Discussion Paper No. 18

FOUNDATIONS FOR A GENERAL THEORY OF INFORMATION

I. Technical Exposition
II. Relationship to Extant Works
III. Commentary

by

David W. Peterson

October 12, 1972

Part of the material presented in this paper was developed while the author was a Research Fellow at the International Institute of Management, Berlin.

This work was supported by Northwestern University, Evanston, and the International Institute of Management, Berlin. Its development took place over a span of several years, during which the author has benefited from the advice of and examples set by Jacob Marshak and Stanley Reiter; from the advice of several reviewers, known and unknown; and from the atmospheres of encouragement and support which the author encountered at the above-named institutions.

FOUNDATIONS FOR A GENERAL THEORY OF INFORMATION

I. Technical Exposition
II. Relationship to Extant Works
III. Co[illegible]

David W. Peterson
International Institute of Management
Science Center, Berlin
Northwestern University
Graduate School of Management
Evanston

ABSTRACT

In this work we have attempted to define precisely the term 'information', and to do so in a way sufficiently general to be applicable to the myriad situations in which people apply the term. We have identified a collection of basic objects as antecedents, herein called noumena, and argued that these are the only topics which humans use as a basis for information. A noumenon coupled with a meaning function is then defined to be information. The remainder of the paper probes some of the consequences of this structure, relating it to physics, computation, inferential logic, communication, knowledge, the entropy function, value theory; to theories of Marschak, Shannon, Brillouin, Moles, Bar-Hillel and Carnap, and others; and to informal everyday situations involving the information in a book and the translation of information from one language to another.

In all, we find the proposed definition at once satisfyingly simple and comprehensive. It clarifies the relation-

ship between information theory and physics, taking a position quite different from that of Brillouin and Peters. Within the proposed structure, the symbol processing theories of Marschak and Shannon emerge as prominent examples of the kinds of specific models that need to be analyzed to effect understanding of particular information processing situations. Perhaps most satisfying is the extent to which aesthetic aspects are describable in our terminology: The information content of art and the enjoyment it evokes are as readily described in terms of noumena, meaning functions and value as are the workings of a mechanistic information system.

FOUNDATIONS FOR A GENERAL THEORY OF INFORMATION

I. Technical Exposition
II. Relationship to Extant Works
III. Commentary

David W. Peterson

INTRODUCTION TO THE FOUNDATIONS

Background

The past 40 years has seen the development of several theories pertaining to information, for it has been recognized that information is one of the fundamental entities with which humans must deal. In contrast, theories concerning the equally fundamental entity energy have been developing for at least 300 years, with the result that we are today in possession of such useful and complimentary concepts as conservation of momentum, conservation of energy, equivalence of mass and energy, the principle of uncertainty, and many other general laws which enable one to deliberate quantitatively and objectively problems involving energy.

The theories of information have not yet attained this unity. The statistical communication theory of Shannon treats coding and transmission problems in a statistical setting and introduces a convenient measure of volume of information, ignoring philosophical issues relative to the fundamental nature of information, and practical issues relating to economics of information processing. A physical theory of information, germinated by Szilard and developed by Brillouin, relates information to physics also within a statistical setting.

The relationship is accomplished in a way which successfully explains the paradox of Maxwell's demon, though the explanation could as well be accomplished without their formal definition of information. This theory too fails to capture certain fundamental aspects of information, and it avoids questions of economics.

The theory of Marschak does introduce some economic considerations in the obtaining and processing of information for decision purposes; his work represents a significant enrichment and extension of Shannon's theory. It does fail, however, to explain what information is, a shortcoming that does not impair its intended use. The theory of Moles introduces and expounds in a qualitative way the aesthetic aspects of information. The works of Fisher and Hurwicz contain references to measures of information which they developed for particular purposes of their own, and while useful within certain settings, the measures are not associated with full fledged attempts at the development of an information theory. The semantic information theory of Bar-Hillel and Carnap deals with some problems of meaning, concentrating particularly on the logical implication relation, but the technical viewpoint they adopt does not appear broad enough to explain meaning as it is experienced in everyday affairs.

While each of these theories points up some aspects of information which are important within the particular domains for which the theories were developed, none of the theories appears to be broadly enough based to be able to accommodate all of the others. Furthermore, none of the theories defines in a satisfactory way the very entity with which they deal, information. Indeed, many authors seem assiduously to avoid the attempt.

Shannon recognizes and admits this limitation of his theory, though many enthusiasts of the Shannon theory have tended to overlook this. Brillouin (among others) meets the issue head-on and defines information to be one of its properties, a logical error comparable to that of defining a marble to be its mass. Another effort is that of Yovits and Ernst, which defines information as data which is useful in making decisions. This definition is vague in that it leaves information defined in terms of the elusive and undefined words data and decision.

Thus the situation regarding the theory of information is at this time one of disarray. There are several partial theories poorly related; none of the theories appears to be constructed on what could be considered elementary or fundamental principles. The diversity of theories, lack of common ground, and lack of fundamental principles, are a hindrance to clear, precise, quantitative thinking concerning information, the sort of thinking which is necessary to cope with the multivarious problems of information processing.

Purpose of the Present Work

It is the purpose of this paper to provide a precisely defined general framework within which to accommodate the further study of information. A secondary objective is to interrelate relevant existing works in order to appreciate more fully their significance.

An Overview of the Foundations

Information is defined in Part I in terms of those things which are universally perceived by human beings, and in terms of set membership functions representing meanings. The definition appears sufficiently general to label as information anything anyone would ever be inclined to call information, and it possesses a satisfyingly simple mathematical preciseness. This definition is then interpreted in the contexts of natural laws, computation, meaning, inference, knowledge, communication, value, and, in closing, in the contexts of several colloquial usages.

In Part II: Relationship to Extant Works, we indicate how our viewpoint contrasts with or reinforces previously given viewpoints developed by physicists, economists, psychologists, philosophers, statisticians, and engineers. In the Commentary (Part III), we appraise the new epistemology with respect to its comprehensiveness, philosophical limitations, and its significance for future research.

FOUNDATIONS FOR A GENERAL THEORY OF INFORMATION

I. Technical Exposition

David W. Peterson

ABSTRACT

Set forth herein is a mathematical framework for the explication of information and its processing. The proposed epistemology for information is analyzed in relation to natural laws, computation, meaning, inference, knowledge, communication, value and colloquial usage, as each is viewed by the author.

FOUNDATIONS FOR A GENERAL THEORY OF INFORMATION

I. Technical Exposition

David W. Peterson

Introduction

In this first part of a three-part work, we present a mathematical viewpoint of the universe and the aspect thereof that appears to correspond to the colloquial usage of the word information. The presentation is purposely cryptic because it is desired to assemble here a concise collection of ideas which can then be discussed in total in subsequent parts.

Definition of Information

The fundamental viewpoint of the universe we take herein is given by the following axiom:

> Axiom. The universe is describable in terms of a set function η defined on the set S of subsets of E^4.

We use E^4 to represent all of space-time. The range of the function η is in some space appropriate for describing the contents of any subset of space-time. For example, if it is assumed that the universe is composed solely of energy, then the range of η is contained in E^1, and the value of η on a subset of E^4 indicates the amount of energy contained in that subset.

> Definition 1. A set of subsets of E^4 is called a noumenon.

A noumenon, as defined here, is quite similar to the 'thing-in-itself' noumenon discussed by Kant. Any arbitrary region in E^4 may be thought to consist of or be filled with an object. One such region corresponds exactly to me during this minute as I sit writing; another to my chair; yet another to this room and its contents. Another region corresponds to, for example, the lower halves of the two left legs of this table along with the rim of my lampshade, the former enduring from five minutes ago to five minutes hence, the latter from noon tomorrow until next week sometime. We take a noumenon to be a set of such regions of E^4. We admit the empty set as a noumenon, for mathematical convenience. Letting c be the cardinality of E^1, it follows that

> Remark 1. The set U of noumena is of cardinality 2^{2^c}.

Some noumena correspond not to specific people or chairs or such things, but to something more abstract. For example, the set of regions in E^4 which are occupied by spherical objects is a noumenon with which we may conveniently identify the property of being spherical. Formally, we define *property* the same as *noumena*:

> Definition 2. A set of subsets of E^4 is called a *property*.[1]

In application, one pragmatically designates only certain noumena as properties.

> Definition 3. A noumenon N contained in a property P is said to *possess property* P.

Thus if a noumenon is a subset of the set of spherical noumena, it is spherical.

In some contexts a noumenon acts as a sign, symbol or designation of something. The meaning of such a noumenon appears to be representable in terms of a meaning function.

> Definition 4. Let D denote a subset of U. A function $M_N : D \rightarrow [0,1]$ associated with a noumenon N is called a *meaning function*.

Consider for example the noumenon associated with the word "tree". This noumenon consists of all regions in space-time wherein the uttering or writing of the word "tree" has taken or will take place. The meaning function associated with this noumenon captures the meaning of the word "tree" by assigning to any noumenon corresponding to a real tree a value near 1, and to noumena which do not correspond to real trees, values near 0. In this case D may be taken equal to U. In other cases, D may be much smaller, as in the case of a meaning function associated with the phrase "this is true": It doesn't make sense to apply this meaning function to a noumenon corresponding to an apple, so such a noumenon would be excluded from D.

Noumena and their meaning functions are the basic ingredients of information. Specifically, we define the following:

> Definition 5. A tuple $< N, M_N >$, where N is a noumenon and M_N its meaning function, is called an *information item*.[2]

Definition 6. A nonempty collection of information items is called <u>information</u>.

For mathematical and epistemological purposes, noumena may be regarded as pre-existing man, machines and information. We may consider any object to have at least one associated noumenon, namely that set whose sole element is the subset of E^4 occupied by the object. A noumenon may correspond to something abstract, as for example the word "tree", as previously illustrated. Even an idea or a mental outlook has one or more noumena associated with it: an idea resides within some physical boundaries (such as the human brain or a portion thereof), and the set of E^4- subsets each of which contains that idea may be considered as the noumenon representing that idea. Meaning functions are for the most part characteristics of a living person -- just as are subjective probabilities and preferences. We consider the meaning of a noumenon to a person to be represented by his meaning function, recognizing that to a given person, some (in fact most) noumena will have no particular meaning (i.e., the domain of the meaning function will be nearly or totally empty). We do not exclude the possibility that animals and computers may be thought to possess meaning functions, nor do we exclude the possibility that upon occasion a noumenon may have associated with it two or more meaning functions. When a noumenon is accorded meaning through the application of a meaning function, we regard the set consisting of the noumenon and meaning function as information. Also calling unions of such sets information, we preserve the grammatically singular usage of the word.

Laws and Computation

A law, as used here, corresponds to a natural physical, social, behavioral, mathematical or economic law, but not in general to the type of law from which lawyers derive their livlihood.

> Definition 7. A law is triple $< P, B., Q >$, where P and Q are properties and $B_N \subset Q$ for each $N \subset P$.

Thus, if N is a noumenon with property P, the law permits the assertion that the noumenon B_N (which depends on N) has property Q. The property P might be that associated with the words "a marble rolling freely over a large, flat smooth surface", the noumenon N might be a particular occurance of such an event, B_N might be a similar occurance displaced into the future, and Q might be identical with P. In this case the law would summarize the bit of human wisdom that generally predicts a marble rolling freely on a flat smooth surface will continue to do so. As with meaning functions, we regard laws to be personal entities. Our laws are not immutable unknown equations according to which the universe evolves: They are the individually held models of how space-time phenomena are related. That different people may construct quite different laws is emphasized by the following qualitative definition:

> Definition 8. A realistic law is a law which is apparently rarely violated.

One realistic law might postulate that η is discretely valued, an assumption consistent with quantum mechanics. If one assumes further that the structure of η is so coarse that (as suggested by the uncertainty principle) it is not possible to distinguish pairs of co-ordinate sets which differ but slightly, one concludes that the distinguishable disjoint sub-sets of E^4 are countable, that it is reasonable to use only the σ-algebra built on these sub-sets (instead of S, all of the sub-sets of E^4) to define information, and that, therefore, the number of noumena in the universe is at most a mere 2^c. The further assumption that the number m of distinguishable disjoint sub-sets of E^4 is finite leads to the conclusion that the number of noumena in the universe is 2^{2^m}.

A realistic law can be used for making inferences about events standing in any space-time relation to the noumenon N of Definition 7. In order for communication to take place between individuals, however, it is necessary that the agent receiving the communication employ a law which is not only realistic but causal. Adopting a relativistic set of laws as governing the behavior of the universe, we could specify in precise terms the idea that causality requires the noumenon P_N of Definition 7 to have only elements contained in the intersection of backward-time (or else forward-time) light-cones of points in elements of N. Sufficient to exhibit our concern for the matter is the following more qualitative statement, which permits one to postpone commitment to relativistic laws.

Definition 9. A <u>causal law</u> is one which relates noumena suitably spaced in time.

What constitutes a suitable time displacement of noumena in the determination of the causality of a law may vary widely with circumstances. We have suggested an example involving light-cones, based on a typical relativistic analysis. Social laws or physical laws applied to highly non-viscous areas of η may impose more stringent restrictions on the causality of a physical or social law; the 'propagation velocity' may be considerably reduced, thus producing a time cone much narrower than the cone of the relativistic example.

The following definitions express the very general way in which we view the process of computation:

Definition 10. A realistic causal law is called an <u>algorithm</u>.

Definition 11. Let $A = \langle P, B., O \rangle$ be an algorithm. Let N be a noumenon such that $N \subset P$. Let R be a subset of E^4 containing the elements of N and of B_N. We call the pair $\langle R, A \rangle$ a <u>computer</u>.

Thus we regard as a computer any region in space-time, properties of which are related through a realistic causal physics. A radio plus the understanding by someone that it converts radio waves into audible sound constitute a computer. An economy along with some understanding by someone of some aspects of its evolution constitutes a computer. With this viewpoint, one sees that all of the universe and all of the realistic causal laws are available to him as raw materials for calculation and the processing of information. In addition to our digital computers and electronic analog computers, any chunk of the universe stands ready to act as a computer if we can but describe its behavior through a realistic causal physics.

The Algebra of Meaning

The meaning of the word sofa is taken to be, roughly, the set of all things which are sofas. Associated with a noumenon which is this particular writing of sofa is a meaning function M which assigns to every noumenon in the universe a degree of membership in the class of sofas. A noumenon N is a sofa if $M(N) \approx 1$, it is not if $M(N) \approx 0$, and its classification is uncertain if $M(N) \approx 1/2$.

In like manner, the meaning of 'liberty' is taken to be the meaning function which assigns values near one to noumena which are examples of liberty, near zero to those which are not. An especially appealing feature of this formalization of the concept of meaning is that humans appear to learn meanings largely by example. A child learns the meaning of the word yellow through the examination of many examples: dandelion blossoms and ears of corn are in the set, healthy grass and chocolate cake are out, sometimes the sun is in, sometimes its membership is tenuous, sometimes it is out altogether.

The characterization of meaning through meaning functions is suggestive of a calculus of meaning. Consider for example a noumenon A consisting of all printed sequences of the words

it is cold and it is raining (A)

and a noumenon B consisting of all printed sequences of the words

it is cold or it is raining. (B)

Related to these noumena are two other noumena, C and D, sequences of printed words

it is cold (C)
it is raining. (D)

Associated with these four noumena, in their current context, are meaning functions M_A, M_B, M_C, and M_D. It is of interest to examine the extent to which these functions are related. We assume the domains of the functions to be U.

If a noumenon $E \in U$ is a good example of something cold, we should expect that

$M_A(E)$ is undetermined

$M_B(E) \simeq 1$

$M_C(E) \simeq 1$

$M_D(E)$ is undetermined.

In all, we would anticipate that for any $E \in U$

$M_A(E) \simeq 1$ implies $M_B(E) \simeq 1$, $M_C(E) \simeq 1$, $M_D(E) \simeq 1$

$M_A(E) \simeq 0$ implies nothing

$M_B(E) \simeq 1$ implies that either $M_C(E) \simeq 1$ or $M_D(E) \simeq 1$, or both

$M_B(E) \simeq 0$ implies $M_A(E) \simeq 0$, $M_C(E) \simeq 0$, $M_D(E) \simeq 0$

$M_C(E) \simeq 1$ implies $M_B(E) \simeq 1$

$M_C(E) \simeq 0$ implies $M_A(E) \simeq 0$

$M_D(E) \simeq 1$ implies $M_B(E) \simeq 1$

$M_D(E) \simeq 0$ implies $M_A(E) \simeq 0$.

A collection of relationships among the meaning functions can be devised to reflect these implications. Letting $\wedge$ and $\vee$ denote the conjunctive operations 'and' and 'or', and ~ denote the negation operation, we note that

$$A = C \wedge D \tag{1}$$
$$B = C \vee D, \tag{2}$$

and postulate that

$$M_{C \vee D}(E) = M_C(E) + M_D(E) - M_{C \wedge D}(E)$$

$$M_{\sim C}(E) = 1 - M_C(E)$$

$$M_{\phi}(E) = 0,$$

where ϕ is a self-contradictory noumenon, e.g. $C \wedge (\sim C)$.

By these postulates we ensure that the meaning functions associated with an algebra α of noumena are well defined and have the properties of a probability function when applied to a fixed $E \varepsilon U$, namely:

1) $M_D(E) \geq 0$ for each $D \varepsilon \alpha$

2) If $D_1, \ldots, D_n \varepsilon \alpha$ are such that $M_{D_i \wedge D_j}(E) = 0$ for $i \neq j$ and $M_{D_1 \vee \ldots \vee D_n}(E) = 1$, then

$$\sum_{i=1}^{n} M_{D_i}(E) = 1.$$

It is critical to note that the operations $\wedge$, $\vee$, $\sim$ are conceived as logical operations on meanings, not primarily as physical operations on noumena. It is incidental that the operations on the meanings can, in this instance, be so closely related to operations on the noumena. Generally, meaning functions associated with various noumena and their combination are quite unrelated. Thus the meanings of the letters of the alphabet combine in no particular way to yield the meaning of a word, nor do the meaning functions of the 50,000 noumena associated one-to-one with the 50,000 scraps of paper into which this page might be torn combine in an orderly way to produce the meaning of this page. The problem is analogous to that of trying to determine a joint probability function given only the marginal functions: in general, it cannot be done. In the case of probability, the construction may not be possible because the variables involved are not statistically independent; in the case of meaning functions, the construction may fail because the noumena used as symbols have meanings dependent on the context in which they are used.

Thus we conclude that while it is possible to construct simple relationships among the meaning functions of some special algebras of noumena, it is generally not possible to do so given an arbitrary algebra of noumena.

Knowledge

The knowledge of an agent, if we accept the present framework, is exactly describable in terms of the distribution of elements comprising the agent. If an agent could be examined in sufficient detail and if one knew the meanings of the noumena encountered in the examination, presumably he could determine just what knowledge the agent possesses.

On a somewhat more operational level, it seems natural to describe the knowledge possessed by an individual as the awareness that certain noumena have certain properties. We capture this by characterizing the knowledge of an agent as a set of meaning functions:

> Definition 12. The knowledge of an agent, A, is the set of meaning functions $\{M\}_A$ which he employs.

The properties that such an agent recognizes we assume are of the form $M^{-1}[T]$, where $T \subset [0,1]$ and $M \varepsilon \{M\}_A$, and unions, compliments and intersections of such forms. Knowledge of a law might be equated to knowledge of three meaning functions: One each for the properties P and Q of Definition 7, and one which identifies noumena of the form NUR_N.

A measure of the amount of knowledge of an individual is the number of noumena he can distinguish through application of his meaning functions. Let P_M be a partition induced on U by the meaning function M, as it generates cells of the form $M^{-1}[\{x\}]$, x ranging over $[0,1]$. The product of all such P_M as M ranges over $\{M\}_A$ is a partition P dividing U into a set C_A of cells.

> Definition 13. The cardinality m_A of C_A is the amount of knowledge associated with $\{M\}_A$.

As to the quality of the knowledge possessed by an agent, nothing can be said within the present context except ruefully to acknowledge that it is possible to know much of little value.

Remark 2. The cardinality of C_A is at most that of U.

Suppose that at the beginning of his life, an individual's collection of meaning functions consists of a large number of functions, each with empty domains.[3] As he matures, the individual modifies these functions by communicating, thinking and forgetting. If an elementary modification of a meaning function is to assign it a new value on a region of U, (thus enlarging its domain or altering its value on a portion of its domain), and if such a modification requires, for an individual, at least some positive amount of time ε_A, then after a finite amount of time only a finite number m of such modifications will have been made, and the cardinality m_A of C_A will be at most the finite number 2^m.

Remark 3. If the cardinality of U is infinite and if the mechanism of learning is as just described, an individual cannot learn all there is to know.

Let two sets of meaning functions $\{M\}_A$ and $\{M\}_{A'}$ be related to each other in the following way. For each $M' \varepsilon \{M\}_{A'}$ let there exist a function G such that for each $N \varepsilon U$,

$$M'(N) = G[\{M(N)\}_A].$$

Thus the set of meaning functions $\{M\}_{A'}$ can be "derived" from the set $\{M\}_A$. A function G might, for example, correspond to the supremum operator on $\{M(N)\}_A$ or some other simple operation.

Definition 14. If two sets of meaning functions $\{M\}$ and $\{M\}_{A'}$ are related as above, we say the set $\{M\}_{A'}$ is <u>introspectively dependent</u> on the set $\{M\}_A$.

Remark 4. The partition P induced on U by $\{M\}_A$ is (weakly) finer than the partition P' induced by $\{M\}_{A'}$.

Thus, the amount of knowledge associated with a set of meaning functions is no greater than that of a set of meaning functions on which it is introspectively dependent.

An example of two sets of meaning functions one of which is introspectively dependent on the other arises in connection with computing. The set of meaning functions associated with the output symbols should be (in any useful case) introspectively dependent on the set of meaning functions associated with the set of input symbols. It may also be that the meaning functions representing the knowledge of an agent after a period of introspective thought (i.e., no communication or fresh observation of his environment) are introspectively dependent on his meaning functions at each earlier instant during this period. In this context, Remark 4 claims that the quantity of knowledge of an individual is increased only through communication and experimentation. Of course, introspection can greatly increase the quality of a set of meaning functions while reducing the quantity of knowledge, for it may happen that many unimportant distinctions are cast out in the process.

Communication

Among the elements of communication between two agents A and B are the following:

1. A must know an algorithm < P, B.,Q > for communicating with B.
2. A must believe B knows the meaning function to use to properly interpret the forthcoming signal.
3. A must initiate a noumenon N with property P, which he believes will cause a noumenon P_N (the signal) of type Q to impinge on B.
4. B must detect from whatever noumenon N impinges on him that a communication is intended.
5. B must apply the meaning function M_K to the noumenon which is A to determine what A is trying to communicate.

Evidently, something can go wrong at any step:

1. A may fail to speak loud enough, or be unaware that a telephone circuit is faulty.
2. A may think B understands only French, and not knowing French, A may not even attempt communication.
3. A may forget to call B.
4. B may doze through the discourse.
5. B may in fact understand only French, and be unable properly to interpret A's English.

Recognition of such possibilities comes easily from the detailed description of the communication process. Steps 2, 4 and 5 are major stumbling blocks in the effort to communicate with things in other worlds - or with animals.

Consider two agents A and B as characterized by sets $\{M\}_A$ and $\{M\}_B$ of meaning functions, respectively. Communication between A and B may be regarded as introspection on the set $\{M\}_A \cup \{M\}_B$ of meaning functions, generating new meaning functions $\{M\}_{A'}$ and $\{M\}_{B'}$.

Remark 5. If the partitions P_A, P_B, $P_{A'}$ and $P_{B'}$ are induced on U by $\{M\}_A$, $\{M\}_B$, $\{M\}_{A'}$, and $\{M\}_{B'}$, then

$P_{A'}$ is no finer than $P_A P_B$

$P_{B'}$ is no finer than $P_A P_B$.

Thus the amount of knowledge agent A can possess after conversation with B is limited from above by the cardinality of the set of cells into which $P_A P_B$ partitions U, and hence by the product $m_A m_B$ of the cardinalities of C_A and C_B.

For Agent A to communicate with Agent B, Agent A must cause certain kinds of noumena to appear near B, and B must apply the proper meaning functions. If noumena D_1 and D_2 have meanings M_{D_1} and M_{D_2}, corresponding perhaps to

D_1 - red light near B is illuminated (A turned it on to indicate that there is a telephone call for B)

D_2 - blue light near B is illuminated (A turned it on to indicate that there is a visitor waiting to see B)

then the noumenon $D_1 \cup D_2$ cannot have a separate meaning (arbitrarily chosen meaning function) such as

$D_1 \cup D_2$ - both lights on (A turned them on because the coffee is ready).

Remark 6. The maximal number of noumena which may be assigned independent meanings permitting unambiguous interpretation is c, the cardinality of space-time.

Thus, though the things about which one might wish to speak, the noumena comprising U, are of cardinality 2^{2^c}, one's vocabulary consists of at most c noumena. This limitation on communication is fundamental, and independent of the assumption that the cardinality of space-time is that of the set of real numbers. If m is the cardinality of space-time, m finite or infinite, the set of noumena has cardinality 2^{2^m} which is strictly greater than the maximal number m of noumena which can be assigned arbitrary meaning functions and still be unambiguously decoded.

In fact the limitation may be more severe than suggested above when one considers the number of noumena which will ever be successfully used as symbols. For an individual to assign a meaning to a symbol is an act requiring a non-zero amount of time, during which, we assume, he cannot assign meanings to other noumena. Thus an individual, in a finite lifetime, can decode but a finite number of noumena; and a race of individuals enduring forever can ascribe appropriate meanings to at most a countably infinite number of noumena. Hence our ability to communicate is grossly inadequate to accommodate the totality of things which we might wish to communicate.

In the communication of knowledge from generation to generation of mankind, it appears reasonable to suppose that the amount of knowledge of an individual human, measured as in Definition 13, is finite. This may be because of physical limitations on the diversity of meaning functions which the brain can accommodate, or simply because of the finiteness of life and the discrete nature of the knowledge acquisition process. With a finite human population, the knowledge of mankind is thus also finite by our measure. The bound on human knowledge thus obtained can be increased if

i) the elemental learning process can be speeded up
ii) the forgetting process can be slowed down, or
iii) the population increases,

but the bound remains finite and a lid effective in suggesting that ultimately the knowledge possessed by man in ages past can be retained only at the expense of forgoing new knowledge.

Thus even mankind as a group can differentiate through communication among only a finite number of the 2^{2^c} noumena - and this sharp limitation on the number of communicable distinctions is not likely to be significantly altered. The few meaning functions mankind may possess must be carefully selected to ensure their utmost usefulness. The progress of the development of human knowledge should be measured in terms of the quality of these functions, not of their associated amount of knowledge.

Contemplation

Because every identifiable object (including ideas thought, words used, and sofas built) is presumed to be a noumenon, 2^{2^c} is an upper bound for the number of these entities. If one makes the pragmatically appealing assumption that anything worthy of being described as an idea or thought exists over some open interval in space-time, then the assumption of a countably infinite population of thinkers leads to the conclusion that the number of thoughts past, present and future is, at most, countably infinite.

This restriction suggests that though one may think about sets containing elements numbering well in excess of $\aleph_0$, one cannot possibly contemplate all the individual elements of such sets. Hence there are several classifications of sets:

- Finite sets can be contemplated element by element, though not all finite sets will be contemplated since the set of finite sets is of cardinality c.
- Countably infinite sets can be contemplated element by element, but no one human of finite life span will accomplish the entire task.
- Sets of cardinality greater than $\aleph_0$ can be conceived as entities themselves, and some of their elements can be contemplated. However, not all elements can be contemplated even by a finite community of thinkers.
- Sets of cardinality greater than 2^{2^c} can be regarded as being composed largely of elements which cannot be identified with reality, since the cardinality of the set of noumena is at most 2^{2^c}. Indeed, this restriction may be considered in force even for sets of cardinality greater than c, for the latter is the maximal number of noumena which can be assigned independent meanings.

These restrictions on what sorts of sets can be conceived pose interesting problems for decision makers who are wont to consider sets of all possible outcomes, all possible actions and all possible observations: Unless these sets are finite or unless they are describable in terms not requiring contemplation of each individual member of each set, these sets cannot be imagined, constructed or specified (we say they are unimaginable). This barrier may partly explain why the mathematical theory of decision is not more widely used.

The Value of Information, the Quality of Knowledge

Though economists, engineers and statisticians have utilized measures of the value associated with information in certain circumstances, there appears to be no way of formulating a useful value or quality theory on the level of generality of the information theory we have thus far developed. The major obstacles to the formulation of such a general theory are

i) the considerable practical difficulty of assigning a numerical value to an information item or a collection of items, and

ii) the lack of any general rules relating the values of individual information items to values of sets of such items.

The assignment of a value to an item of information involves considerations of the use to which the information will be put and the rewards gained thereby. The value must depend on the time at which the information is brought to the attention of that agent. There is no particular reason to suppose that value can, in all circumstances, be expressed as a real number: Receipt of an item of information may benefit an agent both financially and psychologically, or its effects may be spread over time or space in a way necessitating that value be vector-valued.[4]

Sample Interpretations in the Theory

A book is a noumenon which acts as a symbol to which many people would assign meanings similar to those assigned it by its author. The book itself, being a noumenon, is an appropriate topic for discussion quite independent of its function as a symbol. Usually, the reader of a book is interested primarily in its role as a symbol, and has but passing interest in the book itself. The book contains, by itself, no information in any sense.

The value of the information a reader derives from a book may be economic, aesthetic, or both. This value may arise because of the symbolic aspect of the book (e.g., it tells one how to improve his economic condition, or it elicits a feeling of satisfaction through the depicting of a heart-warming sequence of events), or because of the utility of the book itself as a rarity, an item of court evidence or a door jamb. The reader of a book may be thought to employ a great many meaning functions. Functions may be associated with noumena corresponding to individual words, phrases, paragraphs, chapters or the book as a whole. The domains of the meaning functions may be exterior to the author (in case of a factual report) or interior to the author (in case of a personal outpouring of emotion). In the process of reading the book, a reader's meaning functions undergo change and hence his knowledge alters.

Two books, one written in English and the other its French translation, represent the same information in the sense that an English person reading the English book, and a French person reading the French version, will perceive, through their meaning functions, nearly identical distinctions among the noumena in U pertinent to the topic of the books.

A digital computer is programmed to differentiate among the symbols or noumena in an input set, and to respond in accordance with an algorithm to yield noumena in the output set. As far as the computer is concerned, the input noumena need have no meaning. The output noumena derive meaning by virtue of their relation through the algorithm to the input noumena; and they derive additional meaning if meanings are assigned to the input symbols.

Summary and Conclusions

In this work we have attempted to define precisely the term 'information', and to do so in a way sufficiently general to be applicable to the myriad situations in which people apply the term. We have identified a collection of basic objects as antecedents, herein called noumena, and considered these as the only topics which humans use as a basis for information. A noumenon coupled with a meaning function is defined to be an information item, and information is identified with sets of information items. The remainder of the paper probes some of the consequences of this structure, relating it to natural laws, computation, meaning, knowledge, communication, contemplation and value theory, and to informal everyday situations involving the information in a book, the translation of information from one language to another, and the general processing of information by a computer.

In all, we find the proposed definition at once satisfyingly simple and comprehensive. In II. Relationship to Extant Works, we shall show how it relates to previous developments in philosophy, engineering, economics, aesthetics, computer science, statistics, psychology, linguistics, and physics.

Footnotes

1. Another definition of "property" which might be convenient in some applications would employ fuzzy sets or grades of membership of individual subsets of E^4 in the set (= property). Our simpler treatment is adequate for the development of several conclusions regarding computation and communication.

2. More explicit notation would be $\langle N, M_N, D\rangle$, where D is the domain of M_N.

3. A more realistic starting point might be one in which a standard set of biologically inherited meaning functions is postulated. It is interesting to speculate on the variability of such a standard set in the evolution of man.

4. That information can possess value other than that measured in terms of dollars or similar units associated with aggrandizement seems first to have been clearly expounded by Moles. An exploration of the relationship of his work and the works of others to the present one is presented in II.

Foundations for a General Theory of Information

II. Relationship to Extant Works

David W. Peterson

Abstract

Set forth herein are references to many works in the areas of philosophy, engineering, economics, statistics, psychology, linguistics, computation theory, aesthetics, logic and physics which are related to the present work on general information theory. It is shown that the viewpoint expressed in Part I. Technical Exposition satisfactorily accommodates many previously expressed views, and improves, in the author's opinion, on some others.

Foundations for a General Theory of Information

II. Relationships to Extant Works

David W. Peterson

Introduction

In this second part of a three-part essay, we present a survey of some of the work which has preceeded our own. Pertinent previous works are very numerous and are from such a diversity of disciplines that we can, within the scope of this paper, only hope to indicate through selected citations that the theory presented in Part I. Technical Exposition is a consistent step in the development of a general theory of information. With this aim, we discuss the relationship to our own of previous developments in each of several diverse disciplines.

Philosophy

Substantial portions of the viewpoint expressed in I. are to be found in the philosophical literature. Our Axiom, for example, is a specific statement of Planck's first fundamental cannon of natural philosophy, namely that there is a real outer world (which exists independently of our act of knowing), a view similar to that of Plato [34,138-9]. The word "noumenon" has previously been used to refer to "an object that is conceived by reason and consequently thinkable but is not knowable by the senses: (a) thing-in-itself" [47,1545] - a meaning which Kant developed [22,267-285] and which conveys something of our view of how primitive are the sets of subsets of space-time.

The identification of things with their spatial coordinates was suggested by Descartes [6,144], the inventor of the rectangular coordinate system. By taking as noumena sets of subsets of space-time, we appear at odds with Whitehead's assertion that "in a sense, everything is everywhere at all times" [48,87]. The resolution of this disparity may lie in the recognition that through application of natural laws and meaning functions (which we choose to define separately from objects), information concerning an object may be located anywhere in space-time. At worst, the disparity can be attributed to a difference in models, with the consequence that in some situations one model may be more appropriate than the other.

In surveying the literature on knowledge, we were dismayed by the general lack of precisely defined terms. We felt that many assertions could have been more forcefully made (or perhaps would never have been made at all) if initially some basic terminology had been precisely defined. Though many examples are available, we cite Chisholm's definition of truth and compare it with one of our own, as this indicates how pragmatically we regard the concept of truth in addition to demonstrating a use of our terminology:

Chisholm [7,103]: A _belief_ or _assertion is true_ provided, first, that it is a belief or assertion with respect to a certain state of affairs that that state of affairs exists, and provided, secondly, that that state of affairs does exist; and a _belief_ or _assertion is false_ provided, first, that it is a belief or assertion with respect to a certain state of affairs that that state of affairs exists, and provided, secondly, that that state of affairs does not exist. _It is true that_ a given state of affairs exists provided that that state of affairs exists; and _it is false that_ a given state of affairs exists provided that that state of affairs does not exist. And _a truth_, finally, is a state of affairs that exists.

Peterson: "True" is a word. Almost every English-speaking person has a non-trivial meaning function associated with this word. A noumenon is generally true if a sufficiently large number of these meaning

> functions indicate sufficiently strong membership in individuals' sets of true noumena. A noumenon is true for an individual if his meaning function assigns it sufficiently strong membership in his set of true noumena.

Our definition involves terms which are generally unambiguous and widely understood, or which have been precisely defined in such terms. It also avoids the difficulties associated with having to determine if a state of affairs really does exist, and, for that matter, having to define what a state of affairs is if it does not exist.[1] Our concept of truth is pragmatic and individual-specific: "true" is a word and like all words it belongs to the people. "Ultimate truth" is different from our "truth", and the former seems to be a less useful concept. Our concept of truth permits flat-earthers to be regarded as truthful, their model of the world being satisfactory to them for their purposes. This view may be similar to the one held by Parmenides [6 ,19] when he wrote

> "All things that mortals have established, believing in their truth, are just a name ...."

That inductive inference is the only process by which new knowledge comes into the world is an assertion of Mill [27]. In our view, all induction is done introspectively: meaning functions are altered in some fashion, perhaps "logical", perhaps unexplainable. Mathematical induction, using the standard schema of showing that a proposition $P(o)$ is true and that if for any $n \geq o$, $[P(n)$ is true $\rightarrow$ $P(n+1)$ is true$]$, is a very specific kind of introspection.

The meaning of the word "true" in this context is that held by a circle of mathematicians for such contexts. A proposition proved by induction is often said to be true (by these mathematicians), but the truth of the proposition does not automatically follow from the proof: rather, the proof makes the proposition plausable, and for lack of any substantial evidence to the contrary, the proposition is designated as true. Because induction is introspective, it does not increase the amount of knowledge (as measured by our Definition 13) of an individual, though it may well result in better organization of his knowledge. Concerning the less formal types of induction discussed by Hintikka [17] and Popper [35] we have little to say except to note that these too are introspections.

The interpretation we place on a natural law regarding its fallability and transitory nature is well-established in the literature; Duhem [14] for example articulates our view. Natural laws being personal beliefs, it is folly for example to debate the issue of determinism in hopes of "settling" the matter. We should admire the beauty of the words (Tennyson [43])

"The stars," she whispered, "blindly run."

not argue them. If a man doesn't have a realistic physics through which he can predict some aspect of the evolution of the universe, then to him the universe, in that aspect, blindly runs.

A precursor to our combination of noumena, natural laws and meaning functions to transmit information can be found in Bishop Berkeley's writings [48,67]:

"Is it not plain, therefore, that neither the castle, the planet nor the cloud, which you see here, are those real ones which you suppose exist at a distance?"

This suggested distinction is maintained in our terminology.

The adequacy of language as a vehicle for exchanging information has undergone study from ancient times, and our quantitative conclusion concerning the shortage of uniquely decodeable communication symbols was clearly anticipated in qualitative forms:

Cassirer [5,107-108] "If we took an exact reproduction as our norm, we should be driven to an attitude of fundamental skepticism toward the value of the sign as such. If for example, we regarded it as the true and essential function of language to express once again, but merely in a different medium, the very same reality that lies ready-made before us in particular sensations and intuitions - we should be struck at once by the vast inadequacy of all languages. Measured by the limitless richness and diversity of intuitive reality, all linguistic symbols would inevitably seem empty ..."

Cassirer [5,123] "In his youth he (Plato) studied with Cratylus who, in opposition to the Sophists, represented the positive side in Heraclitean thought, since he looked upon words as the true and authentic instruments of knowledge, expressing and encompassing the essence of things. Heraclitus had asserted an identity between the whole of language and the whole of reason; Cratylus transferred this identity to the relation between the particular word and its conceptual content. (...) With surpassing irony Plato tears down the thesis that there is a naturally correct term for every existing thing ..."

Gorgias seems to have anticipated this inadequacy of language and shared our (Planck's) view of the outside existing but not necessarily knowable world when writing [5,187]

> "if there is being, it is inaccessible and unknowable; if it is knowable, it is inexpressible and incommunicable."

In our definition of noumena, we have given equal importance to all noumena, even though most noumena correspond to items of little or no use. In practice, there are relatively few noumena that we assign names to, and also few which we do the honor of designating as a property. Just how we happened to name the noumena that we have, or how we happen to work only with the properties we do seems to be the result of a long series of arbitrary choices. Russell [37,33] writes

"...from a logical point of view, a proper name may be assigned to any continuous portion of space-time."

Agreeing with this, we have left noumena formally undifferentiated.

Donald MacKay's work on information theory we mention in this section not because he is a professional philosopher, but because his work, arising in a scientific and engineering environment, is of an unusual degree of generality. MacKay [24] introduced his basic ideas at the time when many of his colleagues were over-enthusiastically welcoming Shannon's great work. MacKay and Bar-Hillel (of whom more will be said later) were notable in continuing to develop their worthy ideas in spite of the fact that for a long time general interest in information theory centered only within the narrow confines unintentionally fostered by Shannon. MacKay argued strongly for regarding meaning as personal, noting that meaning is always meaning to someone. He recognized that there are differences among intended, received and conventional meanings, and that it is not reasonable to suppose "...that only what can be expressed precisely in words has precise meaning." In gently chastising the enthusiastic users of Shannon's entropy function, MacKay remarks that information is not any of its properties, but something else. Although we do not, in I., adopt all of MacKay's models, we have quantified meaning and knowledge in a way which appears to us consistent with his qualitative models, and we regard his work as a close ancestor to our own.

The Shannon Model

A simple form of the Shannon [39] model is as follows:

A Source, consisting of an alphabet $A = \{a_1, \ldots, a_n\}$ of symbols, one of which is emitted by the source each second. The symbol emitted is chosen according to the fixed probabilities $P(a_i)$, each selection independent of all preceeding selections.

A Channel, consisting of an input alphabet A and an output alphabet $P = \{b_1, \ldots, b_m\}$ and a transition probability matrix with elements $p(b_j|a_i)$. The channel can accept one input per second.

It is assumed the source can be connected to the input of the channel and that at the output of the channel, an attempt is made to determine which symbol or symbols the source recently emitted.

For this general model, we may regard as relevant only those noumena associated with the alphabets A and B. In a specific case, an a_i might mean "send help" or "buy bonds," and in such cases it might be reasonable to assign a meaning function to each a_i. With the present general statement of the model, we see no reason to associate meanings with elements of A. The meaning of a b_j is in a sense no more definite than that of an a_i. However, because of the mechanical system being modelled, interest attaches to the fact that an a_i begets a b_j, and that knowledge of a particular b_j which has occurred enables one to assess the liklihood with which each a_i might have caused that b_j. Consequently the meaning function M_{b_j} associated with a particular b_j might reasonably be defined only on A, and $M_{b_j}(a_i)$ might measure the liklihood with which a_i caused b_j. A particularly convenient way of assigning values to $M_{b_j}(a_i)$ is to let it be $p(a_i|b_j)$, the probability that given b_j, a_i occurred. The information which is associated with an output symbol is the single element set $\{<b_j, M_{b_j}>\}$.

We note incidentally that the meaning of "a symbol from A is selected" might conveniently be identified with the function $p(a_i)$ as a_i ranges over A, and that the meaning of "a symbol from B has been selected" might be similarly identified with $p(b_j)$.

An observer watching the output terminal of the channel and using the meaning functions M_{b_j} given above would have knowledge $\{M_{b_j}\}_{j=1}^{m}$ relative to the output. The amount of knowledge he has (about A given B) would be the number of equivalence classes into which the functions $\{M_{b_j}\}$ partition A.

Shannon's entropy functions, used to measure uncertainty, surprise, average information volume, and channel capacity are appropriate to this type of model but not necessarily to other types. They are good examples of measures which must be introduced in certain circumstances to augment the functions and terminology developed in I.

Economic Works

Jacob Marschak [25] describes a symbol processing system of more complex structure than that of Shannon. Marschak's intent is to examine questions concerning the economics of symbol processing, and he explicitly introduces such effects as payoffs resulting from decisions made on the basis of information received, and the costs of data collecting, processing and transmitting. The system he analyzes, like Shannon's, can be described in terms of noumena and their relations through physical laws, and, as in Shannon's system, the noumena need have no meaning outside the system for the analysis to proceed.

In a different economic context, work on information quantification has developed from the observation that the proper functioning of an economy requires the constant interchange of information among agents, for example in advertising their wares and making their needs and desires known. The free market and price mechanism serve as powerful means of summarizing and communicating information concerning the availability of goods and services. However, there are many variations of economic organization through which such information might be diseminated, and considerable recent interest has been shown in examining in detail the efficiencies of some of these organizations.

Hurwicz [21] examines an economic system in which are identified individual agents characterized by their desires, production and communication capabilities, and possessions.

The agents communicate their desires as though through a general message clearing house, and eventually actions are agreed upon and effected. The informational efficiency of this system is measured in terms of the size of the set of messages or symbols which are required to convey the information necessary eventually to achieve appropriate action.

Reiter [36] studies a somewhat generalized version of Hurwicz's system, measuring efficiency in terms of the reduction of uncertainty in the agents' knowledge of their environment. Reiter defines the meaning of a symbol in a mathematically precise way, as the set of states of the environment consistent with the occurance of that symbol. If S is the set of possible states of environment in such a model and if the occurance of a symbol b permits the inference that some subset S' of S contains the present state of the environment, then the meaning function associated with b may be defined to have the power set of S as domain, taking values 1 on the power set of S', 0 elsewhere. Thus our terminology accomodates itself to Reiter's usage.

Aesthetic Information

Aside from its association with decisions and payoffs, information is associated with pleasure, dismay, boredom, excitement - those states of human minds which can occur in response to observation of noumena related to music, painting or one's general surroundings. Moles [28] suggests that information dissociated with decision making but giving rise to human emotion be labeled aesthetic information, and recognized as inherently different from the type usually studied.

The present theory provides a sufficiently general definition of information to include both Moles' aesthetic information and the information associated with decision making. Any item of information presumably has some aesthetic value, as well as some degree of use in a decision process. An item to which Moles might refer as aesthetic information would be one for which the aesthetic value is high, though perhaps not to the exclusion of its value in decision making. Thus while we appreciate Moles' recognition of aesthetic aspects of information, we do not view "aesthetic information" as inherently different from any other type of information.

Though it is often difficult to assign a number to the economic value of information, it is perhaps more difficult to assign a number to the aesthetic value. Indeed, it is difficult to put forth any reasonable axioms of aesthetic information value, but we are saved from a demoralizing impass by recognizing that there is no apparent purpose served in putting forth such axioms. I know, yet, of no reason why we should want to quantify the aesthetic aspects of information.

One might, of course, be interested in quantitatively analyzing the content of theater works, objects of art and production line job requirements to make sure that there is variety enough to keep the participants from falling asleep, but this is a quantification activity we regard as within the domain of psychologists, and not so much a problem in aesthetics. Moles observes, however, that good art is often complex in the sense that it cannot be easily transmitted to the human mind in its entirety. One may contemplate a good painting many times and still see new things; a good symphony may be heard again and again, each time revealing new, interesting facets to the listener: Good art tends to be composed of noumena with richly varied meaning functions. With our way of thinking of information, we note that though the artist prepares a noumenon, it is the beholder who supplies the meaning functions and any value assessment.[2]

Computers and Computing

The rise of digital computers and the vast activity surrounding it have had a self-reinforcing effect of the type that has manifested itself throughout history with unimagineable consequences: the Greek identification of lines and circles as elementary objects obscured for centuries the analytical results obtainable from slightly different axioms; the mass choice by the Chinese of a non-alphabetic writing system has surely hindered their scientific development; and even now our view of the world as being composed of matter-energy in a four-dimensional expanse may be blocking our efforts to develop more realistic natural laws. So it is that digital computers have swung attention away from the general problem of computing, and rigidified thinking about the process of computation. The concepts of algorithm and computability are formally developed (see, e.g., Davis [11]) only within a digital computing framework. So synonymous has "digital computing" become with "computing", that efforts such as Cobham's [9] to show that multiplication is intrinsically more complicated than addition often fail to note that their proofs are in support of the proposition

> multiplication is intrinsically more complex than addition, <u>on a digital computer</u>.

On a slide rule, the situation is otherwise, by most reasonable measures; and in general, computational complexity must <u>always</u> be measured with respect to a given computer.

Computation, as we define it, includes the operation of digital computers as well as electronic analog computers, and virtually all processes of simulation and analogical development. Most striking about this viewpoint is the clarity with which it points out the relative lack of theoretical and physical development in areas of computing other than digital.

Models of digital computers, with special regard to their technical capabilities, have been developed by Turing [44] and in more recent times by Hartmanis and Stearns [16] and by Vairavan [45]. With regard to the economics of computing, the model proposed by Marschak [25] represents some initial progress. Even the Shannon channel may be regarded as a computer model - perhaps a stochastic computer, although the mechanism of its internal operation is not as explicit as in most models previously cited. Neither the Shannon nor the Marschak models pertain only to digital computers.

It seems quite reasonable that in some sense the information entering a computer must limit the information leaving it. Shannon shows that this holds between the average uncertainty $H(A)$ of the input and the average uncertainty $H(A|B)$ of the input given the output. Generally, $H(A) \geq H(A|B) \geq 0$. If the amount of information out of the channel (or computer) is measured as $H(A) - H(A|B)$, and the amount of information entering the channel is measured as $H(A)$, then $0 \leq H(A) - H(A|B) \leq H(A)$.

Another sense in which this relationship between input and output information holds is the following. Let the meaning functions $\{M_b\}_{b\varepsilon B}$ at the output be defined, as before, only on the elements of A. If we take as the meaning of "a_i has occurred" the meaning function

$$M_{a_i}(a_j) = \begin{cases} 1, & j = i \\ 0, & j \neq i, \end{cases}$$

then $\{M_a\}_{a\varepsilon A}$ partitions A to its ultimate fineness, hence at least as fine as the partition induced by $\{M_b\}_{b\varepsilon B}$. Thus with these meaning functions, an observer of the output knows no more about the input, by our Definition 13 measure, than an observer of the input. This conclusion obtains for arbitrary sets A and B.

Statistical Works

Several important concepts concerning information have been developed within statistical settings. R. A. Fisher [15] proposed a measure of the amount of information a valuation of a real random variable may be anticipated to give, on average, about its mean. In our terminology, Fisher regarded the meaning of "x is a normal random variable with (unknown) mean μ and a variance σ^2" as the normal probability function on the power set of the real line, with parameters μ and σ^2. This meaning function is characterized by the normal probability density function. $\Psi(x;\mu,\sigma^2)$, having parameters μ and σ^2. The average amount of information an observation of x gives about μ is then defined as

$$\int_{-\infty}^{\infty} \left| \frac{\partial \Psi}{\partial \mu} \right|^2 / \Psi \, dx,$$

which from the foregoing may be regarded as a formula for mapping meaning functions into the real line. S. Kullback [23] also introduced some statistical measures of information volume. These, along with the measures investigated by Blachman [3], are quite similar to Shannon's H function. They represent, within our framework, further examples of specific measures of meaning function properties that can be introduced advantageously for specialized situations.

Information value has been treated explicitly by Howard [18,19] within a contract bidding context, and a great deal of related work has been published on value without specific reference to information: the utility theory of von Neumann and Morgenstern [46], with augmentation by

Savage [38] and Debreu [12] being especially prominent. Though utility theory is useful in some situations to describe information value, one can easily construct situations in which the rules given by the theory do not lead to reasonable information values. For example, an agent may have learned that the price of a certain security is due to change abruptly by one dollar, though he knows not the sense of the change. To learn either that (E_1) the price is going up or that $(\sim E_1)$ the price is going down benefits the agent. If he is highly risk averse, to learn that E_1 will occur with probability .4 will not benefit him at all. In this case the values violate the basic and very useful von Neumann-Morgenstern value relationship

$$V(E_1, .4 ; \sim E_1, .6) = .4V(E_1, 1.) + .6V(\sim E_1, 1.),$$

where V measures the value of the information suggested by its argument. Thus our attitude toward utility theory is that it is useful in certain circumstances, that we perceive no direct conflict in terminology or philosophy,[3] and that we have no general improvements to add. We are inclined, for the present, to regard the concepts of information value and quality of knowledge as semi-colloquial, taking explicit form only when one models particular situations.

H. Simon [40] evidently recognized in a qualitative way the problem of unimaginable sets in a statistical decision-making context. Because one cannot, in many cases, enunciate the full range of possible available actions, or the full range of possible states of nature, one is likely to save frustration by recognizing the futility of determining an "optimal" decision, and instead concentrating on choosing satisfactory actions. (A somewhat analogous conclusion can be made in favor of heuristic over exact solutions to mathematical programming problems - but the supporting reasons are generally based on computer economics rather than on the outright impossibility of one of the tasks.)

Psychological Works

The Shannon model generated a great deal of enthusiasm among experimental psychologists, many of whom were anxious to apply the model and measures to humans. As is well recognized now, the Shannon model is not rich enough in several respects to describe certain important features of human information processing. Attneave [1] gives an account of many works by experimental psychologists based largely on application of the Shannon H function. Recently, interest has been shown in the development of more detailed models of human information processing (e.g; memory characteristics, types of word associations), and the carefree use of the H function seems to have decreased. (See e.g. Wickens [49], Norman [31], or for an overview of several recent works, Newell [30].)

A recent viewpoint on the nature of language and thought, expressed by the psychologist D.R. Olson [32], provides a qualitative model with which to compare some features of our own quantitative model. Olson argues that such entities as "categories, meaning, meaning components and dimensions that correspond to the recurrent features of the world" comprise, in his terms, the semantic system. External to this are the referents, objects and events in the environment. An individual, Olson argues, does not select words for use only based on conditions within the semantic system (as might be suggested by linguistic models analyzed by Davidson [10], Chomsky [8] or Bar-Hillel and Carnap [2]), but also on the basis of conditions external to the semantic system: the referents, objects and events in the environment. A supporting example cited involves the possible answer to the question "who threw the ball through my window?":

The boy threw the ball.

If only one boy were in immediate view, this answer would suffice; if two or more boys were in view, it would not.

We agree that word choice is more realistically regarded as extra-semantic. Moreover, we regard meaning as being extra-semantic, in the sense that it may be considered to pre-exist any rules or structure: it is only under fortunate circumstances that meaning functions are interrelated according to some simple rule. We also agree that the meaning of a word is generally sensitive not only to the verbal context within which it appears, but to the surrounding environment, as illustrated in the example of the boy(s) and the window. With our meaning functions, one may in principle describe not only the meanings of individual words within their verbal and environmental contexts, but also the meaning of a whole group of words within a verbal and environmental context, or even the meaning of the environment, including the words. Thus it is possible, without major shifts in terminology, to pass easily from consideration of the meaning of "boy" as just used, to the meaning of "The boy threw the ball" as just used, to the meaning of the environment consisting of my desk and these papers scattered upon it. And there is no reason to suppose that any one of these meaning functions is simply related to the others.

Language and Meaning

Only a few selected ties of our views to the vast literature on language and meaning can be noted here. (A good initial source for a more extensive comparative analysis is provided by Dixon [13] who preceeds exposition of his own views with an extensive literature review.)

The Wittgenstein [34,127] assertion that a sentence is either true or else false or else meaningless can be captured within our terminology by letting the domain D of the meaning function M associated with the word "true" contain only those noumena which are either true or false, leaving as meaningless the noumena in $U \sim D$. To conform strictly to Wittgenstein's dictate, one would presumeably want M to take value 1 for true noumena in D, the value 0 for false noumena.

We agree with MacKay [24,110] that the meaning of a word or sentence need not be exactly expressible in terms of other words: our meaning functions are primitive, and we offer no precise explanation of how they came into being, nor how they are interrelated. As Suppes [41] wrote concerning the impossibility of ever finding out what energy really is, so think we that energy is only what people refer to when they use the word, and that no ultimate energy can be determined because the meaning of "energy" cannot be exactly communicated. As MacKay [24], Olson [32] and Shannon [39] among many others have recognized, meaning is associated with selection, and inasmuch as meaning functions differentiate among noumena, we concur. More restrictive than our conception of meaning is that of Suri [42,290], who states that "meaning is an anticipated response to a significant symbol." However, this definition apparently could be cast into our terminology by a suitable choice of meaning functions.

Generally, we perceive most of the activity in the area of grammar and semantics as a study of formal interrelationships among meaning functions. Of Chomsky [8], Morris [29] and Bar-Hillel and Carnap [2], some leading authors on these topics, we extract only from the latter:

> Let a language system L_n^π denote a collection of n individuals, $a_1,\ldots,a_n$ and π properties, $P_1,\ldots,P_\pi$. Let an atomic statement be $P_i a_j$, which rendered in words is "a_j has property P_i." Let a content element be any disjunction which for each of the $n\pi$ atomic statements contains either its statement or else its negation. Atomic statements are statements, as are finite logical combinations of atomic statements. The class of content elements logically implied by any statement (i) in L_n^π is called the content of this statement, denoted Cont(i).

In our terminology, the content of a statement is synonymous with its meaning. The meaning function M_i associated with a statement (i) has as domain D the power set S of the set of content elements; the function takes the value 1 on the subsets of Cont(i), 0 elsewhere. The characterization of implication in terms of meaning functions thus becomes

$$[(i) \Rightarrow (j)] \Leftrightarrow [M_i(s) \leq M_j(s), \forall s \in S].$$

Furthermore, an algebra of meaning functions results directly:

$$\forall s \epsilon S, \quad M_{i \wedge j}(s) = M_i(s) M_j(s)$$

$$M_{\sim i}\ (s) = 1 - M_i(s)$$

$$M_{i \vee j}(s) = \max\{M_i(s), M_j(s)\} = M_i(s) + M_j(s) - M_i(s) M_j(s),$$

where $\vee$, $\sim$ and $\wedge$ represent, respectively, disjunction, negation and conjunction. Thus we regard the Bar-Hillel and Carnap model as one involving binary-valued meaning functions which submit to an algebra, a characterization which suggests several avenues along which the model might be extended.

Physics

Attempts have been made by Brillouin [4], Peters [33] and others to link information and physics, largely through the suggestive similarity between the Shannon H function and thermodynamic entropy. Our view of the matter is that this is a tenuous link at best, and that a more viable relationship is that based on noumena, meaning functions and natural laws, as described in I. Our view of the relationship between the H function and thermodynamic entropy is developed below.

The entropy function H is a real-valued function defined on the set of descrete probability spaces[4] as follows:

$$H(P) = - \sum_{p \in P} p \log p$$

where P is any countable set of probabilities summing to 1. Like a mean or a variance, H is a function which depends on an entire probability function, and a value of H can be thought a summary description of the properties of P. The value of H for a particular set P of probabilities measures, roughly speaking, the variability, unpredictability or generalized dispersion of events governed by P. In this respect it is analogous to the variance function, except that the latter is defined only in case P describes the behavior of a random variable, while H is defined even if the events occurring in accordance with P do not involve numbers.

A measure of the variability associated with a probability function is in some sense a measure of the amount of information gained by a person when he is told which specific event has occurred, having previously known only the event would occur in accordance with that probability function. This is an interpretation which R. A. Fisher [15] placed upon the variance function, and an interpretation which Shannon so successfully pursued with this entropy function. Shannon was able to show that not only does the entropy function have good qualitative properties for measuring information volume, but also it can be related in a precise way to the average volume of space required to house noumena representing which of the events has occurred.

Thermodynamic entropy is a property of an energy configuration. It happens that the entropy function H when applied to a particular probabilistic description of an energy configuration yields a measure of the thermodynamic entropy of that system. Thermodynamic entropy is a carefully defined physical quality in the same sense that heat and temperature are well defined, and the thermodynamic entropy of a system is not dependent on whether or not there exists a being who has knowledge of the precise configuration of the system in question: a system has but one entropy though it be regarded both by a person with complete knowledge of its structure, or by a casual observer.

The fact that the entropy function was born in thermodynamics and weaned in communication theory has misled some theorists into more closely identifying thermodynamics with communication theory than is warranted. The critical observation delineating these two areas seems to be that in the former the entropy function is applied to probabilities obtained in a mechanical, impersonal way, while in the latter, probabilities may be chosen to reflect one's knowledge.

Several authors (including Brillouin [4], Peters [33], McCulloch [26]) have likened the Shannon H measure of information to thermodynamic entropy, and then proceeded to invoke the second law of thermodynamics on H. Because these authors have not distinguished carefully between information on the one hand and a probabilistic model involving information processing on the other, it is very difficult to understand from them the significance of the second law of thermodynamics for general information theory.

Summary

To provide an appropriate context for the expository work in I. Technical Exposition, we have briefly surveyed various areas of literature which relate to information and information processing. Results from each of these areas have been related to our own, and we find that our epistemology appears quite adaptable, in some cases lending insights and precision of thought heretofore lacking.

[1]This is not meant to degrade Chisholm's work or that of his predecessors: for their purposes their constructions may have been entirely satisfactory. Our ultimate purpose is the rather mundane one of developing a mathematical language useful in the design and analysis of a variety of information systems. For higher philosophical purposes our approach may be quite unsatisfactory.

[2]This thought is anticipated deftly in an unlikely reference [20], "Beauty is in the eye of the beholder."

[3]Indeed, our views are perhaps even more subjective than those of Savage, long a champion of subjective probability.

[4]Functions somewhat analogous to the entropy function have been defined on more general probability spaces, but they do not concern us at present.

(1) Attneave, F., Applications of Information Theory to Psychology, Holt, Rinehart, and Winston, New York, 1959.

(2) Bar-Hillel, Y. and Carnap, R., "Semantic Information", The British Journal for the Philosophy of Science, Vol. IV, Aug. 1953 to Feb. 1954, pp. 147-157.

(3) Blachman, N.M., "The Amount of Information that Y Gives about X", IEEE Transactions on Information Theory, IT-14, no. 1, 1968.

(4) Brillouin, L., Scientific Uncertainty, and Information , New York and London: Academic Press, 1964.

(5) Cassirer, E., The Philosophy of Symbolic Forms, Volume One: Language, New Haven and London: Yale University Press, 1966, tr. by Ralph Manheim.

(6) Cassirer, E., The Philosophy of Symbolic Forms, Volume Three: The Phenomenology of Knowledge, New Haven and London: Yale University Press, 1st printing, 1966, tr. by Ralph Manheim.

(7) Chisholm, R.M., Theory of Knowledge, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1966.

(8) Chomsky, N., Aspects of the Theory of Syntax, M.I.T. Press, 1965.

(9) Cobham, A., "The Intrinsic Computational Difficulty of Functions", Logic, Methodology and Philosophy of Science, Y. Bar-Hillel, ed., North Holland Publishing Company, Amsterdam, 1965, pp. 24-30.

(10) Davidson, D., "Theories of Meaning and Learnable Languages", Logic, Methodology and Philosophy of Science, Y. Bar-Hillel, ed., North Holland Publishing Company, Amsterdam, 1965, pp. 383-394, Proceedings of the 1964 International Congress, Stanford, California.

(11) Davis, M., Computability and Unsolvability, McGraw-Hill, New York, 1958.

(12) Debreu, G., Theory of Value, Wiley, New York, 1959.

(13) Dixon, R.M.W., What is Language?: A New Approach to Linguistic Description, Longmans, Green and Co. Ltd., London, 1965.

(14) Duhem, P.M.M., "The Aim and Structure of Physical Theory", Princeton, 1954, pp. 165-179, tr. by P.P. Wiener, as quoted in Philosophy of Science, J.J. Kockelmans, ed., Collier-MacMillan Ltd., London, 1968, pp. 299-313.

(15) Fisher, R.A., The Design of Experiments, Oliver & Boyd, Ltd., London, 1935.

(16) Hartmanis, J., and Stearns, R.E., "Automata-Based Computational Complexity", Information Science, Vol. 1, No. 2, April 1969, pp. 173-184.

(17) Hintikka, J., "Towards a Theory of Inductive Generalization", Logic, Methodology and Philosophy of Science, Y. Bar-Hillel, ed., North Holland Publishing Co., Amsterdam, 1965, pp. 274-288, Proceedings of the 1964 International Congress, University of Helsinki, Helsinki, Finnland.

(18) Howard, R.A., "Information Value Theory", IEEE Transactions on Systems Science and Cybernetics, Vol. SSC-2, No. 1, August, 1966, pp. 22-26.

(19) Howard, R.A., "Value of Information Lotteries", IEEE Transactions on Systems Science and Cybernetics, Vol. SSC-3, No. 1, 1967, p. 54.

(20) Hungerford, M., Molly Bawn, Smith, Elder, & Co., London, 1908, p. 142.

(21) Hurwicz, L., "Optimality and Informational Efficiency in Resource Allocation Processes", Mathematical Methods in the Social Sciences, 1959, K.J. Arrow, S. Karlin, and P. Suppes, eds., Stanford: Stanford University Press, 1960, pp. 27-46.

(22) Kant, I., Kritik der reinen Vernunft, Wiesbaden, Inal-Verlag Zweigstelle, 1956, pp. 267-285.

(23) Kullback, S., Information Theory and Statistics, John Wiley and Sons, Inc., 1959.

(24) MacKay, D.M., Information, Mechanism and Meaning, The M.I.T. Press, Cambridge, Mass., 1969.

(25) Marschak, J., "Economics of Information Systems," in Frontiers of Economics, M. Intriligator, ed., North Holland Publishing Co., 1971

(26) McCulloch, W.S., "Machines that Think and Want", Comparative Psychology Monographs, Vol. 20, 1950, pp. 39-50.

(27) Mill, J.S., A System of Logic, West Strand, London, John W. Parker, 2nd ed., Vol. 1, Book III, 1846.

(28) Moles, A., Information Theory and Aesthetic Perception, tr. by Joel E. Cohen, Urbana: University of Illinois Press, 1966.

(29) Morris, C.W., "Foundations of the Theory of Signs", International Encyclopedia of Unified Science, Vol. 1, No. 2, Chicago and London: The University of Chicago Press, 1938.

(30) Newell, A., "Remarks on the Relationship Between Artificial Intelligence and Cognitive Psychology", in Theoretical Approaches to Non-Numerical Problem Solving, R. Banerji and M.D. Mesarovic, eds., Carnegie Mellon University, Pittsburgh, Lecture notes in Operations Research, Mathematical Systems 28, Springer-Verlag, Berlin, 1970, pp. 363-400.

(31) Norman, D.A., Memory and Attention: an introduction to human information processing, Wiley, 1969.

(32) Olson, D.R., "Language and Thought: Aspects of a Cognitive Theory of Semantics", Psychological Review, Vol. 77, No. 4, 1970, pp. 257-273.

(33) Peters, J., Einführung in die Allgemeine Informationstheorie, Springer-Verlag, Berlin, 1967.

(34) Popper, K.R., "The Nature of Philosophical Problems", The British Journal for the Philosophy of Science, Vol. III, May 1952 - Feb. 1953, pp. 124-156.

(35) Popper, K.R., The Logic of Scientific Discovery, Hutchinson and Company, London, 1959.

(36) Reiter, S., lecture in Feb. 1970.

(37) Russell, B., An Inquiry Into Meaning and Truth, Allen and Unwin, London, 1940.

(38) Savage, L.J., Foundations of Statistics, John Wiley & Sons, Inc., 1954.

(39) Shannon, C.E., and W. Weaver, The Mathematical Theory of Communication, Urbana: The University of Illinois Press, 1964.

(40) Simon, H.A., The Sciences of the Artificial, M.I.T. Press, Cambridge, Mass., 1969, pp. 55-83.

(41) Suppes, P., "Kinematics and Dynamics of Concept Formation", Logic, Methodology and Philosophy of Science, Y. Bar-Hillel, ed., North Holland Publishing Company, 1965, pp. 405-414.

(42) Suri, S., "The Communicative Process: An Interpretive Study of George H. Mead's Theory of Language", Synthese, Vol. IX, Issue 3, Nos. 3-5, pp. 289-295.

(43) Tennyson, A., "In Memoriam A.H.H." III, The Life and Works of Tennyson, Vol. VII, MacMillan and Co., Ltd., London, 1899, p. 31.

(44) Turing, A.M., "On Computable Numbers, with an Application to the Entscheidungs Problem", Proceedings of the London Mathematical Society, (series 2), Vol. 24, 1936, pp. 230-265.

(45) Vairavan, K., "Input-Output Relations of Finite Memory Systems", Information and Control, Vol. 16, 1970, pp. 52-65.

(46) von Neumann, J. and O. Morgenstern, Theory of Games and Economic Behavior, Princeton: Princeton University Press, 1947.

(47) Webster, M., Webster's Third New International Dictionary of the English Language Unabridged, Philip Babcock Gove, Ph.D., ed., G. & C. Merriam Company, Springfield, Mass., 1966.

(48) Whitehead, A.N., Science in the Modern World, MacMillan paperback, Mentor New American Library, Free Press, orig. 1926.

(49) Wickens, D.D., "Encoding Categories of Words: An Empirical Approach to Meaning", Psychological Review, Vol. 77, No. 1, 1970, pp. 1-15.

(50) Yovits, M.C., and R.L. Ernst, "Generalized Information Systems: Some Consequences for Information Transfer", Department of Computer and Information Science, The Ohio State University, Columbus, Ohio, Sept. 15, 1968.

Foundations for a General Theory of Information

III. Commentary

David W. Peterson

Abstract

The foundations set forth in I. Technical Exposition are discussed herein with regard to their scope, limitations and essential features, to the spirit with which they should be applied, to the areas of application forseen, and to alternative formulations. The opinion is expressed that these foundations may be useful to psychologists, linguists, computer scientists, behavioral scientists, designers of information systems, economists, and philosophers.

Foundations for a General Theory of Information

III. Commentary

David W. Peterson

Introduction

In this third part of a three-part work, we discuss several general aspects of the foundations presented in I. Technical Exposition. Discussed first are several philosophical points; subsequently are offered some thoughts on possible applications.

Philosophical Aspects

The foundations set forth in I. are intended primarily as an epistemology to be useful in describing information and its processing in a wide variety of situations. They are presented in mathematical form to enhance their usefulness in applications, where computer-like exactness is sometimes a necessity. Thus it is only a secondary matter that the foundations have relationships to topics of a philosophical nature, and we see these relationships as assymetrical in the sense that our work seems to follow from certain well-established lines of philosophical thought, while we neither perceive nor require that the work contribute to the development of such thought. Nevertheless, the foundations do raise interesting questions which do not appear to bear directly on their usefulness in practical situations, and some of these are taken up below.

We note first that the foundations leading to the definition of information given in I. are not a theory, for there is nothing to be proved or disproved. Identified there are certain entities which seem to combine in a natural manner to give explications of both technical and colloquial information handling scenarios. The importance of the foundations rests essentially on this capability. No claim is made that the epistemology is uniquely dictated from a set of broad specifications, nor that it represents some form of ultimate truth: It is merely an identification and labeling of certain elements, done in a way which shows promise of being useful.

Among the assumptions passed over lightly in I. is the one that some form of coordinate system in space-time exists. But such a coordinate system must be bound to the irregular features of the matter-energy distribution described by the function η, and is not something which could exist for any arbitrary distribution, nor does it necessarily exist over the whole of the domain of η. If the distribution is nearly homogeneous over a very large expanse, it is difficult to imagine how a coordinate system might be constructed: Persons on a ship at sea on a cloudy night without radio, compass or clock have limited means of distinguishing the ship's position at one time relative to that at another. Continuing with this pathology, we note that according to the Second Law of Thermodynamics, the matter-energy distribution is becoming more homogeneous with the passage of time, suggesting that many years hence it will be generally more difficult to establish and use coordinate systems. However, for the present, people appear to be able to construct coordinate systems which work well enough, and it is this with which we justify the reference to a coordinate system in I.

In choosing the set U, the second-power set of space-time, as the set of noumena we were guided by the following considerations. Simple objects such as particular trees, chairs and people can be identified with the one or more subsets of space-time they occupy. A meaning function associated with "this tree" could be defined on the subsets of space-time in such a way as to assign values near 1 to those subsets of space-time corresponding to the tree, values near 0 to other subsets. Now what is the thing with which that meaning function is associated? A collection of words "this tree" - but these words as an object are not a single subset of space-time: they are better taken as a set of subsets of space-time, for their physical dimensions are not exact. Thus the object with which we associate meaning is an element of the power set of space-time, and the set of such objects is U. We have not encountered any insurmountable difficulties in assuming that any object, thing or entity to which we might wish to accord meaning can be regarded as an element of U.

With regard to the domain of a meaning function, the situation appears less clear. In I. we have taken the domain as U, the second-power set of space-time; in the above paragraph, the meaning of "this tree" was identified with a meaning function having as domain the first-power set of space-time. Other possibilities exist: a convenient meaning function to associate with the object "number, according to Russell" has as domain the third-power set of space-time, and is developed as follows:

> Let $A_{n\alpha}$ be a subset of space-time containing n things, where n is an integer.
>
> Let A_n be the set of subsets of space-time, each of which contains n things.
>
> Let A be the set of sets of subsets of space-time, each of which is an A_n for some n.
>
> Let the meaning function M associated with "number according to Russell" have as domain the third-

power set of space-time, and be defined as .

$$M(B) = \begin{cases} 1 \text{ if } B = A \\ 0 \text{ otherwise.} \end{cases}$$

An alternative but less mechanistic meaning function assignable to "number, according to Russell" is one which has as domain the first-power set of space-time, and takes values near 1 on those subsets corresponding to Russell as he thought about this concept of number, and lesser values elsewhere.

Evidently situations in which the domain of the meaning function is the first-power set of space-time can also be treated with a meaning function having as domain the subset of the second-power set of space-time consisting of single element sets. Our choice in I. of the second-power set of space-time to contain meaning function domains was made because

i) it is at least as comprehensive a formulation as that in which the first-power set contains the meaning function domains;

ii) it appears that for applications, having as domain the first-power set is usually sufficient;

iii) the adoption of the second-power set introduces a pleasing symmetry, in that the objects with which meanings are associated also constitute the domain of the meaning functions.

In practice, the identification of a thought, word, idea or thing with a noumenon is likely to be a trivial matter, as when in II, we made such identifications for the Shannon and Bar-Hillel and Carnap models. It is a sobering suggestion, however, that all things, words, ideas, and entities that ever were or will be, or which one can contemplate, are identifiable with so small a set as U, the power set of the power set of space-time. Such a statement appears at odds with the fact that mathematicians from time to time deal with sets of cardinality much larger than that of U; but such sets are always defined intensively (not extensively, i.e; exhaustively), and require for their characterization a number of noumena considerably less than their cardinality.

The choice of the interval [0,1] as the range of meaning functions is arbitrary and somewhat cavalier, for it imposes a restriction on the number values a meaning function may take, and it imputes an order to meaning function values, neither of which may be desireable in some situations. The choice of the interval is convenient in cases where a meaning function is also a probability function, as might be the case in a Shannon-type model. In case the meaning function is binary valued, as in the Reiter and Bar-Hillel and Carnap models, the range may be simplified to be {0,1}. Primarily for concreteness and in preparation for applications which come to mind, we have settled on [0,1] as a reasonable range. The possible advantages to be gained by regarding a meaning function more abstractly, perhaps as a vector-valued function or as a relation on UxU, are not obvious to us and they have not been seriously investigated.

To characterize the knowledge of a person as a collection of meaning functions is vague in the respect that a person's knowledge constantly undergoes change, and it is not clear physically what it means to have a meaning function at an instant of time when it is not put to use, and when that function is different at previous and subsequent times. Counter to this observation is the observation that our meaning functions are not meant to suggest any physical mechanisms by which a human retains knowledge; we intend them only to describe in an overall fashion some information handling characteristics of a human. Thus it is not too surprising that the physical location, association or significance of certain individual features of a collection of meaning functions may be difficult to determine. We persist in this characterization of knowledge believing it both reasonable and useful in at least specialized domains, such as those involving knowledge of a computer's input given its output, or involving the alteration of knowledge through a well-defined logical process.

Our choice of definition of a natural law, though couched in quite general terms, is not to be defended either on the ground that every known natural law can be cast into this form, or that it is an especially useful form with which to work.[1] Rather, the formulation permits clarification of our assertions that natural laws are subjective, that computing is a much more general activity than often thought, and that it is necessary for purposes of communication to employ a realistic causal law. Of our definition of a law, we require nothing else.

Practical Aspects

The terminology presented in I. is intended primarily to permit introduction into myriad practical situations a terminology which is at once precise and convenient. The users of information theory have struggled too long under the encumbrances imposed by partial and imprecise epistemologies. It is our hope that our terminology will make absolutely evident the distinctions sometimes required among the terms data, information, knowledge, information value, amount of information and so forth. We hope also that the definitions posed will specialize naturally to situations of interest, and have indicated in II. how some of these specializations might be undertaken.

There are several areas where we foresee that the foundations may be applied. To psychologists, it might be interesting to determine what number of non-trivial meaning functions are characteristic of a human being, and how complex these functions are. It might be that some people are best modeled as having a few, very complicated meaning functions (perhaps representing expert knowledge), while others would be better modeled as having many simple ones. Corresponding to long-term and short-term memory, there might be some reasonable decay model stated in terms of slow and fast erosions of meaning function domains and complexities.

There are dynamic communication models to be built and analyzed, with the intent of imitating the human learning process and devining the relative merits of experimentation, conversation and introspection at various stages in life. There are models (perhaps now of more interest to sociologists or linguists) to be built describing how two or more cultures, each with its own set of meaning functions, merge to form a new culture.

With respect to the area of computer science, our world view indicates the possibility of seriously studying computers other than the usual digital types, in order to develop new theories of computation, a deeper understanding of what constitutes an economically efficient computation, and ultimately, new types of efficient computers.

Applications are forseen as well in the design and modification of information systems in organizations such as firms and economies. In these circumstances many of the flexibilities provided by the foundations, e.g., continuously variable meaning function values, personal, subjectively defined meaning functions and a large, rich domain of referents (noumena), should be useful. Extensions of our work may help illuminate the essential features of an informationally efficient organization; they will hopefully also shed some light on the problems of the maintenance of historical data, whose meaning and physical recoverability tend to decrease with time.

Summary

The foundations set forth in I. Technical Exposition and discussed in II. Relationship to Extant Works and in the present paper represent an attempt to quantify at an abstract level the elemental concept of information. The selection of entities to be named and their naming has been done with the primary purpose of creating a terminology useful in a wide variety of information processing situations. The terminology proposed and analyzed appears to be satisfactory in respect to its general applicability, and it appears to fit naturally into an evolution of thought which has been underway for centuries. But its ultimate justification rests with the accomplishments it facilitates.

[1]We suspect the former may be true, and doubt the latter.