball by its "feel"? This possibility of sense replacement is closely dependent upon the amounts of information with which the various senses operate. At first the eye would seem to involve vastly the greatest amount (being, for one thing, "two-dimensional"); however, it appears that the amount with which the brain has to deal is considerably reduced by the ability of vision to centre the perceived object, by its possessing only a very restricted range of sharp focus and by its accommodation (saturation over areas of uniform colour or brightness, which are "steady-state" and give limited information,* in contrast to edges or boundaries). The possibilities of assisting the sense-deficient sufferer are increasing and perhaps may soon lead to the construction of devices giving greater "rates of information" than those at present in use.

* In the United Kingdom work on the design of reading devices and guiding devices is carried on at St. Dunstan's. In the United States work on sense-substitution devices is co-ordinated under the National Research Council.

† The earliest raised-character systems for the blind were invented in Spain and France in the seventeenth century.

(4) SCIENTIFIC METHOD

In this brief history, we have attempted to trace how the idea of information has existed in early times and has gradually entered into a great variety of sciences, to a certain extent integrating them together. Nowadays the concept of information would seem to be essential to all research workers, and as universal and fundamental as the concepts of energy or entropy. Speaking most generally, every time we make any observation, or perform any "experiment," we are seeking for information; the question thus arises: How much can we know from a particular set of observations or experiments? The modern mathematical work, at which we have glanced, seeks to answer in precise terms this very question which, in its origin, is an epistemological one. But first a word of caution: the term "information" has been used by different authors as having different meanings. In previous Sections we have considered the sense in which it applies to communication engineering ("selective" information) and to analogous fields. The meaning is somewhat different when applied to problems of extraction of "information" from nature, by experiment.

In a classic work, "The Design of Experiments" (1935) Fisher[42] considered these problems, largely from the point of view of the application of correct statistical methods[43] and with the subsequent extraction of valid conclusions. The experimenter always assumes "that it is possible to draw valid inferences from the results of an experimentation; that it is possible to argue from consequences to causes, from observation to hypothesis . . . or, as a logician might put it, from the particular to the general." That is, inductive reasoning is involved, essentially, after an experiment has been made: "inductive inference is the only process . . . by which new knowledge comes into the world." The experimental method essentially implies uncertainty, and the subsequent inductive reasoning raises the thorny question of inverse probability.

In the case of an elementary experiment in, say, mechanics the uncertainty is usually neglected—a body moving in a certain way, under given conditions, will always repeat this motion under the same conditions. But we cannot be so certain for an experiment in, for example, agriculture, where the controlling factors are vastly more complex and less well understood. If we are to understand how to draw the maximum information† from the experiment then, as Fisher stresses, "the nature and degree of the uncertainty [must] be capable of rigorous expression." The information supplied by an experiment may perhaps be thought of as a ratio of *a posteriori* to the *a priori* probabilities (strictly, the logarithm of this ratio).

The importance of the method of inductive reasoning, for arguing from observational facts to theories, seems first to have been recognized (at least in Europe), by the Rev. Thomas Bayes (1763), who considered the following problem: if H_1, H_2, . . . H_i . . . represent various mutually exclusive hypotheses which

* An analogy is a television signal, passed through a high-pass-filter, which gives an outline picture, perfectly recognizable.

† The word "information" is commonly used by statisticians in a special sense. If $P_\theta(x)$ is a distribution function, with some parameter θ (e.g. a mean value) then, by writing $L(x/\theta) = \Sigma \log P_\theta(x)$, where x_1, x_2, x_3 . . . are independent samples, the "information" about θ which these samples give is defined as the mean value of $\partial^2 L/\partial\theta^2$. However, the definition such as is used in communication theory, as we have seen, is $-\Sigma P_\theta(x) \log P_\theta(x)$, where θ is known. This is the "information" given by x and is the mean value of $\log P_\theta(x)$.

can explain an event, what are their relative probabilities of being correct? He assumed certain data to be known before the event happens, and he let E represent some additional data after the event. Then Bayes's theorem gives:

$$p(\mathrm{H}_i/E) = \frac{p(E/\mathrm{H}_i)p(\mathrm{H}_i)}{\sum_i p(E/\mathrm{H}_i)p(\mathrm{H}_i)}$$

where $p(\mathrm{H}_i/E)$ = Probability of H_i after the event.
$p(\mathrm{H}_i)$ = Probability of H_i before the event.
$p(E/\mathrm{H}_i)$ = Probability of obtaining the data E if the hypothesis H_i be assumed.

Although this theorem is generally accepted, its applicability is questioned by some mathematicians on the grounds that the prior probabilities $p(\mathrm{H}_i)$ are, strictly speaking, unknown. Bayes put forward an axiom, in addition: if there are zero prior data, then all hypotheses are to be assumed equally likely, i.e. $p(\mathrm{H}_i) = 1/n$.

However, an exact knowledge of these probabilities $p(\mathrm{H}_i)$ is unimportant, as has been stressed by I. J. Good.[45] This author expresses the above equation logarithmically:

$$\log p(\mathrm{H}_i/E) - \log p(\mathrm{H}_i) = \log p(E/\mathrm{H}_i) - \log \sum_i p(E/\mathrm{H}_i)p(\mathrm{H}_i)$$

the "log Σ" term being a constant. If expected values are taken (i.e. averages) this formula may be shown to give Shannon's expression for the rate of transmission of information, R, through a noisy communication channel, in terms of $H(x)$, the (selective) entropy of the input signal, and $H_y(x)$, the conditional entropy of the input, knowing the output:

$$-H_y(x) + H(x) = R$$

or the alternative form

$$-H_x(y) + H(y) = R$$

The analogy is thus: an observer receives the distorted output signals (the posterior data E) from which the attempts to reconstruct the input signals (the hypotheses H_i), knowing only the language statistics (the prior data).

A recent work by MacKay[46] seeks to obtain a logical quantitative definition of the information given by an experiment or scientific proposition. He observes: "Many scientific concepts in different fields have a logically equivalent structure. One can abstract from them a logical form which is quite general and takes on different particular meanings according to the context. . . . It is suggested that the fundamental abstract scientific concept is 'quantal' in its communicable aspects."

MacKay applies this formal logical view to scientific concepts, observing that they are based on limited data given by sets of observations, and concluding that a scientific statement may be dissected into elementary ("atomic") propositions,[47] each of which may be answered by "true" or "false"; a "unit of information" is then defined as that which induces us to add one elementary proposition to the logical pattern of the scientific statement. MacKay then draws attention to two complementary aspects of information. First, the *a priori* aspect, related to the structure of the experiment; for example, a galvanometer may perhaps have a response time of 0·01 sec; to describe readings at closer intervals is impossible, for the instrument is capable only of giving information in terms of these small, but finite (quantal), intervals. This structural aspect corresponds to the logon concept of Gabor,[15] originally framed to define the response characteristics of a communication channel (see Section 2). Experimentation abounds with similar uncertainties: "each time that a compromise has to be struck, say between the sensitivity and response time of a galvanometer, or the noise level and bandwidth of an amplifier, or the resolving power and aperture of a microscope. . . ." Secondly, the *a posteriori* aspect, related to the "metrical information-content" of the experiment; for example, a galvanometer may be used to record a set of values of a magnitude, each reading representing a certain "amount of metrical information." These recordings being capable of only a certain accuracy, the amount of metrical information obtained may be thought of as a dimensionless measure of precision, or weight of evidence. It is related to Fisher's definition[42] of amount of statistical information, as the reciprocal of the variance of a statistical sample.

Referring to his book "Cybernetics"[37] Wiener observes: "one of the lessons of the book is that any organism is held together by the possession of means for the acquisition, use, retention and transmission of information"; in this short history we have collected some evidence. Such means exist in human society in terms of their spoken language, their Press, telephone system, etc., enabling it to operate as an integrated organism; an insect society is linked by its language of smells, postures or sounds. The human body can operate as a unit inasmuch as it possesses means of communicating information, in the brain and nervous system. Most generally, any physical experiment or proposition is an entity, in that it involves communication of information between the observer and the observed.

Thus the work of communication engineers is capable of the broadest interpretation; the mathematical theory that is developing is likely to assist not only electrical communication (especially telephony) but, perhaps to a far greater extent, other branches of science by providing concepts and theorems giving a quantitative basis to the measurement of information.

(5) ACKNOWLEDGMENT

The author would like to acknowledge, with gratitude, the assistance given by numerous friends, especially Dr. D. Gabor, in compiling this short history.

(6) BIBLIOGRAPHY

Languages and Codes.

(1) DIRINGER, D.: "The Alphabet" (Hutchinson, 1948).
(2) BODMER, F.: "The Loom of Language" (Allen and Unwin, 1944).
(3) PRATT, F.: "Secret and Urgent" (Blue Ribbon Books, 1939) [for letter frequency tables].
(4) Encyclopaedia Britannica (Encyclopaedia Britannica Co., Ltd., 1929, Edition 14).
(5) D'AGAPAYEFF, A.: "Codes and Ciphers" (Oxford University Press, 1939).
(6) PLEDGE, H. T.: "Science Since 1500" (H.M. Stationery Office, 1939). See also references 24, 29 and 33.

Communication Theory.

(7) COLPITTS, E. H., and BLACKWELL, O. B.: "Carrier Current Telephony and Telegraphy," *Transactions of the American I.E.E.*, 1921, **40**, p. 205.
(8) PERO MEADE, S.: "Phase Distortion and Phase Distortion Correction," *Bell System Technical Journal*, 1928, **7**, p. 195.
(9) CARSON, J. R.: "Notes on the Theory of Modulation," *Proceedings of the Institute of Radio Engineers*, 1922, **10**, p. 57.
(10) NYQUIST, H.: "Certain Factors Affecting Telegraph Speed," *Bell System Technical Journal*, 1924, **3**, p. 324.
(11) KÜPFMÜLLER, K.: "Über inschwiEngvorgänge in Wellenfiltern," *Elektrische Nachrichtentechnik*, 1924, **1**, p. 141.
(12) HARTLEY, R. V. L.: "Transmission of Information," *Bell System Technical Journal*, 1928, **7**, p. 535.
(13) VAN DER POL, B.: "Frequency Modulation," *Proceedings of the Institute of Radio Engineers*, 1930, **18**, p. 1194.
(14) ARMSTRONG, E. H.: "A Method of Reducing Disturbances in Radio Signalling by a System of Frequency Modulation," *ibid.*, 1936, **24**, p. 689.

(15) GABOR, D.: "Theory of Communication," *Journal I.E.E.*, 1946, **93**, Part III, p. 429.
(16) VILLE, J. A.: "Théorie et Applications de la Notion de Signal Analytique." *Câbles et Transmission*, 1948, **2**, p. 61.
(17) PAGET, R.: "Human Speech" (Kegan Paul, Trench, Trubner and Co. Ltd., 1930).
(18) FLETCHER, H.: "Speech and Hearing" (Van Nostrand, 1929).
(19) DUDLEY, H., RIESZ, R. R., and WATKINS, S. S. A.: "A Synthetic Speaker," *Journal of the Franklin Institute*, 1939, **227**, p. 739.
(20) DUDLEY, H.: "Remaking Speech," *Journal of the Acoustical Society of America*, 1939, **11**, p. 165.
(21) HALSEY, R. J., and SWAFFIELD, J.: "Analysis-Synthesis Telephony, with Special Reference to the Vocoder," *Journal I.E.E.*, 1948, **95**, Part III, p. 391.
(22) GABOR, D.: "New Possibilities in Speech Transmission," *Journal I.E.E.*, 1947, **94**, Part III, p. 369, and 1948, **95**, Part III, pp. 39 and 412.
(23) TULLER, W. G.: "Theoretical Limits on the Rate of Transmission of Information," *Proceedings of the Institute of Radio Engineers*, 1949, **37**, p. 468.
(24) SHANNON, C. E.: "A Mathematical Theory of Communication," *Bell System Technical Journal*, 1948, **27**, pp. 379 and 623.
(25) COOKE, D., JELONEK, Z., OXFORD, A. J., and FITCH, E.: "Pulse Communication," *Journal I.E.E.*, 1947, **94**, Part IIIA, p. 83.
(26) DELORAINE, E. M.: "Pulse Modulation," *Proceedings of the Institute of Radio Engineers*, 1949, **37**, p. 702.
(27) FANO, R. M.: "The Transmission of Information" (M.I.T. Technical Report No. 65, 1949).

See also:

(*a*) GABOR, D.: "Communication Theory and Physics," *Philosophical Magazine*, 1950, **41**, p. 1161.
(*b*) References 3, 37 and 48.
(*c*) SHANNON, C. E., and WEAVER, W.: "The Mathematical Theory of Communication" (University of Illinois Press, 1949).
(*d*) SHANNON, C. E.: "Communication Theory of Secrecy Systems," *Bell System Technical Journal*, 1949, **28**, p. 656.
(*e*) SHANNON, C. E.: "Recent Developments in Communication Theory," *Electronics*, April 1950, **23**, p. 80.
(*f*) SHANNON, C. E.: "Communication in the Presence of Noise," *Proceedings of the Institute of Radio Engineers*, 1949, **37**, p. 10.
(*g*) GABOR, D.: "Investigations in Communication Theory," *Union Radio Scientifique Internationale*, 1950, **6**.
(*h*) RICE, S. O.: "Communication in the Presence of Noise—Probability of Error for Two Encoding Schemes," *Bell System Technical Journal*, 1950, **29**, p. 60.

(28) "Proceedings of a Symposium on Large-Scale Digital Calculating Machinery," The Annals of the Computation Laboratory of Harvard University, vol. 16 (Oxford University Press).
(29) "Report of a Conference on High Speed Automatic Calculating-Machines," University Mathematical Laboratory (Ministry of Supply, 1950). (See this for extensive bibliography.)
(30) HARTREE, D. R.: "Calculating Instruments and Machines" (University of Illinois Press, 1949).
(31) HARTREE, D. R.: "A Historical Survey of Digital Computing Machines," *Proceedings of the Royal Society*, 1948, **195**, p. 265.
(32) BABBAGE, H. P.: "Babbage's Calculating Engines" (Spon Ltd., 1889). *See also* references 33, 34, 36 and 37.

Aspects of "Mechanized Thinking."

(33) HAMMING, R. W.: "Error Detecting and Error Correcting Codes," *Bell System Technical Journal*, 1950, **29**, p. 147.
(34) MCCULLOCH, W. S., and PITTS, W.: "A Logical Calculus of the Ideas Immanent in Nervous Activity," *Bulletin of Mathematical Biophysics*, 1943, **5**, p. 115.
(35) VON NEUMANN, J., and MORGENSTERN, O.: "Theory of Games" (Princeton, 1947).
(36) SHANNON, C. E.: "Programming a Computer for Playing Chess," *Philosophical Magazine*, 1950, **41**, p. 256.

See also:

(*a*) SHANNON, C. E.: "Communication Theory of Secrecy Systems," *Bell System Technical Journal*, 1949, **28**, p. 656.
(*b*) BERKELEY, E. C.: "Giant Brains, or Machines that Think" (J. Wiley and Sons, 1949).
(*c*) References 6, 29, 37 and 41.

Feedback, Controlled Systems, Cybernetics.

(37) WIENER, N.: "Cybernetics" (J. Wiley, 1948).
(38) BLACK, H. S.: "Stabilized Feedback Amplifiers," *Electrica Engineering*, 1934, **53**, p. 114.
(39) NYQUIST, H.: "Regeneration Theory," *Bell System Technical Journal*, 1932, **11**, p. 126.
(40) BODE, H. W.: "Network Analysis and Feedback Amplifier Design" (Van Nostrand, 1945).
(41) WALTER, G.: "The Functions of Electrical Rhythms in the Brain" (24th Maudsley Lecture), *Journal of Mental Science*, 1950, **96**, p. 1.
See also Reference 4, under "Behaviourism."

Information Theory and Scientific Method.

(42) FISHER, R. A.: "The Design of Experiments" (Oliver and Boyd, 1935).
(43) FISHER, R. A.: "Statistical Methods for Research Workers" (Oliver and Boyd, 1925).
(44) BAYES, T.: "An Essay towards Solving a Problem in the Doctrine of Chances," *Philosophical Transactions of the Royal Society*, 1763, I, iii, p. 370.
(45) GOOD, I. J.: "Probability and the Weighing of Evidence" (Griffen and Co., 1950).
(46) MACKAY, D. M.: "Quantal Aspects of Scientific Information," *Philosophical Magazine*, 1950, **41**, p. 289.
(47) WITTGENSTEIN, L.: "Tractatus Logico-Philosophicus" (Kegan Paul, 1922).
(48) SZILARD, L.: "Über die Entropieverminderung in einem Thermodynamischen System bei Eingriffen intelligenter Wesen," *Zeitschrift für Physik*, 1929, **53**, p. 840.

See also:

GABOR, D.: "Communication Theory and Physics," *Philosophical Magazine*, 1950, **41**, p. 1161.

Noise.

(49) EINSTEIN, A.: "Theory of Brownian Movement" (Methuen, 1926).
(50) SCHOTTKY, W.: "Über spontane Stromschwankungen in verschiedenen Elektrizitätsleitern," *Annalen der Physik*, 1918, **57**, p. 541.
(51) SCHOTTKY, W.: "Zur Berechnung und Beurteilung des Schroteffektes," *Annalen der Physik*, 1922, **68**, p. 157.
(52) JOHNSON, J. B.: "Thermal Agitation of Electricity in Conductors," *Physical Review*, 1928, **32**, p. 97.
(53) NYQUIST, H.: "Thermal Agitation of Electric Charge in Conductors," *ibid.*, 1928, **32**, p. 110.

See also:

(*a*) RICE, S. O.: "Mathematical Analysis of Random Noise," *Bell System Technical Journal*, 1944, **23**, p. 282, and 1945, **24**, p. 46.
(*b*) MACDONALD, D. K. C.: "Some Statistical Properties of Random Noise," *Proceedings of the Cambridge Philosophical Society*, 1949, **45**, p. 368.