THE MATHEMATICS OF COMMUNICATION

An important new theory is based on the statistical character of language. In it the concept of entropy is closely linked with the concept of information

by Warren Weaver

HOW do men communicate, one with another? The spoken word, either direct or by telephone or radio; the written or printed word, transmitted by hand, by post, by telegraph, or in any other way—these are obvious and common forms of communication. But there are many others. A nod or a wink, a drumbeat in the jungle, a gesture pictured on a television screen, the blinking of a signal light, a bit of music that reminds one of an event in the past, puffs of smoke in the desert air, the movements and posturing in a ballet—all of these are means men use to convey ideas.

The word communication, in fact, will be used here in a very broad sense to include all of the procedures by which one mind can affect another. Although the language used will often refer specifically to the communication of speech, practically everything said applies equally to music, to pictures, to a variety of other methods of conveying information.

In communication there seem to be problems at three levels: 1) technical, 2) semantic, and 3) influential.

The technical problems are concerned with the accuracy of transference of information from sender to receiver. They are inherent in all forms of communication, whether by sets of discrete symbols (written speech), or by a varying signal (telephonic or radio transmission of voice or music), or by a varying two-dimensional pattern (television).

The semantic problems are concerned with the interpretation of meaning by the receiver, as compared with the intended meaning of the sender. This is a very deep and involved situation, even

when one deals only with the relatively simple problems of communicating through speech. For example, if Mr. X is suspected not to understand what Mr. Y says, then it is not possible, by having Mr. Y do nothing but talk further with Mr. X, completely to clarify this situation in any finite time. If Mr. Y says "Do you now understand me?" and Mr. X says "Certainly I do," this is not necessarily a certification that understanding has been achieved. It may just be that Mr. X did not understand the question. If this sounds silly, try it again as "Czy pan mnie rozumie?" with the answer "Hai wakkate imasu." In the restricted field of speech communication, the difficulty may be reduced to a tolerable size, but never completely eliminated, by "explanations." They are presumably never more than approximations to the ideas being explained, but are understandable when phrased in language that has previously been made reasonably clear by usage. For example, it does not take long to make the symbol for "yes" in any language understandable.

The problems of influence or effectiveness are concerned with the success with which the meaning conveyed to the receiver leads to the desired conduct on his part. It may seem at first glance undesirably narrow to imply that the purpose of all communication is to influence the conduct of the receiver. But with any reasonably broad definition of conduct, it is clear that communication either affects conduct or is without any discernible and provable effect at all.

One might be inclined to think that the technical problems involve only the

engineering details of good design of a communication system, while the semantic and the effectiveness problems contain most if not all of the philosophical content of the general problem of communication. To see that this is not the case, we must now examine some important recent work in the mathematical theory of communication.

THIS is by no means a wholly new theory. As the mathematician John von Neumann has pointed out, the 19th-century Austrian physicist Ludwig Boltzmann suggested that some concepts of statistical mechanics were applicable to the concept of information. Other scientists, notably Norbert Wiener of the Massachusetts Institute of Technology, have made profound contributions. The work which will be here reported is that of Claude Shannon of the Bell Telephone Laboratories, which was preceded by that of H. Nyquist and R. V. L. Hartley in the same organization. This work applies in the first instance only to the technical problem, but the theory has broader significance. To begin with, meaning and effectiveness are inevitably restricted by the theoretical limits of accuracy in symbol transmission. Even more significant, a theoretical analysis of the technical problem reveals that it overlaps the semantic and the effectiveness problems more than one might suspect.

A communication system is symbolically represented in the drawing on pages 12 and 13. The information source selects a desired message out of a set of possible messages. (As will be shown, this is a particularly important func-

tion.) The transmitter changes this message into a signal which is sent over the communication channel to the receiver.

The receiver is a sort of inverse transmitter, changing the transmitted signal back into a message, and handing this message on to the destination. When I talk to you, my brain is the information source, yours the destination; my vocal system is the transmitter, and your ear with the eighth nerve is the receiver.

In the process of transmitting the signal, it is unfortunately characteristic that certain things not intended by the information source are added to the signal. These unwanted additions may be distortions of sound (in telephony, for example), or static (in radio), or distortions in the shape or shading of a picture (television), or errors in transmission (telegraphy or facsimile). All these changes in the signal may be called noise.

The questions to be studied in a communication system have to do with the amount of information, the capacity of the communication channel, the coding process that may be used to change a message into a signal and the effects of noise.

First off, we have to be clear about the rather strange way in which, in this theory, the word "information" is used; for it has a special sense which, among other things, must not be confused at all with meaning. It is surprising but true that, from the present viewpoint, two messages, one heavily loaded with meaning and the other pure nonsense, can be equivalent as regards information.

In fact, in this new theory the word information relates not so much to what you *do* say, as to what you *could* say. That is, information is a measure of your freedom of choice when you select a message. If you are confronted with a very elementary situation where you have to choose one of two alternative messages, then it is arbitrarily said that the information associated with this situation is unity. The concept of information applies not to the individual messages, as the concept of meaning would, but rather to the situation as a whole, the

unit information indicating that in this situation one has an amount of freedom of choice, in selecting a message, which it is convenient to regard as a standard or unit amount. The two messages between which one must choose in such a selection can be anything one likes. One might be the King James version of the Bible, and the other might be "Yes."

THE remarks thus far relate to artificially simple situations where the information source is free to choose only among several definite messages—like a man picking out one of a set of standard birthday-greeting telegrams. A more natural and more important situation is that in which the information source makes a sequence of choices from some set of elementary symbols, the selected sequence then forming the message. Thus a man may pick out one word after another, these individually selected words then adding up to the message.

Obviously probability plays a major role in the generation of the message, and the choices of the successive symbols depend upon the preceding choices. Thus, if we are concerned with English speech, and if the last symbol chosen is "the," then the probability that the next word will be an article, or a verb form other than a verbal, is very small. After the three words "in the event," the probability for "that" as the next word is fairly high, and for "elephant" as the next word is very low. Similarly, the probability is low for such a sequence of words as "Constantinople fishing nasty pink." Incidentally, it is low, but not zero, for it is perfectly possible to think of a passage in which one sentence closes with "Constantinople fishing," and the next begins with "Nasty pink." (We might observe in passing that the sequence under discussion *has* occurred in a single good English sentence, namely the one second preceding.)

As a matter of fact, Shannon has shown that when letters or words chosen at random are set down in sequences dictated by probability considerations alone, they tend to arrange themselves in
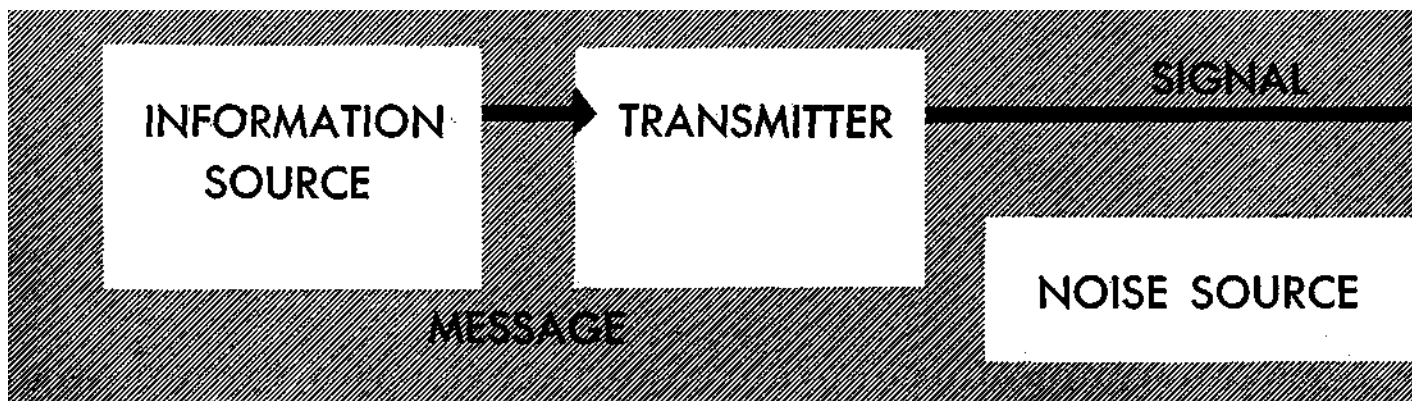
meaningful words and phrases *(see illustration on page 15).*

Now let us return to the idea of information. The quantity which uniquely meets the natural requirements that one sets up for a measure of information turns out to be exactly that which is known in thermodynamics as entropy, or the degree of randomness, or of "shuffledness" if you will, in a situation. It is expressed in terms of the various probabilities involved.

To those who have studied the physical sciences, it is most significant that an entropy-like expression appears in communication theory as a measure of information. The concept of entropy, introduced by the German physicist Rudolf Clausius nearly 100 years ago, closely associated with the name of Boltzmann, and given deep meaning by Willard Gibbs of Yale in his classic work on statistical mechanics, has become so basic and pervasive a concept that Sir Arthur Eddington remarked: "The law that entropy always increases—the second law of thermodynamics—holds, I think, the supreme position among the laws of Nature."

Thus when one meets the concept of entropy in communication theory, he has a right to be rather excited. That information should be measured by entropy is, after all, natural when we remember that information is associated with the amount of freedom of choice we have in constructing messages. Thus one can say of a communication source, just as he would also say of a thermodynamic ensemble: "This situation is highly organized; it is not characterized by a large degree of randomness or of choice—that is to say, the information, or the entropy, is low."

We must keep in mind that in the mathematical theory of communication we are concerned not with the meaning of individual messages but with the whole statistical nature of the information source. Thus one is not surprised that the capacity of a channel of communication is to be described in terms of the amount of information it can

A COMMUNICATION SYSTEM may be reduced to these fundamental elements. In telephony the signal is a varying electric current, and the channel is a wire. In speech the signal is varying sound pressure, and the

transmit, or better, in terms of its ability to transmit what is produced out of a source of a given information.

The transmitter may take a written message and use some code to encipher this message into, say, a sequence of numbers, these numbers then being sent over the channel as the signal. Thus one says, in general, that the function of the transmitter is to encode, and that of the receiver to decode, the message. The theory provides for very sophisticated transmitters and receivers—such, for example, as possess "memories," so that the way they encode a certain symbol of the message depends not only upon this one symbol but also upon previous symbols of the message and the way they have been encoded.

We are now in a position to state the fundamental theorem for a noiseless channel transmitting discrete symbols. This theorem relates to a communication channel which has a capacity of C units per second, accepting signals from an information source of H units per second. The theorem states that by devising proper coding procedures for the transmitter it is possible to transmit symbols over the channel at an average rate which is nearly C/H, but which, no matter how clever the coding, can never be made to exceed C/H.

VIEWED superficially, say in rough analogy to the use of transformers to match impedances in electrical circuits, it seems very natural, although certainly pretty neat, to have this theorem which says that efficient coding is that which matches the statistical characteristics of information source and channel. But when it is examined in detail for any one of the vast array of situations to which this result applies, one realizes how deep and powerful this theory is. How does noise affect information? Information, we must steadily remember, is a measure of one's freedom of choice in selecting a message. The greater this freedom of choice, the greater is the uncertainty that the message actually selected is some particular one. Thus

greater freedom of choice, greater uncertainty and greater information all go hand in hand.

If noise is introduced, then the received message contains certain distortions, certain errors, certain extraneous material, that would certainly lead to increased uncertainty. But if the uncertainty is increased, the information is increased, and this sounds as though the noise were beneficial!

It is true that when there is noise, the received signal is selected out of a more varied set of signals than was intended by the sender. This situation beautifully illustrates the semantic trap into which one can fall if he does not remember that "information" is used here with a special meaning that measures freedom of choice and hence uncertainty as to what choice has been made. Uncertainty that arises by virtue of freedom of choice on the part of the sender is desirable uncertainty. Uncertainty that arises because of errors or because of the influence of noise is undesirable uncertainty. To get the useful information in the received signal we must subtract the spurious portion. This is accomplished, in the theory, by establishing a quantity known as the "equivocation," meaning the amount of ambiguity introduced by noise. One then refines or extends the previous definition of the capacity of a noiseless channel, and states that the capacity of a noisy channel is defined to be equal to the maximum rate at which useful information (*i.e.*, total uncertainty minus noise uncertainty) can be transmitted over the channel.

Now, finally, we can state the great central theorem of this whole communication theory. Suppose a noisy channel of capacity C is accepting information from a source of entropy H, entropy corresponding to the number of possible messages from the source. If the channel capacity C is equal to or larger than H, then by devising appropriate coding systems the output of. the source can be transmitted over the channel with as little error as one pleases. But if the channel capacity C is less than H, the entropy
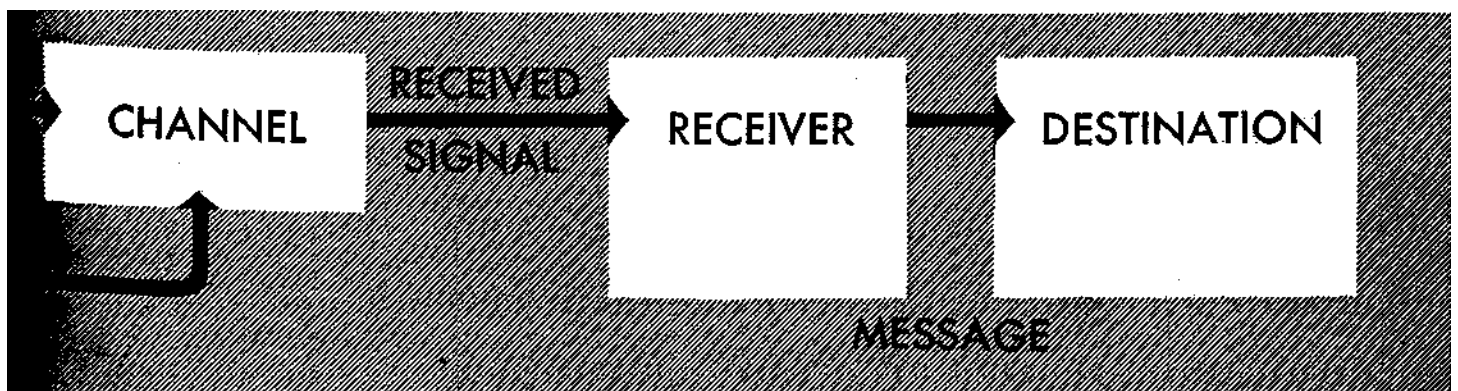
of the source, then it is impossible to devise codes which reduce the error frequency as low as one may please.

However clever one is with the coding process, it will always be true that after the signal is received there remains some undesirable uncertainty about what the message was; and this undesirable uncertainty—this noise or equivocation—will always be equal to or greater than H minus C. But there is always at least one code capable of reducing this undesirable uncertainty down to a value that exceeds H minus C by a small amount.

This powerful theorem gives a precise and almost startlingly simple description of the utmost dependability one can ever obtain from a communication channel which operates in the presence of noise. One must think a long time, and consider many applications, before he fully realizes how powerful and general this amazingly compact theorem really is. One single application can be indicated here, but in order to do so, we must go back for a moment to the idea of the information of a source.

Having calculated the entropy (or the information, or the freedom of choice) of a certain information source, one can compare it to the maximum value this entropy could have, subject only to the condition that the source continue to employ the same symbols. The ratio of the actual to the maximum entropy is called the relative entropy of the source. If the relative entropy of a certain source is, say, eight-tenths, this means roughly that this source is, in its choice of symbols to form a message, about 80 per cent as free as it could possibly be with these same symbols. One minus the relative entropy is called the "redundancy." That is to say, this fraction of the message is unnecessary in the sense that if it were missing the message would still be essentially complete, or at least could be completed.

It is most interesting to note that the redundancy of English is just about 50 per cent. In other words, about half of the letters or words we choose in writing or speaking are under our free choice,

CHANNEL → RECEIVED SIGNAL → RECEIVER → DESTINATION

MESSAGE

channel the air. Frequently things not intended by the information source are impressed on the signal. The static of radio is one example; distortion in telephony is another. All these additions may be called noise.

and about half are really controlled by the statistical structure of the language, although we are not ordinarily aware of it. Incidentally, this is just about the minimum of freedom (or relative entropy) in the choice of letters that one must have to be able to construct satisfactory crossword puzzles. In a language that had only 20 per cent of freedom, or 80 per cent redundancy, it would be impossible to construct crossword puzzles in sufficient complexity and number to make the game popular.

Now since English is about 50 per cent redundant, it would be possible to save about one-half the time of ordinary telegraphy by a proper encoding process, provided one transmitted over a noiseless channel. When there is noise on a channel, however, there is some real advantage in not using a coding process that eliminates all of the redundancy. For the remaining redundancy helps combat the noise. It is the high redundancy of English, for example, that makes it easy to correct errors in spelling that have arisen during transmission.

THE communication systems dealt with so far involve the use of a discrete set of symbols—say letters—only moderately numerous. One might well expect that the theory would become almost indefinitely more complicated when it seeks to deal with continuous messages such as those of the speaking voice, with its continuous variation of pitch and energy. As is often the case, however, a very interesting mathematical theorem comes to the rescue. As a practical matter, one is always interested in a continuous signal which is built up of simple harmonic constituents, not of all frequencies but only of those that lie wholly within a band from zero to, say, W cycles per second. Thus very satisfactory communication can be achieved over a telephone channel that handles frequencies up to about 4,000, although the human voice does contain higher frequencies. With frequencies up to 10,000 or 12,000, high-fidelity radio transmission of symphonic music is possible.

The theorem that helps us is one which states that a continuous signal, T seconds in duration and band-limited in frequency to the range from zero to W, can be completely specified by stating 2TW numbers. This is really a remarkable theorem. Ordinarily a continuous curve can be defined only approximately by a finite number of points. But if the curve is built up out of simple harmonic constituents of a limited number of frequencies, as a complex sound is built up out of a limited number of pure tones, then a finite number of quantities is all that is necessary to define the curve completely.

Thanks partly to this theorem, and partly to the essential nature of the situation, it turns out that the extended

theory of continuous communication is somewhat more difficult and complicated mathematically, but not essentially different from the theory for discrete symbols. Many of the statements for the discrete case require no modification for the continuous case, and others require only minor change.

The mathematical theory of communication is so general that one does not need to say what kinds of symbols are being considered—whether written letters or words, or musical notes, or spoken words, or symphonic music, or pictures. The relationships it reveals apply to all these and to other forms of communication. The theory is so imaginatively motivated that it deals with the real inner core of the communication problem.

One evidence of its generality is that the theory contributes importantly to, and in fact is really the basic theory of, cryptography, which is of course a form of coding. In a similar way, the theory contributes to the problem of translation from one language to another, although the complete story here

clearly requires consideration of meaning, as well as of information. Similarly, the ideas developed in this work connect so closely with the problem of the logical design of computing machines that it is no surprise that Shannon has written a paper on the design of a computer that would be capable of playing a skillful game of chess. And it is of further pertinence to the present contention that his paper closes with the remark that either one must say that such a computer "thinks," or one must substantially modify the conventional implication of the verb "to think."

The theory goes further. Though ostensibly applicable only to problems at the technical level, it is helpful and suggestive at the levels of semantics and effectiveness as well. The formal diagram of a communication system on pages 12 and 13 can, in all likelihood, be extended to include the central issues of meaning and effectiveness.

Thus when one moves to those levels it may prove to be essential to take account of the statistical characteristics of the destination. One can imagine, as an addition to the diagram, another box labeled "Semantic Receiver" interposed between the engineering receiver (which

changes signals to messages) and the destination. This semantic receiver subjects the message to a second decoding, the demand on this one being that it must match the statistical semantic characteristics of the message to the statistical semantic capacities of the totality of receivers, or of that subset of receivers which constitutes the audience one wishes to affect.

Similarly one can imagine another box in the diagram which, inserted between the information source and the transmitter, would be labeled "Semantic Noise" (not to be confused with "engineering noise"). This would represent distortions of meaning introduced by the information source, such as a speaker, which are not intentional but nevertheless affect the destination, or listener. And the problem of semantic decoding must take this semantic noise into account. It is also possible to think of a treatment or adjustment of the original message that would make the sum of message meaning plus semantic noise equal to the desired total message meaning at the destination.

ANOTHER way in which the theory can be helpful in improving communication is suggested by the fact that error and confusion arise and fidelity decreases when, no matter how good the coding, one tries to crowd too much over a channel. A general theory at all levels will surely have to take into account not only the capacity of the channel but also (even the words are right!) the capacity of the audience. If you overcrowd the capacity of the audience, it is probably true, by direct analogy, that you do not fill the audience up and then waste only the remainder by spilling. More likely, and again by direct analogy, you force a general error and confusion.

The concept of information developed in this theory at first seems disappointing and bizarre—disappointing because it has nothing to do with meaning, and bizarre because it deals not with a single message but rather with the statistical character of a whole ensemble of messages, bizarre also because in these statistical terms the words information and uncertainty find themselves partners.

But we have seen upon further examination of the theory that this analysis has so penetratingly cleared the air that one is now perhaps for the first time ready for a real theory of meaning. An engineering communication theory is just like a very proper and discreet girl at the telegraph office accepting your telegram. She pays no attention to the meaning, whether it be sad or joyous or embarrassing. But she must be prepared to deal intelligently with all messages that come to her desk. This idea that a communication system ought to try to deal with all possible messages, and that the intelligent way to try is to

base design on the statistical character of the source, is surely not without significance for communication in general. Language must be designed, or developed, with a view to the totality of things that man may wish to say; but not being able to accomplish everything, it should do as well as possible as often as possible. That is to say, it too should deal with its task statistically.

This study reveals facts about the statistical structure of the English language, as an example, which must seem significant to students of every phase of language and communication. It suggests, as a particularly promising lead, the application of probability theory to semantic studies. Especially pertinent is the powerful body of probability theory dealing with what mathematicians call the Markoff processes, whereby past events influence present probabilities, since this theory is specifically adapted to handle one of the most significant but difficult aspects of meaning, namely the influence of context. One has the vague feeling that information and meaning may prove to be something like a pair of canonically conjugate variables in quantum theory, that is, that information and meaning may be subject to some joint restriction that compels the sacrifice of one if you insist on having much of the other.

Or perhaps meaning may be shown to be analogous to one of the quantities on which the entropy of a thermodynamic ensemble depends. Here Eddington has another apt comment:

"Suppose that we were asked to arrange the following in two categories—*distance, mass, electric force, entropy, beauty, melody.*

"I think there are the strongest grounds for placing entropy alongside beauty and melody, and not with the first three. Entropy is only found when the parts are viewed in association, and it is by viewing or hearing the parts in association that beauty and melody are discerned. All three are features of arrangement. It is a pregnant thought that one of these three associates should be able to figure as a commonplace quantity of science. The reason why this stranger can pass itself off among the aborigines of the physical world is that it is able to speak their language, *viz.,* the language of arithmetic."

One feels sure that Eddington would have been willing to include the word meaning along with beauty and melody; and one suspects he would have been thrilled to see, in this theory, that entropy not only speaks the language of arithmetic; it also speaks the language of language.

Warren Weaver is Director for the Natural Sciences in the Rockefeller Foundation.

1. Zero-order approximation
XFOML RXKHRJFFJUJ ZLPWCFWKCYJ FFJEYVKCQSGXYD QPAAMKBZAACIBZLHJQD

2. First-order approximation
OCRO HLI RGWR NMIELWIS EU LL NBNESEBYA TH EEI ALHENHTTPA OOBTTVA NAH BRL

3. Second-order approximation
ON IE ANTSOUTINYS ARE T INCTORE ST BE S DEAMY ACHIN D ILONASIVE TUCOOWE AT TEASONARE FUSO TIZIN ANDY TOBE SEACE CTISBE

4. Third-order approximation
IN NO IST LAT WHEY CRATICT FROURE BIRS GROCID PONDENOME OF DEMONSTURES OF THE REPTAGIN IS REGOACTIONA OF CRE

5. First-Order Word Approximation
REPRESENTING AND SPEEDILY IS AN GOOD APT OR COME CAN DIFFERENT NATURAL HERE HE THE A IN CAME THE TO OF TO EXPERT GRAY COME TO FURNISHES THE LINE MESSAGE HAD BE THESE.

6. Second-Order Word Approximation
THE HEAD AND IN FRONTAL ATTACK ON AN ENGLISH WRITER THAT THE CHARACTER OF THIS POINT IS THEREFORE ANOTHER METHOD FOR THE LETTERS THAT THE TIME OF WHO EVER TOLD THE PROBLEM FOR AN UNEXPECTED

ARTIFICIAL LANGUAGE results when letters or words are set down statistically. 1. Twenty-six letters and one space are chosen at random. 2. Letters are chosen according to their frequency in English. 3. Letters are chosen according to the frequency with which they follow other letters. 4. Letters are chosen according to frequency with which they follow two other letters. Remaining examples do the same with words instead of letters.