



THE REDUNDANCY OF ENGLISH

CLAUDE E. SHANNON

*Bell Laboratories,
Murray Hill, N. J.*

THE CHIEF subject I should like to discuss is a recently developed method of estimating the amount of redundancy in printed English. Before doing so, I wish to review briefly what we mean by redundancy. In communication engineering we regard information perhaps a little differently than some of the rest of you do. In particular, we are not at all interested in semantics or the meaning implications of information. Information for the communication engineer is something he transmits from one point to another as it is given to him, and it may not have any meaning at all. It might, for example, be a random sequence of digits, or it might be information for a guided missile or a television signal.

Carrying this idea along, we can idealize a communication system, from our point of view, as a series of boxes, as in Figure 22, of which I want to talk mainly about the first two. The first box is the information source. It is the thing which produces the messages to be transmitted. For communication work we abstract all properties of the messages except the statistical properties which turn out to be very important. The communication engineer can visualize his job as the transmission of the particular messages chosen by the information source to be sent to the receiving point. What the message means is of no importance to him; the thing that does have importance is the set of statistics with which it was chosen, the probabilities of various messages. In general, we are usually interested in messages that consist of a sequence of discrete symbols or symbols that at least can be reduced to that form by suitable approximation.

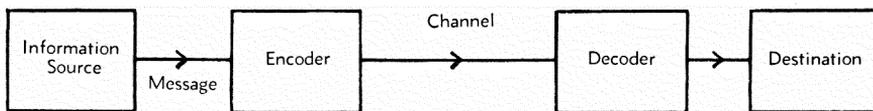


FIGURE 22

[124] | The second box is a coding device which translates the message into a form suitable for transmission to the receiving point, and the third box has the function of decoding it into its original form. Those two boxes are very important, because it is there that the communication engineer can make a saving by the choice of an efficient code. During the last few years a theory has been developed to solve the problem of finding efficient codes for various types of communication systems.

The redundancy is related to the extent to which it is possible to compress the language. I think I can explain that simply. A telegraph company uses commercial codes consisting of a few letters or numbers for common words and phrases. By translating the message into these codes you get an average compression. The encoded message is shorter, on the average, than the original. Although this is not the best way to compress, it is a start in the right direction. The redundancy is the measure of the extent to which it is possible to compress if the best possible code is used. It is assumed that you stay in the same alphabet, translating English into a twenty-six-letter alphabet. The amount that you shorten it, expressed as a percentage, is then the redundancy. If it is possible, by proper encoding, to reduce the length of English text 40 per cent, English then is 40 per cent redundant. The redundancy can be calculated in terms of probab-

ities associated with the language; the probabilities of the different letters, pairs of letters; probabilities of words, pairs of words; and so on. The formula for this calculation is related to the formula of entropy, as no doubt has appeared in these meetings before. Actually, to perform this calculation is quite a task. I was interested in calculating the redundancy of printed English. I started in by calculating it from the entropy formulas. What is actually done is to obtain the redundancy of artificial languages which are approximations to English. I pointed out that we represent an information source as a statistical process. In order to see what is involved in that representation, I constructed some approximations to English in which the statistics of English are introduced by easy stages. The following are examples of these approximations:

1. XFOML RXXHRJFFJUI ZLPWCFWKCYJ FFJEYVKCQSGHYD
2. OCRO HLI RGWR NMIELWIS EU LL NBNSEBYA TH EEL
3. ON IE ANTSOUTINYS ARE T INCTORE ST BE S DEAMY ACHIN D ILONASIVE TUCCOOWE AT TEASONARE FUSO | [125]
4. IN NO IST LAT WHEY CRATICT FROURE BIRS GROCID PONDENOME OF DEMONSTURES OF THE RETAGIN IS REGIACIONA OF CRE.
5. 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. 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.

In the first approximation we don't introduce any statistics at all. The letters are chosen completely at random. The only property of English used lies in the fact that the letters are the twenty-seven letters of English, counting the space as an additional letter. Of course this produces merely a meaningless sequence of letters. The next step (2) is to introduce the probabilities of various letters. This is constructed essentially by putting all the letters of the alphabet in a hat, with more *E*'s than *Z*'s in proportion to their relative frequency, and then drawing out letters at random. If you introduce probabilities for pairs of letters, you get something like Approximation 3. It looks a bit more like English, since the vowel consonant alternation is beginning to appear. In Approximation 3 we begin to have a few words produced from the statistical process. Approximation 4 introduces the statistics of trigrams, that is, triplets of letters, and is again somewhat closer to English. Approximation 5 is based on choosing words according to their probabilities in normal English, and Approximation 6 introduces the transition probabilities between pairs of words. It is evident that Approximation 6 is quite close to normal English. The text makes sense over rather long stretches. These samples show that it is perhaps reasonable to represent English text as a time series produced by an involved stochastic process.

The redundancies of the languages 1 to 5 have been calculated. The first sample (1) is a random sequence and has zero redundancy. The second, involving letter frequencies only, has a redundancy of 15 per cent. This language could be compressed 15 per cent by the use of suitable coding schemes. The next approximation (3), based on digram[!] structure, gives a redundancy of 29 per cent. Approximation 4, based on trigram structure, gives a redundancy of 36 per cent. These are all the tables that are available on the basis of letter frequencies. Although cryptographers | have tabulated frequencies of letters, digrams, and trigrams, so far as I know no one has obtained a complete table of quadrugram frequencies. However, there are tables of word frequencies in English which are quite extensive, and it is possible to calculate from them the amount of redundancy (Approximation 5) due to unequal probability of words. This [126]

came out to be 54 per cent, making a few incidental approximations; the tables were not complete and it was necessary to extrapolate them.

In this case the language is treated as though each word were a letter in a more elaborate language, and the redundancy is computed by the same formula. To compare with the other figures, it is reduced to the letter basis by dividing by the average number of letters in a word.

PITTS: Do other languages have the same frequency or the same degree of redundancy?

SHANNON: I have not calculated them; but according to the work of Zipf,¹ who has calculated the frequency of words in various languages, for a large number of them the falling off of frequency against rank order of the word, plotted on log-log coordinates, is essentially a straight line. The probability of the n th-most probable word is essentially a constant over n for quite a large range of n :

$$pn = \frac{k}{n}.$$

PITTS: But presumably the constants are different for different languages?

SHANNON: That I don't know. They are not vastly different in the examples which Zipf gave in his books, but the difference in constants would make some difference in the calculated redundancy. The equation $pn = \frac{k}{n}$ cannot hold indefinitely. It sums to infinity. If you go out to infinite words, it must tail off. That was one of the approximations involved here which makes the figure somewhat uncertain.

TEUBER: Your probabilities are based on predicting from one letter to the next, or one word to the next?

MCCULLOCH: No, upon the word.

[127] SHANNON: In the particular case (4) this refers to a language in which words are chosen independently of each other, but each | has the probability that it has in English. »The« has a probability of .07 in the English language. So we have seven of them in a hat of 100.

TEUBER: As you led up to it you said, first of all, you take the probability that any one letter will occur in that particular system, and then, given that letter, and the next one that occurs, you predict the one that follows immediately after that?

SHANNON: Yes.

TEUBER: Do you go further than that and say that any one letter will follow the one you took beforehand, regardless of what letter was in between?

SHANNON: That is true in the calculation from 3. This is based on probabilities of groups of three letters. Number 4 goes on a new tack and starts over with the word as a new unit. The words are independently chosen. It is a better approximation to English than 3, since each group of letters forms a word, but the words don't hang together in sentences. At this point there seemed to be no way to go any further, because no one had tabulated any frequencies for pairs of words. Of course, such a table would be impractically large because of the enormous number of possible pairs of words. However, the thought occurs that every one of us who speaks a language has implicitly an enormous statistical knowledge of the structure of the language if we could only get it out. That is to say, we know what words follow other words; we know the standard clichés, grammar, and syntax of English. If it were possible, for example, to translate that information into some form adapted to numerical analysis,

1 ZIPF, G. K.: *Human Behavior and the Principle of Least Effort. An Introduction to Human Ecology.* Cambridge, Mass.: Addison Wesley Press, Inc. (1949)

we could get a further estimate of the redundancy. It turns out that there is a way to do this.

The method is based on a prediction experiment. What you do is shown by the following typical experiment. Take a sample of English text:

(A) T H E R E I S N O R E V E R S E O N A M O T O R C Y C L E ...
 (B) 1 1 1 5 1 1 2 1 1 2 1 1 1 5 1 1 7 1 1 1 2 1 3 2 1 2 2 7 1 1 1 1 4 1 1 1 1 1 ...

and it goes on from there. Take a subject who does not know what the text is, and ask him to guess the text letter by letter. As soon as he arrives at the correct letter, he is told so and goes on to guess the next letter of the text. In this case he guessed *T* as the first letter, which was right. We put down 1 because he guessed right on the first guess. He was also right on the first guess for *H* and *E*. For the letter *R* he guessed four wrong letters and | finally got it on the fifth. In general the numbers in the lower row represent the guesses at which he finally obtained the right answer. The figures may look surprising, starting off with three right guesses. It is actually very reasonable, since the most common initial letter is the *T*. The most common letter to follow that is *H* and the most common trigram is »the«. In this particular sample, which went on for a total of 102 letters, the score obtained by this subject was as follows:

[128]

Right on guess	1	2	3	4	5	> 5
Occurrences	79	8	3	2	2	8
						Total, 102

The number of ones out of 102 letters is 79. He was right 79 times on first guess, that is, about 78 per cent of the time. He was right on the second guess eight times, three times on the third guess, and on four and five, twice each. He required more than five guesses only eight times out of 102. This is clearly a good score. It is more or less typical for literary English. The scores vary. With newspaper English scores are poorer, mainly because of the large number of proper names, which are rather unpredictable.

I should like to point out that in a certain sense we can consider the second line of such an experiment to be a »translation« of the first line into a new »language.« The second line contains the same information as the first line. We have operated on the first line with a device, our predicting subject, and obtained the second. Now the crucial question is: Could we, knowing the second line, obtain the first by a suitable operation? I would say that this property is actually the central characteristic of a translation: it is possible to go from *A* to *B* and from *B* back to *A*, and nothing is lost either way. In the case at hand, it is possible to go from *B* back to *A*, at least conceptually, if we have an identical twin of the person who made the first record. When I say »identical,« I mean a mathematically identical twin who will respond in exactly the same way in any given situation. Having the same information, he will make the same response. If we have available the line *B*, we ask this twin what the first letter is. He guesses it correctly because he guesses as the original subject guessed; and we know that is correct because it is the first guess he made. This process is continued, working through the text. At the letter *R*, | for example, we ask him to guess five times, and at the fifth guess we say that that is right.

[129]

Of course, we don't have available mathematically identical twins, but we do have mathematically identical computing machines. If you could mechanize a reasonably good predicting process in a computing machine, you could mechanize it a second time and have the second machine perform precisely the same prediction. It would be

possible to construct a communication system based on this principle in which you sent as the signal the second line B from one point to another. This would be calculated by the computing machine which is doing the predicting. At the second point the second computer recovers the original text.

From the data in the second line B , it is possible to set upper and lower bounds for the entropy of English. There is a theorem on stochastic processes that the redundancy of a translation of a language is identical with that of the original, if it is a reversible translating process going from the first to the second. Consequently, an estimation of the redundancy of the line B gives an estimate of the redundancy of the original text, that is, of English. Line B is much easier to estimate than line A , since the probabilities are more concentrated. The symbol 1 has a very high individual probability, and the symbols from 6 to 27 have very small probabilities.

WIENER: This is very interesting to me. In my prediction theory I start from a series with a correlation and I build a series which is equivalent; each is the past linear function of the past of the other. In addition, however, in my new series the choices are completely independent, whereas in my earlier series the series are partially dependent. In fact, the way I built the linear prediction theory was by the reduction of my dependent choices to independent choices. My method has some parallelisms to this. Excuse me for interrupting.

SHANNON: That is perfectly cogent, I think. I was going to say in connection with this that the successive symbols in line B are not yet statistically independent, but they are much closer to independence than they are in the original text. It is approaching the sort of thing we are talking about, an uncorrelated time series.

WIENER: Yes.

SHANNON: To continue the analysis: it is possible to estimate an upper bound and lower bound for the amount of redundancy for time series B . The two bounds for redundancy are based upon the frequencies of these various numbers in the second line. If we carry out this experiment with a very long sample of text, we will obtain a good estimate of the frequencies of ones, two, threes, and so forth. Let these be q_1 , q_2 , q_3 , respectively.

The lower bound for the redundancy is given by

$$R \geq \log_2 27 + \sum_{i=1}^{27} q_i \log_2 q_i.$$

This follows from the fact that if the line B were an uncorrelated time series, its redundancy would be that given by the right-hand member. Any correlation present increases the redundancy. Since the redundancy of line B equals that of line A , the result follows.

There is also an upper bound of a sort, but it is not quite as secure. It is given by

$$\log_2 27 - \sum_{i=1}^{27} (q_{i+1} - q_i) \log_2 i \geq R.$$

This is a provable upper bound if we assume the prediction to be ideal; that is, that the subject guesses first the most probable next letter, second the next most probable, and so forth. Actually, of course, human subjects will not guess quite this accurately, although in the actual experiments I believe they were close to ideal prediction. They were supplied with various statistical tables concerning English to aid in the prediction. All in all it is probably safe to use the upper bound given above.

SAVAGE: Does the theory say that if a subject is an ideal predictor, his pattern of integers will be uncorrelated?

PITTS: Will he exhibit the maximum correlation?

SAVAGE: No, I suspect not. Suppose, for example, that the subject guesses the first letter immediately; it takes him two guesses to think up the second letter and five to think up the third. Does that information imply anything about how many guesses the next letter is going to take him?

SHANNON: Well, normally we don't start from the beginning of a sentence. Most of the actual experiments were done by giving a subject N letters of text, asking him to guess the next letter. This | was done 100 times with each value of N for 16 different values of N . [131]

I am not sure I have really answered your question. At any point the subject knows the text up to this point, or at least he knows N letters of it. That will influence his next guess, because he will try to continue the test.

SAVAGE: Will it influence how quickly he will? That will certainly influence the probability that his next guess will be right; there will be certain circumstances on which the guess is sure to be right.

BIGELOW: Like in your previous example. After the fellow got the letter C in the word »motorcycle« he knew at that moment the whole word and got therefore every letter after one successful guess.

SAVAGE: However, if you look simply at a sequence of integers generated by such a performance, do they have implications in probability for the next integer to be generated?

SHANNON: I think they might if you were clever enough.

SAVAGE: I see. There is no general theoretical reason why they might not?

SHANNON: No, there isn't. There are cases in which you rather expect a series of right ones. By an analysis of this, you could say quite a bit, probability-wise, about the next numbers.

WIENER: There should be ways, I think, of sorting this out. That would be more absolute, like codings. I think there would be a reduction of the choices to completely independent choices. That could be worked out.

SHANNON: Well, there certainly is in principle, if you allow the encoding of long sections of text into long sections of uncoded text, but as a practical matter I don't know how to get at it.

WIENER: To do it is difficult, because you would need more complete tables.

PITTS: Very extensive.

WIENER: Very extensive tables.

SHANNON: When I evaluated the upper and lower bounds for redundancy from the experiment, the following results were obtained:

N	8	10	15	100
Lower bound for D_n	74%	75%	75%	93%
Upper bound for D_n	50%	57%	60%	72%

D_n is the redundancy owing to the statistical structure of English | extending over N letters of text. [132]

SAVAGE: Is this only some lower bound or the greatest lower bound?

SHANNON: The only provable bound in case it were an ideal prediction but as I say, I suspect it still does bound the actual value because I think these people were close

enough to ideal prediction too, so that the other things involved in this lower bound in this discrepancy more than compensate.

SAVAGE: What mathematical properties of the new language do you use to introduce these bounds?

SHANNON: The lower bound is rather trivial. It follows from the fact that the least redundancy possible with a certain set of letter frequencies would occur if they were independent. The upper bound is more difficult to prove. It involves showing how the ideal predictor would predict. An ideal predictor lines up the conditional probabilities in the order of decreasing magnitudes. You line those up and find the worst set of those that could occur. This will produce the highest possible value of entropy. Then you calculate this value in terms of the q_i .

Of course, these values are not only subject to the conditions given but also to statistical fluctuation, since the samples involve only 100 trials.

PITTS: That is, even where words are considered as being made of letters, where they are not treated ideographically?

SHANNON: Everything is reduced to a letter basis.

PITTS: It might be quite different if one simply carried the parallel through with respect to ideographic words much as you could imagine carrying out the same sort of translation and the same sort of estimation?

STROUD: Didn't some joker do that? He exposed one-, two- and three-word samples and asked the subjects to guess the next word. He did it the other way around. He didn't ask them to predict the text. He simply reported the texts that were created by this method, placed certain other restrictions on it as to subject matter, and then gave you the samples as merely samples.

MCCULLOCH: Miller. His name was Miller?

LICKLIDER: He did do that sort of experiment.

PITTS: Have you carried through exactly the same procedure for words as a whole that you have for letters, and do you have bounds?

SHANNON: In the first place, difficulties arise concerning the number of words there are in the language.

[133] PITTS: You can ask a man for the first guess, second guess. |

SHANNON: Prediction of words as a whole is certainly possible, but it would take a long time to obtain a reasonable sample.

PITTS: There would be so many more words and letters that it would take much longer to guess the right one.

LICKLIDER: There is a problem that bothers me. I am sure you have taken care of it, but I don't see quite how, in having the prediction made, compact prose is always generated.

SHANNON: This is taken from standard text chosen at random out of a book chosen at random. It represents what a literary man would write. When he chooses an improbable letter, the subject usually must guess many times before he gets it right.

BROSIN: I don't know the Zipf evidence. It is actually astonishing how the different languages, the different types of prose, literary, and so forth, follow the graph, the straight line.

MCCULLOCH: Isn't the Zipf evidence for the newspaper the same as that for James Joyce's work?

BROSIN: Yes.

STROUD: Both for James Joyce's work and for newspaper print. Joyce had a fantastic vocabulary, something like 15,000 words.

MCCULLOCH: It seems the law holds approximately for Joyce as for newspapers.

HUTCHINSON: The number of moths of a given species plotted against the number of species containing that number of individuals, in the catch of a moth trap, or any similar statistics, behaves in very much the same way [FISHER, R. A., CORBET, A. S., and WILLIAMS, C. B.: The relationship between the number of species and the number of individuals in a random sample of an animal population. *J. Animal Ecology*, 12, 42 (1943).] The constant can be used as an index of diversity of the population.

LICKLIDER: My criticism was against the other method of doing it, not the one you used at all; it was against having the sample generated by presenting N items to a person and having him give you the $N+1$ st.

SHANNON: I don't think that a statistical sample of statistical letters is true.

MCCULLOCH: May I ask whether it would be possible for you to look up, or to recognize among printed letters, queues to phonemes, and see what the frequency for phonemes, more constricted than the sequence of letters, really is?

STROUD: That is a very important question because of the multiplicity of representation.

MCCULLOCH: Yes.

SHANNON: I think it would be quite easy to do most of these | experiments with phonemes in place of letters. They go through rather rapidly after you get into the swing of it. I don't know that choosing phonemes would come as naturally, though, to the experimental subject as choosing letters. [134]

MCCULLOCH: I was thinking that one might link it with the intelligibility of speech of sentences as opposed to that of words, of that of words as opposed to that of nonsense syllables, and so forth. If you knew the sequence of phonemes represented by these letters, you might be able to link it to the intelligibility or to the increased intelligibility.

LICKLIDER: Miller has completed some work on the learnability of speech that ties in with this analysis. The difficulty of learning a sample of synthetic prose is roughly proportional to its information content.

TEUBER: By this logic, would baby talk be more predictable or less predictable than the talk of an adult?

SHANNON: I think more predictable, if you are familiar with the baby.

TEUBER: Do you know that linguists claim that baby talk is highly similar, that is, similarly patterned from baby to baby in different language systems, in terms of its phoneme structure? Wouldn't it be easier anyway to go from one phoneme to the next in predicting speech?

STROUD: In some of the latest orations from my youngest there is only one pair of phonemes.

GERARD: If you had taken simplified spelling, would you have decreased redundancy?

SHANNON: Yes, that is right.

LICKLIDER: Conversely, if you used the International Phonetic Alphabet to do the phonetics, you would find high redundancy.

SHANNON: There is one other experiment that we performed in connection with this work which did not have much bearing on anything but which proved to be rather interesting. We asked a couple of subjects to predict in reverse, that is, to start in at the end of the sentence and to guess backward letter by letter. It turned out that the scores were almost as good as in prediction in the forward direction. The problem was much more difficult, however, from the psychological point of view. The subject was really tired out after he had worked through a sentence in reverse, whereas in going forward it is quite easy.

GERARD: As a matter of curiosity, how did the guesses go? Looking at those actual guess numbers, the first one is perfectly clear. Then you get to the *I*, and I suppose the first guess was an *I*. | The subject decided it was not *A*, so he guessed *I*. Now how did [135] he get the *N* on the second guess? That I don't see.

STROUD: »There is a,« »there is the,« »there is no,« are some of the most common progressions.

BIGELOW: Assertionary denials.

STROUD: Either followed by the negative or by the article.

SHANNON: I think *A*, *T*, and *N* would be the first three guesses; since »the« sounds a little strange in that construction, he probably guessed *A* and then *N*.

SAVAGE: Then he has *O* for sure. He got space without the *T*.

STROUD: No, most probably, and not the most probable.

PITTS: Or *L* less.

SAVAGE: No, reverse *EHT* there.

HUTCHINSON: Isn't five rather high?

SHANNON: The *R* took 15 guesses, and one for the *E*.

STROUD: How does that appear compared to the probability of *R* as a first letter of a new word?

SHANNON: I think this particular subject was not using tables and probably guessed his initial letters improperly. In later experiments, the ones I got these estimates with, the people were supplied with all the tables we had on the statistics of English and used them in any way in which they saw fit to aid their guessing.

SAVAGE: How did you pick a book? At random?

SHANNON: I just walked over to the shelf and chose one.

SAVAGE: I would not call that random, would you?

GERARD: Unless you were blindfolded.

SAVAGE: There is the danger that the book might be about engineering.

BAVELAS: The book would be.

KLÜVER: I wonder whether Sievers' *Schallanalyse* is in some way related to the problems discussed here. The *Schallanalyse* or sound analysis of Sievers was concerned with »translating« auditory sequences as found in human speech into motor sequences [SIEVERS, E.: *Ziele und Wege der Schallanalyse*. Heidelberg: Carl Winter's Universitätsbuchhandlung (1924), pp. 65-111. Cf. also Vol. 35 of the *K. Sächs. Gesellschaft der Wiss., philol.-histor. Klasse*.] Sievers called attention to the fact that speech, no matter whether we are dealing with poetry or prose, tends to be accompanied by certain movements, postures, and tonus regulations. He held that the same is true for any written text, since all texts represent potential speech. He published numerous »curves« (Becking curves, time curves, and signal curves) supposedly involved in [136] form of auditory reproduction. These motor »curves,« let us say, | a circle or the figure eight, may be produced by movements of the hands and arms while reciting, for instance, a poem. Sievers also made use of optical signals, such as brass figures lying before the speaker. These figures were supposed to influence auditory reproduction if viewed by the speaker while talking. However, the chief contention of the *Schallanalyse* was that out of all possible motor curves it is always only one particular curve that »goes with« a particular auditory sequence. It was said that the voice becomes inhibited if any attempt is made to produce curves that do not go with the poem or text in question.

WERNER: The *Schallanalyse* is an extremely interesting but somewhat controversial method. Before Sievers, Rutz worked out a system of tones of melodic rhythm patterns.

KLÜVER: As I remember the story, Sievers really never succeeded in teaching his methods to others. The exception seems to have been a student who flunked all examinations, although he was extraordinarily gifted for *Schallanalyse* in Sievers' sense.

PITTS: I suspect philologists denied the acceptability of the method except for reasons in which discontinuity of the authorship came in.

SAVAGE: G. U. Yule's efforts were bona fide and not mysterious. They were primarily based on English, and as I recall, in one or two cases on the Latin language, which he knows and is therefore able to deal with. He examines the frequency of occurrence of various obvious sorts of things, computes the average length of word, makes other like calculations, and finally makes a judgment based on standard statistical principles and his own very extensive statistical experience of whether the two works in question do or do not have style similar enough to be attributed to a common author.

KLÜVER: As far as Sievers' work is concerned, it rests on certain correlations between motor-kinesthetic and auditory phenomena. Sievers apparently was very gifted in »expressing« auditory sequences in kinesthetic patterns. He always insisted that only one type of motor curve was adequate whenever he recited or read a certain text. If he sensed that certain movements representing a particular motor curve were no longer adequate, his voice simply gave out until he replaced the old curve by one that »fitted.« He was able in such a way to assign a particular motor curve or a sequence of different motor curves to a given auditory reproduction.

WERNER: That's about it as I recall. Looking at a poetic line, Sievers transforms it into a kinesthetic pattern; if various other | lines are produced by the same author they will also fit the pattern. If then he comes to a part which does not go with the kinesthetic pattern, he infers that a foreign element has been introduced. [137]

PITTS: What sort of kinesthetic patterns?

WERNER: Though I do not remember very clearly, he contended that there were a restricted number of patterns. These patterns could be objectified by visual symbolic representations; a triangular pattern was one of them; a pendulum pattern was another.

PITTS: A particular phenomenon would be translated into the triangular?

WERNER: Not quite. The visual representation is an aid for *Schallanalyse*. A poem has a certain rhythm which is repeated and which is represented by a certain visual signal, such as a triangle.

PITTS: How do you carry out the translation? It is not obvious how you would carry out the translation of the poem into the kinesthetic pattern.

KLÜVER: It looks as if the »translation« of a poem into specific kinesthetic patterns has remained a secret of Sievers.

PITTS: It is not an objective method?

WERNER: No. Statistics have been applied, though, and they contend that these statistics bear them out.

KLÜVER: It may be argued that Sievers' »curves« were more than subjective motor-kinesthetic patterns. There seems to be some evidence that they represented objective indicators. If I recall correctly, Sievers was tested by some experimental psychologists who showed him, for example, a text he had not previously seen. This text had been written by two different authors. He read the text while at the same time producing the movements and curves that, according to him, »went with« the text, but then suddenly insisted that he could not go on reading. His voice gave out and he could continue only after having discovered the right kind of motor curve. The point thus

located by Sievers was the point that actually separated the texts of two authors as previously determined by other philological methods.

PITTS: Nobody ever ascertained the rule by which he obtained this?

[138] KLÜVER: William Stern used to say that he would give a doctor's degree in psychology to anybody who could throw light on the psychological mechanisms involved in Sievers' performances. Incidentally, I am sure that Sievers' own »explanations« do not suffice or are wrong. C. K. Ogden of Basic English fame told me just before the war that he had used the methods of *Schallanalyse* | to settle the old question of how the Romans really did pronounce Latin. In fact, I spent part of my last night in Europe making a gramophone record of Ogden's recitation of a Latin poem. And I took this application of Sievers' principles, that is, the gramophone record, to New York.

PITTS: Does this agree, on philological evidence, with the most powerful of the schools with respect to the actual pronunciation?

KLÜVER: I understand that the pronunciation of Latin as practiced by the Romans remains a controversial matter.

WERNER: I may add that Sievers was one of the outstanding German philologists who wrote classic works on old German grammar and texts. *Schallanalyse* was obviously a hobby for him at first which he later extended and included among his methods for textual analysis. Later he felt the *Schallanalyse* was much more satisfactory than the traditional philological methods.

PITTS: Philologists have explored more eccentric methods of argumentation than most people have.

STROUD: Every individual shows a reluctance to repeat himself in the use of a single word. Sometimes there is reluctance to place verbs too close to the noun, and there are other idiosyncrasies of sentence structure which are to some degree characteristic. I wonder if such landmarks could not be guessed at reasonably well with a fair length of text.

PITTS: The only trouble is that the serious questions are practically never concerned with two texts of any considerable length, whether by the same author or not. Usually both texts are very short.

BROSIN: I don't know how relevant it is, but the Masserman-Bolk chromatic analyses of parts of speech in the Murray Thematic Apperception Tests with relevance to psychiatric reactions are an effort in this direction. I don't know how seriously you would consider this.

PITTS: You might decide the Shakespeare controversy in this fashion but probably very few others.

BROSIN: I don't know how useful it is, but obsessed people use words in certain orders and quantities. Surely the scope of word patterns and, let us say, the kinesthetic rhythms of one type of schizophrenic, say a hebephrenic, will surely have distinctive patterns. Whether these are sufficient to be computed for absolute values, I don't know.

[139] MEAD: Milton Eric[k]son has a whole series of texts taken down from different diagnostic psychiatric types. In these the clear formal properties of the language and the types of balance that recur can be distinguished. It would be such patterns, I think, | that you would have to deal with. You would have to use a good many abstractions, such as balance and types of repetition and inversion, in such an analysis.

FRANK: Eliot Chapple, who has done the same thing at the Massachusetts General Hospital, has developed a machine for recording the pattern of speech.

MEAD: But his records contain the interaction with another person within it. That would be the difference here, would it not?

FRANK: Probably.

MEAD: If you introduce two people into the picture, then Eliot Chapple's chronograph gives a very diagnostic speech pattern when one individual's responses to another's are analyzed.

MCCULLOCH: May I ask one clumsy question? I don't see quite how to put it yet. Most of the things you have been going at, Mr. Shannon, have been on a small scale. Can you work out anything that would handle affairs as large as your ordinary grammatical units, phrases, clauses, sentences, and so on?

SHANNON: I think the 100-letter approximation is beginning to bite into phrases. The person with 100 letters has a fraction of a sentence, perhaps a full sentence, to work on. He makes use of all that information, perhaps finds a key word 100 letters previous. More in the spirit of what you are asking, I feel there is a point at which this statistical approach is going to break down. It is very questionable to me that the very long-range structure of language can be represented by a statistical process, even that there is any meaning of speaking of the probabilities of sequences which are so rare that they never have occurred. Certainly the frequency concept of probability begins to weaken at some point. Also, when you are considering very long-range structure there are questions of whether the stochastic process is stationary or not. The process may not have the stationary properties that are implied by most of this analysis.

TEUBER: Isn't it true – for long passages at least – that it is easier to make predictions after some preliminary acquaintance with the idiosyncrasies of a particular author, or even of a particular group of people with whom you have been in contact?

PITTS: Probably after the passages exceed a certain length, and if you are concerned with the question of deciding whether or not some particular person or anybody at all could possibly have said that. Your chances would be much better if you were to analyze on the basis of whether or not it is a kind of notion or idea that could have been expressed at that time, rather than on the basis of the statistics of a series of words in it. | [140]

TEUBER: That could be based on a single experience. Just say the word »Gestalt« in this group, and you can just about predict what will happen and who will say what. You make your prediction in terms of past experience, but one previous exposure to that interchange of points of view will suffice. This is of course irrelevant to what Mr. Shannon is trying to do. For him, the important thing is to get rid of all idiosyncrasies.

PITTS: There you would find probabilities of notions of the man rather than of the symbols used to denote them. I don't see why this should be necessarily impossible a priori.

BATESON: I was thinking about the extraordinary difficulty of reconstructing stenotypic transcript. If a word like »ratio« becomes »ration,« as it cannot do on the stenotype but may on the typewriter, it may be very hard to get back to the original alternative. Meaningful distortion has crept in where there has been a deformation.

PITTS: You can correct it astoundingly well, probably much better than you think on the basis of statistics about series of words. If you tried correcting it solely on the basis of the statistics of a possible series of words, and extended it, say, to ten-word sequences, even you would probably do it much worse.

MCCULLOCH: If you can find anybody who knows the speech of the person in question and can imitate it, and if you make him read aloud the stenotypic notes, you can reconstruct the speech again and again.

BATESON: Have you ever tried it with a long text in which 10 per cent of the letters, say, are deleted at random? What percentage of letters do you have to delete before unintelligibility sets in?

SHANNON: I tried a few experiments with a 27-letter alphabet, again with the space a letter, and found that you can reconstruct, say, about 70 per cent of the text when about 50 per cent of the letters are deleted at random. The trouble is that the random letters, the deletions, pile up in certain places. If you delete every other letter, you can do quite well with almost 50 per cent missing. You can delete all the vowels in a passage and have no difficulty in reconstructing it. Only very infrequently will you miss a word. Vowels constitute 40 per cent of the letters. You can also delete the spaces, which is another 20 per cent.

FRANK: Is it correct to infer from your remarks that the tables you build up are predicated upon knowledge of the uniformities of English language on the part of your ideal informants, who have then interpreted the text with these deletions according to
[141] | the structure of English language and ideas? Is that the way it operates in your procedures?

SHANNON: I would say that all of this work simplifies the complexity of English a great deal in that we say there is one kind of English and there is one set of statistics for English. Actually, English is really many different languages, each with different statistics. If a person knew who wrote the text he is predicting and was familiar with the author's habit patterns, he could certainly do better than if he were just taking it as blind English.

HUTCHINSON: Isn't it like the library catalogue which provides you an English book full of information by giving a single number that can be provided from Washington or New York?

SHANNON: Provided you wanted to send that book; but suppose I write a book. That book is not in the catalogue, yet I want to send that; but you don't have any number for it.

HUTCHINSON: That is true.

SHANNON: You should have numbers for all books and the one that might be written by this information source.

MCCULLOCH: What is known about the frequency of the various parts of speech in the ordinary grammatical sense? To what extent can you guess what is coming next?

SHANNON: I have not seen any tables on that, but it is possible to guess surprisingly well in this kind of experiment. When subjects obtain scores like 60 per cent right on the first guess, they must have known what word was coming.

LICKLIDER: There is obviously a close relation between this and what Rudolph Flesch says in his little book, *The Art of Plain Talk*. The latter is on a less precise and more intuitive level, but I think it could be translated into terms of information and redundancy. Flesch says that if you really want to communicate with someone, you have to make your speech (and more especially your writing) even more redundant than it naturally is. You have to repeat two or three times, then say the same thing in different words. This gives us, perhaps, a very dismal outlook for verbal communication. I suppose it does not make the outlook for conferences like this very hopeful. They have to be long.

MARQUIS: There might be another implication of the same thing: for example, the best communication occurs if you say what the listener expects you to say.

LICKLIDER: There should be an optimum degree of correlation between the talker and the listener. If the correlation is zero, the listener has no expectations and understands
[142] | nothing. If the correlation is unity, he doesn't need to listen. |

MEAD: If you take a form of verbal communication like the one Harold Laswell uses, involving a vocabulary from about six different disciplines simultaneously in places where such usage is not expected, most people, while they may know all six vocabu-

laries, will find following him exceedingly difficult. He has also adopted the device in his usual communications of interlarding endless redundant clauses, such as, »It is unnecessary to specify.« One can, of course, learn to recognize these interpolations.

PITTS: You mean this is deliberate?

MEAD: He does it all the time. I tried the experiment once of asking him to cut them out. After I had asked him to cut them out, I found it about impossible to listen.

PITTS: On the book page you can find the significant part of the sentence because it is padded with these things.

MEAD: If you are used to the style, you knock these phrases out, and the pauses give you a chance to adjust to the shifts in vocabulary. If you get *him* to knock them out when you are trying to listen, then you realize how difficult it is to make these shifts all the time. This is the opposite of the point you were making about the condition in which a person is saying exactly what is expected. As you move away from expectancy, even in type of vocabulary, you need this padding which is not necessarily saying the thing over but just permitting a slight shift from one frame to another.

PITTS: He could probably speak more slowly and do nearly as well.

GERARD: Friends rarely finish sentences with each other, since one knows what is coming and picks it up.

MCCULLOCH: Mr. Shannon, will you say a word about the assurance of getting a message across as it affects the redundancy? Have you any actual evidence on that?

SHANNON: No, I don't have any numerical evidence. Do you mean into a human being or through a noisy communication channel?

MCCULLOCH: Getting information through noise.

PITTS: Then the redundancy reduces automatically. By the way you calculate the redundancy it reduces if you have to put it through noise. I mean, the way you calculate information means that in case you put it through noise then repetition becomes less redundant than it would otherwise be.

MCCULLOCH: That is right. Have you any quantitative work on it at all?

SHANNON: We have some work, for example, done by Rice at the Bell Laboratories on White thermal type of noise and various | methods of encoding for transmission through it. Rice has developed formulae which show roughly how much delay is required in the encoding operation to introduce redundancy properly so as to overcome the effects of noise and enable correction of errors. It appears that a rather large delay is usually required if you wish to approach the ideal encoding with, say, one error in a hundred transmitted symbols. [143]

VON FOERSTER: Is there knowledge of redundancy of different languages, or only of English?

SHANNON: The only work I know of is in English.

VON FOERSTER: What do you expect for the other languages, the same figure or different ones?

SHANNON: The Zipf curves suggest that the redundancy for other Indo-European languages may be of the same order as that of English.

VON FOERSTER: This has certainly something to do with the closely related grammars among the different Indo-European languages. The grammar of a language is probably more or less an expression of its structure. With respect to the redundancy of a language, it is certainly true that the more freedom of choice the grammar leaves the less redundant the language becomes. On the other hand, a language with an extremely highly developed grammar would be a language with a big redundancy. As for instance, mathematics or symbolic logic are languages with 100 per cent redundancy. I see here two tendencies operating against each other to develop the optimum of a lan-

guage. The one tendency tries to decrease the redundancy in order to transmit as much information as possible; the other one tends to increase the redundancy by establishing a highly structural order within the language. That means we have to expect certain values for the redundancy in an optimized language. Perhaps the numbers of letters – or perhaps the number of phonemes – play a very important role in optimizing a language. I would like to remind at this point that the first attempts in writing are usually solved by an idiographic system which does not know any letter but has a symbol for every word. I am thinking of the old Maya texts, the hieroglyphics of the Egyptians or the Sumerian tables of the first period. During the development of writing it takes some considerable amount of time – or an accident – to recognize that a language can be split in smaller units than words, e.g., syllables or letters.

[144] I have the feeling that there is a feedback between writing and speaking. After writing freed itself from the archaic rigidity of | idiographs and became more fluent and versatile due to the elastic letter system, I would expect a certain adaptation of the possibility of making words and making sentences. In other words, I believe there should be a connection between the redundancy of a language with respect to its word and with respects to its single letters. These figures must give a certain knowledge about the structure of a language.

SHANNON: I am sure it does. I believe there are a large number of compromises in constructing a language, one of them being that the language ought to be pronounceable on reading it. This requires certain constraints about how the vowels and consonants separate each other. This already implies a certain amount of redundancy. I believe that there are many other desirable features that you require of a language which force the redundancy to be fairly high in order to satisfy these requirements.

WERNER: Of course there is also the phonological aspect of language, which has been studied so extensively by phonologists and which enters into the problem of redundancy. Every language uses a limited number of diacritical phonic signs; certain combinations of sounds exist in one language but do not exist in another. The phonic unit *kn* does not occur in English but occurs in German, and so forth. Because languages differ in phonological structure, there is a qualitative and quantitative difference in redundancy between languages.

PITTS: In simple ways. You know, for example, that one language has common words on the average much longer than another. Then it will be more redundant in that sense. Thus, translation from English into Latin increases the length of the book. As a matter of fact, that is very clear in the Bible, in which the English translation is the shortest, except for the Chinese one. That, I think, almost certainly implies that the redundancy of English is less, on a syllabic basis, than that of most other languages.

If you are interested, however, in the capacity of language, in your sense you probably mean to a greater extent the redundancy on the basis of entire words taken as units.

If you were to take the article »the« and add four symbols to it, in one sense you may increase the redundancy. Another language that has articles may suffer more redundancy than a language that has none at all.

[145] TEUBER: Mightn't that possibly give us a definition of a »primitive« language? Some of the American Indian languages seem to us to »overqualify« all the time. They have to keep going for a | long time before they get where we would get in a few words. At the same time, would there be greater predictability of phonemes, on the basis of their sequence, that is, to be able to predict which phoneme follows which? I raised the question about predictability in baby talk for that same reason. If you have a very redundant language, you could argue that in that language one makes the same noises for a longer time, and one could soon tell what noise tends to follow what noise, so

that there is greater predictability. I don't know whether that is so, and whether it would lead to a precise definition of »primitiveness« of a language.

MEAD: I don't think you could do that. With a great knowledge of phonetic structure you might be able to construct an ideal, such as working on your ideal redundancy point, an ideal language of an ideal degree of primitiveness, especially using the child as a model. But with actual primitive languages you cannot do anything of the sort. You get extraordinary variations; therefore you would not be able to make any kind of sequence.

MCCULLOCH: Some are extremely redundant and some are not?

MEAD: Yes. Some are not.

VON FOERSTER: This situation might be illustrated by the problem of translation, where the same thought has to be expressed in different languages. It very often happens that a certain phrase, a poem or a thought can be beautifully expressed in one language and sounds impossible in another one.

For instance, the beauty and clearness of Aristotle in its own language becomes in German tedious and clumsy, whereas in English Aristotle regains his sharpness and conciseness. On the other hand, to read Goethe's *Faust* in French is almost ridiculous. But these two examples don't say anything against German in the first case or French in the second one. It seems to me that different languages are able to express things with different results. Isn't it so, that a language is a symbol for an intellectual world like any other human expression, architecture for instance? I am considering the Greek temples and the Gothic domes. Each of them is perfect in itself, serving the same purpose – and how different they are. I think word-redundancy is not a sufficient key to judge the value of a language.

KLÜVER: Dr. Von Foerster, it has been said that a German can understand Kant's *Critique of Pure Reason* only if he reads it in English. Obviously, an English translation makes it necessary to transform the long Kantian sentences into simple, short sentences.

VON FOERSTER: Yes, certainly. According to Pitts' statement about the Bible, Kant should have written his *Critique* in Chinese. |

[146]

SAVAGE: It should be emphasized again that Shannon has talked about redundancy at the presemantical level. Redundancy of printed speech really refers exactly neither to the spoken language nor to the very difficult problem of semantical redundancy. I think it would require special and difficult experimentation to measure the semantical redundancy of the language, though the experiment last reported by Shannon does bear on the subject to some extent in that the guesser knows English and is utilizing that knowledge in his guesses. But still, to isolate the semantical redundancy, to separate it from the phonemes would probably be very difficult. Yet it is the thing you are all talking about now. It is the thing that refers perhaps to what you would call the spirit or essence of the language.

STROUD: I wonder if you would care to consider such highly artificial languages, if you like, as symbolic logic, or to consider mathematical notations as being perhaps among the least redundant symbols that we have. My reason for bringing this up is really very simple. I planned at one time to use a sample of quite readable text by simply reading off the rules by which you extract the roots of the cubic equation. This can be read in English; and yet only those of you who are mathematicians would have recognized it for what it was, I am sure, and could have reproduced the equation from the instructions. A majority of us might have recognized that it was the process, but we could not have reproduced the process; some of us might not even have recognized it for what it was. I had thought of using this as an example that was perfectly pro-

nounceable at the good standard elocution rate of one phoneme per tenth second, though insufficiently redundant to be absorbed by the listener. I hoped thereby to indicate that in this case that amount of information which I was trying to convey to the subject in question, lacking in many cases the full knowledge of the statistics involved, was too much for our abilities to handle.

PITTS: No, I don't think you can say in general that the artificial languages of symbolic logic express a minimum of redundancy. I suspect that anyone would agree with that who had read Bell's book on mathematical methods in biology in which he explains the elementary principles, or the *principia mathematica*.

SAVAGE: I would say they are highly redundant. They risk nothing.

PITTS: Exactly.

[147] SAVAGE: Freedom from redundancy is a desideratum of mathematical and, to a somewhat lesser extent, of logical notation. It is not, however, foremost among the desiderata. To appreciate this it is important to recall that redundancy means here the sort of statistical redundancy which has been defined, that is, the frequent use of long or otherwise awkward expressions. Thus, redundancy includes more than what we call redundancy in ordinary usage; namely, actually saying the same thing twice or using expressions that might well be dropped. The business of trying to make the frequent symbols simple, though more consciously pursued in mathematics than in ordinary speech, has been going on for a much shorter time.

STROUD: I shall have difficulty in believing this until I have seen fair samples of fairly long expositions in symbolic logic subject to this sort of analysis of the diagram, trigram type in which the knowledge of the person reading the material of the rules whereby these were established was ruled out. If you include the rules in a complete knowledge of them, hypothetically, at least, unless the person has been deliberately redundant, you should be able to reproduce the entire text once you got one-half of an equation of it.

LICKLIDER: I think there is a possibility of studying the semantic aspects of language profitably in simple languages that grow up in special situations. I have some friends in the Human Resources Research Laboratories in Washington. They are very much interested in what they call Airplanese. This is the language of the control towers that get the airplanes down at airports, that control takeoffs, and so forth. They have kept track of the words and messages that pass between the towers and the planes. I remember this preliminary result: at an early stage they had 10,000 tokens, 10,000 words recorded; of these, 5,000 were either numerals or place names, such as Washington, Bolling, Andrews. The remaining 5,000 tokens included just 400 types. By any way of figuring, that is quite redundant. The most frequently occurring actions are landings and takeoffs, requests for wind directions, and so forth. The physical problems determine the messages. This suggests, since the semantic aspect of language has to do with the relation between the signs and their referents, that the signs are autocorrelated and redundant in large part because the world we live in is. When we get into routines, we find stretches of language that are extremely redundant. They are redundant because they are trying to describe actual situations that are redundant.

PITTS: Every natural law expresses a great redundancy in nature.

[148] LICKLIDER: That may be pursued farther in that direction. I think it is important to bring the referents into the picture. For example, the rate at which information flows over a Ground Control or Approach link is dependent upon the type of plane that is being controlled. When a jet plane comes in, the rate of direction goes up. When a training plane comes in, the rate goes down. The idea, then, is that you may be able to

trace much of the structure of language to the structure of the actual situations in which language is used.

MCCULLOCH: What comes in, then, is mere padding or actual repetitions of directions.

LICKLIDER: In the G.C.A. example.

MCCULLOCH: It sounds like an American radio advertisement. Is that the kind?

LICKLIDER: The G.C.A. operator never goes off the air for any length of time during a talk-down. He talks, keeps talking; the idea is not to let the pilot get the notion that the radio link is dead. The pilot would be much distressed, flying blind, coming into the ground with nobody talking. If the operator is talking to a slow plane, he has to keep on saying the same thing over and over. It is extremely redundant. If the plane is fast, the situation changes rapidly enough for the operator to make every second or third phrase a new instruction.

SAVAGE: Does it suffice to maintain some inanimate acoustical contact or is actual talking preferable?

LICKLIDER: I have heard that the pilots don't like the operators to be overchatty. They would just as soon not have that.

PITTS: He might get a lick in for security to make perfectly sure he has been perfectly understood, since he has to –

WIENER: The direction might be repeated half a dozen times.

GERARD: Such as, »Keep coming.«

LICKLIDER: I think the G.C.A. operator does not talk about the wind direction. It is more like: »You are on course, you are on course, you are doing fine, you are on course; now two degrees left, two degrees left; the heading is so-and-so, the heading is so-and-so; hold that heading, hold that heading.« It goes on at that rate; the number of words said is purposely high. Actually, the whole problem of coding in military communication comes into our picture. There are two rival philosophies. One says that you want to set up a restricted set of messages and enforce the restriction, hoping that an emergency does not arise for which it will not be adequate –

STROUD: And »hope« is the right word.

LICKLIDER: The other says that you want people to use their heads about the phrasing of messages. I don't think there has been a decision between the two yet. |

[149]

BAVELAS: What may seem redundant from one point of view may not be redundant at all in terms of what one might call second-order information. I am reminded of a little study, done several years ago, in which college students were asked to tell what kinds of things people like themselves could do to help the war effort but which would very probably invite criticism from their neighbors. Also, they were asked to tell what people like themselves could do to help the war effort which would evoke praise from their neighbors. On the basis of that information, two leaflets were prepared. Both leaflets were purported to be statements by a public official urging college students to help the war effort and suggesting what they might do. One of them suggested only those activities which the interviewed group said would be criticized; the other leaflet suggested only those things which the interviewed group said would be praised. These leaflets were distributed to entirely new samples of college students. When these students were asked what they thought about them, it was clear that they attributed to the »public official« favorable or unfavorable characteristics, depending upon the extent to which the suggested behaviors were of the one kind or the other. In other words, the text not only bore information with respect to the activities in which the college student might take part but also information about the author.

BATESON: Communication about relationship between you and the other person.

PITTS: You could code them and send them all at the beginning very quickly.

SAVAGE: I say you could not.

PITTS: Not communication of information. That is the important point, even of second-order information.

FRANK: That is the difference between machine and man.

PITTS: If I could say you are alive in half an hour and continue to say it –

SAVAGE: When you marry, tell your wife on the wedding morning, »I love you, darling; I love you eternally, no matter what I say or do from now on. I love you eternally, remember.« Then you tell her again. If you never refer to the subject again, see what happens.

PITTS: That shows exactly that it is not proper to speak of it as a communication of second-order information in the same sense in which primary information is communicated, because the assumption is reduced to absurdity by exactly that remark.

[150] SAVAGE: She can't decode that. |

STROUD: This concept of redundancy abstracts the information from the date. If for any reason the date of origin is part of information, there isn't any such thing as a redundant signal, as near as I can make out.

SAVAGE: I sympathize with Mr. Pitts. It is something of a tour de force to call these things information. The obvious differences, which are brought out by these examples as reassurance, emotional contact, and so forth, though there is communication in them, are connections between individuals which are not just transmission of statements of fact.

BAVELAS: I think they are. If we could agree to define as information anything which changes probabilities or reduces uncertainties, such examples of changes in emotional security could be seen quite easily in this light. A change in emotional security could be defined as a change in the individual's subjective probabilities that he is or is not a certain kind of person or that he is or is not »loved.«

MCCULLOCH: Verily, verily I say unto you.

BAVELAS: What I am saying is this: I want to avoid technical language, but if a man walks in and pats you on the back or winks at you across the table, this may be information in the very same sense that any other message is information if it reduces your uncertainty as to your present state among a possible number of states, your position in the group.

FRANK: Can't you refine that by saying that this may consist of information if you want to call it such, but that primarily the communicated sign or signal or gesture keeps the recipient of your communication tuned to the meaning of the information that you want to convey? In a sense it is getting the person ready with the right expectation so that the information you want to convey will be accepted, received, and interpreted in the terms you want it to be.

BAVELAS: That is one function, but there is also the function of informing his present state in, for instance, a group relationship.

FRANK: Yes.

BATESON: In the straight telegraphic situation you have a whole series of conventional signs for, »Please repeat; I received the last word; talk louder,« and so on. Now those are very simplified analogues of the thing you are talking about, aren't they?

BAVELAS: Of other things too.

BATESON: I mean giving commands.

[151] BAVELAS: Let us suppose that I am a stranger in this group. As I sit here, a gentleman whom I don't know but whom I assume is | one of the leading members of the group

smiles across the table at me. When I make a comment, he nods in agreement. Now it seems to me that his behavior bears information in the sense that it makes it possible for me to select certain ones from all the possible relations that I might conceive myself as having to this group.

PITTS: Certainly there is always an informative component in these emotional communications. I think our point was mostly that that is not all; very often it is not the essential part of it, nor is it the reason why people spend so much time at it as they do.

FRANK: There is another aspect of the situation.

PITTS: It is the birthday telegram with us.

FRANK: Often it is a reaffirmation, saying, »I mean,« and then repeating in other words because the facial expression and response that you are watching on the other individual indicates that you must make another attempt at communication because you see you are not getting through. Therefore you restate it, reaffirm it, put it in another way while watching that individual, until you believe that you perhaps have made a communication. So I think that is another aspect, face-to-face conversation.

GERARD: It is about time to tell a story that has been on my mind for a while. Speaking about the stranger in the group makes it relevant. A guest spoke to a group that was intimately knit. One who preceded him said a few words and ended with, »72.« Everybody roared. Another person said a few words, then, »29,« and everybody roared.

The guest asked, »What is this number business?«

His neighbor said, »We have many jokes but we have told them so often that now we just use a number to tell a joke.«

The guest thought he'd try it, and after a few words said, »63.« The response was feeble. »What's the matter, isn't this a joke?«

»Oh, yes, that is one of our very best jokes, but you did not tell it well.«

PITTS: Of course, in a certain sense there is much more information in this kind of communication than one would suppose. If a man tells his wife every morning for thirty years that he loves her, the declaration may convey very little information if she is in no real doubt. If he omits it, however, the omission gives her considerable information. You must consider the possibility of communication not occurring if it is not a message at all.

STROUD: Something that bothers me is the way in which we make wisecracks about this thing we call noise. For many very practical considerations it seems often very wise to consider the whole message as a message, and the noise, if it has any meaning at all, merely as that portion of the message about which you do not wish to be informed. This has some very practical applications in that you have two ears: there is nothing to prevent you from hearing my voice as it comes to you from several paths, each of which repeats substantially the same pressure pattern, but with a slight time delay. From one point of view, if you say merely that it is sufficient for you to hear one version of this time-pressure pattern and not the other time-pressure patterns, then hearing with two ears is a highly redundant affair. If, however, you begin to find out that because people have two ears they hear these same pressure patterns in several different versions with slight modifications of amplitude and phase relations, you discover that they are able to take the range and bearing of the speaker with a considerable degree of accuracy. The information, then, was not redundant at all. It is perfectly true that they did not use the information to verify that I said »is« when I said »is,« but they did use it for the very useful purpose of informing themselves about where the joker was who was talking to them.

LICKLIDER: And what sort of room he is in.

STROUD: And where he is in the room, such questions as these.

SAVAGE: None the less, it is highly redundant to answer these questions a thousand times over.

STROUD: They are highly redundant in the sense that your past experience tells you that you will not move without having some other source of information, or that the room will not change without some other comparison of information; but since in these matters one requires a very high degree of security, they are not over redundant if you want to attain high orders of probability in not making a mistake.

[153] There is another case: Suppose I want to take an electrocardiogram on a man who is sick in bed in his own home. I try it, but what I get out on the trace is practically pure 60-cycle hum. This was not the information that I came for. I already knew that the local power company put out at that rate. However, it is a very practical stunt to be quite well informed about this, and, incidentally, independently informed. This I do by putting on an aerial that picks up this same information about the power system. In a certain sense this is rather stupid, but I can use this as a minus message in the same sense that Dr. Licklider meant when he was talking about transmitting the noise on one channel and the mixed message and noise on the other, then using the pure noise as a minus on the mixed signal, and coming out with nothing but the pure information. I use the information about the same self | power supply as a minus message to the confusion I get from the patient and conclude with a fairly respectable electrocardiogram. So I often suspect that it might be considerably more profitable, at least in many contexts, not to be too quick to define what is the signal and what is the noise. I know that I am of ten very amply rewarded by stopping to find out what the stuff I call noise in a message is. Sometimes it turns out to be more valuable than I had implied when I said it was noise, when I said that the »not« signal I was interested in was noise.

SAVAGE: Your mistake was in defining noise as part of the signal you were interested in. The distortion of speech that comes to you from the walls of the room is not in itself noise. It is distortion, that is to say, recoding.

STROUD: How, for example? I am sorry.

LICKLIDER: Then you'd never know about it if there were really some 60-cycle stuff in the cardiogram.

STROUD: That is the difficulty. Perhaps the fellow with past experience with cardiographs would expect me to pick out the 60-cycle.

I study the presence of a message. This message is defined on an a priori basis in the presence of white noise. If this is realized in the experimental situation, do you know what my white-noise generators tell me? They tell me about random movement, charges in the gas tube, in the magnetic field. I must admit that in the vast majority of cases this information is nothing for me to be particularly concerned about, but I would remind you that this restriction upon what is a signal in a message, and what is a »not« signal is often highly arbitrary. Very frequently it is not stated, much to the detriment of the discussion that follows, even in your treatment of it.

[154] FRANK: May we go on from what you said earlier about noise to consider the situation that man and his mammalian ancestors grew up in a world of communication that consisted chiefly of noise? Everything was going on at once and producing a myriad of to-whom-it-may-concern messages. Man gradually evolved the ability to select and orient himself, as you say, with the two ears. Thus we get a conception of language as a codification of events which we have learned to pay attention to and to interpret in specific ways. Each culture has picked out what it pays attention to and what it will ignore among the innumerable messages from events. We think and speak in terms of selective awareness patterned by our cultural traditions, the *eidōs* and *ethos* of each culture. Out of the noise patterns it is not *the* message that is received | but rather *a* message that our selective awareness, readiness, or expectation filters out and gives mean-

ing to as a communication which may or may not be authentic or valid. In other words, we may be said to create the communication, each in his own image.

STROUD: I have a suspicion that all of our sensory inputs are capable of supplying us with a tremendous number of items of information, by the billion in the case of the eye, in very short periods of time. The limiting factor in the thing is the computer, and into that we are only able to insert a relatively small number.

PITTS: We discussed at length last time how the optic nerve is greatly reduced as compared to the retina. As a matter of fact there is an immense loss of information.

STROUD: We speak of these as losses, but I suspect they are not simple losses. They are well planned, programmed.

PITTS: It is done by the aid of natural laws generally.

STROUD: The fact that we are able to attend only to a limited number of bits per second in our abstraction leads us to quite unconscious and implicit statements as to what the signal and the »not« signal are in any given set of sensory inputs. It also leads to a similar arbitrariness in handling a set of information which we get over an instrumental, extrasense system, which most of these gadgets are.

PITTS: I don't think, in the sense of the exact theory of information which we have been discussing this evening, that there is any vagueness or ambiguity at all in what is meant by the message, in what is meant by the noise. It is perfectly true that in loose everyday use of the term there may be.

SAVAGE: There is arbitrariness.

STROUD: Yes.

SAVAGE: Consider a telephone. It seems to me that from the engineering point of view it is an arbitrary question whether the telephone is to communicate, say, the meaning of English spoken into it, or the affect – or whatever the psychological word is – of the English as well, whether it is to communicate not only the English spoken into it but also other sounds which happen to be occurring in the room. Practical considerations might dictate the discussion of a telephone from any of these points of view. It is therefore these practical considerations which would determine what is to be considered signal and what, noise.

I think Shannon ought to say a word now. His ideas have been discussed for half an hour, therefore I suppose he is interested in | expressing his own views about the discussion. [155]

SHANNON: There are several comments I want to make on the last point. I never have any trouble distinguishing signals from noise because I say, as a mathematician, that this is signal and that is noise. But there are, it seems to me, ambiguities that come in at the psychological level. If a person receives something over a telephone, part of which is useful to him and part of which is not, and you want to call the useful part the signal, that is hardly a mathematical problem. It involves too many psychological elements. There are very common cases in which there is a great mass of information going together. One part is information for *A* and another part is information for *B*. The information for *A* is noise for *B*, and conversely. In fact, this is the case in a radio system in which one person is listening to WOR and another to WNBC. You can also have situations in which there is joint information, something of this general nature. You can have a device with information going in at one point *A*, part of it coming out at *B*, and part of it coming out at *C*. It is possible to set up the device in such a way that it is not possible to transmit any information whatever from *A* to *B* alone or to *C* alone, but if two of these people get together and combine their information then you can transmit information from *A* to the pair of them. This shows that information is not always additive. In this case the information at *C* is essentially a key for the infor-

mation at *B*, and vice versa. Neither is sufficient by itself. If two of them get together, they can combine and find out exactly what the input was.

STROUD: The ideal minus message case. I didn't mean to cast any shadows or doubts. I merely wish to make the point that Mr. Shannon is perfectly justified in being as arbitrary as he wishes. We who listen to him must always keep in mind that he has done so. Nothing that comes out of rigorous argument will be uncontaminated by the particular set of decisions that were made by him at the beginning, and it is rather dangerous at times to generalize. If we at any time relax our awareness of the way in which we originally defined the signal, we thereby automatically call all of the remainder of the received message the »not« signal and noise. This has many practical applications.

LICKLIDER: It is probably dangerous to use this theory of information in fields for which it was not designed, but I think the danger will not keep people from using it. In psychology, at least in the psychology of communication, it seems to fit with a fair approximation. When it occurs that the learnability of material is roughly proportional [156] to the information content calculated | by the theory, I think it looks interesting. There may have to be modifications, of course. For example, I think that the human receiver of information gets more out of a message that is encoded into a broad vocabulary (an extensive set of symbols) and presented at a slow pace, than from a message, equal in information content, that is encoded into a restricted set of symbols and presented at a faster pace. Nevertheless, the elementary parts of the theory appear to be very useful. I say it may be dangerous to use them, but I don't think the danger will scare us off.

MCCULLOCH: May I ask a question out of ignorance? I meant to ask it earlier in the day. Has any work been done on the number of simultaneous speeches that one can hear with noise, without noise, with distortion or without distortion?

LICKLIDER: The only work I know is a preliminary effort, made a couple of years ago, to compare the intelligibility of two talkers talking at once with the intelligibilities of the same talkers talking separately. In one test the two signals were simply superposed. In another they were alternated at various rates. The word lists were read slowly enough so that the listeners had sufficient time to get down both words when there were two talkers going at once. None of the listeners was able to do as well, of course, in that case, but it came out between two-thirds and three-quarters as well. Thus, even though the talkers were trying to enunciate the test words simultaneously, there was only moderate interference. It turned out that one of the two talkers had a big advantage over the other: I was the one with the advantage, and I held it over a friend with a much better voice. It just happened that he had a slightly clipped manner of speech (New York State), and my Midwestern words began sooner and ended later than his words did. So the listeners heard me first and last, and presumably therefore better. We tried putting one talker's signal into one of the listener's ears, the other talker's signal into the other. The isolation thus provided did not help much.

MCCULLOCH: How about alternation? You clipped and alternated?

LICKLIDER: Not clip, but blank. An electronic switch turned on first one talker, then the other.

MCCULLOCH: Regardless of rate of clipping?

LICKLIDER: We tried only a few rates, ones that looked interesting. None of the ones we tried proved useful in separating the two talkers.

TEUBER: In aphasics that have gotten practically well, one of the last symptoms of [157] aphasia you can detect is a difficulty, on the | patient's part, to follow a dialogue between two people in the room, neither of whom is directly addressing the patient. He may grasp pretty complicated things if you talk to him directly.

LICKLIDER: When you figure what the pattern of a pair of superposed vowel sounds must be like, it is really a little puzzling how the auditory system ever sorts the things out, especially if the two talkers are talking at almost the same pitch. There are, of course, a number of possible clues.

McCULLOCH: Have you tried different pitches, and so on?

LICKLIDER: As I said, this was not a very elaborate enterprise. We don't know any more about it.

McCULLOCH: Has anyone any questions bearing directly on Shannon's theme?

LICKLIDER: I have one. At Christmas time, here in New York, Shannon defined the concept of information – not just amount of information, but information itself. When I heard him, I said to myself: »That is wonderful. Why didn't I think of that?« It was simple and very clear. When I got back to Cambridge, however, I got a hint about why I hadn't thought of it. I couldn't even reproduce it. So I'd like to hear it again, and I think it would be of interest to all of us.

SHANNON: Yes. The general idea is that we will have effectually defined »information« if we know when two information sources produce the same information. This is a common mathematical dodge and amounts to defining a concept by giving a group of operations which leave the concept to be defined invariantly. If we have a message, it is natural to say that any translation of the message, say into Morse code or into another language, contains the same information provided it is possible to translate uniquely each way. In general, then, we can define the information of a stochastic process to be that which is invariant under all reversible encoding or translating operations that may be applied to the messages produced by the process. In other words, we define the information as the equivalence class of all such translations obtained from a particular stochastic process. Physically we can think of a transducer which operates on the message to produce a translation of the message. If the transducer is reversible, its output contains the same information as the input. When information is defined in this way, you are led to consider information theory as an application of lattice theory.

PITTS: Can the transducer wait for infinite time before commencing its translation?

SHANNON: There are actually two theories, depending on whether delays are allowed or not. The more general type of equivalence allows delays approaching infinity, while the restricted type demands that the translation and its inverse occur instantaneously, with no delay. Either type leads to a set of translations of a given information source, each containing the same information. [158]

LICKLIDER: The information is the group that is generated?

SHANNON: Yes. Put it another way: It is that which is common to all elements of the group.

McCULLOCH: With your permission I am going to omit the presentation of semantics by Walter Pitts and hold him as a whip over our heads to keep us lined up semantically from that time on. He and I have agreed that this is probably the most effective way to make use of him. If we approach the semantic problem before we go into the problems of learning languages, he and I both feel it would be somewhat contentless, whereas after we get the problems of learning languages in the open I think it will be extremely useful to us. We shall continue then, as follows: we shall first ask Margaret Mead to give us a picture of how one learns languages if he does not know the languages of that family or the culture of the people, languages for which there is no dictionary. That is a situation in which an adult consciously attempts to break a code.

References

1. SHANNON, C.E., and WEAVER, W.: *The Mathematical Theory of Communication*. Urbana: The University of Illinois Press 1949.
2. SHANNON, C. E.: Prediction and entropy of printed English. *Bell System Technical Journal*, 30, 50 (1951).