AN OUTLINE OF A THEORY OF SEMANTIC INFORMATION

RUDOLF CARNAP
YEHOSHUA BAR-HILLEL

TECHNICAL REPORT NO. 247

OCTOBER 27, 1952

RESEARCH LABORATORY OF ELECTRONICS
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MASSACHUSETTS

The research reported in this document was made possible through support extended the Massachusetts Institute of Technology, Research Laboratory of Electronics, jointly by the Army Signal Corps, the Navy Department (Office of Naval Research) and the Air Force (Air Materiel Command), under Signal Corps Contract No. DA36-039 sc-100, Project No. 8-102B-0; Department of the Army Project No. 3-99-10-022; and in part by a grant from the Rockefeller Foundation.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
RESEARCH LABORATORY OF ELECTRONICS

Technical Report No. 247 October 27, 1952

AN OUTLINE OF A THEORY OF SEMANTIC INFORMATION

Rudolf Carnap
University of Chicago
Yehoshua Bar-Hillel

Abstract

In distinction to current Theory of Communication which treats amount of information as a measure of the statistical rarity of a message, a Theory of Semantic Information is outlined, in which the concept of information carried by a sentence within a given language system is treated as synonymous with the content of this sentence, normalized in a certain way, and the concept of amount of semantic information is explicated by various measures of this content, all based on logical probability functions ranging over the contents. Absolute and relative measures are distinguished, so are D-functions suitable for contexts where deductive reasoning alone is relevant and I-functions suitable for contexts where inductive reasoning is adequate. Of the two major types of amount of information investigated, the one, cont, is additive with respect to sentences whose contents are exclusive, the other, inf, with respect to sentences which are inductively independent. The latter turns out to be formally analogous to the customary information measure function.

Various estimate functions of amount of information are investigated leading to generalized semantic correlates of concepts and theorems of current Communication Theory. A concept of semantic noise is tentatively defined, so are efficiency and redundancy of the conceptual framework of a language system. It is suggested that semantic information is a concept more readily applicable to psychological and other investigations than its communicational counterpart.

LIST OF SYMBOLS

OUTLINES OF A THEORY OF SEMANTIC INFORMATION

§1. The problem

The concepts of information and amount of information are distinguished. The explication of these concepts is attempted only insofar as they apply to (declarative) sentences or, alternatively, to propositions. Prevailing theory of communication (or transmission of information) deliberately neglects the semantic aspects of communication, i.e., the meaning of the messages. This theoretical restraint is, however, not always adhered to in practice, and this results in many misapplications. The theory outlined here is fully and openly of a semantic character and is therefore deemed to be a better approximation to a future theory of pragmatic information. For didactic purposes, the present theory of semantic information may be identified with a theory of pragmatic information for an "ideal" receiver.

It seems desirable that one should be able to say not only what information a message or an experiment has supplied but also how much. Hence we are going to distinguish between information (or content) and amount of information.

We shall deal with these concepts only insofar as they apply to either sentences or propositions, where 'sentences' is short for 'declarative sentences' or 'statements', and propositions are the nonlinguistic entities expressed by sentences. The theory we are going to develop will presuppose a certain language system and the basic concepts of this theory will be applied to the sentences of that system. These concepts, then, will be semantic concepts, closely connected with certain concepts of inductive logic, as we shall show below. Since inductive logic has been treated at length by one of us,* we shall make extensive use of the results achieved there. Relevant definitions and theorems will, however, be repeated to such an extent as to make the present treatment stand almost completely on its own.

The restriction of the range of application of the concepts to be explicated to sentences (or propositions) is probably not serious, since other applications seem to be reducible to this one. Instead of dealing with the information carried by letters, sound waves, and the like, we may talk about the information carried by the sentence, 'The sequence of letters (or sound waves, etc.)...has been transmitted'. The situation is similar to that prevailing with regard to the concept of truth, which is used presystematically as applying not only to sentences or propositions but also to many other entities such as concepts and ideas. There, too, these latter usages seem to be reducible to the former ones.

In recent authoritative presentations of the so-called Mathematical Theory of Communication, or Theory of (Transmission of) Information, great care has been taken to point out that this theory is not interested in the semantic aspects of communication.

---

*Rudolf Carnap: Logical Foundations of Probability, University of Chicago Press, 1950; The Continuum of Inductive Methods, University of Chicago Press, 1952. These works will be referred to hereafter as [Prob.] and [Cont.], respectively.

The following two quotations may be regarded as representative. Claude E. Shannon states in The Mathematical Theory of Communication (of which he is co-author with Warren Weaver), Univ. of Illinois Press, Urbana, 1949, p. 3: "These semantic aspects of communication are irrelevant to the engineering problem." E. Colin Cherry in "A History of the Theory of Information", Proc. Inst. Elec. Engrs. 98, 383, 1951, says: "It is important to emphasize, at the start, that we are not concerned with the meaning or the truth of messages; semantics lies outside the scope of mathematical information theory."

It has, however, often been noticed that this asceticism is not always adhered to in practice and that sometimes semantically important conclusions are drawn from officially semantics-free assumptions. In addition, it seems that at least some of the proponents of communication theory have tried to establish (or to reestablish) the semantic connections which have been deliberately dissevered by others.

In 1948 Donald MacKay conceived a theory of information that should be broad enough to cover both theory of communication and theory of scientific information, the latter dealing with the formation of representations or their rearrangement in the representational space of the observer, the former dealing with the replication of representations in the mind of the receiver, which were already present in the mind of the sender of a message. (Cf. "Quantal Aspects of Scientific Information", Phil. Mag. 41, 289, 1950, and "The Nomenclature of Information Theory", prepared for the Symposium on Information Theory held in London in September 1950, and printed in revised form as Appendix I of a talk, "In Search of Basic Symbols", given before the Eighth Conference on Cybernetics held in New York in March 1951, and published in the Transactions of this Conference, pp. 181-235.)

Jean Ville, in a talk before the 18th International Congress of Philosophy of Science in Paris, 1949 (published in Actualites Scientifiques et Industrielles, No. 1145, pp. 101-114, Paris, 1951), also treats information as a basically semantic concept and develops functions and theorems which stand in close correspondence to some of the functions and theorems with which we deal in this report. A more thorough evaluation of these and other contributions to the foundations of information theory, as well as a comparison between the theory presented here and the theory of communication, will be undertaken elsewhere.

Our theory lies explicitly and wholly within semantics.

It does not deal, however, with what has been termed by Weaver in his contribution to the afore-mentioned book "the semantic problem" of communication, which, as defined by him, is "concerned with the identity, or satisfactorily close approximation, in the interpretation of meaning by the receiver, as compared with the intended meaning of the sender." We would rather prefer to consider an investigation in which sender and receiver are explicitly involved as belonging to pragmatics.

We shall talk about the information carried by a sentence, both by itself and relative to some other sentence or set of sentences, but not about the information which the sender intended to convey by transmitting a certain message nor about the information a receiver obtained from this message. An explication of these usages is of paramount importance, but it is our conviction that the best approach to this explication is through an analysis of the concept of semantic information which, in addition to its being an approximation by abstraction to the full-blooded concept of pragmatic information, may well have its own independent values.

Anticipating later results, it will turn out, under all explications envisaged by us, that the amount of information carried by the sentence '$17 \times 19 = 323$' is zero and that the amount of information of 'The three medians of the sides of a plane triangle intersect in one point', relative to some set of sentences serving as a complete set of axioms for Euclidean geometry, is likewise zero. This, however, is by no means to be understood as implying that there is no good sense of 'amount of information', in which the amount of information of these sentences will not be zero at all, and for some people, might even be rather high. To avoid ambiguities, we shall use the adjective 'semantic' to differentiate both the presystematic senses of 'information' in which we are interested at the moment and their systematic explicata from other senses (such as "amount of psychological information for the person P") and their explicata. This adjective will, however, be dropped in those cases where ambiguities are unlikely to arise.

The following comparison might be of value for pointing out one of the services which a clarification of the semantic concept of information should render for a future theory of pragmatic information. The theory of so-called ideal gases is of great importance in physics despite the fact that no actual gas is ideal and that many gases are very far from being ideal. The semantic information carried by a sentence with respect to a certain class of sentences may well be regarded as the "ideal" pragmatic information which this sentence would carry for an "ideal" receiver whose only empirical knowledge is formulated in exactly this class of sentences. By an "ideal" receiver we understand, for the purposes of this illustration, a receiver with a perfect memory who "knows" all of logic and mathematics, and together with any class of empirical sentences, all of their logical consequences. The interpretation of semantic information with the help of such a superhuman fictitious intellect should be taken only as an informal indication. We shall not refer to this fiction in the technical part of this paper.

Our task can now be stated much more specifically. We intend to explicate the presystematic concept of information, insofar as it is applied to sentences or propositions and inasmuch as it is abstracted from the pragmatic conditions of its use. We shall then define, on the basis of this systematic concept of semantic information, various explicata for the presystematic concept (or concepts) of amount of semantic information and shall investigate their adequacy and applicability.

§ 2. General explanations

The language-systems relative to which the present theory of information is developed are described as containing a finite number of individual constants and primitive one-place predicates. The following fundamental syntactic and semantic concepts are explained: atomic sentence, molecular sentence, basic sentence, molecular predicate, L-true, L-false, factual, L-implies, L-equivalent, L-disjunct, L-exclusive, Q-predicator, Q-property, Q-sentence, state-description, and range. Some terms and symbols of class-theory (set-theory) are introduced, mainly complement, sum, and product.

The language-systems relative to which our theory of information will be developed are very simple ones, so simple indeed that the results to be obtained will be of only restricted value with regard to language-systems complex enough to serve as possible languages of science. The restriction, however, was partly imposed by the fact that inductive logic – on which we shall have to rely heavily – has so far been developed to a sufficiently elaborate degree only for languages that are not much richer than those treated here, ([Prob.] §§15, 16), and partly for the sake of simplicity of presentation. It is hoped that in spite of this the results will be immediately applicable to certain simple situations and will be suggestive with respect to more complex ones.

Our language-systems $\mathscr{L}_n^\pi$ contain n different individual constants which stand for n different individuals (things, events, or positions) and π primitive one-place predicates which designate primitive properties of the individuals. (n and π are finite numbers; under certain assumptions, however, it is easy to extend the results obtained here to systems with a denumerably infinite number of individual constants.) In an atomic sentence, for example, 'Pa' ('the individual a has the property P'), a primitive property is asserted to hold for an individual. Other molecular sentences are formed out of atomic sentences with the help of the following five customary connectives:

$\sim$	not	negation
$\vee$	or	disjunction
$\cdot$	and	conjunction
$\supset$	if...then	(material) implication
$\equiv$	if, and only if (written iff)	(material) equivalence

All atomic sentences and their negations are called basic sentences. Analogously, other molecular predicates or predicators are formed out of primitive predicates with the help of the (typographically) same connectives (for example, 'M.~P' standing for 'M and not P'). A sentence consisting of a predicator and an individual constant is called a full sentence of this predicator. Though our systems do not contain individual variables, quantifiers, or an identity sign, their expressive power is thereby not essentially affected. Sentences like 'there are exactly three individuals having the property P' can still be rendered in these systems, though only in the form of a rather clumsy disjunction of conjunctions of basic sentences. Hence absolute frequencies (cardinal numbers of classes or properties) and relative frequencies can be expressed in these systems (but not measurable quantities like length and mass).

Any sentence is either L-true (logically true, analytic, e.g., 'PaV~Pa') or L-false (logically false, self-contradictory, e.g., 'Pa.~Pa') or factual (logically indeterminate, synthetic, e.g., 'PaV[M.~N]b'). Logical relations between sentences i and j can be defined:

i L-implies j	$=_{Df}$	$i \supset j$	is L-true
i is L-equivalent to j	$=_{Df}$	$i \equiv j$	is L-true
i is L-disjunct with j	$=_{Df}$	$i \vee j$	is L-true
i is L-exclusive of j	$=_{Df}$	$i . j$	is L-false

We shall use 't' as the name of a particular L-true sentence, a "tautology", say, of '$Pa \vee \sim Pa$'.

A Q-predicator is a conjunction (of predicates) in which every primitive predicate occurs either unnegated or negated (but not both) and no other predicate occurs at all. The property designated by a Q-predicator is called a Q-property. A full sentence of a Q-predicator is a Q-sentence. A state-description is a conjunction of n Q-sentences, one for each individual. Thus a state-description completely describes a possible state of the universe of discourse in question.* For any sentence j of the system, the class of those state-descriptions in which j holds, that is, each of which L-implies j, is called the range of j. The range of j is null if, and only if, j is L-false; in any other case, j is L-equivalent to the disjunction of the state-descriptions in its range.

The following theorems will be of use later:

T2-1.

a. The number of atomic sentences is $\beta = \pi n$.

b. The number of Q-predicators is $\kappa = 2^{\pi}$.

c. The number of state-descriptions is $z = 2^{\beta} = 2^{\pi n} = (2^{\pi})^{n} = \kappa^{n}$.

In our metalanguage, that is, the language in which we talk about our language-systems $\mathscr{L}_n^{\pi}$ (in our case a certain unspecified sublanguage of ordinary English enriched by a few additional symbols), we shall use some customary terms and symbols of the theory of classes (or sets). The class of all those entities (of a certain type) which do not belong to a certain class K will be called the complement (-class) of K and denoted by '-K'. The class of those entities which belong either to a class K or to a class L (or to both) will be called the (class-theoretical) sum of these classes and will be denoted by '$K \cup L$'. The class of those entities which belong to each of the classes K and L will be called the (class-theoretical) product of these classes and denoted by '$K \cap L$'.

Those readers who might not be familiar with abstract logical concepts and terms will profit from the following illustration, which will be carried through this whole monograph. Let a census be taken in a small community of only three inhabitants, and let the census be interested only in whether the inhabitants counted are male or non-male (female) and young or non-young (old), respectively. Let the three individuals be designated by 'a', 'b', 'c', and the properties by 'M', '~M' (or 'F'), 'Y', and '~Y' (or 'O'), respectively. The language-system, in which the outcome of the census can be exhaustively described is therefore a $\mathscr{L}_3^2$ - system, in our notation. 'Ma' is an atomic

*This holds, strictly speaking, only if the primitive properties are logically independent. For a discussion of the problems involved here, see R. Carnap: Meaning Postulates, Phil. Studies 3, 65-73, 1952 and the literature mentioned there.

sentence, 'Ma. Fb. (Mc⊃Oc)' another molecular sentence, 'F. ~Y' a Q-predicator, '[F. ~Y]b' a Q-sentence, '[M. Y]a.[~M. Y]b.[~M. ~Y]c' a state-description. For later references, the list of all 64 state-descriptions is given in Table I, in abbreviated form. Line 10 of this table, for example, is to be interpreted as short for the state-description '[M. Y]b. [M. Y]c. [F. Y]a'. Later (§4), however, a different interpretation of the same table will be given.

Table I

	M, Y	M, O	F, Y	F, O		M, Y	M, O	F, Y	F, O
1.	a, b, c	-	-	-	33.	b	-	-	a, c
2.	-	a, b, c	-	-	34.	a	-	-	b, c
3.	-	-	a, b, c	-	35.	-	c	-	a, b
4.	-	-	-	a, b, c	36.	-	b	-	a, c
5.	a, b	c	-	-	37.	-	a	-	b, c
6.	a, c	b	-	-	38.	-	-	c	a, b
7.	b, c	a	-	-	39.	-	-	b	a, c
8.	a, b	-	c	-	40.	-	-	a	b, c
9.	a, c	-	b	-	41.	a	b	c	-
10.	b, c	-	a	-	42.	a	c	b	-
11.	a, b	-	-	c	43.	b	a	c	-
12.	a, c	-	-	b	44.	b	c	a	-
13.	b, c	-	-	a	45.	c	a	b	-
14.	c	a, b	-	-	46.	c	b	a	-
15.	b	a, c	-	-	47.	a	b	-	c
16.	a	b, c	-	-	48.	a	c	-	b
17.	-	a, b	c	-	49.	b	a	-	c
18.	-	a, c	b	-	50.	b	c	-	a
19.	-	b, c	a	-	51.	c	a	-	b
20.	-	a, b	-	c	52.	c	b	-	a
21.	-	a, c	-	b	53.	a	-	b	c
22.	-	b, c	-	a	54.	a	-	c	b
23.	c	-	a, b	-	55.	b	-	a	c
24.	b	-	a, c	-	56.	b	-	c	a
25.	a	-	b, c	-	57.	c	-	a	b
26.	-	c	a, b	-	58.	c	-	b	a
27.	-	b	a, c	-	59.	-	a	b	c
28.	-	a	b, c	-	60.	-	a	c	b
29.	-	-	a, b	c	61.	-	b	a	c
30.	-	-	a, c	b	62.	-	b	c	a
31.	-	-	b, c	a	63.	-	c	a	b
32.	c	-	-	a, b	64.	-	c	b	a

The reader will easily verify that the range of the sentence 'Ma. Ya. Fb. Yb', which might profitably be rewritten in the form '[M. Y]a. [F. Y]b', contains exactly 4 state-descriptions, namely, 9, 25, 42, and 53. The range of 'Fa' contains 32 state-descriptions. The range of 'MaVYaVFbVYbVFcVOc' contains 63 state-descriptions, that is, all state-descriptions except 52. A reader with some training in propositional logic will see immediately that this last sentence is L-equivalent to '~(Fa. Oa. Mb. Ob. Mc. Yc)', hence to the negation of state-description 52.

§ 3. The presystematic concept of semantic information

> A requirement of adequacy for any proposed explication of semantic information – In – is stated: In(i) includes In(j) if and only if i L-implies j. From this requirement various theorems are derived. In addition to the absolute information carried by a sentence, the information carried by a sentence j in excess to that carried by some other sentence i is often of importance. This concept of relative information is defined by: In(j/i) = In(i. j) - In(i). One of the theorems is: if t is any L-true sentence, In(j/t) = In(j). Two concepts that fulfill the requirement but differ in some aspect are investigated but none of them accepted as an explicatum for In.

To disperse, at least partially, the haziness which envelops the inevitably vague discussions of the adequacy of the explication to be offered later for the concept of semantic information, let us state a requirement which will serve as a necessary condition for this adequacy.

Whenever i L-implies j, i asserts all that is asserted by j, and possibly more. In other words, the information carried by i includes the information carried by j as a (perhaps improper) part. Using 'In(...)' as an abbreviation for the presystematic concept 'the information carried by ...', we can now state the requirement in the following way:

R3-1. In(i) includes In(j) iff i L-implies j.

By this requirement we have committed ourselves to treat information as a set or class of something. This stands in good agreement with common ways of expression, as for example, "The information supplied by this statement is more inclusive than (or is identical with, or overlaps) that supplied by the other statement."

We shall now state some theorems which hold for 'In' and, therefore, also for that concept which we shall offer, in the following section, as the explicatum for 'In'. These theorems follow from R1* and well-known theorems of the theory of classes.

T3-1. In(i) = In(j) iff i is L-equivalent to j.

If a class K of classes contains a class which is included in every member of K, this class will be called "the minimum class (of K)". If K contains a class which includes every member of K, this class will be called "the maximum class (of K)". The minimum class and the maximum class may coincide with the null-class and the universal-class (of the corresponding type), respectively, but need not do so.

Since an L-true sentence is L-implied by every sentence, and an L-false sentence L-implies every sentence, we have:

T3-2. In(i) = the minimum class of K (where K is now the class of the In-classes of sentences) iff i is L-true.

T3-3. In(i) = the maximum class of K iff i is L-false.

It might perhaps, at first, seem strange that a self-contradictory sentence, hence

---

*For the sake of brevity, theorems, definitions, or requirements, when found in the same section in which they are first stated, will be referred to only by the corresponding letter 'T', 'D', or 'R' and by their second numbers. Here, for instance, we have 'R1' instead of the longer 'R3-1'.

one which no ideal receiver would accept, is regarded as carrying with it the most inclusive information. It should, however, be emphasized that semantic information is here not meant as implying truth. A false sentence which happens to say much is thereby highly informative in our sense. Whether the information it carries is true or false, scientifically valuable or not, and so forth, does not concern us. A self-contradictory sentence asserts too much; it is too informative to be true.

T3-4. In(i) properly includes In(j) iff i L-implies j but j does not L-imply i.

T3-5. In(i) properly includes the minimum class and is properly included in the maximum class iff i is factual.

T3-6. In(i) includes In(iVj) and is included in In(i. j).

When we use the term 'information' in ordinary language, we often refer to the information carried by a sentence absolutely, so to speak. At least as often, however, we intend to refer to the information carried by a sentence in excess of that carried by some other sentence (or class of sentences). If not otherwise stated or implicitly understood through the context, this other sentence will often be that in which the total knowledge available to the receiver of the information, before he receives the new information, is stated. In contradistinction to the concept of absolute information treated so far, we shall now define, still on the presystematic level, the concept of relative (or additional or excess) information of j with respect to i as the class-theoretical difference of In(i. j) and In(i), that is, the class-theoretical product of In(i. j) with the complement of In(i); in symbols:

D3-1. $\mathrm{In}(j/i) =_{Df} \mathrm{In}(i.j) - \mathrm{In}(i)\ (=\mathrm{In}(i.j) \cap -\mathrm{In}(i))$.

In(j/i) is again a class. Its members belong to the same type as the members of In(i). The following theorems follow immediately from D1, R1, and the previous theorems.

Complete formal proofs will be given only when an indication of the theorems, definitions, and requirements from which a theorem follows will not enable most readers to grasp the proof by inspection. In very simple cases (as in the first 8 theorems), these hints will be omitted.

T3-7. In(j/i) includes the minimum class and is included in the maximum class.

T3-8. If i is L-equivalent to j, then In(k/i) = In(k/j) and $\mathrm{In}(i/\ell) = \mathrm{In}(j/\ell)$.

T3-9. If i L-implies j, then In(j/i) = the minimum class.

Proof: In this case, i. j is L-equivalent to i. The theorem follows from T1 and D1.

T3-10. If j is L-true, then In(j/i) = the minimum class.

T3-11. In(j/i) properly includes the minimum class iff i does not L-imply j.

Proof: In this case, In(i. j) properly includes In(i).

T3-12.

a. If i is an L-true sentence, In(j/i) = In(j).

Proof: In this case, j. i is L-equivalent to j and In(i) = the minimum class.

b. In(j/t) = In(j).

Thus the relative information of j with respect to t equals the absolute information of j. Therefore it would be possible to begin with the relative information as primitive and define the absolute information as the value of the relative information with respect to t. However, it seems more convenient to begin with the simple concept of the absolute information, because it has only one argument and the relative information can be defined on the basis of it. This is the procedure we have chosen here.

So far, we have committed ourselves to treat the information carried by a sentence as a class of something and have stated one requirement which every adequate explicatum will have to meet. This, of course, leaves many possibilities open. With respect to the information carried by an L-true sentence, we were able to state only that it is a minimum and is contained in the information carried by any sentence. It might perhaps seem plausible to require, in addition, that the information carried by an L-true sentence should be empty; hence, the null-class of the appropriate type. But this feeling is due to the fact that we do not always distinguish carefully between information and amount of information. What we really have in mind is rather that the amount of semantic information carried by an L-true sentence should be zero. But this can be achieved by a suitable explicatum even if the information carried by such a sentence is not the null-class.

On the other hand, there also exists no good reason so far why the information of an L-true sentence should not be the null-class. The best procedure is, therefore, to leave this decision open.

There are indeed two plausible explicata for In(i), which differ in exactly this point: according to the one, the information carried by an L-true sentence will be the null-class; according to the other, it will not. Let us denote the first concept by 'Inf_1' and the second by 'Inf_2'. Their definitions are as follows:

D3-2. $\mathrm{Inf}_1(i) =_{Df}$ the class of all sentences (in $\mathscr{L}$) which are L-implied by i and not L-true.

D3-3. $\mathrm{Inf}_2(i) =_{Df}$ the class of all sentences (in $\mathscr{L}$) which are L-implied by i.

We shall not dwell here on an elaborate comparison of the relative merits and faults of these two definitions: first, because such a comparison has already been carried out by one of us in a closely related context*, second, because we shall adopt neither of these definitions for future work but a third one to be explained in the following section.

§4. Content-elements and content

A content-element is defined as the negation of a state-description, and the content of i – Cont(i) – as the class of the content-elements L-implied by i. Cont is taken as the explicatum for In. Cont(j/i) is defined and various theorems derived.

In §2, we defined the range of a sentence i, R(i), as the class of all state-descriptions

*See R. Carnap: Introduction to Semantics, Harvard University Press, 1942, §23, and [Prob.], p. 406.

Z in which i holds or which, in other words, L-imply i. The sentence i says that the state of the universe (treated in $\mathscr{L}$) is one of the possible states which are described by the Z in R(i). Alternatively formulated, i says that the universe is not in one of those states which are described by the Z in V_Z - R(i), where V_Z is the class of all Z. Just as i is L-implied by every Z in R(i), so it L-implies the negation of every Z in V_Z - R(i). We call these negations the content-elements E of i and their class the content of i, in symbols Cont(i). In general, we call the negations of the Z in a given system $\mathscr{L}$ the E of this system. (See [Prob.] §73.)

In our $\mathscr{L}_3^2$, there are, of course, 64 content-elements, namely, the negations of its 64 state-descriptions. These content-elements appear in Table I, when interpreted in a different way from that given before. We can now read line 10, for example, as '[MVY]bV[MVY]cV[FVY]a', a content-element which is L-equivalent to the negation of state-description 37, as the reader will verify for himself.

The content of the sentence 'Ma. Ya. Fb. Yb' contains 60 content-elements, namely, the negations of all state-descriptions except 9, 25, 42, and 53.

The following theorem, T4-1, can be deduced from the theorems concerning state-descriptions and L-concepts in [Prob.] §§18A, 18D, 19, 20, 21B.

T4-1. For every E_i the following holds:

a. E_i is factual ([Prob.] T20-5b, T20-6).

b. If E_j is distinct from E_i, then E_i and E_j are L-disjunct.

Proof: $\sim E_i . \sim E_j$ is L-false ([Prob.] T20-8a). Therefore the negation of this conjunction is L-true. But this negation is L-equivalent to $E_i \vee E_j$.

c. The conjunction of all E_i is L-false.

Proof: Let d be the disjunction of the negations of the E_i, hence L-equivalent to the disjunction of all Z. Therefore d is L-true ([Prob.] T21-8b); hence $\sim$d is L-false. But $\sim$d is L-equivalent to the conjunction of all E_i.

d. If E_i L-implies j, then j is either L-true or L-equivalent to E_i; in other words, E_i is a weakest factual sentence.

Just as a state-description says the most that can be said in the given universe of discourse, short of self-contradiction, so a content-element says the least, beyond a tautology. 'a is male and young, b is female and young, and c is female and old' is a strongest factual sentence in the census; its negation 'a is female or old (or both), or b is male or old, or c is male or young' (where 'or' is always to be understood in its nonexclusive sense) a weakest one.

T4-2.

a. Cont(i) = the null-class of E, Λ_E, iff i is L-true.

b. Cont(i) = the class of all E, V_E, iff i is L-false.

c. Cont(i) = neither Λ_E nor V_E iff i is factual.

d. Cont(i) includes Cont(j) iff i L-implies j.

e. Cont(i) = Cont(j) iff i is L-equivalent to j.

f. Cont(i) and Cont(j) are exclusive (i. e., have no members in common) iff i and j are L-disjunct ([Prob.] T20-1c, d).

The contents of 'Ma' and of 'Fa∨Mb' are exclusive since 'Ma∨Fa∨Mb' is L-true. The reader can verify from Table I, in its second interpretation, that these contents have indeed no members in common.

T4-3.

a. Cont(∼i) = -Cont(i) (short for 'V_E - Cont(i)', [Prob.] T18-1e).

b. Cont(i∨j) = Cont(i)∩Cont(j).

Proof: Let $\overline{R}$(...) be the class of the negations of the members of R(...). Then Cont(i∨j) = $\overline{R}$(∼(i∨j)) = $\overline{R}$(∼i.∼j) = $\overline{R}$(∼i)∩$\overline{R}$(∼j) = Cont(i)∩Cont(j).

c. Cont(i.j) = Cont(i)∪Cont(j).

Proof: Cont(i.j) = -Cont(∼(i.j)) = -Cont(∼i∨∼j) = -(Cont(∼i)∩Cont(∼j)) = -(-Cont(i)∩-Cont(j)) = Cont(i)∪Cont(j).

To verify T3b, c take, for instance, i as 'Ma∨Fb∨[M∨Y]c' and j as 'Fa∨Mb'.

T2d shows that Cont fulfills requirement R3-1. We decide to take Cont as our explicatum for In. The explication of the information carried by a sentence j, as the class of the negations of all those Z which are excluded by j, is intuitively plausible and in accordance with the old philosophical principle, "omnis determinatio est negatio." Our main reason, however, for giving it preference over the two explicata mentioned in the previous section, Inf_1 and Inf_2, lies in the fact that an explication of amount of information will turn out to be rather simple if based on Cont, in accordance with the fourth requirement for a good explication stated in [Prob.] p. 7.

Let us notice that according to T2a, Cont shares with Inf_1 the property that their value for an L-true sentence as argument is the null-class.

We now have to define the relative content of j with respect to i. What has to be done is, of course, simply to replace 'In' in D3-1 by 'Cont'.

D4-1. Cont(j/i) $=_{Df}$ Cont(i.j) - Cont(i).

Clearly, T3-12a, b hold for Cont if 'In' is replaced in them by 'Cont'. Let us state explicitly only the correlate of T3-12b, that is,

T4-4. Cont(j/t) = Cont(j).

The remarks following T3-12b hold also mutatis mutandis for absolute and relative content.

§5. The presystematic concept of amount of information

Requirements of adequacy for the explication of amount of semantic information – in – are stated, and theorems for in derived. No formal requirement of additivity is accepted since the conditions under which additivity is to hold cannot be given unambiguously, so far. in(j/i), the amount of information of j relative to i, is defined and theorems derived.

Our next task is to find an explicatum, or perhaps various explicata, for the presystematic concept of amount of information. This will again be preceded by the statement of some requirements, the fulfillment of which will be a necessary condition for the adequacy of the explicata to be proposed.

We shall use 'in' as the symbol for the presystematic concept of amount of information and distinguish between the absolute amount of information of a sentence i, in(i), and the relative amount of information of the sentence j with respect to i, in(j/i). The relative amount is clearly definable on the basis of the absolute amount:

D5-1. $in(j/i) =_{Df} in(i.j) - in(i)$

(where the '-' sign is this time the symbol for numerical difference and not, as in D3-1 or D4-1, for class-difference). Therefore it is sufficient to state only the requirements with respect to the absolute amount. It seems plausible to require that the amount of information of i should be not less than the amount of information of j, if the content of i includes the content of j; that the amount of information of an L-true sentence should be zero; and, for finite systems, that the amount of information of a factual sentence should be greater than zero. (The qualification 'for finite systems' might perhaps look superfluous. It can, however, be shown that with regard to the explicata envisaged by us, this requirement would not be fulfilled in an infinite system.) More formally:

R5-1. $in(i) \geqslant in(j)$ if (but not only if) Cont(i) includes Cont(j).

R5-2. $in(j) = 0$ if $Cont(j) = \Lambda_E$.

R5-3. $in(j) > 0$ if Cont(j) properly includes Λ_E.

Instead of R3 we might also have required the following somewhat stronger condition from which R3 follows immediately:

R5-4. $in(i) > in(j)$ if Cont(i) properly includes Cont(j).

We could also have stated these requirements directly in terms of 'L-implies' and 'L-true', without recourse to Cont. For the benefit of those who, for some reason, are not satisfied with our explication of 'information' and who therefore might try to explicate 'amount of information' on the basis of some other explicatum for 'information' or perhaps even without reference to any such explicatum (a perfectly reasonable and achievable goal), the following version is given:

R5-1*. $in(i) \geqslant in(j)$ if (but not only if) i L-implies j.

R5-2*. $in(j) = 0$ if j is L-true.

R5-3*. $in(j) > 0$ if j is not L-true.

The following theorems follow from R1 through R3 and the previously stated properties of Cont.

T5-1. If Cont(i) = Cont(j), then in(i) = in(j).

T5-2. If i is L-false, then in(i) has the maximum in-value.

Proof: An L-false sentence L-implies every sentence.

T5-3. $0 < in(i) <$ the maximum in-value iff i is factual.

T5-4. $in(i \vee j) \leqslant in(i) \leqslant in(i.j)$.

The requirements R1 through R3 are clearly rather weak, and one might look for further requirements. One that recommends itself immediately would be that of additivity, that is, to have $in(i.j) = in(i) + in(j)$ if i and j are independent of each other in a certain sense. However, we shall not make this one of our formal requirements because the sense of the independence involved is not clear at this moment. We shall find later that each of our explicata is indeed additive but not all of them in the same sense, because the conditions of independence are not the same in the various cases.

The additivity holds, of course, only under certain conditions, whatever those conditions may be in exact terms. It is clear that, in general, in(i. j) $\neq$ in(i) + in(j). It is further clear that there will be cases where in(i. j) < in(i) + in(j). This will be the case, for example, whenever i L-implies j and j is not L-true, because under these circumstances i is L-equivalent to i. j, so that in(i. j) = in(i), whereas in(j) > 0, and hence in(i) < in(i) + in(j). So far, we can state only a lower limit for in(i. j), viz:

T5-5. in(i. j) $\geqslant$ max[in(i), in(j)].

Does there exist a general upper limit for in(i. j) that is not trivial? No theorem to this effect can be deduced from the requirements of this section. They do not exclude, for instance, the possibility that sometimes in(i. j) > in(i) + in(j). This possibility might perhaps look so implausible that one would like to exclude it by the explicit additional requirement in(i. j) $\leqslant$ in(i) + in(j). However, it seems better not to require this. We shall see later that the second of our explicata (inf) violates this condition, and we shall then make this violation plausible. If someone insists that the requirement just stated has to be fulfilled, then he can accept only the first of our explicata (cont). For this concept the requirement is indeed fulfilled (T6-4m).

The following theorems correspond to T3-7 through T3-12.

T5-7. The maximum in-value $\geqslant$ in(j/i) $\geqslant$ 0.

T5-8. If i is L-equivalent to j, then in(k/i) = in(k/j) and in(i/ℓ) = in(j/ℓ).

T5-9. If i L-implies j, then in(j/i) = 0.

T5-10. If j is L-true, then in(j/i) = 0.

T5-11. in(j/i) > 0 iff i does not L-imply j.

T5-12.

a. If i is an L-true sentence, in(j/i) = in(j).

b. in(j/t) = in(j).

§ 6. The first explicatum: Content-measure (cont)

One way of fulfilling the requirements stated in the previous section is outlined. It consists, essentially, in defining a measure-function over the content-elements, fulfilling certain conditions, and then taking as the measure of the content of a sentence the sum of the measures ascribed to the elements of its content. Since measure-functions over state-descriptions – m-functions – have been treated by one of the authors at length before, a shorter way of introducing content-measures – cont – is chosen, simply by equating cont(i) with $m_P(\sim i)$, where 'm_P' stands for proper m-function, i. e., m-function fulfilling certain conditions. Many theorems for cont(i) are derived, among them theorems for the content-measures of basic sentences, for disjunctions and conjunctions of such, for Q-sentences, and for sentences in disjunctive and conjunctive normal form. cont(j/i) is defined, and among others, the important theorem cont(j/i) = cont(i⊃j) is derived.

We could have defined an adequate explicatum for the amount of information carried by a sentence with the help of measure-functions ranging over the contents of the sentences of $\mathscr{L}$ and fulfilling the conditions laid down in the previous section. Since there exist, however, close relations between contents and ranges (§4), we shall make use of the fact that the definitions of various measure-functions over ranges have already been treated at length in [Prob.] and define the functions in which we are now interested simply on the basis of those measure-functions.

It seems profitable to start with a kind of measure-function ranging over state-descriptions and other sentences which has not been discussed explicitly in [Prob.] or [Cont.], namely, with proper m-functions, to be denoted by 'm_P'.

We define:

D6-1. m is a proper m-function (in $\mathscr{L}$) $=_{Df}$ m fulfills the following nine conditions:

a. For every Z_i, $m(Z_i) > 0$.

b. The sum of the m-values of all Z = 1.

c. For any L-false sentence j, m(j) = 0.

d. For any non-L-false sentence j, m(j) = the sum of the m-values for the Z in R(j).

e. If Z_j is formed from Z_i by replacing the individual constants of Z_i by those correlated to them by any permutation of the individual constants, then $m(Z_j) = m(Z_i)$. (Less strictly but more suggestively: all individuals are treated on a par.)

f. If Z_j is formed from Z_i by replacing the primitive predicates of Z_i by those correlated to them by any permutation of the primitive predicates, then $m(Z_j) = m(Z_i)$ (i.e., all primitive properties are treated on a par).

g. If Z_j is formed from Z_i by replacing any of the primitive predicates of Z_i by their negations (omitting double negation signs), then $m(Z_j) = m(Z_i)$ (i.e., each primitive property is treated on a par with its complement).

The last three conditions could have been stated in a somewhat weaker form, but no attempt was made to reduce redundancy by sacrificing psychological clarity.

h. If i and j have no primitive predicates in common, then $m(i.j) = m(i) \times m(j)$.

i. m(i) is not influenced by the number of individuals of $\mathscr{L}$ not mentioned in i. (This condition will be used only in the derivation of formula (6) in §10.)

An m-function fulfilling conditions (a) through (d) is called regular ([Prob.] p. 295). If it fulfills, in addition, condition (e), it is called symmetrical ([Prob.] p. 485). All theorems that hold for regular m-functions hold a fortiori for any proper m-function.

m_P is believed to be an adequate explicatum for one of the senses in which 'probability' is used, namely that which might be termed 'absolute logical probability', that is, logical probability on no evidence (or tautological evidence or irrelevant evidence). Similarly, c_P, to be defined in D7-3, is believed to be an adequate explicatum of relative logical probability.

Any two sentences (not only state-descriptions) that stand in the relation stated in D1e are called isomorphic.

The following theorem holds for all regular m-functions ([Prob.] §§55A, 57A), hence also for all proper m-functions:

T6-1.

a. $0 \leq m(i) \leq 1$.

b. $m(i) = 1$ iff i is L-true.

c. $m(i) = 0$ iff i is L-false.

d. $0 < m(i) < 1$ iff i is factual.

e. If i L-implies j, then $m(i) \leq m(j)$.

f. If i is L-equivalent to j, then $m(i) = m(j)$.

g. $m(i.j) \leq m(i) \leq m(i \vee j)$.

h. $m(i \vee j) = m(i) + m(j) - m(i.j)$.

i. $m(i \vee j) = m(i) + m(j)$ iff i.j is L-false (i.e., iff i and j are L-exclusive).

j. $m(i.j) = m(i) + m(j) - m(i \vee j)$.

k. $m(i.j) = m(i) + m(j) - 1$ iff $i \vee j$ is L-true (i.e., iff i and j are L-disjunct).

ℓ. $m(\sim i) = 1 - m(i)$.

m. $m(i.j) \leq m(i) + m(j)$.

The measure-function in which we are interested and which we shall call from now on content-measure and denote by 'cont' is defined by

D6-2. $\text{cont}(i) =_{Df} m_P(\sim i)$.

From this definition it immediately follows that the cont-value of any E equals the m_P-value of the corresponding Z.

T6-2. For every Z_i, if E_i is $\sim Z_i$, $\text{cont}(E_i) = m_P(Z_i)$.

D2 and T1ℓ entail

T6-3.

a. $\text{cont}(i) = 1 - m_P(i)$.

b. $m_P(i) = 1 - \text{cont}(i)$.

c. $\text{cont}(\sim i) = m_P(i)$.

The following theorem follows from T1 and T3b:

T6-4.

a. $1 \geq \text{cont}(i) \geq 0$.

b. $\text{cont}(i) = 0$ iff i is L-true.

c. $\text{cont}(i) = 1$ iff i is L-false.

d. $1 > \text{cont}(i) > 0$ iff i is factual.

e. If i L-implies j, then $\text{cont}(i) \geq \text{cont}(j)$.

f. If i is L-equivalent to j, then $\text{cont}(i) = \text{cont}(j)$.

g. $\text{cont}(i.j) \geq \text{cont}(i) \geq \text{cont}(i \vee j)$.

h. $\text{cont}(i \vee j) = \text{cont}(i) + \text{cont}(j) - \text{cont}(i.j)$.

i. $\text{cont}(i \vee j) = \text{cont}(i) + \text{cont}(j) - 1$ iff i and j are L-exclusive.

j. $\text{cont}(i.j) = \text{cont}(i) + \text{cont}(j) - \text{cont}(i \vee j)$.

k. $\text{cont}(i.j) = \text{cont}(i) + \text{cont}(j)$ iff i and j are L-disjunct.

ℓ. $\text{cont}(\sim i) = 1 - \text{cont}(i)$.

m. $\text{cont}(i.j) \leq \text{cont}(i) + \text{cont}(j)$.

T4e, b, and c-d show that cont fulfills the requirements of adequacy R5-1*, R5-2*, and R5-3*, respectively.

The condition under which additivity is stated in T4k to hold for cont appears quite plausible at first glance. If i and j are L-disjunct, then the contents of i and j are exclusive (T4-2f). Nothing in that which is asserted by i is simultaneously asserted by j; in other words, there is no factual sentence which is L-implied both by i and by j. However, we shall later (§7) make certain considerations which will raise some doubts with respect to this special condition of additivity.

The relative content-measure of j with respect to i is meant as the increase of the value of cont by adding j to i. Hence, in conformance with D5-1:

D6-3. $\text{cont}(j/i) =_{Df} \text{cont}(i.j) - \text{cont}(i)$.

T6-5.

a. $\text{cont}(j/i) = \text{cont}(j) - \text{cont}(i \vee j)$ (D3, T4j).

b. $\qquad\qquad = \text{cont}(j)$ iff i and j are L-disjunct ((a), T4b).

T6-6. $\text{cont}(j/i) = \text{cont}(i \supset j)$.

Proof: j is L-equivalent to $(i \vee j).(\sim i \vee j)$. The components of this conjunction are L-disjunct. Therefore $\text{cont}(j) = \text{cont}(i \vee j) + \text{cont}(\sim i \vee j)$ (T4k). Hence, with T5a, $\text{cont}(j/i) = \text{cont}(\sim i \vee j)$. But $\sim i \vee j$ is L-equivalent to $i \supset j$.

The last theorem is especially interesting. It shows that the relative content-measure of j with respect to i is the same as the absolute content-measure of the (material) implication $i \supset j$. If an "ideal" receiver possesses the knowledge i and then acquires the knowledge j, his possession of information is only increased in the same amount as if $i \supset j$ were added instead of j. This is, indeed, highly plausible since j is a logical consequence of the sentences i and $i \supset j$, and an "ideal" receiver, by definition, is able to draw such consequences instantaneously.

From T6 we also see that if i L-implies j, $\text{cont}(j/i) = 0$. We know this already since it holds for all our explicata for the relative amount of information in virtue of T5-9.

The following inequality, an immediate consequence of T5a, is of interest:

T6-7. $\text{cont}(j/i) \leqslant \text{cont}(j)$.

We can express cont(j/i) directly in terms of m_P in various ways:

T6-8.

a. $\text{cont}(j/i) = m_P(i) - m_P(i.j)$ (D3, T3a).

b. $\qquad\qquad = m_P(i.\sim j)$ (T6, $i \supset j$ is L-equivalent to $\sim(i.\sim j)$, T3a).

c. $\qquad\qquad = m_P(i \vee j) - m_P(j)$ (D3, T5a).

Two sentences, i and j, that fulfill the condition, $m_P(i.j) = m_P(i) \times m_P(j)$, are called inductively independent (or initially irrelevant, in the terminology of [Prob.] - p. 356) with respect to that m_P. We get

T6-9. If i and j have no primitive predicate in common, then

$$m_P(i \vee j) = m_P(i) + m_P(j) - m_P(i) \times m_P(j) \text{ (T1h, D1h).}$$

T6-10.

a. For any basic sentence B, $m_P(B) = 1/2$.

Proof: B∨~B is L-true. Therefore, by T1b, $m_P(B \vee \sim B) = 1$. Hence the assertion with D1g.

b. For any conjunction, C_n, of n basic sentences with n distinct primitive predicates, $m_P(C_n) = (1/2)^n$ (D1h, (a)).

c. If i and i' are isomorphic, then $m_P(i) = m_P(i')$ (D1e).

We now get

T6-11. If i and j have no primitive predicate in common, then

a. cont(~(i. j)) = cont(~i) × cont(~j) (D1h, T3c).

b. cont(i∨j) = cont(i) × cont(j).

Proof: i∨j is L-equivalent to ~(~i. ~j). ~i and ~j have no primitive predicate in common since i and j do not. Hence the assertion from (a).

c. cont(i. j) = cont(i) + cont(j) - cont(i) × cont(j) (T4j, (b)).

In our $\mathscr{L}_3^2$, cont('Ma∨Yb') = cont('Ma∨Ya') = 1/4 and cont('Ma.~Yb') = 3/4.

T6-12. Let D_n be a disjunction of n (⩾2) sentences with no primitive predicate occurring in more than one of these sentences. Then $\text{cont}(D_n)$ = the product of the cont-values of the n components (T11b).

T6-13.

a. For any basic sentence B, cont(B) = 1/2 (T3a, T10a).

b. For any disjunction, D_n, of n basic sentences with n distinct primitive predicates, $\text{cont}(D_n) = (1/2)^n$ (T3a, T12, (a)).

c. For any conjunction, C_n, of n basic sentences with n distinct primitive predicates, $\text{cont}(C_n) = 1 - (1/2)^n$ (T3a, T10b, (a)).

d. For any Q-sentence i, $\text{cont}(i) = 1 - (1/2)^\pi$ ((c)) $= 1 - 1/\kappa$ (T2-1b) $= (\kappa-1)/\kappa$.

$$\text{cont}('[M.\sim Y]a') = 3/4 \text{ (since } \pi = 2, \kappa = 4 \text{ (T2-1b))}.$$

e. Let i have the form $C_1 \vee C_2 \vee \ldots \vee C_m$, where each C is a conjunction of n basic sentences with n distinct primitive predicates, the same n atomic sentences occurring in all conjunctions. (Under these circumstances, i has disjunctive normal form. See [Prob.] p. 94 or any textbook on Symbolic Logic.) Then

$$\text{cont}(i) = 1 - \frac{m}{2^n}.$$

Proof: Any two distinct conjunctions are L-exclusive. Therefore, from T4i, $\text{cont}(i) = \text{cont}(C_1) + \text{cont}(C_2) + \ldots + \text{cont}(C_m) - (m-1)$. Hence the conclusion with (c).

$$\text{cont}('(\text{Ma. Yb})\vee(\sim\text{Ma. Yb})\vee(\text{Ma.}\sim\text{Yb})') = 1 - \frac{3}{2^2} = \frac{1}{4}.$$

Notice that this disjunction is L-equivalent to 'Ma∨Yb', that is, a disjunction fulfilling (b).

f. Let i have the form $D_1 . D_2 . \ldots . D_m$, where each D is a disjunction of n basic sentences with n distinct primitive predicates, the same n atomic sentences occurring in all disjunctions. (Under these circumstances, i has conjunctive normal form. See [Prob.] p. 95.) Then

$$\text{cont}(i) = \frac{m}{2^n}.$$

Proof: Any two distinct disjunctions are L-disjunct. Therefore, from T4k, $\text{cont}(i) = \text{cont}(D_1) + \text{cont}(D_2) + \ldots + \text{cont}(D_m)$. Hence the assertion with (b).

$$\text{cont}(\text{'(Ma} \vee \text{Yb)}.(\sim\text{Ma} \vee \text{Yb)}.(\text{Ma} \vee \sim\text{Yb)'}) = \frac{3}{2^2} = \frac{3}{4}.$$

Notice that this conjunction is L-equivalent to 'Ma. Yb', that is, a conjunction fulfilling (c).

T6-14. If i and i' are isomorphic and j and j' are isomorphic on the basis of the same permutation of the individual constants, then cont(j'/i') = cont(j/i) (T8b, T10a).

T6-15.

a. For any two basic sentences, B_i and B_j, with different primitive predicates, $\text{cont}(B_j/B_i) = 1/4$ (T13c, T10a) $= 1/2\ \text{cont}(B_i)$ (T13a).

$$\text{cont}(\text{'Ya'/'Ma'}) = \frac{1}{4}.$$

b. Let $B_1, B_2, \ldots, B_n$ be basic sentences with n distinct primitive predicates. Let C_m be the conjunction of the first m of them. Then, for every m (m=2, ..., n-1),

$$\text{cont}(B_{m+1}/C_m) = \frac{1}{2^{m+1}}.$$

Proof: $C_m . B_{m+1} = C_{m+1}$. Hence

$$\text{cont}(B_{m+1}/C_m) = \text{cont}(C_{m+1}) - \text{cont}(C_m) = 1 - \frac{1}{2^{m+1}} - \left(1 - \frac{1}{2^m}\right) \text{(T13c)} = \frac{1}{2^{m+1}}.$$

T6-16. Let i and j be molecular sentences with no primitive predicate in common. Then $\text{cont}(j/i) = \text{cont}(j) - \text{cont}(i) \times \text{cont}(j)$ (T11c) $= \text{cont}(j) \times (1 - \text{cont}(i)) = \text{cont}(j) \times \text{cont}(\sim i)$ (T4ℓ) $= \text{cont}(j) \times m_P(i)$ (T3c).

§ 7. The second explicatum: Measure of information (inf)

One of the theorems derived in the previous section states that if i and j are basic sentences with different primitive predicates, then cont(j/i) = 1/2 cont(i). Since basic sentences with different primitive predicates are inductively independent, this result makes cont look inadequate as an explicatum for in. It turns out that no explicatum fulfilling all our intuitive requirements for an amount-of-information function is possible, indicating a certain inconsistency between these requirements. cont fulfills a partial set of these requirements, and a different, though overlapping, partial set is fulfilled by another function, called measure of information, denoted by 'inf', and defined as

$$\text{inf}(i) = \text{Log}\, \frac{1}{1-\text{cont}(i)}.$$

It is shown that

$$\text{inf}(h, e) = \text{Log} \frac{1}{c_P(h, e)}$$

where $c_P(h, e)$ is the degree of confirmation of the hypothesis h on the evidence e, defined as

$$\frac{m_P(e.h)}{m_P(e)} .$$

The last-but-one theorem of the preceding section (T6-15) may not appear entirely plausible. According to this theorem, if an "ideal" receiver with no previous knowledge receives a sequence of n basic sentences with n different primitive predicates, the amount of information he gets from the first sentence is 1/2, from the second only 1/4, from the third 1/8, from each only half as much as from the preceding one. And this will be the case despite the fact that these basic sentences are independent from each other not only deductively but also inductively. One has the feeling that under such conditions the amount of information carried by each sentence should not depend upon its being preceded by another of its kind.

An inconsistency in our intuitions, at which we already hinted above (§6), becomes now even more prominent. The feeling to which we referred in the preceding paragraph may be expressed also as a requirement that additivity should hold for the amount of information carried by the conjunction of two sentences if these sentences are inductively independent. We saw, however, that additivity holds for cont only if these sentences are L-disjunct and have no content in common. Now, it is clear that two basic sentences, B_1 and B_2, with different primitive predicates, have content in common: the factual sentence $B_1 \vee B_2$, for instance, is L-implied by each. Nevertheless, this condition of additivity looked plausible in its context.

It seems best to solve this conflict of intuitions by assuming that there is not one explicandum "amount of semantic information" but at least two, for one of which cont is indeed a suitable explicatum, whereas the explicatum for the other still has to be found.

Let us now state the additional requirement in a formal way:

R7-1. If i and j are inductively independent, then $\text{in}(i.j) = \text{in}(i) + \text{in}(j)$.

From R1 and D5-1 follows immediately:

T7-1. If B_i and B_j are two basic sentences with distinct primitive predicates, then $\text{in}(B_j/B_i) = \text{in}(B_j)$.

Let us also decide, for the sake of normalization, to assign to each basic sentence an in-value of 1.

R7-2. For any basic sentence B, $\text{in}(B) = 1$.

We have now

T7-2. For a conjunction of n basic sentences, C_n, with n distinct primitive predicates, $\text{in}(C_n) = n$ (R1, R2).

T6-13c stated that $\text{cont}(C_n) = 1 - (1/2)^n$, hence

$$2^n = \frac{1}{1-\text{cont}(C_n)}$$

hence

$$n = \text{Log}\,\frac{1}{1-\text{cont}(C_n)}$$

(where 'Log' is short for 'logarithm on the base 2'). This, combined with T2 yields

T7-3. For a conjunction of n basic sentences, C_n, with n distinct primitive predicates,

$$\text{in}(C_n) = \text{Log}\,\frac{1}{1-\text{cont}(C_n)}.$$

T3 gives us the lead for defining the second explicatum for "amount of information". This new function will be called measure of information and denoted by 'inf'. Extending the relationship stated in T3 to hold for all sentences, we define

D7-1. For any sentence i,

$$\text{inf}(i) = \text{Log}\,\frac{1}{1-\text{cont}(i)}.$$

D1 may be usefully transformed into

T7-4.

a. $$\text{inf}(i) = -\text{Log}(1-\text{cont}(i)) = -\text{Log cont}(\sim i) = \text{Log}\,\frac{1}{\text{cont}(\sim i)}.$$

b. $\text{inf}(\sim i) = -\text{Log cont}(i)$.

c. $\text{cont}(\sim i) = 2^{-\text{inf}(i)}$.

d. $\text{cont}(i) = 1 - 2^{-\text{inf}(i)}$.

T7-5.

a. $$\text{inf}(i) = \text{Log}\,\frac{1}{m_P(i)} \text{ (D1, T6-3)}$$

b. $$= -\text{Log}\, m_P(i).$$

The form of T5a is analogous to the customary definition of amount of information in communication theory. In the place of the concept of probability in the statistical sense (relative frequency) used in that definition, we have here the logical (inductive) probability m_P. For a detailed discussion of the relation between these two concepts, see [Prob.] §§ 3, 10.

T7-6. $m_P(i) = 2^{-\text{inf}(i)}$.

A host of other theorems for inf can easily be derived. We shall mention only a few of them.

T7-7. $$\text{inf}(\sim i) = \text{Log}\,\frac{1}{1-m_P(i)} = -\text{Log}(1-m_P(i)).$$

T7-8.

a. $0 \leq \text{inf}(i) \leq \infty$ (T6-4a).

b. $\inf(i) = 0$ iff i is L-true (T6-4b).

c. $\inf(i) = \infty$ iff i is L-false (T6-4c).

d. $\inf(i)$ is positive finite iff i is factual (T6-4d).

e. If i L-implies j, then $\inf(i) \geqslant \inf(j)$ (T6-4e).

f. If i is L-equivalent to j, then $\inf(i) = \inf(j)$ (T6-4f).

g. $\inf(i.j) \geqslant \inf(i) \geqslant \inf(i \vee j)$ (T6-4g).

h. $\inf(i.j) = -\mathrm{Log}\ \mathrm{cont}(\sim i \vee \sim j)$ (T4a)

$$= -\mathrm{Log}(\mathrm{cont}(\sim i) + \mathrm{cont}(\sim j) - \mathrm{cont}(\sim i. \sim j)) \text{ (T6-4h)}$$

$$= -\mathrm{Log}\ 2^{-\inf(i)} + 2^{-\inf(j)} - 2^{-\inf(i \vee j)} \text{ (T4c)}.$$

i. If i and j are L-disjunct, then $\inf(i.j) = -\mathrm{Log}\left(2^{-\inf(i)} + 2^{-\inf(j)} - 1\right)$ ((h), (b)).

j. $\inf(i \vee j) = -\mathrm{Log}\ \mathrm{cont}(\sim i. \sim j)$ (T4a) $= -\mathrm{Log}\left(1 - 2^{-\inf(\sim i. \sim j)}\right)$ (T4d).

k. If i and j are L-exclusive (hence $\sim i$ and $\sim j$ L-disjunct), then

$\inf(i \vee j) = -\mathrm{Log}\ (\mathrm{cont}(\sim i) + \mathrm{cont}(\sim j))$ ((j), T4k) $= -\mathrm{Log}\left(2^{-\inf(i)} + 2^{-\inf(j)}\right)$ (T4d).

ℓ. $\inf(\sim i) = \inf(i) - \mathrm{Log}\left(2^{\inf(i)} - 1\right)$ (T4b, d).

Whereas the correspondence between T8a through g and T6-4a through g is straightforward, T8h through ℓ are much more complicated and much less convenient for computation than their corresponding theorems T6-4h through ℓ.

As against the complicated formula T8i, we have, however,

T7-9. Additivity. $\inf(i.j) = \inf(i) + \inf(j)$ iff i and j are inductively independent.

Proof: $\inf(i.j) = -\mathrm{Log}\ m_P(i.j)$ (T5b)

$$= -\mathrm{Log}(m_P(i) \times m_P(j)) \text{ (by hypothesis)}$$

$$= -\mathrm{Log}\left(2^{\inf(i)} \times 2^{-\inf(j)}\right) \text{ (T6)}$$

$$= \mathrm{Log}\ 2^{\inf(i)} + \mathrm{Log}\ 2^{\inf(j)} = \inf(i) + \inf(j).$$

To T6-13 corresponds

T7-10.

a. For any basic sentence B, $\inf(B) = 1$.

b. For any disjunction, D_n, of n basic sentences with n distinct primitive predicates,

$$\inf(D_n) = \mathrm{Log}\ \frac{1}{1-(1/2)^n} = \mathrm{Log}\ \frac{2^n}{2^n-1} = n - \mathrm{Log}(2^n - 1).$$

c. For any conjunction, C_n, of n basic sentences with n distinct primitive predicates, $\inf(C_n) = n$.

d. For any Q-sentence i, $\inf(i) = \pi$.

e. Let i have disjunctive normal form: $C_1 \vee C_2 \vee \ldots \vee C_m$. Let every C be a conjunction of n basic sentences with n distinct primitive predicates, the same n atomic sentences occurring in all conjunctions. Then $\inf(i) = n - \mathrm{Log}\ m$.

$\inf('(Ma.Yb) \vee (\sim Ma.Yb) \vee (Ma. \sim Yb)') = 2 - \mathrm{Log}\ 3$ (=0.412).

f. Let i have conjunctive normal form: $D_1.D_2. \ldots .D_m$. Let every D be a disjunction of n basic sentences with n distinct primitive predicates, the same n atomic sentences occurring in all disjunctions. Then $\text{inf}(i) = n - \text{Log}(2^n - m)$.

$$\text{inf}('(Ma \vee Yb).(\sim Ma \vee Yb).(Ma \vee \sim Yb)') = 2 - \text{Log}(2^2 - 3) = 2.$$

T8e, b, and d show that inf fulfills R5-1* through R5-3*. T9 corresponds to R1, and T10a to R2. Thereby it is shown that inf fulfills all our requirements for the second explicatum for amount of information.

The following table gives approximative inf-values for D_2 (T10b) through D_{10}:

Table II

n	$\text{inf}(D_n)$
2	0.412
3	0.192
4	0.093
5	0.046
6	0.023
7	0.0113
8	0.0056
9	0.0028
10	0.0014

We define now the relative measure of information in the already familiar way:

D7-2. $\text{inf}(j/i) =_{Df} \text{inf}(i.j) - \text{inf}(i)$.

T7-11.

a. For any two basic sentences, B_i and B_j, with distinct primitive predicates, $\text{inf}(B_j/B_i) = 1$ (D2, T10c, a) $= \text{inf}(B_i)$ (T10a).

$$\text{inf}('Ma'/'Yb') = \text{inf}('Ma'/'Ya') = 1.$$

b. Let B_1, B_2, ..., B_n be basic sentences with n distinct primitive predicates. Let C_m be the conjunction of the first m of them. Then, for every m (m=2, ..., n-1),

$$\text{inf}(B_{m+1}/C_m) = 1 \text{ (D2, T10c, (a))}.$$

T7-12.

a. $\text{inf}(j/i) = \text{inf}(j)$ iff i and j are inductively independent (D2, T9).

b. If i and j have no primitive predicates in common, $\text{inf}(j/i) = \text{inf}(j)$ (T6-9a, (a)).

In [Prob.] §55, the concept of degree of confirmation of an hypothesis h on the evidence e, on the basis of a given range measure m, is defined as follows:

$$c(h, e) = \frac{m(e.h)}{m(e)}.$$

e L-implies h if, and only if, the range of e is wholly contained in the range of h. If, however, only a part of R(e) is contained in R(h), then none of the customary relations of deductive logic holds between e and h. If, say, that part of R(e) which is contained in R(h) is three fourths of R(e), as measured by m, if, in other words,

$$\frac{m(e.h)}{m(e)} = \frac{3}{4},$$

then we shall say that the hypothesis h is confirmed by the evidence e to the degree 3/4 and write this relation, which is fundamental for inductive logic, as 'c(h, e) = 3/4'. c is meant as an explicatum for (relative) inductive probability.

Figure 1 might be of some help for a visualization of the difference between L-implication and degree of confirmation as dependent upon the relations between the ranges of the hypothesis and the evidence.

For an m_P-function, we have more specifically

D7-3. $c_P(h, e) = \dfrac{m_P(e.h)}{m_P(e)}$.

T7-13. If inf and c_P are based on the same m_P, then

$$\inf(h/e) = \text{Log}\,\frac{1}{c_P(h, e)} = -\text{Log}\,c_P(h, e).$$

Proof: $\inf(h/e) = \inf(e.h) - \inf(e) = \text{Log}\,m_P(e) - \text{Log}\,m_P(e.h)$

$$= \text{Log}\,\frac{m_P(e)}{m_P(e.h)} = \text{Log}\,\frac{1}{c_P(h, e)}.$$

This theorem shows the strong connection that exists between the relative measure of information of a new message h with respect to the knowledge e and the degree of confirmation of an hypothesis h on the evidence e, in other words, the relative inductive probability of an hypothesis h on the evidence e. The characterization of h as message and e as knowledge, on the one hand, or as hypothesis and evidence, on the other, has didactical value only; h and e are, strictly speaking, simply any sentences of the given system. T13 shows that inf(h/e) is the greater the more improbable h is on the evidence e. That the relative amount of information carried by a sentence should increase

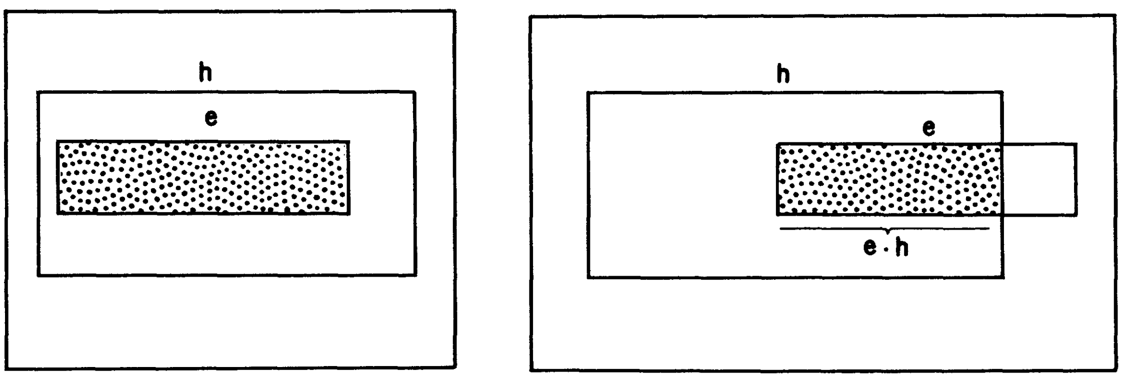

Fig. 1

Deductive Logic

'e L-implies h' means that the range of e is entirely contained in that of h.

Inductive Logic

'c(h, e) = 3/4' means that three-fourths of the range of e is contained in that of h.

with its degree of improbability seems plausible. This holds also for cont, if e remains fixed, as shown by the following theorem.

T7-14. If cont and c_P are based on the same m_P, then

$$\text{cont}(h/e) = m_P(e) \times (1 - c_P(h, e)) = m_P(e) \times c_P(\sim h, e).$$

Proof: $\text{cont}(h/e) = m_P(e) - m_P(e.h)$ (T6-8a)

$$= m_P(e) \times \frac{m_P(e) - m_P(e.h)}{m_P(e)}$$

$$= m_P(e) \times (1 - c_P(h, e)) \text{ (D3)}.$$

Notice, however, that for variable e it need not be the case that the smaller $c_P(h, e)$ is, the larger cont(h/e) will be, because of the factor $m_P(e)$. (See end of §10, below.)

§ 8. Comparison between cont and inf

The cont and inf measures are compared in greater detail. Both exhibit properties which look intuitively plausible and others which look intuitively implausible. The formally most striking comparison is given by the following pair of theorems:

$$\text{cont}(h/e) = m_P(e) - m_P(e.h),$$
$$\text{inf}(h/e) = \text{Log}\, m_P(e) - \text{Log}\, m_P(e.h).$$

We are now ready for a comparison between the two explicata for amount of information. Let us begin with stating some corresponding theorems one beside the other for better confrontation.

T6-4k. cont(i. j) = cont(i) + cont(j) iff i and j are L-disjunct.	T7-9. inf(i. j) = inf(i) + inf(j) iff i and j are inductively independent.
T6-4m. cont(i. j) $\leqslant$ cont(i) + cont(j).	
T6-13a. For any basic sentence B, cont(B) = 1/2.	T7-10a. For any basic sentence B, inf(B) = 1.
T6-13c. For any conjunction, C_n, of n basic sentences with n distinct primitive predicates, $\text{cont}(C_n) = 1 - (1/2)^n$.	T7-10c. For any conjunction, C_n, of n basic sentences with n distinct primitive predicates, $\text{inf}(C_n) = n$.
T6-15a. For any two basic sentences, B_i and B_j, with distinct primitive predicates, $\text{cont}(B_j/B_i) = 1/4 = 1/2\ \text{cont}(B_i)$.	T7-11a. For any two basic sentences, B_i and B_j, with distinct primitive predicates, $\text{inf}(B_j/B_i) = 1 = \text{inf}(B_i)$.
T6-15b. Let B_1, B_2, ..., B_n be basic sentences with n distinct primitive predicates. Let C_m be the conjunction of the first m of them. Then, for every m (m = 2, ..., n - 1), $\text{cont}(B_{m+1}/C_m) = \frac{1}{2^{m+1}}$.	T7-11b. Let B_1, B_2, ..., B_n be basic sentences with n distinct primitive predicates. Let C_m be the conjunction of the first m of them. Then, for every m (m = 2, ..., n - 1), $\text{inf}(B_{m+1}/C_m) = 1$.
T6-6h. cont(j/i) = cont(j) iff i and j are L-disjunct.	T7-12a. inf(j/i) = inf(j) iff i and j are inductively independent.
T6-7. cont(j/i) $\leqslant$ cont(j).	
T6-4*l*. cont(~i) = 1 - cont(i).	T7-8*l*. $\text{inf}(\sim i) = \text{inf}(i) - \text{Log}(2^{\text{inf}(i)} - 1)$.

We see that the conditions of additivity for cont and inf are entirely different. This divergence is not surprising at all. On the contrary, dissatisfaction with the condition of additivity stated for cont in T6-4k was one of the reasons for our search for another explicatum of amount of information. It is of some psychological interest to notice that common sense would probably prefer T7-9 to T6-4k, whereas inf has no property comparable to that exhibited by cont in T6-4m, a theorem that looks highly intuitive.

The counter-intuitiveness of the lack of a counterpart to T6-4m might be reduced by the following example. Consider a system with 6 primitive predicates, P_1 to P_6, hence with $2^6 = 64$ Q-properties. All proper m-functions have equal values for the 64 Q-sentences with the same individual constant, hence the value 1/64 for each. Let i be 'P_1a'. P_1 is the disjunction of the first 32 Q's. Hence $m_P(i) = 1/2$. Therefore $inf(i) = -Log\,(1/2) = 1$. Let M be a disjunction of 32 Q's, that is, of Q_1 and the last 31 Q's. Let j be 'Ma'. Then $m(j) = 1/2$ and $inf(j) = 1$. i.j is L-equivalent to 'Q_1a'; hence, it is a very strong sentence. $m(i.j) = 1/64$. $inf(i.j) = -Log\,(1/64) = 6$. This is three times as much as the sum of the inf-values of the two components. This result becomes plausible if we realize that i says merely that a has one of certain 32 Q's, but that by the addition of j, which by itself also says no more than that a has one of 32 Q's, our information about the situation is at once made completely specific; that is, it is specified as saying that a has one particular Q.

Continuing the comparison, we may dismiss the difference between T6-13a and T7-10a as inessential, the number 1 in T7-10a being only a matter of normalization. However, the differences between T6-13c and T7-10c, T6-15a and T7-11a, and T6-15b and T7-11b are decisive. Whereas the cont-value of a basic sentence relative to a conjunction of basic sentences with different primitive predicates is always less than its absolute cont-value and decreases, moreover, with the number of components in the conjunction, the inf-value of a basic sentence relative to such a conjunction is equal to its absolute inf-value and is therefore also independent of the number of components in this conjunction.

The relation between cont and inf [illegible] in the perhaps simplest and most striking fashion by the following pair of formulas which appear in the proofs of T7-13 and T7-14:

$$inf(h/e) = Log\, m_P(e) - Log\, m_P(e.h) \quad (1)$$

$$cont(h/e) = \quad m_P(e) - \quad m_P(e.h). \quad (2)$$

For the tautological evidence t, we get

$$inf(h/t) = inf(h) = -Log\, m_P(h) \quad (3)$$

and

$$cont(h/t) = cont(h) = 1 - m_P(h), \quad (4)$$

formulas that are nothing else than variants of T7-5b and T6-3a but look now much more akin, especially if we write (3) as

$$\inf(h/t) = \inf(h) = \text{Log } 1 - \text{Log } m_P(h). \tag{3'}$$

Let us illustrate the relation between cont and inf also in the following numerical example.

Let B_1, B_2, ... be basic sentences with distinct primitive predicates. Let C_1 be B_1, C_2 be $B_1.B_2$, ..., C_n be $B_1.B_2. \ldots .B_n$, ... Then cont and inf have the following values for these C, according to T6-13c and T7-10c:

Table III

C_i	cont(C_i)	inf(C_i)
C_1	1/2	1
C_2	3/4	2
C_3	7/8	3
C_4	15/16	4
.	.	.
.	.	.
.	.	.
C_n	(n-1)/n	n
.	.	.
.	.	.
.	.	.

§9. D-functions and I-functions

Not all m_P-functions can be regarded as equally adequate explicata of initial inductive probability. It seems that only those which fulfill the additional requirement of instantial relevance – m_I-functions – are adequate for ordinary scientific purposes, whereas that m_P-function which exhibits instantial irrelevance – m_D – has properties which make it suitable for situations where inductive reasoning is of minor importance. Computations with m_D and the in-functions based upon it, are relatively easy due to the fact that m_D assigns equal values to all state-descriptions. One consequence is, for instance, that a sufficient condition for $m_D(i.j) = m_D(i) \times m_D(j)$ is already that i and j should have no atomic sentences in common, whereas only the much stronger condition that i and j should have no primitive predicates in common is sufficient for the corresponding theorem concerning m_I.

Not every m_P can serve as a basis for an inductive method that is in agreement with customary scientific procedures (cf. [Cont.] §2). There is at least one additional requirement for the adequacy of an m-function to serve as an explicatum for (absolute, initial) inductive probability. This is

R9-1. Requirement of instantial relevance. Let 'M' be a factual, molecular predicate. Let e be any non-L-false molecular sentence. Let i and h be full sentences of 'M' with two distinct individual constants which do not occur in e. Then

$$\frac{m(e.i.h)}{m(e.i)} > \frac{m(e.h)}{m(e)}.$$

(This can be formulated more simply in terms of 'c' as

$$c(h, e.i) > c(h, e).)$$

The requirement says, in effect, that one instance of a property is positively relevant to (the prediction of) another instance of the same property. This seems a basic feature of all inductive reasoning concerning the prediction of a future event.

We therefore define inductive m-function (in the narrower sense), to be denoted by 'm_I', as

D9-1. m is an inductive m-function $=_{Df}$ m is an m_P and fulfills R1.

Among the proper m-functions which do not fulfill R1, there is one which fulfills, so to speak, a requirement of instantial irrelevance. For this m-function, to be denoted by 'm_D' ('D' for 'deductive' since this function plays a special role in deductive logic), observed instances of a molecular property have no influence on the prediction of future instances of this property. Experience cannot teach us anything about the future if this function is applied. It has, nevertheless, great importance: its definition is of extreme simplicity, calculations on its basis are relatively easy, and results obtained by its use may have at least approximative value in cases where experience is estimated to be of little or no influence.

The definition of m_D incorporates a principle which looks very plausible to untrained common sense, viz. the principle of assigning equal m-values to all state-descriptions. It is of some psychological interest that this rather obvious procedure should lead to an inductive method that is unacceptable as a final method. (The function designated here by 'm_D' has been denoted by '$m^{\dagger}$' in [Prob.] §100A and by 'm_∞' in [Cont.] §13.)

We define

D9-2.

a. For every Z_i, $m_D(Z_i) =_{Df} 1/z$.

b. For every L-false sentence j, $m_D(j) =_{Df} 0$.

c. For every non-L-false sentence j, $m_D(j) =_{Df}$ the sum of the m_D-values for the Z in R(j); this is r(j)/z, where r(j) is the number of the state-descriptions in R(j).

It can easily be verified that m_D fulfills conditions D6-1a through D6-1g and D6-1i. That m_D also fulfills D6-1h and is therefore an m_P-function, follows from the much stronger theorem T9-1 below.

T9-1. If i and j have no atomic sentences in common, then

$$m_D(i.j) = m_D(i) \times m_D(j).$$

Proof: Let K_1 be the class of those atomic sentences which occur in i, K_2 the class of those atomic sentences which occur in j, K_3 the class of all other atomic sentences. Let C_1 be the class of those conjunctions which contain, for each atomic sentence in K_1, either it or its negation, but not both nor any other component. Let C_2 and C_3 be determined analogously with respect to K_2 and K_3. Let c_1 be the number of the conjunctions in C_1. Let c_2 and c_3 be determined analogously with respect to C_2 and C_3. (However, if C_3 is empty, let $c_3 = 1$.) Each Z is a conjunction of three conjunctions (disregarding the order) belonging respectively to C_1, C_2, and C_3. Therefore

$$z = c_1 \times c_2 \times c_3. \tag{1}$$

Let $c_1(i)$ be the number of those conjunctions in C_1 which L-imply i, and let $c_2(j)$ be the number of those conjunctions in C_2 which L-imply j. (Notice that i cannot be L-implied by any conjunction of C_2 or C_3, nor can j be L-implied by any conjunction of C_1 or C_3.) Therefore

$$r(i) = c_1(j) \times c_2 \times c_3 \tag{2}$$

and

$$r(j) = c_2(j) \times c_1 \times c_3. \tag{3}$$

But for the same reason we have also

$$r(i.j) = c_1(i) \times c_2(j) \times c_3. \tag{4}$$

From (2) and (3) we get

$$\begin{aligned} r(i) \times r(j) &= c_1(i) \times c_2 \times c_3 \times c_2(j) \times c_1 \times c_3 \\ &= r(i.j) \times c_1 \times c_2 \times c_3 \quad \text{(from (4))} \\ &= r(i.j) \times z \quad \text{(from (1))}. \end{aligned} \tag{5}$$

Dividing by z^2 we get finally

$$\frac{r(i)}{z} \times \frac{r(j)}{z} = \frac{r(i.j)}{z} \tag{6}$$

from which the assertion follows, by D2c.

Since m_D is a m_P-function, all theorems stated in §6 for m_P hold also for m_D. But some of them, having conditional form, can be strengthened by weakening the antecedent. We get, for instance, in analogy to T6-9,

T9-2. If i and j have no atomic sentence in common, then

$$m_D(i \vee j) = m_D(i) + m_D(j) - m_D(i) \times m_D(j)$$

and in analogy to T6-10b,

T9-3. For any conjunction, C_n, of n basic sentences with n distinct atomic sentences, $m_D(C_n) = (1/2)^n$.

For that cont-function which is based on m_D according to D6-2, $cont_D$, we have

T9-4.

a. For every E_i, $cont_D(E_i) = 1/z$.

b. For every sentence j, $cont_D(j) = n/z$, where n is the number of E which belong to Cont(j).

$cont_D$ has advantages and disadvantages similar to those of m_D. T1b points to the extreme simplicity, at least in principle, of its computation.

All theorems on cont stated in §6 also hold, of course, for $cont_D$ and all $cont_I$, the cont-functions defined on the basis of the m_I, analogously to D3. With respect to $cont_I$, no additional theorems in the form of equalities can be derived from R1. We shall not care to derive some inequalities from R1 and the previous theorems, especially since we shall treat later (§10) at some length a numerical example based on a specific

cont_I-function.

With respect to cont_D, however, various theorems holding for cont_P can be strengthened by weakening the condition in the antecedent, in complete analogy to the relation between m_D and m_P. T6-9, T6-10b, T6-11, T6-12, T6-13b, c, e, f, T6-15, and T6-16 hold for cont_D, even if the expression 'primitive predicate(s)' in their antecedents is replaced by 'atomic sentence(s)'. That this should be so is plausible in view of T1. But it is also easy to check the truth of our general assertion by inspecting the proofs of these theorems.

Let us state, however, also one theorem which is not a counterpart of a previous theorem:

T9-5.

a. For every conjunction, C_n, of n distinct E, $\text{cont}_D(C_n) = n/z$ (T4b).

b. For every conjunction, C_n, of n distinct E, different from E_i,

$$\text{cont}_D(E_i/C_n) = 1/z \text{ (D6-2, (a)).}$$

The relation between inf_D, defined on the basis of cont_D following D7-1, and inf is the same as that between cont_D and cont. We shall therefore state only those theorems which are based on T4 and T5.

T9-6.

a. For every E, $\text{inf}_D(E) = \beta - \text{Log}(z-1)$.

Proof: $\text{inf}_D(E) = \text{Log}\,\frac{1}{1-\frac{1}{z}}$ (T4a) $= \text{Log}\,\frac{z}{z-1} = \beta - \text{Log}(z-1)$ (T2-1c).

b. For every conjunction, C_n, of n distinct E, $\text{inf}_D(C_n) = \beta - \text{Log}(z-n)$.

Proof: $\text{inf}_D(C_n) = \text{Log}\,\frac{1}{1-\frac{n}{z}}$ (T5a) $= \text{Log}\,\frac{z}{z-n} = \beta - \text{Log}(z-n)$ (T2-1c).

c. For every conjunction, C_n, of n distinct E, different from E_i,

$$\text{inf}(E_i/C_n) = \text{Log}(z-n) - \text{Log}(z-n-1) \text{ (D7-2, (b)).}$$

According to the correlate of T7-10e, $\text{inf}_D(i) = n - \text{Log}\, m$, where i has disjunctive normal form: $C_1 \vee C_2 \vee \ldots \vee C_m$, each C being a conjunction of basic sentences with n distinct atomic sentences, the same n atomic sentences occurring in all conjunctions. According to a well-known theorem in the sentential calculus, there exists for every molecular sentence a sentence in disjunctive normal form L-equivalent to it (see, for instance, [Prob.] D21-2). It follows that for every molecular sentence i, $\text{inf}_D(i)$ has the form n - Log m, where both n and m are integers. Hence it is easy to calculate the inf_D-value of any molecular sentence. Such a sentence has to be transformed into one of its disjunctive normal forms, according to some standard procedure available for this purpose. Then the number of its components has to be counted as well as the number of atomic sentences in one of these components. Finally, a table for Log m, for integer m, will have to be consulted and a simple subtraction performed. For purposes of reference, such a table is given here for some selected integral values of m.

Table IV

m	Log m	m	Log m	m	Log m
1	0.0000	39	5.2853	57	5.8328
2	1.0000	40	5.3219	58	5.8579
3	1.5849	41	5.3575	59	5.8826
4	2.0000	42	5.3923	60	5.9068
5	2.3219	43	5.4262	61	5.9307
6	2.5849	44	5.4594	62	5.9541
7	2.8073	45	5.4918	63	5.9772
8	3.0000	46	5.5235	64	6.0000
9	3.1699	47	5.5545	100	6.6438
10	3.3219	48	5.5849	128	7.0000
16	4.0000	49	5.6147	250	7.9657
32	5.0000	50	5.6438	251	7.9715
33	5.0443	51	5.6724	252	7.9772
34	5.0874	52	5.7004	253	7.9829
35	5.1292	53	5.7279	254	7.9886
36	5.1699	54	5.7548	255	7.9943
37	5.2094	55	5.7813	256	8.0000
38	5.2479	56	5.8073	1000	9.9657

Let the E in our $\mathscr{L}_3^2$ be E_1, E_2, ..., E_{64}. Let C_m be the conjunction of the first m E. Then $cont_D(C_m) = m/64$ (T2a) and $inf_D(C_m) = 6 - Log(64-m)$ (T3b). Table V gives the values of the absolute and relative $cont_D$ for the first six values of m and for the last six values of m.

We see from this that if a series of messages is received, each being an E, then $cont_D$ grows by every one of these messages by the same amount, namely, by 1/64, from 0 to 1. inf_D, however, behaves in a different way. It grows from 0 to ∞ by unequal amounts. The first message contributes only a small fraction. Every further message contributes a little more than the preceding one. The last-but-three message contributes less than 1/2. The last-but-two contributes more than 1/2. The last-but-one contributes 1. And the last message contributes ∞. This behavior of inf_D becomes plausible when we realize that the different messages, although each of them is an E, nevertheless play different roles in the series of messages. When we have received sixty messages (in other words, when we have the knowledge C_{60}), then we know that sixty of the sixty-four possible states of the universe are excluded. There still remain four possible states; that is, our knowledge C_{60} means that the universe is in one of the four remaining states. The sixty-first message excludes among these four possible states a further one; hence, the range of those that are still open decreases from four to three. By the sixty-second message the range is further decreased from three to

Table V

m	$cont_D(C_m)$	$cont_D(E_m/C_{m-1})$	$inf_D(C_m)$	$inf_D(E_m/C_{m-1})$
1	1/64	1/64	0.0228	0.0228
2	1/64	2/64	0.0459	0.0231
3	1/64	3/64	0.0693	0.0234
4	1/64	4/64	0.0931	0.0238
5	1/64	5/64	0.1174	0.0242
6	1/64	6/64	0.1421	0.0247
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
59	1/64	59/64	3.6781	0.2630
60	1/64	60/64	4.0000	0.3219
61	1/64	61/64	4.4151	0.4151
62	1/64	62/64	5.0000	0.5849
63	1/64	63/64	6.0000	1.0000
64	1/64	1	∞	∞

two, and this may well be regarded as a stronger addition to our knowledge than the decrease from four to three. At this moment, only two possibilities are left open. The sixty-third message gives us information concerning which of these two remaining sentences is the actual one and hence completes our knowledge of the universe. Thus, this step has a great weight, more than any of those prior to it. After this step, nothing can be added to our knowledge in a consistent way. The sixty-fourth message is incompatible with the conjunction of the sixty-three preceding ones. If this message is nevertheless added, then this is a still more weighty step which leads to contradiction. The strongest factual message, which is a state-description, a conjunction of 6 basic sentences, carries 6 units of information as measured by inf_D. The only messages that carry more units of information, and then by necessity infinitely many such units, are the messages that contradict either themselves or prior messages.

§10. $cont^*$ and inf^*

Two special $cont_I$ and inf_I functions, $cont^*$ and inf^*, are defined and theorems regarding them developed. These functions are based upon that m_I — m^* — which assigns equal values to all structure-descriptions, i.e., disjunctions of isomorphic state-descriptions. Since m^* seems to have a special status among the various m_I-functions, $cont^*$ and inf^* are deemed to be of special importance. Various computations and tables regarding these functions are presented.

We shall now define and investigate two special I-functions that might turn out to be of special importance. They are based on the function m^* defined in [Prob.] §110 essentially as a proper m-function which has the same values for all structure-descriptions, that is, disjunctions of isomorphic state-descriptions.

Recalling the definition of 'isomorphic sentences' given in §6, the reader will easily see that our $\mathscr{L}_3^2$ has exactly 20 structure-descriptions. Let the Z of $\mathscr{L}_3^2$, as presented in Table I, be $Z_1, Z_2, \ldots, Z_{64}$. Then the structure-descriptions 'T_1', 'T_2', ..., 'T_{20}', are:

T_1: Z_1
T_2: Z_2
T_3: Z_3
T_4: Z_4
T_5: $Z_5 \vee Z_6 \vee Z_7$
T_6: $Z_8 \vee Z_9 \vee Z_{10}$
T_7: $Z_{11} \vee Z_{12} \vee Z_{13}$
T_8: $Z_{14} \vee Z_{15} \vee Z_{16}$
T_9: $Z_{17} \vee Z_{18} \vee Z_{19}$
T_{10}: $Z_{20} \vee Z_{21} \vee Z_{22}$
T_{11}: $Z_{23} \vee Z_{24} \vee Z_{25}$
T_{12}: $Z_{26} \vee Z_{27} \vee Z_{28}$
T_{13}: $Z_{29} \vee Z_{30} \vee Z_{31}$
T_{14}: $Z_{32} \vee Z_{33} \vee Z_{34}$
T_{15}: $Z_{35} \vee Z_{36} \vee Z_{37}$
T_{16}: $Z_{38} \vee Z_{39} \vee Z_{40}$
T_{17}: $Z_{41} \vee Z_{42} \vee Z_{43} \vee Z_{44} \vee Z_{45} \vee Z_{46}$
T_{18}: $Z_{47} \vee Z_{48} \vee Z_{49} \vee Z_{50} \vee Z_{51} \vee Z_{52}$
T_{19}: $Z_{53} \vee Z_{54} \vee Z_{55} \vee Z_{56} \vee Z_{57} \vee Z_{58}$
T_{20}: $Z_{59} \vee Z_{60} \vee Z_{61} \vee Z_{62} \vee Z_{63} \vee Z_{64}$

For all i, $m^*(T_i) = 1/20$, hence $m^*(Z_1) = m^*(Z_2) = m^*(Z_3) = m^*(Z_4) = 1/20$, $m^*(Z_5) = \ldots = m^*(Z_{40}) = 1/60$, $m^*(Z_{41}) = \ldots = m^*(Z_{64}) = 1/120$.

In [Cont.] §18, an argument is presented which shows that the function c^* based on m^* is in a certain sense simpler than other c_P-functions. Explicata for amount of information based on m^* would share this special status.

$c^*(h, e)$, $cont^*(e)$, $cont^*(h/e)$, $inf^*(e)$, and $inf^*(h/e)$ can all be expressed as simple functions of $m^*(e)$ and $m^*(e.h)$:

$$c^*(h, e) = \frac{m^*(e.h)}{m^*(e)} \text{ (D7-3)}. \quad (1)$$

$$cont^*(e) = 1 - m^*(e) \text{ (T6-3a)}. \quad (2)$$

$$cont^*(h/e) = m^*(e) - m^*(e.h) \text{ (T6-8a)}. \quad (3)$$

$$\inf^*(e) = -\mathrm{Log}\, m^*(e) \text{ (T7-5b)}. \tag{4}$$

$$\inf^*(h/e) = \mathrm{Log}\, m^*(e) - \mathrm{Log}\, m^*(e.h) \text{ (formula (1) in §8)}. \tag{5}$$

Let e be 'Ma. Mb' and h be 'Mc'. Then, by inspection of Table I, we see that

$$m^*(\text{'Ma. Mb'}) = 2 \times \frac{1}{20} + 10 \times \frac{1}{60} + 4 \times \frac{1}{120} = 0.3.$$

Notice that $m_D(e) = 0.25$. The larger value of m^* is due to instantial relevance. We also have

$$m^*(\text{'Ma. Mb. Mc'}) = 2 \times \frac{1}{20} + 6 \times \frac{1}{60} = 0.2.$$

Hence

$$c^*(\text{'Mc', 'Ma. Mb'}) = \frac{0.2}{0.3} = \frac{2}{3}.$$

$$\mathrm{cont}^*(\text{'Ma. Mb'}) = 0.7.$$

$$\mathrm{cont}^*(\text{'Mc'/'Ma. Mb'}) = 0.3 - 0.2 = 0.1;$$

On the other hand,

$$\mathrm{cont}_D(\text{'Mc'/'Ma. Mb'}) = 0.125.$$

$$\inf^*(\text{'Ma. Mb'}) = -\mathrm{Log}\, 0.3 = \mathrm{Log}\, 10 - \mathrm{Log}\, 3 = 1.7370,$$

as against an $\inf_D$-value of 2. Finally,

$$\inf^*(\text{'Mc'/'Ma. Mb'}) = \mathrm{Log}\,\frac{0.3}{0.2} = \mathrm{Log}\, 3 - \mathrm{Log}\, 2 = 0.5849$$

whereas the corresponding relative $\inf_D$-value is 1.

It might perhaps be worthwhile to investigate now another sample language, this time with only one primitive predicate and n distinct individual constants. In this case, c^* yields the same values as Laplace's rule of succession. (See [Prob.] §110E.) Let e be a conjunction of s < n basic sentences with s distinct individual constants, among them s_1 atomic sentences with 'P' and $s-s_1$ negations of such. Let h be ' ' where 'b' is an individual constant not occurring in e. Then the following holds (according to formula (4), [Prob.] p. 566, cf. remark to D6-1i).

$$m^*(e) = \frac{s_1!\,(s-s_1)!}{\quad} = \frac{1}{(s+1)\binom{s}{s_1}}. \tag{6}$$

$$m^*(e.h) = \frac{(s_1+1)!\,(s-s_1)!}{(s+2)!} = m^*(e) \times \frac{s_1+1}{s+2}. \tag{7}$$

$$c^*(h,e) = \frac{s_1+1}{s+2} \text{ ((1), (6), (7))}. \tag{8}$$

$$\mathrm{cont}^*(h/e) = m^*(e) \times \left(1 - \frac{s_1+1}{s+2}\right) \text{ ((3), (6), (7))} = m^*(e) \times \frac{s-s_1+1}{s+2}. \tag{9}$$

To have a numerical example, assume s=10. We get

$$m^*(e) = \frac{1}{11\binom{10}{s_1}}. \quad (10)$$

$$m^*(e.h) = m^*(e) \times \frac{s_1+1}{12}. \quad (11)$$

$$c^*(h, e) = \frac{s_1+1}{12}. \quad (12)$$

$$cont^*(h/e) = m^*(e) \times \frac{11-s_1}{12}. \quad (13)$$

$$inf^*(e) = Log \frac{1}{m^*(e)}. \quad (14)$$

$$inf^*(h/e) = Log \frac{12}{s_1+1}. \quad (15)$$

The values given in Table VI are calculated according to these formulas.

Table VI

s_1	$m^*(e)$	$m^*(e.h)$	$c^*(h, e)$	$cont^*(e)$	$cont^*(h/e)$	$inf^*(e)$	$inf^*(h/e)$
0	0.09091	0.0076	0.0833	0.90909	0.08333	3.459	3.585
1	0.00909	0.0015	0.1667	0.99091	0.00758	6.781	2.585
2	0.00202	0.0005	0.2500	0.99798	0.00152	8.751	2.000
3	0.00076	0.0003	0.3333	0.99924	0.00051	10.366	1.585
4	0.00043	0.0002	0.4167	0.99957	0.00025	11.174	1.263
5	0.00036	0.0002	0.5000	0.99964	0.00018	11.437	1.000
6	0.00043	0.0003	0.5833	0.99957	0.00018	11.174	0.778
7	0.00076	0.0005	0.6666	0.99924	0.00025	10.366	0.585
8	0.00202	0.0015	0.7500	0.99798	0.00051	8.751	0.415
9	0.00909	0.0076	0.8333	0.99091	0.00152	6.781	0.263
10	0.09091	0.0833	0.9167	0.90909	0.00758	3.459	0.126

In addition to these formulas, we have, of course, also

$$m^*(h) = \frac{1}{2}. \quad (16)$$

$$cont^*(h) = \frac{1}{2}. \quad (17)$$

$$inf^*(h) = 1. \quad (18)$$

A few comments on Table VI might be indicated. The columns for $m^*(e)$ and $m^*(e.h)$ show that this m_I-function, as is to be expected from any adequate m_I-function, puts a premium on homogeneity; that is, those states for which the absolute difference between the individuals having P and those not having P is higher, are treated as initially more probable. When the evidence states that 5 individuals have P and 5 others do not have P, the last column shows that the inf^*-value of our hypothesis, which states that an eleventh individual has P, is just 1. Hence it is the same as the absolute inf^*-value of this hypothesis. The greater the number of individuals having P, according to the evidence, the larger $c^*(h, e)$ and the smaller $inf^*(h/e)$. $cont^*(h/e)$, however, behaves differently. It reaches its minimum for intermediate values of s_1 but increases both when s_1 increases from 6 to 10 and when it decreases from 5 to 0.

§ 11. Estimates of amount of information

A scientist is often interested in the expectation-value of the amount of information conveyed by the outcome of an experiment to be made. If the various possible outcomes can be expressed by $h_1, h_2, \ldots, h_n$, such that these sentences are pairwise exclusive and their disjunction L-true on the given evidence e, in short, when $H = \{h_1, h_2, \ldots, h_n\}$ is an exhaustive system on e, the estimate of the amount of information carried by H with respect to e is given by the formula

$$est(in, H, e) = \sum_{p=1}^{n} c(h_p, e) \times in(h_p/e).$$

Various formulas for est(cont, H, e), est(inf, H, e), and other functions based upon them are derived. The concepts of posterior estimate of amount of information, amount of specification, estimate of the posterior estimate, and estimate of the amount of specification are defined, and various theorems concerning them proved. A simple illustrative application is given.

If an experiment is performed, the possible results of which are expressed in n sentences $h_1, \ldots, h_n$ (or in n sentences L-equivalent to them), we can compute the amounts of information which each possible outcome would convey, assuming that an m-function has been defined for all the sentences of the language in which the h's are formulated. So long as the actual outcome is not known, the amount of information it carries is also unknown. But, for certain purposes, it is important to have a good estimate of this amount. The situation is analogous to that existing very often in scientific investigations, where a certain magnitude is unknown and one has to work instead with an estimate of this magnitude.

To give a crude but sufficiently illustrative example: Imagine a thermometer which is divided rather unconventionally into three regions so that in region 1 the pointer indicates Warm, in region 2 Temperate, and in region 3 Cold. Let the thermometer be read in a place where, according to available evidence, most past readings indicated Cold, some Temperate, and only very few Warm. Since the same distribution (approximately) is expected for future readings, an adequate measure of information will assign to the sentence 'Cold(t_1)' (where t_1 is a time-point in the future, that is, one not mentioned in the evidence) a lower value, relative to the evidence, than to 'Temperate(t_1)' which again will have a lower value than 'Warm(t_1)'. Let these sentences be h_1, h_2, and h_3, respectively. What would be a reasonable estimate of the amount of information a future observation is expected to carry? One might at first think of taking the arithmetic mean of the three amounts of information, that is,

$$\frac{in(h_1/e) + in(h_2/e) + in(h_3/e)}{3},$$

but a little reflection will show that this would be utterly inadequate. The amounts have to be weighted differently. It seems rather natural to take as appropriate weights here, as well as in general, the degrees of confirmation which the sentences h_1, h_2, and h_3 have on the available evidence. (For a more thorough discussion of this procedure, see [Prob.] Chap. ix.) We arrive, therefore, at the value

$$c(h_1, e) \times in(h_1/e) + c(h_2, e) \times in(h_2/e) + c(h_3, e) \times in(h_3/e),$$

or, in the convenient customary shorthand,

$$\sum_{p=1}^{3} c(h_p, e) \times in(h_p/e).$$

Expressions of this type are well known in the theory of probability and statistics (with the degree-of-confirmation subformula usually replaced by a corresponding relative-frequency formula) under the name 'the mathematical expectation (or hope) of ...', in our case, '... of the amount of information carried by the observation to be made at t_1'.

In general, whenever we have a class of sentences $H = \{h_1, \ldots, h_n\}$ such that the available evidence e L-implies $h_1 \vee h_2 \vee \ldots \vee h_n$ as well as $\sim(h_i . h_j)$, for all $i \neq j$, we shall say that H is an exhaustive system relative to e, and the expression

$$\sum_{p=1}^{n} c(h_p, e) \times in(h_p/e)$$

will be called 'the (c-mean) estimate of the amount of information carried by (the members of) H with respect to e', symbolized by 'est(in, H, e)'.

So far, our discussion has been proceeding on a partly presystematic, partly systematic level. To switch to a completely systematic treatment, we obviously have only to replace the explicandum 'in' by one or the other of its explicata. We define:

D11-1. Let H, h_p, e be as above. Then

$$est(cont, H, e) =_{Df} \sum_{p} c(h_p, e) \times cont(h_p/e).$$

D11-2. Let H, h_p, e be as above. Then

$$est(inf, H, e) =_{Df} \sum_{p} c(h_p, e) \times inf(h_p/e).$$

E (example) 11-1. Let, for example, with respect to our $\mathscr{L}_3^2$, h_1 = 'Mc', h_2 = 'Fc', H = $\{h_1, h_2\}$, e = 'Ma. Mb'. On the basis of Table I, some formulas in the preceding section, and the two following formulas which the reader will easily be able to derive for himself, namely,

$$\text{cont}^*(\text{'Mc'}/\text{'Ma. Mb'}) = 0.1$$

and

$$\text{cont}^*(\text{'Fc'}/\text{'Ma. Mb'}) = 0.2,$$

we obtain now

$$\text{est}(\text{cont}^*, H, e) = \frac{2}{3} \times 0.1 + \frac{1}{3} \times 0.2 = 0.133$$

and

$$\text{est}(\text{inf}^*, H, e) = -\left(\frac{2}{3} \times \text{Log}\,\frac{2}{3} + \frac{1}{3} \times \text{Log}\,\frac{1}{3}\right) = 0.918.$$

$\text{est}(\text{inf}_D, H, e)$, on the other hand, equals 1, of course.

E11-2. Let h_1, h_2, and H be as before, but let now

e = 'Ma. Mb. Ya. Yb. Yc'.

Then

$$m^*(e) = \frac{1}{20} + \frac{1}{60} = \frac{1}{15},$$

$$m^*(e.h_1) = \frac{1}{20},$$

$$m^*(e.h_2) = \frac{1}{60}.$$

Hence

$$\text{cont}^*(h_1/e) = \frac{1}{60},$$

$$\text{cont}^*(h_2/e) = \frac{1}{20},$$

$$\text{inf}^*(h_1/e) = 0.4151,$$

$$\text{inf}^*(h_2/e) = 2,$$

$$c^*(h_1, e) = \frac{3}{4},$$

and

$$c^*(h_2, e) = \frac{1}{4}.$$

Hence

$$\text{est}(\text{cont}^*, H, e) = \frac{3}{4} \times \frac{1}{60} + \frac{1}{4} \times \frac{1}{20} = \frac{1}{40},$$

and

$$\text{est}(\text{inf}^*, H, e) = \frac{3}{4} \times 0.4151 + \frac{1}{4} \times 2 = 0.811$$

(whereas $\text{est}(\text{inf}_D, H, e)$ equals 1).

For the following theorems it is always assumed that H, h_p, and e fulfill the above-mentioned conditions.

T11-1.

$$\text{est}(\text{cont}, H, e) = \sum_p \frac{m(h_p . e)}{m(e)} \times m(e) \times (1 - c(h_p, e)) \quad (\text{T7-14})$$

$$= \sum m(h_p . e) \times (1 - c(h_p, e))$$

$$= \sum m(h_p . e) \times c(\sim h_p, e)$$

$$= \frac{1}{m(e)} \sum m(h_p . e) \times m(\sim h_p . e)$$

$$= \frac{1}{m(e)} \sum m(h_p . e) \times (1 - m(h_p . e))$$

$$= \sum c(h_p, e) \times m(\sim h_p . e)$$

$$= c(e, t) \sum c(h_p, e) \times c(\sim h_p, e).$$

Let $K = \{k_1, \ldots, k_n\}$ be an exhaustive system with respect to e. Then from well-known theorems in the theory of inequalities, the following theorem can be derived:

T11-2. Let $c(k_i, e) = c(k_j, e)$ for all i and j (hence $= 1/n$), and let there be at least one pair i and j such that $c(h_i, e) \neq c(h_j, e)$. Then

$$\text{est}(\text{cont}, K, e) > \text{est}(\text{cont}, H, e).$$

T11-3. For fixed n, $\text{est}(\text{cont}, H_i, e)$ is a maximum for those H_i all whose members have the same c-values on e. Hence

$$\max_i[\text{est}(\text{cont}, H_i, e)] = m(e) \times \frac{n-1}{n}.$$

(This is, of course, also the cont-value of each h_p^i belonging to these H_i.)

T11-4. For fixed n, $\text{est}(\text{cont}, H_i, e)$ is a minimum for those H_i one member of which has the c-value 1 on e (and hence all the other members the c-value 0 on e); hence

$$\min_i[\text{est}(\text{cont}, H_i, e)] = 0.$$

Theorems similar to T2, T3, and T4 can be obtained for the second explicatum inf. Let us first state a transformation of D2, according to T7-13:

T11-5.

$$\text{est}(\text{inf}, H, e) = \sum c(h_p, e) \times \text{Log} \frac{1}{c(h_p, e)}$$

$$= -\sum c(h_p, e) \times \text{Log}\, c(h_p, e).$$

We now get

T11-6. Let $c(k_i, e) = c(k_j, e)$ for all i and j (hence $= 1/n$), and let there be at least one pair i and j such that $c(h_i, e) \neq c(h_j, e)$. Then

$$\text{est}(\text{inf}, K, e) > \text{est}(\text{inf}, H, e).$$

T11-7. For fixed n, $\text{est}(\text{inf}, H_i, e)$ is a maximum for those H_i all whose members have the same c-values on e; hence

$$\text{max}_i\left[\text{est}(\text{inf}, H_i, e)\right] = \text{Log}\, n.$$

(This is, of course, also the inf-value of each h_p^i belonging to these H_i.)

T11-8. For fixed n, $\text{est}(\text{inf}, H_i, e)$ is a minimum for those H_i one member of which has the c-value 1 on e (and hence all the other members have the c-value 0 on e). Hence

$$\text{min}_i\left[\text{est}(\text{inf}, H_i, e)\right] = 0.$$

An expression analogous to

$$\text{'} -\sum c(h_p, e) \times \text{Log}\, c(h_p, e)\text{'}$$

but with degree of confirmation replaced by (statistical) probability, plays a central role in communication theory, as well as in certain formulations of statistical mechanics, where the probability concerned is that of a system being in cell p of its phase space. In statistical mechanics, in the formulation given it by Boltzmann, this expression is said to measure the entropy of the system. In analogy to this, some communication-theoreticians call the corresponding expression, which arises when the probabilities concerned are those of the (expected) relative frequencies of the occurrence of certain messages, the entropy of this system of messages. Other terms, used synonymously, though unfortunately without any real effort for terminological clarification, were uncertainty, choice, and even simply as well as confusingly, information.

Let H and K be exhaustive systems with respect to e, let H contain n members, and K contain m members. Let 'H.K' be short for '$\{h_1.k_1, h_1.k_2, \ldots, h_1.k_m, h_2.k_1, \ldots, h_n.k_m\}$'. Then we define

D11-3.

$$\text{est}(\text{in}, H.K, e) =_{Df} \sum_{q=1}^{m} \sum_{p=1}^{n} c(h_p . k_q, e) \times \text{in}(h_p . k_q / e).$$

With respect to the explicatum inf, the following theorem can be proved:

T11-9. est(inf, H.K, e) $\leq$ est(inf, H, e) + est(inf, K, e), where equality holds only if, for all p and q, $c(h_p . k_q, e) = c(h_p, e) \times c(k_q, e)$, in other words, when the h's and the k's are inductively independent on e (with respect to that m-function on which c is based).

E11-3. Let e = 'Ma. Mb. Ya. Yb', h_1 = 'Mc', h_2 = 'Fc', k_1 = 'Yc', k_2 = 'Oc', $H = \{h_1, h_2\}$, and $K = \{k_1, k_2\}$.

Then,

$$H.K = \{h_1.k_1,\ h_1.k_2,\ h_2.k_1,\ h_2.k_2\}.$$

We have

$$m^*(e) = \frac{1}{10},$$

$$m^*(h_1.k_1.e) = \frac{1}{20},$$

$$m^*(h_1.k_2.e) = m^*(h_2.k_1.e) = m^*(h_2.k_2.e) = \frac{1}{60},$$

$$c^*(h_1.k_1, e) = \frac{1}{2},$$

$$c^*(h_2.k_2, e) = c^*(h_2.k_1, e) = c^*(h_2.k_2, e) = \frac{1}{6},$$

$$\text{cont}^*(h_1.k_1/e) = \frac{1}{20},$$

$$\text{cont}^*(h_1.k_2/e) = \ldots = \frac{1}{12},$$

$$\text{inf}^*(h_1.k_1/e) = 1,$$

$$\text{inf}^*(h_1.k_2/e) = \ldots = 2.585.$$

Hence

$$\text{est}(\text{cont}^*, H.K, e) = \frac{1}{15}$$

and

$$\text{est}(\text{inf}^*, H.K, e) = 1.792.$$

$$\text{est}(\text{inf}^*, H, e) = \text{est}(\text{inf}^*, K, e) = 0.918.$$

We verify that

$$\text{est}(\text{inf}^*, H.K, e) < \text{est}(\text{inf}^*, H, e) + \text{est}(\text{inf}^*, K, e),$$

the h's and the k's not being inductively independent on this e with respect to m^*. They are, however, independent with respect to m_D, and indeed

$$\text{est}(\text{inf}_D, H.K, e) = 2 = \text{est}(\text{inf}_D, H, e) + \text{est}(\text{inf}_D, K, e).$$

In general, est(in, H, e) will be different from est(in, H, e. k), where k is a sentence that has been added to the prior evidence e. Since 'est(in, H, e. k)' and similar expressions are of great importance, it is worthwhile to give it a special name. We shall call it the posterior estimate (of the amount of information carried by H on the evidence comprised of e and k). The expression 'est(in, H, e)' will then be called, for greater clarity, the prior estimate (of ...). It is often important to investigate how such a prior estimate has been changed through some additional evidence. We shall therefore give also to the difference between the prior and the posterior estimate a special name, the amount of specification of H through k on e, and denote this function by a special symbol 'sp(in, H, k, e)':

D11-4. $\text{sp}(\text{in}, H, k, e) =_{Df} \text{est}(\text{in}, H, e) - \text{est}(\text{in}, H, e.k)$.

E11-4. Let e, H, and k_1 be as in E3. Then $e.k_1$ is the e of E2. Therefore $\text{est}(\text{inf}^*, H, e.k_1) = 0.811$. Since $\text{est}(\text{inf}^*, H, e) = 0.918$ (from E3), we have $\text{sp}(\text{inf}^*, H, k_1, e) = 0.918 - 0.811 = 0.107$.

It can be easily seen that sp(in, H, k, e) = 0 if (but not only if) k is inductively independent of the h's on e. Otherwise sp can be either positive or negative. Its maximum value is obviously equal to est(in, H, e) itself. This value will be obtained when e. k L-implies one of the h's. In this case H is maximally specified through k on e.

Situations often arise in which the event stated in k has not yet occurred or, at least, in which we do not know whether or not it has occurred but know only that either it or some other event belonging to an exhaustive system of events will occur or has occurred. In such circumstances, it makes sense to ask for the expectation value of the posterior estimate of the amount of information carried by H on e and (some member of the exhaustive system) K. We are led to the (c-mean) estimate of this posterior estimate which we shall denote by 'est(in, H/K, e)' and define as

D11-5.

$$\text{est}(\text{in}, H/K, e) =_{Df} \sum_{q=1}^{m} c(k_q, e) \times \text{est}(\text{in}, H, e.k_q).$$

E11-5. Let e, H, and K be as in E3. Then

$$\text{est}(\text{inf}^*, H/K, e) = \frac{2}{3} \times 0.811 + \frac{1}{3} \times 1 = 0.874.$$

The stroke-notation has been chosen instead of a more neutral comma-notation because the following theorem, which stands in a certain analogy to the definitions of relative amounts of information, holds.

T11-10. est(in, H/K, e) = est(in, H.K, e) - est(in, K, e).

Proof:

$$\text{est}(\text{in}, H/K, e) = \sum_q c(k_q, e) \sum_p c(h_p, e.k_q) \times \text{in}(h_p/e.k_q)$$

$$= \sum_q \sum_p c(k_q, e) \times c(h_p, e.k_q) \times \text{in}(h_p/e.k_q)$$

$$= \sum_q \sum_p c(h_p.k_q, e) \times \text{in}(h_p/e.k_q)$$

$$= \sum_q \sum_p c(h_p.k_q, e) \times \left[\text{in}(h_p.k_q/e) - \text{in}(k_q/e)\right]$$

$$= \text{est}(\text{in}, H.K, e) - \sum_q c(k_q, e) \times \text{in}(k_q/e)$$

$$= \text{est}(\text{in}, H.K, e) - \text{est}(\text{in}, K, e).$$

Indeed, est(inf*, H.K, e) - est(inf*, K, e) = 1.792 (from E3) - 0.918 (from E3) = 0.874 (as in E5).

One will often be interested in an estimate of the amount of specification of H on e through K. This function will be symbolized by 'sp(in, H, K, e)'. Its definition is

D11-6.

$$\text{sp}(\text{in}, H, K, e) =_{Df} \sum_q c(k_q, e) \times \text{sp}(\text{in}, H, k_q, e).$$

E11-6. Let e, H, and K be as in E3. Then

$\text{sp}(\text{inf}^*, H, K, e) = \frac{2}{3} \times 0.107$ (from E4) $+ \frac{1}{3} \times (-0.082)$ (computed in the same way)

$= 0.044.$

We see immediately that the following theorem holds:

T11-11. sp(in, H, K, e) = est(in, H, e) - est(in, H/K, e).

Indeed, est(inf*, H, e) - est(inf*, H/K, e) = 0.918 (E3) - 0.874 (E5) = 0.044 (as in E6).

Though it may happen that, for some q, $\text{sp}(\text{in}, H, k_q, e)$ is negative, it can be proved that sp(in, H, K, e) is never negative, in other words, that the estimate of the posterior estimate is at most equal to the prior estimate.

T11-12. $\text{sp}(\text{in}, H, K, e) \geqslant 0$, with equality holding iff the h's and the k's are inductively independent.

Combining T10 and T11, we get

T11-13. sp(in, H, K, e) = est(in, H, e) + est(in, K, e) - est(in, H. K, e).

From T13 follows immediately the following theorem of the symmetricity or mutuality of specification:

T11-14. sp(in, H, K, e) = sp(in, K, H, e).

To illustrate the importance and use of the functions defined in this section, let us work out a different numerical example, albeit an artificially simplified one, for ease of computation. Let h_1 be 'Jones is bright', h_2 be 'Jones is average (in intelligence)', and h_3 be 'Jones is dull'. Somebody who is interested in Jones's intelligence makes him undergo a certain test. Let now k_1 be 'Jones achieves more than 80 percent (in this test)', k_2 be 'Jones achieves between 60 percent and 80 percent', and k_3 be 'Jones achieves less than 60 percent'. Let the following degrees of confirmation hold on the available evidence, according to some m-function:

$$c(h_1, e) = c(h_3, e) = \frac{1}{4}$$

$$c(h_2, e) = \frac{1}{2}$$

$$c(k_1, e.h_1) = c(k_2, e.h_1) = c(k_2, e.h_2) = c(k_2, e.h_3) = c(k_3, e.h_3) = \frac{1}{2}$$

$$c(k_1, e.h_2) = c(k_3, e.h_2) = \frac{1}{4}\cdot$$

(All other $c(k_q, e.h_p) = 0$.) Figure 2 might help to visualize the situation.

For the following computations, the explicatum inf will be used. First we compute with the help of T5 the value of est(inf, H, e) in our example.

$$est(inf, H, e) = \frac{1}{4}\text{ Log } 4 + \frac{1}{2}\text{ Log } 2 + \frac{1}{4}\text{ Log } 4 = 1.5.$$

To evaluate est(inf, K, e) we have first to find the various $c(k_q, e)$. These can be easily read off the diagram.

$$c(k_1, e) = c(k_3, e) = \frac{1}{4},$$

$$c(k_2, e) = \frac{1}{2}.$$

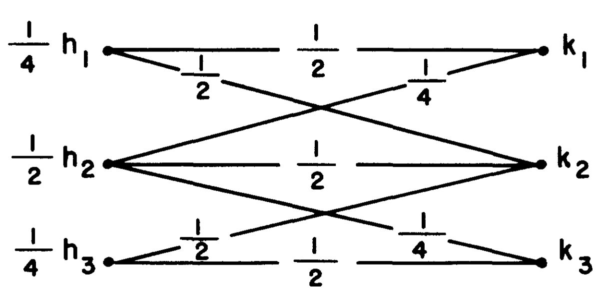

Fig. 2

Since $c(k_i, e) = c(h_i, e)$ for all i (this is pure coincidence), we have

$$est(inf, K, e) = 1.5.$$

For est(inf, H. K, e) we get, again by simple inspection of the diagram,

$$est(inf, H.K, e) = 6 \times \frac{1}{8} \text{Log } 8 + 1 \times \frac{1}{4} \text{Log } 4 = 2.75.$$

This verifies T9. It is obvious that not all h's and k's are inductively independent.

To find the various $est(inf, H, e.k_q)$, we compute first all $c(h_p, e.k_q)$. We get

$$c(h_1, e.k_1) = c(h_2, e.k_1) = c(h_2, e.k_2) = c(h_2, e.k_3) = c(h_3, e.k_3) = \frac{1}{2},$$

$$c(h_1, e.k_2) = c(h_3, e.k_2) = \frac{1}{4}.$$

(All other $c(h_p, e.k_q) = 0$.) Hence we have

$$est(inf, H, e.k_1) = est(inf, H, e.k_3) = 1,$$

$$est(inf, H, e.k_2) = 1.5.$$

Hence we get, according to D4,

$$sp(inf, H, k_1, e) = sp(inf, H, k_3, e) = \frac{1}{2},$$

$$sp(inf, H, k_2, e) = 0.$$

The last result is of special importance. And indeed, if Jones achieves between 60 percent and 80 percent in his test, we are "so klug als wie zuvor", we know exactly as much as we knew before. The addition of k_2 to our evidence left the c-values of the h's unchanged, k_2 is inductively irrelevant to the h's, and our knowledge has not become more specific through this addition. The situation is different with respect to the two other outcomes of the test. In both other cases, our knowledge has become more specific. This appears even on the qualitative level: Before the test, Jones could have been bright, average, or dull. After the test, we know that he is not dull if the outcome is k_1, and that he is not bright, if the outcome is k_3. But one has to be careful with this argument. A reduction of the number of possibilities does not always entail an increase of specifity of the situation. If the probability distribution of the remaining possibilities is much more evenly spread than that of the initial possibilities, the situation may become, in a certain important sense, less specific. Examples could be easily constructed. In our case, however, there is a real increase in specificity, though not a large one.

It seems reasonable to measure one aspect of the effectiveness of this intelligence test by the estimate of the amount of specification. One might compare the effectiveness of various proposed tests in this way. In our case, according to D6,

$$sp(inf, H, K, e) = \frac{1}{4} \times \frac{1}{2} + \frac{1}{2} \times 0 + \frac{1}{4} \times \frac{1}{2} = \frac{1}{4},$$

a result that could, of course, also have been obtained from T13. Incidentally, it follows that the test was a pretty poor one. Whereas a direct measurement of Jones's intelligence, were it only possible, could be expected to yield 1.5 units of information, the mentioned test can be expected to give us only 0.25 of a unit of information on Jones's intelligence. The difference between 1.5 and 0.25, i.e., 1.25, is the value of est(inf, H/K, e), according to T11. The same value would be obtained by using either D5 or T10. We may say that by applying the test instead of measuring the intelligence directly, we must content ourselves with expecting a "loss" of 1.25 units of information. The correlate of this function within communication theory has been called by Shannon* the equivocation. With fixed H, that test is more efficient whose K (the class of possible outcomes) yields the higher value for the estimate of the amount of specification of H on e through K, or the lower value for the estimate of the posterior estimate of the amount of information carried by H on e and K.

§12. Semantic noise, efficiency of a conceptual framework

Two usages of 'semantic noise' are distinguished and a more general concept of distortion through noise defined. Efficiency of the conceptual framework of a language is introduced, both with respect to some given evidence and absolutely. The symmetrical treatment of a predicate and its negation maximizes initial efficiency. With increasing evidence, the efficiency of a language generally decreases.

Whenever a receiver of a message is unable to reconstruct immediately the message as originally sent, the communication engineer describes the situation by saying that the message has been distorted by noise. To combat noise is one of his principal tasks.

Sometimes the receiver of a message, in spite of a reception which is physically free of distortion, reacts to it in a way which is different from that expected by the sender. Attempts have been made to formulate this situation in terms of semantic noise. Indeed, the same sentence (more exactly, two tokens of the same sentence-type) may convey different informations (with different or equal amounts of information) to two people (e.g. the sender and the receiver of a message) and this in at least two different ways: first, the two tokens, which are physically alike, are interpreted as belonging to different languages**, and second, probably more common and interesting, the information carried by them is evaluated with respect to different evidences. Misunderstandings may be due either to physical mishearing or to semantic misevaluation (or to both).

In addition to the two metaphorical usages of 'noise' mentioned above, which seem pretty straightforward and should cause no confusion if properly distinguished among themselves and from the engineer's noise, it seems natural to use this term also in the

*The Mathematical Theory of Communication, p. 36.

**See Charles F. Hockett: An Approach to the Quantification of Semantic Noise, Phil. Sci. 19, 257-260, 1952.

following general situations. Whenever one is interested in knowing whether a certain event out of an exhaustive system of events, H, has happened (or is going to happen) but is unable, for some reason, to observe directly the occurrence of these events and has to content oneself with the observation of some event out of another exhaustive system, K, where not all of the k_q are irrelevant to the h_p on e (cf. [Prob.] §65), one can regard K as a distortion or a transformation through noise of H.

Following this usage, we may not only say that the system of sounds coming out of a telephone receiver is a distortion through noise of the systems of sounds coming out of the mouth of the speaker and that the system of symbol printings at the output of a teletypewriter is a distortion of the system of symbol printings at the input, but also that the system of positions of a thermometer at a certain time is a distortion of the system of the temperature situations at those times (for somebody who is interested in the temperatures), that the system of weather predictions of a certain weather bureau is a distortion of the system of weather situations at the times for which the predictions are made (for somebody who is interested in the weather), and that the system of IQ-test results is a distortion of the system of intelligence characteristics (for somebody interested in these characteristics).

Whether it is worthwhile, in the three last examples and in similar cases, to talk about nature communicating with us and about our receiving nature's messages in a noise-distorted fashion in order to drive home a useful analogy, is questionable. Some heuristic value to such a form of speech can hardly be denied, but the strain such usage would put upon terms like 'communication' or 'message' might well be too high.

The twin concepts of code-efficiency and code-redundancy play an important role in communication theory. We shall not discuss here the definitions given these concepts nor dwell on their various applications (and misapplications) but give instead definitions for certain semantic correlates which seem to have some importance.

By the efficiency of (the conceptual framework of) the language L_1, with respect to (the amount-of-information function) in and (the evidence) e, in symbols: $ef(L_1, in, e)$, we understand the ratio of $est(in, H_1, e)$, where H_1 is the class of the full sentences of all Q-predicators (§2) with an argument not mentioned in e, to $\max_i[est(in, H_i, e)]$, where the H_i are the corresponding classes in other languages L_i covering, intuitively speaking, the same ground.* It seems to us that the choice of the class of the Q-sentences as the class relative to which the efficiency of a language is defined is a natural one, though it certainly is not the only plausible one. The efficiency of a language, as defined here, changes, as a function of e, with a change of the evidence taken into account. A language

*This loose statement is in need of much elaboration. This is expected to be achieved at a later stage. We have in mind that the languages L_i refer to the same physical magnitudes without, however, there existing a sentence-by-sentence translatability between them.

may become, in a sense, more or less efficient with a change in experience.

> For an inhabitant of New York, a language with the predicates 'W', 'T', and 'C', designating Warm (above 75° F.), Temperate (between 40° and 75°), and Cold (below 40°), respectively, would be quite efficient. Should he move to San Francisco, however, its efficiency would be highly reduced, because 'T' occurs here much more frequently than the other two.

We would, therefore, like to have also a concept of efficiency that is independent of experience. Such a concept is, of course, readily available. We have only to consider the efficiency relative to the tautological evidence, i.e., $ef(L_1, in, t)$. Let us call this concept the initial efficiency and denote it also by '$ef_t(L_1, in)$'. A language will accordingly have maximum initial efficiency if and only if each of the mentioned Q-sentences will be initially equiprobable, that is, if and only if the m-function upon which it is based ascribes equal values to all Q-sentences with the same argument, which will be the case when (but not only when) this m-function treats each primitive predicate and its negation on a par, as do, for instance, m_D and all m_I.

The symmetrical treatment of a predicate and its negation loses somewhat the arbitrariness with which it has often been charged; it turns out that this treatment, based psychologically upon some principle of indifference and methodologically upon considerations of simplicity, maximizes the initial efficiency of the language.

With an increase in experience and the establishment of empirical laws, which in their simplest form are equivalent to the statement that certain Q-properties are empty ([Prob.] §38), the efficiency of the respective language generally decreases. The greater the number of laws which are established, and the stronger they are, the less efficient the language becomes. It is plausible that with a continuing decrease of the efficiency of a language, a stage may be reached where this language will be altogether abandoned and replaced by another which, on the same evidence, shows a higher efficiency, mainly through the fact that the (or at least some) empirical laws of the first language have led to a modification of the conceptual framework of the second.

> The New Yorker, in our previous illustration, would do well, after having stayed for some time in San Francisco, to adopt a new language in which 'W'' would stand for More-Than-60°, 'T'' for Between-50°-And-60°, and 'C'' for Less-Than-50°, for instance, to save him from making the inefficient and uninteresting statements about the weather in San Francisco, which had before in almost all cases the form 'T(x)', i.e., 'It is temperate at time x'.

It might be sometimes useful to talk about the inefficiency of a language. The definition is obvious:

$$inef(L_1, in, e) =_{Df} 1 - ef(L_1, in, e).$$

It is advisable to avoid the term 'redundancy' – the term used for the correlate of our 'inefficiency' in communication theory – since the expression 'redundancy of a conceptual framework' is usually understood in a different sense.

§ 13. Conclusions

The concepts of information and information measures explicated here should be of value in various theories, as in the Theory of Design of Experiments and in the Theory of Testing. Various extensions are outlined. One of these would take into account the linear arrangements of the individuals.

The Theory of Semantic Information outlined here is nothing more than a certain ramification of the Theory of Inductive Probability presented in [Prob.]. The explication of the presystematic concept of the information carried by a sentence, which has been attempted here, should be of value for a clarification of the foundations of all those theories which make use of this concept and the measures connected with it. The impact of the concepts presented here for the Theory of Design of Experiments or for the Theory of Testing should be obvious.*

The present theory requires extension into various directions. One extension has already been mentioned: no great difficulties are involved in treating language systems with denumerably many individuals. Nor would introduction of individual variables and quantification over them present problems of which we do not know the solution. Language systems of these types have already been treated in [Prob.].

Other extensions, however, will have to be postponed until the corresponding theories of inductive probability are developed. One of us (R. C.) is engaged, at the moment, in developing concepts with the help of which language systems that allow for taking account of linear arrangements of their individuals can be investigated. This being accomplished, the erection of a Theory of Semantic Information for sequential universes should prove to be a fairly straightforward task.

*See, for example, L. J. Cronbach: A Generalized Psychometric Theory Based on Information Measure; A Preliminary Report, University of Illinois, March 1952.

Corrigenda

1. Replace 'minimum' by 'null' in T3-7, 9, 10, and 11.
2. Cancel T3-12 and the paragraph following it.
3. Replace the text after D4-1 to the end of section 4 by the following:

   Let us state only one theorem on the relative content:

   T4-4.

   a. If i is an L-true sentence, Cont(j/i) = Cont(j).

      Proof: In this case, i.j is L-equivalent to j. The theorem follows from D1 and T2a.

   b. Cont(j/t) = Cont(j).

Thus the relative content of j with respect to t equals the absolute content of j. Therefore it would be possible to begin with the relative content as primitive and define the absolute content as the value of the relative content with respect to t. However, it seems more convenient to begin with the simple concept of absolute content, because it has only one argument and the relative content can be defined on the basis of it. This is the procedure we have chosen here.

Additional Corrigenda - Technical Report No. 247

1. page 26, Table III. Instead of $(n-1)/n$, read $1 - (1/2^n)$.
2. page 28, Eq. 2. Instead of $c_1(j)$, read $c_1(i)$.
3. page 28, line 7 from below. Instead of T1b, read T4b.
4. Page 31, Table V. Transpose the values of columns 2 and 3.