LANGUAGE, INFORMATION, AND ENTROPY

FRANK A. TILLMAN AND B. R. RUSSELL

Communication theorists and physicists frequently claim that the relation between information and entropy is one of reducibility or equivalence. The striking difficulty facing anyone who wishes to assess these claims is the lack of terminological agreement in information theory. The purpose of this paper is to survey some of the principle interpretations of information and entropy and, without presuming to set terminology, to show that information is not reducible to statistical mechanical considerations.

I. Introduction

From its inception the term «information» as employed in communication engineering has been fraught with a most persistent ambiguity. So powerful are its multiple associations that even those who have had a hand in framing a precise definition of the term have been let to make mistakes ([1]). A careful examination of some of its misuses has already been carried out by Bar-Hillel. The most notorious misapplications have arisen from its presumed connection with meaning, even though most theorists have been anxious to avoid this association and have repeatedly warned against extending to this new concept any semantical connotations. (10, 19).

Even after the term has been precisely defined and freed from any reference to semantics (except by explicit reformulation (7),) there is another line of association which has had a direct bearing on the issues involved in the suggested relation of information to entropy.

As early as 1894, the term 'information' was used by Boltzman to suggest that entropy is related to «missing information» (3). Here information is used in reference to some aspect of a physical system. In a similar context, Szilard, in 1929 (21), proposed, in connection with a study of Maxwell's «demon», that information is transformed into negative entropy. In an entirely different context and for different

([1]) The very choice of the term «information» may itself have been engendered by Hartley's confusing the selection of symbols with the selection of words which refer to things (2:93).

reasons Hartley, in 1928 (13), presented a logarithmic measure which he called 'information', i.e., the number of binary decisions that must be made to reconstruct a symbol sequence of some message.

In the hands of Shannon (19 : 19-20) this information measure was to bear a striking formal similarity to Boltzman's statistical function for entropy. Shannon uses the terms 'entropy' and 'information' as interchangeable equivalents (19 : 20). His decision to equate them is no doubt based on the fact that there is a formal similarity between the entropy and information measure. Brillouin, in a series of brilliant papers (3, 4, 5, 6) on entropy and information has indentified information with the negative of entropy. While superficially this involves only a change in sign there are indications that information has been interpreted in a way radically different from Shannon's concept. On the other hand Gabor has warned that limits set up by thermodynamics are of only very remote interest to the electrical engineers employing the information measure (10). Elsewhere (9) he has asserted that the properties of information follow from mathematical form rather than from any intrinsic relation between information and entropy. Other claims are that communication theory is a branch of physics with the implication, at least, that the former is reducible to the latter (8 : 652).

It is recognized that extension of usage of certain concepts like information is normal procedure in the empirical sciences and that such a practice has a high heuristic value in suggesting lines of fruitful research. But after such extended usage is effected and information is interpreted as measuring the microscopic aspect of a physical system (as recent research seems to suggest), we wonder whether, in such contexts, the distinctive features which make the information measure eminently successful in dealing with certain communication problems, are not being lost in the rapid expansion of this field of research. Moreover, unless the distinction is clarified and the terminology is examined carefully in context there is no hope that further claims of reducibility will lead to anything but confusion. It is certain that unless the task of clarification is performed, the highly suggestive claims about the relation between or identity of information and entropy will be opaque and unenligthening.

Our plan is to survey briefly the mathematical apparatus involved in the relevant concepts of information theory and statistical mechanics. We will then review the principal interpretations of this formalism in defining information and entropy. An examination will then be made of the typical claims concerning the relations and differences between these concepts and our purpose will be to determine whether these claims are correct or not.

II. *Formal Apparatus and Concepts*

A. *Formalism*

The concepts of information and entropy result from interpreting certain functions of the calculus of probability and analysis, i.e., providing coordinating definitions. The requisite formal techniques have been set forth in a number or recent publications (12, 14, 22, 25). Since the purpose of this section is to indicate that the concepts of information and entropy are parts of different branches of applied mathematics which employ similar formal techniques, we will be concerned only with functions relevant to making this point.

Suppose a set of probabilities $p = p_1, p_2 \ldots p_n$ satisfies the condition:

$$\sum_{i=1}^{n} p_i = 1.$$

There will be many functions which may be used to measure various properties of the distribution of these probabilities. One of special interest to physicists and communication engineers satisfies three conditions thought to be convenient:
1. when all the probabilities are zero except for one, the function must have the value of zero; 2. when all the probabilities are equal, the function yields a maximum; 3. it must be a function which increases monotonically with n.

A logarithmic function satisfying these conditions is:

$$-\sum_{i=1}^{n} p_i \operatorname{Log} p_i .$$

It does not make sense to call this an information (or entropic) measure until it is appropriately interpreted. It is misleading to speak of a calculus of information (at least at this stage of reasearch) for it my lead one to assume that there is some intrinsic relation between the above function and concepts contained in the statistical theory of information (2). There are those who maintain that the theory of information does not depend exclusively on the choice of this function

(2) Bar-Hillel (2) seems to make this mistake in referring to a «calculus of information». There is nothing wrong with this terminology except that it may be misleading when the same function is employed for other purposes.

(18 : 19). It is also possible to construct alternative entropy functions which appear to be quite different in form from the one to which information is usually compared. (15 : 283)

B. *Information*

A communication system depends on the fact that certain distinguishable things; e.g., letters like 'a', 'b', 'c' etc. are held in common by a transmitter and receiver, human, electronic or otherwise. Following Hartley's suggestion (13) that the communication process is essentially one of a selection of letters out of a set of letters, a certain numerical measure can be given for the number of decisions required to specify a particular letter. For example, if we consider that for each selection we must make a series of «yes» or «no» decisions to specify which letters has been selected, then, if we possess a 32-letter alphabet, a minimum of five decisions is required to identify the letter selected. Thus generally a selection of one among N letters will require H_N decisions:

(i) $H_N = [\mathrm{Log}_2 N]$ decisions (binary decisions or bits) for selecting one letter (sign).

where '[]' indicate that the next integral value above any fraction should be taken. This measure is due to Hartley and is the number of decisions required to identify one letter from other possible letters.

If we consider the case in which the letters (or signs) occur with different probabilities then, we obtain a measure of the expected number of decisions that must be made provided that we know these various probabilities. Suppose there are N possible letters that may be transmitted and to each of which can be assigned a probability of occurrence, then as a measure of the expected number of decisions required per sign (or letter) we have

(ii) $$H = -\sum_{i=1}^{n} p_i \log p_i \text{ bits/sign}, \qquad \sum_{i=1}^{n} p_i = 1,$$

where p_i is the relative frequency with which the i^{th} symbol turns up on the average in an ordered series. This measure is due to Shannon, who called it the «entropy of the source» and in other places «amount of information.» (19 : 18)

It is an unsettled issue as to which of the above are to be taken as measures of information. Each has very different uses and implications. According to (i) it makes good sense to talk about the number of bits in a limerick or a printed book for we merely multiply H_N by n, the number of letters contained in these samples. But according to (ii) it would be ludicrous to say that a limerick contains so many bits, since the probabilities are calculated on the basis of taking very long samples. No little confusion is involved in calling them both «information» even though (ii) reduces to (i) in the special case in which the p_i's need not be calculated; for example, when we have a set of letters in a row such as «a, b, c, d, e, f, g, h» and we are required by the conditions of the problem to calculate the minimum number of decisions required for finding one of the letters that some one has previously selected. In this case the prohability in terms of the relative frequency of each letters is irrelevant.

In accordance with most frequent technical usage and because much fruitful research has resulted from its use, we take, as a measure of information, formula (ii) interpreted appropriately. Information, then, is the *expected* minimum number of binary decisions per sign that must be made to reconstruct a given message. It is clear that information in this sense has nothing to do with the meaning of any possible message or, indeed, with semantic information as such.

This particular function is only one of a growing number that are available (12, 18). Moreover recent work has shown that the problem of reception is a special problem of statistical inference, game theory, and system design (16, 24). In addition other measures relevant to communication problems have been introduced and may be expected to exhibit different properties than those discussed in connection with (ii) above. Nevertheless, the choice of (ii) may be readily justified as a suitable measure useful for measuring the information capacity of a communication channel ([3]).

For a given language, e.g., French or English, the distribution of probabilities p_i is regulated by informal rules of syntax and word formation of the particular language. These rules are relatively fixed in natural languages since they undergo, at most, a slow evolution; i.e. while new words are introduced the rules pertaining to the way sentences are made and the conditions under which one word sequence or letter sequence follows another are relatively stable. However, H varies from one natural language to another (1) and would be expected

([3]) a. «It is a fair measure of what a customer pays for if he buys information capacity». Gabor (10).

b. Certain parameters «...such as time, bandwidth, number of relays, etc., tend to vary linearly with the logarithm of the number of possibilities». Shannon (19:4).

to vary also between the so-called «artificial» languages: esperanto, the algebras, etc. As to whether there are regular changes in H for one language, there are at present no known laws which state that there is a relation between H and the general direction which the course of a language will take. From this it appears that at present the statistical measure has no theoretical import for the study of language change, though it has import in the theory of communication with far reaching implications for designers of communication channels.

C. *Entropy*

The term 'entropy' was introduced originally by Clausius about a century ago in the empirical science of thermodynamics. It designates a quantity of great importance because of its relation to the celebrated second law of thermodynamics. Indeed, it was because of its close relation to the second law that this physical quantity was investigated and given a separate name. In terms of the more familiar quantities, heat and temperature, the variation of entropy dS may be defined as the ratio of the variation of heat energy added to the system dQ and the absolute temperature T thus:

$$dS = dQ/T$$

Using this term one can state the second law quite simply as a tendency for the entropy of closed systems to increase [4]. No exceptions to this law have been demonstrated so far, and it is repeatedly referred to in all the physical sciences. (14, 22)

In statistical mechanics the properties of idealized physical systems are described by means of various mathematical functions and correlations between average values of these functions and ordinary thermodynamic variables have been investigated. The function

$$S = k \log P$$

has been shown to have all the properties of the entropy, P being the number of possible equivalent states of the system all having equal *a priori* probability and K a universal constant (Boltzman's constant). When the *a priori* probabilities are not equal, then the appropriate calculus indicates that the proper average is obtained by weighting the logarithm of the probability for any one state p_i according to the probability of occurrence thus:

(4) It is to be understood in this definition that the change in heat energy is computed for a reversible process and that the initial and final states of the system are equilibrium states.

$$S = -k \sum_{i=1}^{n} p_i \log p_i .$$

The negative sign is necessary on the assumption that the p_i are normalized probabilities having the range $0 \leqslant p_i \leqslant 1$. This statistical function may be shown to possess all the properties which the entropy of empirical thermodynamics displays; in particular, its value can be shown to increase with the time as required by the second law.

Thus the term 'entropy', in its original sense, refers to a well defined physical quantity. Its variation in any physical system of practical interest can be measured with precision, and, in idealized systems to which the methods of statistical mechanics apply, its value can be computed. With the aid of the second law of thermodynamics one can predict with confidence the variation of this quantity to be expected in the future.

III. *Information and Entropy*

Immediately evident is the fact that the information function H differs from the entropy function S in statistical mechanics by a constant factor only. This obvious formal similarity, together perhaps with previous uses of the term 'information' in connection with entropy (5), have suggested to many that the concepts are related in various ways. We shall examine three alternative formulations: 1. the case in which 'entropy' and 'information' are used interchangeably, 2. the case where information and entropy are thought to be derivable from statistical mechanics, 3. the case in which information is equivalent to the negative entropy. While these alternative formulations are not exhaustive, they are at least typical of recent studies in this rapidly growing field.

1. An example of the first is Shannon's suggestive substitutions of 'entropy' for 'information'. After presenting a theorem which contains his H function he notes (19:19) that «the form of H will be recognized as that of entropy as defined in certain formulations of statistical mechanics where p_i is the probability of a system being in cell i of its phase space. H is then, for example, the H in Boltzman's famous H theorems. He shall call $H = p_i \log p_i$ «the entropy of the set of probabilities $p_i, \ldots, p_n$.» He then proceeds to interpret the p_i's in accordance with the requirements of a communication system handling

(5) See introduction.

symbols, i.e. letters or other units. The only connection which appears to be implied is that Boltzman's function and Shannon's information are formally identical except for an arbitrary constant which amounts to a convenient unit of measure. Thus there is implied no *necessary* connection between the concepts of information and entropy. Moreover, as we have seen, the Shannon function is not even necessary to communication engineering, but is justified solely on the grounds of its convenience and plausibility. We conclude, then that Shannon's choice of the term «entropy» to replace «information» is based on a formal analogy and the relation is no more than a formal one, although it is suggestive.

2. The suggestion that the concepts of information and entropy are both derivable from statistical mechanical considerations is a more serious one because it suggests that the relation between the terms is perhaps more than formal; i.e. that there is a comprehensive theory which embraces both the concept of entropy and information. Those who have suggested the view (8 : 652) have usually maintained that communication theory is a branch of physics. This suggestion may be interpreted in at least two ways.

Admittedly, the field of communication engineering is a large one covering the special theory of information as well as techniques and concepts based on the laws of physics. There are certain concepts such as information or inter-symbol influences which are not merely instances of applied physics. If these are considered to be important to the field of communication, as we believe they are, then the existence of these concepts places a burden on the upholders of the reductionist view of showing that these are derivable from statistical physics. It is not presumed that all expressions associated with communication theory need be defined with the help of expressions in statistical mechanics, but perhaps it is not unwarranted to suppose that if the reduction is to be more than partial then such concepts as information or inter-symbol influence must enter into such a reduction. Yet there are no known physical laws by which the variation of information can be predicted and there is no indication that such a variation would be related to the law of entropy or any other physical concept. Thus the existence of these hitherto irreducible concepts would, in any case, stringently limit the reduction of communication theory to physics (6).

(6) There is the possibility, to be considered later, of demonstrating a correlation between an extended concept of information and the physical law of entropy.

On the other hand it is possible that the suggestion that communication theory is a branch of statistical physics was intended only to indicate that certain uninterperted functions of one field are reducible to another by the usual logical and mathematical methods. If this is the intended meaning of reduction then we must agree that the statement asserting that such a reduction is possible is trivially true (and so are a great number of assertions of reducibility in this sense). It should be recognized that such a reduction is not what is usually intended when two empirical sciences are said to be reduced one to the other (17).

Thus the claim concerning the relation of information to entropy does not reveal that they are related or that they are mutually deducible from a more comprehensive theory.

3. Another formulation of the relation of entropy and information has been developed by Brillouin in a recent set of related papers (3, 4, 5, 6).

Brillouin's approach is far more ambitious than the last one considered in that he actually presents a reconstruction of the concept of information. Using the same functions he interprets the p_i's in terms of the coordinates of a physical system in such a way that the new interpretation results in a simple relation between his concept of information and entropy. Without raising the question as to whether Shannon's function or Brillouin's reconstruction deserves the name «information», we shall call Shannon's interpretation 'information'$_{Sh}$ and Brillouin's 'information'$_{Br}$. According to Brillouin's definition there is a strict proportionality between information and a negative portion of the entropy of the physical system employed, and further, information$_{Sh}$ is correlated with this information$_{Br}$ in certain situations, e.g. where information$_{Sh}$ is recorded in a physical system such as a magnetic tape. The question we shall try to answer is just how the two are correlated. We will then be in a position to determine whether on this new formulation there is any relation between information$_{Sh}$ and entropy or its negative value.

That there is a significant difference between information$_{Sh}$ and information$_{Br}$ may be seen in the uses to which Brillouin puts the latter.

a. Brillouin speaks of information as if it is «about» the structure of a system (3), a use that has semantical overtones which are irrelevant to Shannon's concept.

b. For Brillouin, information of a printed book is supposed to increase in direct ratio to the number of copies printed. According to

Shannon, information is not increased by such a factor, for if we can talk at all about the information of a book, the amount of information would remain the same no matter how many copies were made.

c. Brillouin sets up a proportionality between information (in his sense) and the negative of the entropy. Shannon, as we have seen, substitutes information for positive entropy.

The apparent contradictions or disparities between the two uses warrant the distinction we have made in terms of information_{Sh} and information_{Br}.

The mixing of terms is all too easy; in one place Brillouin mentions that a scientist in performing an observation «transforms negative entropy into information», (3 : 337) thus apparently referring to information_{Sh} although for the most part he is primarily concerned with information_{Br}.

Let us turn to Brillouin's formulation. The difference between his formulation and Shannon's may be expressed very concisely in terms of the interpretation made of the probability distribution functions. A physical system may be viewed from many standpoints. From the communication standpoint, Shannon (following Hartley) intended that the probability of occurrence of the i^{th} event p_i be interpreted in accordance with some convenient sign system. The conditions for communication are, of course, that there is an agreed language system common to source and receiver and that the sign events be distinguishable. From this standpoint only the symbol aspect of the system is considered even though a physical medium is involved; e.g. magnetic tape, channel, etc. It is just the physical medium to which Brillouin's formulation is relevant. He interprets the probabilities as the probability of the occurrence of distinguishable states of the physical medium involved (7). An example will illustrate the difference between his and Shannon's procedures.

Suppose a long message is recorded on a magnetic tape. From Shannon's standpoint we are concerned with the statistical properties of the message; i.e. the sample probabilities. The tape might be considered as a source and information the expected number of binary decisions per sign that must be made to reconstruct the source. On the other hand, Brillouin's analysis indicates that the recording of a message on magnetic tape produces a calculable change in the physical entropy of the tape. The average change in the entropy per impressed

(7) This is obviously closer to the interpretation made by Boltzman and Szilard (see introduction to this paper).

symbol may be calculated using the same function which Shannon uses but obviously the interpretation is different. When a proportionality constant is used the change in entropy and information$_{Sh}$ can be shown to have the same numerical value in fact. Thus information$_{Sh}$ is identical in numerical value, in this special case, to the entropy change per symbol of the physical system (after the negentropy has been adjusted by appropriate units). Brillouin calls the latter 'information'; i.e. what we have called 'information'$_{Br}$ ([8]).

Since the physical system involves many other aspects than symbols, the total entropy of the whole system is quite different from the change due to the presence of signs, and in fact the impression of signs generally involves restriction on its possible states and therefore leads to a reduction of entropy of the part of the system bearing the signs. This is why information$_{Br}$ is correlated with the negative of the entropy. From the microscopic point of view even the simplest systems of communication involve an enormous number of components all of which can have different states; the physical aspects of the employed symbols themselves generally affect only a small number of possible distributions of these states among the components. With complete knowledge of all the constituents of the system one can calculate the probability of occurence of any given distribution of the states and hence determine the entropy. The introduction of macroscopically observable changes associated with recording or transmission produces alterations of the possible microscopic distributions, thereby changing the probabilities, and the entropy. Thus a structural message recorded in or transmitted by a physical system results in a change in its entropy which may be computed and correlated with information$_{Sh}$. The total negentropy of a physical system would include all the orderedness of the system and therefore include as a part that which Brillouin correlates with, but which is not the same as or, in general, reducible to information$_{Sh}$. Thus the negentropy, which Brillouin calls 'information' (information$_{Br}$) includes that negentropy pertaining to distinguishable signs. Information$_{Br}$ becomes a calculable and in some cases a directly measurable physical quantity: negentropy/sign. New ways of using this concept have accordingly seemed natural and the apparent contradictions or disparities, noted earlier, arise because information has been redefined.

([8]) In his latest book Brillouin has named these concepts «free information» and «bound information» respectively. *Science and Information Theory* (New York, 1956), 152.

a. Having defined the proportionality between negentropy and information$_{Br}$, Brillouin uses 'negentropy' and 'information' interchangeably, and since the negative entropy is directly traceable to the elements of the structure of a system which limits its freedom, information$_{Br}$ may be regarded as information about the system. This makes sense because knowledge of the negative entropy enables one to construct models having certain structural relations analogous to those of the system oberved and it is, therefore, natural to speak of information received as revealing something of the structure of the system. Thus there is no inconsistency here even though it is unheard of to talk about information$_{Sh}$ as being about a system. Shannon's H applied to the output of a source yields a statistical expectation value, but for Brillouin information$_{Br}$ may be regarded as yielding knowledge about the source's structure.

b. The statement that an increase in the number of books printed constitutes on increase in information, simply means that the information$_{Br}$ or total negentropy has increased by bringing more order into a part of the total system, including the printing press, the blank paper, etc. It is to be noted that information is used here in still a different sense, meaning total number of bits rather than the number of bits per sign.

c. The difference in sign of the entropy correlation given by Brillouin follows as a direct consequence of the restriction of information$_{Br}$ to apply to physical systems. The usage of Shannon, identifying positive entropy with information$_{Sh}$ is formally correct but misleading with regard to physical implications. A brief explanation of the origin of this difference may be seen in the special case in which the probabilities are found to be equal. According to Shannon, one would compute the entropy of information or information$_{Sh}$ as $\log_2$ B bits/sign when there are N different signs employed with equal frequency. This quantity cannot be associated with a positive entropy of the physical medium utilized in recording or transmission because the N different signs are necessarily distinct and are directly correlated with states of the physical systems which are macroscopically distinct and distinguishable. The entropy of a system reduces to this simple form only when the N different states are distinguishable microscopically but completely equivalent and indistinguishable macroscopically. Thus, in computing information$_{Sh}$ one is concerned exclusively with the distinguishable states of the source, the transducer or the receiver, but in computing the entropy one is concerned with number of indistinguishable states as well.

To summarize and conclude this section of our paper, we may say that the difference between information$_{Sh}$ and information$_{Br}$ is to be found in the interpretation of two different though related events which occur with probability p_i. Shannon is concerned with the probability of signs, their order, and their relative frequency of occurence in very long samples. Brillouin considers only the case of events of distinguishable states in a physical system. In computing the entropy of a physical system by the method of statistical mechanics, one must consider not only the distinguishable states, but also the microscopically indistinguishable states as well; loosely, one may say the greater the number of indistinguishable states the greater the entropy of the system and vice versa. Thus for a physical source that successively adopts a large number of different distinguishable states associated with signs (letters etc.) there are many restrictions upon the possible states, therefore the entropy is *less* than the maximum. This reduction is what Brillouin has called 'information' and which we have referred to as 'information'$_{Br}$ to distinguish it from Shannon's concept. Thus information$_{Br}$ increases with an increase in the number of distinguishable states. It is also true that information$_{Sh}$ increases with an increase in the number of distinguishable signs and there is a negative relation between information$_{Br}$ and entropy.

IV. *Conclusion*

In the case of information one is dealing with a statistical distribution measure applied to signs and their relative occurrence or non-occurrence in large samples. It is true that a similar statistical measure may be used to define a function analogous to the entropy of a given physical system. In the case of entropy there are interactions within the system and with its environment which are capable of producing large changes in the distribution function in accordance with the laws of thermodynamics. The information$_{Sh}$ measure is applied, on the other hand, to distributions of signs in messages. The assumption that successive changes of these distributions are governed by physical laws would be a completely unwarranted extrapolation from our present knowledge. We can say of an isolated system that its entropy will, on the average, increase with time, but who would be so bold as to attempt to predict how the information of literary English may be expected to change with the next hundred years. The laws that do apply to this case are either vague or simply unknown and in any case the relations of distribution of language probabilities to laws of thermodynamics has not yet been established in any comprehensive theory. Thus the concept of information is *not* reducible to the concept of entropy.

Vassar College
College of Wooster

Frank A. Tillman
and B. R. Russell

REFERENCES

1. Barnard, G. A., «Statistical Calculation of Word Entropies for Four Western Languages» *I.R.E. Transactions*, Professional Group on Information Theory, 49-53, 1955.
2. Bar-Hillel, Y., «An Examination of Information Theory» *Philosophy of Science, 22*, 86-105, 1955.
3. Brillouin, L., «Maxwell's Demon Cannot Operate: Information and Entropy», *J. of Applied Physics, 22*, 334-337, 1951.
4. Brillouin, L., «Physical Entropy and Information II», *J. of Applied Physics, 22*, 338-343, 1951.
5. Brillouin, L., «Negentropy Principle of Communication», *J. of Applied Physics, 24*, 1152-1162, 1953.
6. Brillouin, L., «Negentropy and Information in Telecommunications, and Reading», *J. of Applied Physics, 25*, 595-599, 1954.
7. Carnap, R. and Y. Bar-Hillel, *An Outline of a Theory of Semantic Information*, Technical Report No. 247, Research Laboratory of Electronics, Massachusetts Institute of Technology, 1952.
8. Cherry, E. C., «The Communication of Information», *American Scientist, 40*, 640-664, 1952.
9. Gabor, D., «Communication Theory and Physics», *Philosophical Magazine, 41*, 1161-1187, 1950.
10. Gabor, D., «Lectures on Communication Theory», Technical Report No. 238, Research Laboratory of Electronics, Massachusetts Institute of Technology, 1952.
11. Gabor, D., «A Summary of Communication Theory», *Communication Theory*, ed. Willis Jackson, London, 1953.
12. Goldman, S., *Information Theory*, New York, 1953.
13. Hartley, R. V. L., «Transmission of Information», *Bell Systems Technical Journal, 7*, 535-563, 1928.
14. Lindsay, R. B. and Henry Margenau, *Foundations of Physics*, New York, 1936.
15. Lindsay, R. B., *Introduction to Physical Statistics*, J. Wiley & Son, New York, 1941.

16. MANDELBROT, Benoit, «Simple Games of Strategy Occurring in Communication Through Natural Languages», Transactions of I.R.E. Professional Group in Information Theory, *3*, 124-137, 1954.
17. NAGEL, «The Meaning of Reduction in the Natural Sciences», *Science and Civilization*, ed. Robert C. Stauffer, Madison, 99-138, 1949.
18. POWERS, K. H., *A Unified Theory of Information*, Technical Report No. 311, Research Laboratory of Electronics, Massachutts Institute of Technology, 1956.
19. SHANNON, C. E. and W. WEAVER, *The Mathematical Theory of Communication*, Urbana, 1949.
20. SHANNON, C., «Predication and Entropy of Printed English», *Bell Systems Technical Journal, 30*, 50-64, 1951.
21. SZILARD, L., «Über die Entropieverminderung in einen thermodynamischen System bei Eingriffen intelligenter Wesen», *Zeitschrift fur Physik, 53*, 840-856, 1929.
22. TOLMAN, R. C., *The Principles of Statistical Mechanics*, Oxford, 1938.
23. TORTRAT, A., «Sur la définition de l'entropie en théorie de l'information», *Annales des Télécommunications, 10*, 39-47, 1955.
24. VAN METER, David and David MIDDLETON, «Modern Statistical Approaches to Reception», *Transactions of I.R.E.*, Professional Group on Information Theory, *4*, 119-141, 1954.
25. WOODWARD, P. M., *Probability and Information Theory, with Applications to Radar*, London, 1953.