DOCUMENT RESUME

ED 097 907		IR 001 280

AUTHOR	Yovits, M. C.; Abilock, Judith G.
TITLE	A Semiotic Framework for Information Science Leading to the Development of a Quantitative Measure of Information.
INSTITUTION	Ohio State Univ., Columbus. Computer and Information Science Research Center.
SPONS AGENCY	National Science Foundation, Washington, D.C. Office of Science Information Services.
PUB DATE	Oct 74
NOTE	9p.; Paper presented at the Annual Meeting of the American Society for Information Science (37th, Atlanta, Georgia, October 1974)
EDRS PRICE	MF-$0.75 HC-$1.50 PLUS POSTAGE
DESCRIPTORS	*Information Science; Information Theory; *Mathematical Models; Matrices; Semiotics; Speeches; Statistics
IDENTIFIERS	Binary Choice Unit

ABSTRACT

If information science is to be considered a "science" in the true sense of the word, a set of general concepts and analytical expressions must be developed. Fundamental to this development is a rigorous and quantifiable measure of information. In previous papers a general framework, called a generalized information system, is suggested which permits the development of these concepts and expressions. Through the use of this generalized model, we have been able to define information quantitatively and in a rigorous manner. The formulation depends on the definition that "information is data of value in decision making" and leads to quantitative relationships between information and the value of a decision state. The value of the decision state is defined as the summation of the expected values of all the possible courses of action weighted by the probability of each course of action. A new measure for the information contained in a particular decision state is developed. The information is defined in terms of a two-choice deterministic situation which we call a "binary choice unit." This measure is universally applicable for all information that is concerned with the effectiveness of the data on the recipient. (Author)

Presented at ASIS Meeting, Atlanta, October 1974

A SEMIOTIC FRAMEWORK FOR INFORMATION SCIENCE LEADING TO THE DEVELOPMENT OF A QUANTITATIVE MEASURE OF INFORMATION

M. C. Yovits and Judith G. Abilock
Department of Computer and Information Science, The Ohio State University
Columbus, Ohio

Abstract

If information science is to be considered a "science" in the true sense of the word, a set of general concepts and analytical expressions must be developed. Fundamental to this development is a rigorous and quantifiable measure of information. In previous papers a general framework, called a generalized information system, is suggested which permits the development of these concepts and expressions. Through the use of this generalized model, we have been able to define information quantitatively and in a rigorous manner.

The formulation depends on the definition that *information is data of value in decision-making* and leads to quantitative relationships between information and the value of a decision state. The value of the decision state is defined as the summation of the expected values of all the possible courses of action weighted by the probability of each course of action. A new measure for the information contained in a particular decision state is developed:

$$I = m \sum_{i=1}^{m} \{P(a_i)\}^2 - 1,$$

where m is the number of courses of action available to a decision-maker and $P(a_i)$ is the probability of each course of action. The information is defined in terms of a two-choice deterministic situation which we call a "binary choice unit." This measure is universally applicable for all information that is concerned with the effectiveness of the data upon the recipient.

INTRODUCTION

The authors have been concerned with the development of a theory and framework for establishing a conceptual basis for information science. Much of this work is summarized in references (7), (9), (10), and (11).

If information science is to be considered a "science" in the true sense of the word, a set of general concepts and analytical expressions regarding the flow of information must be developed. Fundamental to this development is a rigorous and quantitative measure of information. In this paper, such a measure is suggested and examined in the framework of a very general decision model. This formulation depends on the definition that *information is data of value in decision-making* and leads to the establishment of quantitative relationships between information and the value of a decision state.

A decision state is defined as a representation of the entire decision situation. It gives a complete description of all of the elements which enter into a particular decision-making situation. It is shown in this paper that the value of the decision state may be defined as the summation of the expected values of all the possible courses of action weighted by the probability of each course of action. In symbolic terms,

$$V(DS) = \sum_{i=1}^{m} P(a_i) EV(a_i).$$

A measure of the amount of information contained in a particular decision state is suggested as follows:

$$I = m \sum_{i=1}^{m} \{P(a_i)\}^2 - 1,$$

where m is the number of possible courses of action available to the decision-maker and $P(a_i)$ is the probability of each course of action. The distribution of the $P(a_i)$'s reflects the decision-maker's overall inclination towards each of the courses of action. Information is that quantity which changes this probability distribution.

The information measure we suggest is defined in terms of a two-choice deterministic situation in which the amount of information the decision state is one. Information is measured in terms of these binary choice units or b.c.u.'s. The measure is universally applicable for all information that is concerned with the effectiveness of the message upon the recipient. It is accordingly called pragmatic information and is similar to that of Weaver's level three (6). Properties of this and related measures are also discussed.

BACKGROUND

The term "information" has a wide variety of meanings. Shannon and Weaver (6) talk about information in the technical sense. That is, they are concerned with the most efficient way to transmit messages. Charles

---

This research was sponsored in part by the National Science Foundation through grants No. [illegible]4 [illegible]N-41628 from the Office of Science Information Service to the Computer and Information Science Research Center, The Ohio State University.

Morris (5) points out that it is now generally recognized that "information theory" is not a rival to, or a substitute for, a general theory of signs. Shannon and Weaver's concern is with the transmission of any message *independent* of its content. Y. Bar-Hillel (2) and D. M. MacKay (4) clarify this situation. MacKay regards information as that which changes our representations, i.e., our signs. Gaining information is thus changing our expectations, i.e., our dispositions to respond, caused by a sign. He distinguishes between selective and semantic information. Selective information gives the information necessary to select the message itself and is not concerned with the content of the message. Semantic information is concerned with the content of the message. Shannon thus deals with selective information problems. Carnap and Bar-Hillel (3) and Winograd (8) are perhaps best known for their work in the area of semantic information.

The aforementioned views of information are two of the three approaches or levels identified by Weaver. The third level is known as the behavioral or effectiveness level and deals with the effect that information has on the person using it. Ackoff (1) has dealt with information problems at this behavioral level. Our work lies in this area.

DECISION-MAKING

Charles Morris (5) has identified three general requirements of action involved in the decision-making process. A decision-maker must obtain information about the situation in which he is to act, he must select among courses of action, and he must execute this alternative by some specific course of behavior. To effect a meaningful analysis of information, one must examine in detail that which makes decision-making such a challenging activity -- uncertainty. After careful examination of decision models described in the literature, it became evident that these existing models do not provide a comprehensive representation of the uncertainty that exists in decision-making. Most of the models are concerned solely with decision-makers who have an advanced state of knowledge about the decision situation in question. The information science aspects of decision theory must, however, cover comprehensively not only those decision-makers who are expert but also those decision-makers who are average or rather poor. It is extremely important in developing a formal role for information science that *all* levels of effectiveness of decision-makers be considered. For this purpose, a very general decision model is proposed.

A decision model consists of a number of decision elements, including a set of courses of action, a set of possible outcomes, a goal or set of goals, a function relating outcomes to goal attainment, and a set of states of nature.

The decision-maker usually views a complex decision situation in terms of courses of action and possible outcomes. He may be uncertain about what outcomes will occur when a particular course of action is executed. This uncertainty associated with the execution of the alternatives is what we call *executional uncertainty*. A second type of uncertainty identified is *goal uncertainty*. The decision maker may have only a vague notion of the goal to which he aspires, and he may also be uncertain as to the degree to which each of the outcomes will satisfy the various goals. The third type of uncertainty which the decision-maker confronts is that concerned with the states of nature. He may not be able to identify all the possible states, but even if he could, he may still be uncertain as to the relationship between the set of states and the other decision elements. This is termed *environmental uncertainty*. A complete model of a complex decision situation must deal explicitly with all of these types of uncertainty. The conceptual decision model suggested explicitly recognizes all of the decision elements as well as the associated sources of uncertainty.

MATHEMATICAL REPRESENTATION OF THE DECISION MODEL

A decision-maker makes a sequence of related choices from among a discrete set of alternatives $A = \{a_1, \ldots, a_i, \ldots, a_m\}$. The number of elements m in this set may not be constant over time since there may be uncertainty with regard to these elements. The execution of a particular course of action results in the occurrence of one of a set of possible outcomes $O = \{o_1, \ldots, o_j, \ldots, o_n\}$. The number n may also vary over time. Executional uncertainty, i.e., uncertainty as to the relationship between a particular course of action a_i and an outcome o_j will be denoted by the subjective probability estimate w_{ij}, the likelihood that the execution of course of action a_i will result in outcome o_j. The set of relevant states of nature will be denoted by $S = \{s_1, \ldots, s_k, \ldots, s_r\}$ where r may also vary over time according to the decision-maker's current understanding of the decision situation. The probabilities of occurrence of the states will be denoted by the subjective estimates $P(s_1), \ldots, P(s_k), \ldots, P(s_r)$. The values assigned to the decision outcomes that reflect the relative value of each outcome with respect to goal attainment will be denoted by $\{v(o_j)\}$.

The decision elements A, O, and $\{v(o_j)\}$ are dependent upon the state of the external environment. For example, courses of action which seem reasonable under one set of conditions may be wrong under other circumstances. These dependencies can be incorporated in the model by defining the sets A and O to reflect the decision-maker's current understanding of the courses of action and the outcomes for each of the states of nature. Also, a set of $\{w_{ij}^k\}$ and $\{v_k(o_j)\}$ can be identified for each state of nature. Figure 1 depicts the suggested mathematical representation for a particular state of nature s_k.

Relative Values						
Outcomes	$v_k(o_1)$	$v_k(o_2)$	...	$v_k(o_j)$	...	$v_k(o_n)$
Courses of Action	o_1	o_2	...	o_j	...	o_n
a_1	w_{11}^k	w_{12}^k	...	w_{1j}^k	...	w_{1n}^k
a_2	w_{21}^k	w_{22}^k	...	w_{2j}^k	...	w_{2n}^k
.	.	.	...	.	...	.
.	.	.	...	.	...	.
.	.	.	...	.	...	.
a_i	w_{i1}^k	w_{i2}^k	...	w_{ij}^k	...	w_{in}^k
.	.	.	...	.	...	.
.	.	.	...	.	...	.
.	.	.	...	.	...	.
a_m	w_{m1}^k	w_{m2}^k	...	w_{mj}^k	...	w_{mn}^k

Figure 1. The decision matrix for the *kth* state of nature

More details can be found in (7).

The mathematical decision model consists of this decision matrix together with some decision criteria which, when applied to the decision matrix, results in the selection of probabilities of courses of action. The actual decision rule that is used is dependent on the decision-maker's own attitude toward uncertainty. For example, if the decision-maker is conservative, he may select that course of action which maximizes his minimum possible gain. There are many different criteria which can be used, but in this analysis we will assume that the decision-maker assigns probabilities to the alternatives which are proportional to their relative expected value. The decision rule that we have recommended is a very reasonable one, and a number of interesting results follow from this rule. The expected value, EV, is defined by

$$EV(a_i) = \sum_{k=1}^{r} P(s_k) \sum_{j=1}^{n} w_{ij}^{k} v_k(o_j). \qquad (1)$$

That is, the expected value of each alternative is the sum of all the possible values weighted by their probabilities of occurrence.

This generalized decision model provides a framework for a formal and comprehensive representation of uncertainty in decision-making. As such, it also provides a suitable framework for examining the role of information in decision-making that is also formal and comprehensive.

The effect of information is to change the decision-maker's representation of the various types of uncertainty. His decision model at time $t + 1$ will be a revised version of his model at time t. The way in which a particular decision-maker utilizes information to revise his representation of the various types of uncertainty is highly individualistic. The generalized decision model permits the application of a large number of possible learning rules.

A QUANTITATIVE MEASURE OF INFORMATION

Although the separate effects of the various types of uncertainty are clearly important, they are only of significance in their combined effect on the decision-maker's understanding of the situation. The amount or value of the information contained in a data set can be meaningfully expressed only in terms of the total effect of the data on the decision-maker's model of the decision situation.

Regardless of what decision rule a decision-maker is utilizing, it is possible to obtain a distribution that reflects the decision-maker's overall inclination toward the various courses of action. We assume, as already suggested, that the decision-maker chooses a course of action with a probability proportional to its relative expected value. Thus, $P(a_i)$ is defined by

$$P(a_i) = \frac{EV(a_i)}{\sum\limits_{i=1}^{m} EV(a_i)}, \qquad (2)$$

where $EV(a_i)$ is given by equation (1).

The decision matrix representation shown in Figure 1 defines completely the entire decision situation. It explicitly relates courses of action to: (1) observable outcomes; (2) values of these various outcomes to the decision-maker; and (3) the states of nature. *This matrix may thus be defined to be the decision state of the decision-maker.* This decision state is a complete description of all of the elements which enter into any decision-making situation. The uncertainty in this decision state can be measured and calculated, and this will be indicated later. The impact of information on reducing the uncertainty in this state function may then serve as a measure of information. Since the information can also be related directly to the values of the various outcomes, it is therefore also possible to calculate directly the *value* as well as the *amount* of the information.

Equation (1) provides the relationship which yields the Expected Value *(EV)* of any course of action a_i. The uncertainty which exists for any decision state will be a function of the mean square variance of the expected values of the various courses of action. For example, if all the *EV's* are the same, the decision-maker will be totally uncertain as to which alternative to choose and the variance will be zero. If the decision-maker is completely certain as to his course of action, then all of the *EV's* will be zero but one which will be finite. For such a situation, the variance can be shown to be a maximum.

The mean square variance, σ^2, is defined as

$$\sigma^2 = \left[\sum_{i=1}^{m} [EV(a_i) - \mu]^2\right] \Big/ m, \qquad (3)$$

and the mean, μ, is defined as

$$\mu = \left[\sum_{i=1}^{m} EV(a_i)\right] \Big/ m, \qquad (4)$$

where m is the number of possible courses of action.

We can now define the *value* of the *decision state* as the summation of the expected values of all the possible courses of action weighted by the probability of each course of action. In symbolic terms,

$$V(DS) = \sum_{i=1}^{m} P(a_i)\, EV(a_i). \qquad (5)$$

The information contained in a decision state is related to the mean square variance of the expected values of the courses of action. To be precise, the information is related not to σ^2 but rather to σ^2/μ^2, since information should be measured by relating it to the variance measured in units of the mean. This must be the case since a given variance of EV will clearly be much less significant when the mean of the *EV's* is large than when the mean small. The quantity σ^2/μ^2, from equations (and (5), is

$$\sigma^2/\mu^2 = \frac{V(DS)}{\mu} - 1. \qquad ($$

It is perhaps more meaningful to view this relationship in terms of the $P(a_i)$*'s*. With t use of equations (2), (4), and (5), one obtai

$$\sigma^2/\mu^2 = m \sum_{i=1}^{m} \{P(a_i)\}^2 - 1. \qquad ($$

This quantity possesses the desired properties for an information measure. The more uncertain the decision-maker is, the les the amount of information in his decision state. Thus, we define this fundamental quantity to be the amount of information in a particular decision state. That is,

$$I = m \sum_{i=1}^{m} \{P(a_i)\}^2 - 1. \qquad ($$

Note that this quantity has a minimum of zer when all the $P(a_i)$*'s* are equal to $\frac{1}{m}$. This is complete uncertainty. The quantity has a max mum of $m - 1$ in the case of complete certaint where one of the $P(a_i)$*'s* is one and the other are zero.

When there are only two possible courses of action, I will assume values from zero to one. It will be equal to one under conditions of certainty, i.e., when the probability of choosing one course of action is one and the other probability is zero. Accordingly, we will define the unit of information in terms a deterministic two-choice situation. This unit is called a binary choice unit, or b.c.u.

When there are m possible courses of action, then the maximum amount of information from Equation (8) is seen to be $m - 1$ b.c.u. This is in agreement with a well-known principle that a minimum of $m - 1$ choices from pairs of alternatives is required when there are m alternatives to consider. This is pointed out in (1). More explicitly, if $m - 1$ choices are required and the maximum amount of information in each choice is one, then the maximum amount of information is $m - 1$. Analogously the minimum amount of information is zero.

In summary, Equations (6) and (8) provide us with a method of obtaining the value of information in addition to the amount.

We can now define a quantity called the index of determinism.

$$D = \sum_{i=1}^{m} \{P(a_i)\}^2 - \frac{1}{m}$$

Note that this is just I/m. This quantity assumes the value zero when all the $P(a_i)$'s are equal (the case of total uncertainty) and the value $1 - \frac{1}{m}$ when the situation is completely deterministic. For large m, it will approach unity. Thus the *index of determinism* is a quantity *varying from zero to one* which measures *the determinism of the decision state.*

PROPERTIES OF THE INFORMATION MEASURE

The suggested information measure possesses a number of desirable properties. It is defined in terms of a fundamental unit of measure which we termed the binary choice unit, or b.c.u..

The consideration of additional but highly unlikely courses of action has a very small effect on the amount of information in the decision state.

Another desirable property of the information measure is "sequential additivity." The amount of information in a decision state can be measured all at once or the process can be broken up into several steps with the consideration of a few alternatives at a time. Regardless of which method is used, the amount of information in the entire decision state is the same.

A measure of the amount of information in a data set or message can be arrived at by computing the difference in the amount of information in the decision state before and after receipt of the data. That is, the amount of information is arrived at by considering the impact this new data has on the decision-maker's decision state. In symbolic terms, $I(D)$, the amount of information in data set D, is

$$I(D) = I_{t+1} - I_t , \qquad (10)$$

where I_{t+1} and I_t are the amounts of information in the decision state after and before receipt of the data set.

It should be noted that the amount of information in a data set may be either positive or negative. In general, positive information sharpens or refines the decision-maker's understanding of the situation in that it either reduces the number of structural components in the model or reduces the dispersion in one or more of the various probability distributions in the model. Negative information, on the other hand, either increases the number of structural components (e.g., the addition to the model of a previously unknown alternative or outcome) or increases the dispersion in the various distributions. Negative information, despite a possible connotation of the term, does represent information that is of significance to the decision-maker.

DISCUSSION AND SUMMARY

A formal measure for the amount and the value of information contained in a data set or message has been suggested. It quantifies information in terms of its effect on the state of the decision-maker, where a decision state is defined so that it represents a complete description of the decision-maker's overall level of understanding about a particular decision situation at a particular point in time. This measure is universally applicable for pragmatic information. This is equivalent to Weaver's level three which is concerned with the effects of the message upon the recipient.

In order to evaluate this measure of information, it is convenient to use a generalized information system model. The use of this model then permits the evaluation of the measure of information in terms of the reduction of uncertainty. This evaluation could be made in terms of any kind of a decision rule. We have suggested a reasonable decision rule that can be used, and we have developed relationships based on this rule. Virtually any other decision rule could be used for evaluating the effects of the various uncertainties referred to. It would also be possible to evaluate the decision state of the decision-maker in a purely descriptive sense.

In summary, the proposed information measure is a function of the effect that a set of data has on a decision-maker's decision state. This decision state is defined in such a way that it reflects the decision-maker's understanding of a particular decision situation at a particular point in time. Hence, it is a situation dependent and time dependent measure. Clearly it must be time and situation dependent since the same data will have different significance to different decision-makers at any point in time or to the same decision-maker at different times.

FUTURE PLANS

The information measure suggested in the previous sections leads to a measure of the pragmatic information content of a data set for a *particular* decision-maker at a *particular* point in time. The data acquired, processed, stored, and disseminated by an information system may be used, however, as a resource by *various* decision-makers at *various* points in time. Hence, in the design and development of information systems, there exists a problem whose level of complexity is an order of magnitude above that of the primary problems addressed in this study — the problem of quantifying the information contained in a data set in terms of its overall usefulness for a *set of decision-makers over a period of time.*

One possible approach to this problem of assigning a number to a data set to indicate its composite information content would be to start by determining the relationship between the effectiveness of a decision-maker and the pragmatic information content of the data set for this decision-maker. Since what is really desired is some indication of the information content of this data set for this decision-maker over a period of time, one may determine some index $\overline{I(D)}$ of the average information contained in data set D over some period of time. Then, if one were to develop a measure of the effectiveness of each of the decision-makers for whom this data set serves as a resource, it would be possible to formulate an *information profile* for the data set. Such an information profile would indicate the average information content of a data set as a function of decision-maker-effectiveness.

If such a profile could be determined for every data set to be stored in an information system, then some number derived from this profile could serve as an index of the composite value of this data set. This method would be of major importance for the development of a sound procedure for the design of more effective information systems.

REFERENCES

(1) Ackoff, R. L. "Towards a Behavioral Theory of Communication." Management Science, 4 (1958), 218-234.

(2) Bar-Hillel, Y. "An Examination of Information Theory." Philosophy of Science, 22 (1955), 86-103.

(3) Carnap, R., and Bar-Hillel, Y. An Outline of a Theory of Semantic Information. Technical Report No. 247, Research Laboratory of Electronics, M.I.T., 1952.

(4) MacKay, D. M. "In Search of Basic Symbols" and "The Nomenclature of Information Theory," in Cybernetics: Transactions of the Eighth Congress, Heinz von Foerster, ed. (New York: Macy Foundation, 1952).

(5) Morris, Charles. Signification and Significance. (Cambridge, Massachusetts: The M.I.T. Press, 1964).

(6) Shannon, Claude, and Weaver, Warren. The Mathematical Theory of Communication. (Urbana, Illinois: University of Illinois Press, 1949).

(7) Whittemore, Bruce J., and Yovits, M. C. "A Generalized Conceptual Development for the Analysis and Flow of Information." Journal of the American Society for Information Science, May-June 1973, Vol. 24 (No. 3), 221-231.

(8) Winograd, T. "Understanding Natural Language." Cognitive Psychology, 3, 1972.

(9) Yovits, M. C. "A Theoretical Framework for the Development of Information Science," in Information Science, Its Scope, Objects of Research and Problems, edited by A. I. Mikhailov, Publication of International Federation of Documentation Committee on Research on the Theoretical Basis of Information, 1974.

(10) Yovits, M. C. "Information Science: Toward the Development of a True Scientific Discipline." American Documentation, 20 (No. 4)(1969), 369-76.

(11) Yovits, M. C., and Ernst, R. L. "Generalized Information Systems: Consequences for Information Transfer," in People and Information, H. B. Pepinsk ed. (New York: Pergamon Press, 1969).