COWLES FOUNDATION DISCUSSION PAPER NO. 70

Remarks on the Economics of Information*

Jacob Marschak

April, 1959

* Research undertaken by the Cowles Commission for Research in Economics under Contract Nonr-358(01) with the Office of Naval Research.

REMARKS ON THE ECONOMICS OF INFORMATION

J. Marschak *

The words "amount of information", "value of information", "cost of information", "piece-wage to the informant" have been much in use recently but the relations between these quantities are not always made clear. Perhaps an attempt at clarification is in order. Though for some listeners it will elaborate the obvious the attempt seems to be justified by the existing state of discussion. **

---

* Presented at the dedication of the Western Data Processing Center, University of California at Los Angeles, January 1959. The paper, completed during the author's year on a rotating visiting research professorship at Carnegie Institute of Technology, was started under a contract of the Office of Naval Research with the Cowles Foundation for Research in Economics. Acknowledgements are due to my colleagues at the Cowles Foundation, M. Beckmann, G. Debreu, C. B. McGuire, J. Harsanyi, and R. Radner, and to a correspondence with C. B. Tompkins. See also the author's earlier paper [1954] and, of course, L. J. Savage's book [1954] especially Chapters 6-7.

** As pointed out to me by George A. Miller of Harvard to whom I owe many of the references to the literature used here.

1. On Criteria. This paper was prepared as an address to a group of engineers, economists, and mathematicians, gathered to inaugurate a scientific center dedicated to the joint work of these three professions. In deciding between alternative devices, instruments, processes it is natural for an engineer to look for purely physical criteria, goals that are independent of the needs of the human user. The economist's criteria are rooted in human needs and tastes, profits and utilities, and this makes him a nuisance. To compute the number of calories produced and consumed by a society is a neat job; the maximization of the energy output is a goal that can be stated clearly enough. The economist messes it all up by reminding the technocrat that some single individuals are happier with a fatless diet. For the society as a whole the economist computes a complicated criterion such as "national income in deflated dollars"; but even this he does only half-heartedly, troubled by his knowledge that the markets in which dollar prices are made are imperfect so that many human wants remain uncomparable.

The engineer designing a privately owned power station strives to achieve a high quotient of kilowatt-hours per ton of coal. The economist reminds him that the stockholders are more interested in profits and these are impaired by the high interest charge on the investment in a physically efficient but costly generator. Charles Hitch [1958], head of RAND's economic division, gave this other example: a handy criterion occasionally suggested to direct decisions in anti-submarine warfare is the ratio of enemy U-boats destroyed to our merchant ships sunk. This ratio is high when our convoys consist of many destroyers and very few (zero if possible) merchant ships. If this criterion were used in the last war we would shuttle destroyers across the ocean but would fail to transport men and supplies to Europe and to win the war.

This Streit der Fakultaeten is, as usual, a conflict of nomenclatures, definitions. The third profession present here today, the mathematicians, specialize in precise and consistent definitions, and may thus help the other two; provided the formal similarity of mathematical expressions does not prevent us from distinguishing between their various possible physical, or economic, interpretations.

A statistical distribution parameter called entropy, $-\sum_{i=1}^{n} p_i \log p_i$ (where the p_i are probabilities defined on a set of n states) measures, in some sense, the "degree of uncertainty". It also has mathematical properties that enabled Shannon [ ] to use it as a measure of the amount of information characterizing a "source of information", and of the capacity (maximum transmission rate) of a "channel of communication." This parameter clearly does not depend on the particular uses to which the information source or channel will be put. If, on the other hand, the user asks "How much is the source, or the channel, worth to me?", "How much am I willing to pay for it?", it is natural that the answer about this quantity (for which the term "value of information" suggests itself) will vary from user to user. This certainly complicates matters, and attempts have been made to show that if the value, or worth, of information to its receiver is defined in some appropriate way it can yet be made independent of the user; more than that, it will be measured precisely by the entropy formula.

In trying to clarify the terms we have met problems that are not merely linguistic, but actually reveal separate classes of important and measurable quantities. It turned out that writers who tried to define a value of information independent of the user have sometimes used the word to denote such quantities as the cost to the seller of information (similar to the cost of production), or the piece-wage to be paid to the informant to increase his reliability.

2. Demand and Supply Price of Information. It is agreed that for technical purposes the word "information" never denotes a single message but rather a set of potential messages associated with a given instrument (source or channel) of information. The man who buys a newspaper does not know beforehand what will be in the news. He acquires access to potential messages belonging to a set called news.

Information is demanded and supplied, and we can speak of a demand price and a supply price of information as we do in the case of commodities.

The demand price of information is the highest price that a given person is willing to pay for it. This price clearly depends on how useful or important are, for that particular person, the messages that he will obtain from the given information instrument. The horse-race fan will pay much for the newspaper with the latest racing results; weather services are of greater value to farmers, airlines and the tourist trade than to the steel industry; and some people don't care even for the highest-fidelity radio sets.

To an economist, it seems natural to call value of information the average amount earned with the help of that information. Moreover, if I already have access to a certain kind of information which earns for me, on the average, the amount V_0 ; and if the access to some other kind of information will help me to a greater average earning, V_1 ; then my demand price for that information is $V_1 - V_0$.

The supply price of information, on the other hand, is the lowest price its supplier is willing to charge. It cannot, in the long run, fall below the cost to the supplier or he will lose the motive to supply it.

In a market of comparable information instruments (e.g., newspapers), with numerous and freely competing buyers and sellers, a market price would emerge. It will be actually paid by all buyers and accepted by all sellers, excepting those (and only those) buyers whose demand price is lower, and those (and only those) sellers whose supply price is higher, than the market price. These ("submarginal") buyers and sellers will abstain from transaction. In the more usual and interesting case when a large and free market for comparable information instruments does not exist this "classical" economic model does not apply and the analysis becomes more difficult. But the concepts of demand price (related to the value of information) and the supply price (related to the cost of information) remain fundamental.

What is the relation of information value and information cost to the concept of information amount, which has been developed for the uses of the industry of communication devices and has been fruitfully transferred to various sciences? The amount of information does not depend on the needs of any particular buyer of information. Hence it is not identical with the value of information just shown to determine the demand price. But it is presumably related to the cost, and hence the supply price, of information as we shall presently see.

3. <u>A Simple Case: Equiprobable Messages.</u> If the set whose information amount is being measured consists of n potential messages that have equal probabilities, $p_i = 1/n$, the information amount $-\sum_{i=1}^{n} p_i \log p_i =$ $-n \cdot (1/n)[\log (1/n)] = \log n$ is simply defined so as to increase with the number of distinct potential messages. It is also usual to say that the larger this number the more precise is the information instrument: this is a definition. And the more precise the instrument, the costlier is the instrument and its operation: this is, or is asserted to be, an empirical fact. The increase of cost with increasing precision is associated with the fact that larger precision means (by the definition just given) a larger number of symbols

needed at a minimum to distinguish the messages: as, for example, when each message consists of a numerical variable given to the nearest tenth, or hundredth, or thousandth of a unit. It takes more time to transmit more symbols.*) And presumably it takes more labor and other resources to produce an instrument capable of transmitting more symbols per unit of time. The minimum number of symbols needed - e.g., the number of digits in the example just given - increases in proportion to the logarithm of the number of potential messages; so that (roughly?) an equal cost increment is added with every additional symbol. Moreover, if every potential message reports, not on a single variable, but on two independent variables (e.g., temperature and humidity), the total number of potential messages is the product of the numbers of potential messages associated with each variable; but the total number of symbols needed is the sum of the number of symbols needed for each variable. And this again corresponds (roughly) to the behavior of the cost: the cost of introducing a second variable is simply added to the cost of giving messages about the single one.

4. <u>An Illustration</u>. **) While still confining ourselves to the simple case of equiprobable messages we shall illustrate in more detail the fact that information value and information amount do not necessarily go together.

Suppose the price of a stock can change from this to the next week, by any amount between +6 and -6 points with equal probability. Suppose you can use the services of either of two informants, each a faultless predictor of

---

*) See, for example, Gilbert [1958].

**) Taken from Marschak and Radner [1958], Chapter 3.

stock prices. Informant A sends only two kinds of messages: (1) stock will rise; and (2), stock will fall. Informant B is more precise in that his messages can be of three kinds: (1) stock will rise by 2 points of more, (2) stock will fall by 2 points or more, (3) stock will move by less than 2 points in either direction. If, as we shall assume for a while, there are no transaction costs (brokerage charges, etc.) a good rule for your action on the basis of information from A is: "buy stock when A predicts rise, sell otherwise"; and, on the basis of information form B: "buy stock when B predicts rise of 2 points or more, sell when he predicts fall of 2 points or more, do nothing otherwise." Thus informant A enables you to take advantage of every price change while B is useless whenever price change is moderate. You will prefer A. Yet A is the less precise of the two: A uses only two equiprobable potential messages while B uses three! The amounts of information are in the ratio log 2 : log 3 = 1 : 1,6 approximately. But the values of information are in the ratio 9 : 8 = 1 : 0, 9 approximately. For, applying to the information from A your good rule of action, you will gain an average of (6+0)/2 points per share on purchases of rising stock, and obtain the same average gain on sales of falling stock; and since both cases have equal probabilities your expected gain will be

$$(\tfrac{1}{2})\ (\tfrac{6+0}{2}) + (\tfrac{1}{2})\ (\tfrac{6+0}{2}) = 3$$

points per share. On the other hand, B will help you to only

$$(\tfrac{1}{3})\ (\tfrac{6+2}{2}) + (\tfrac{1}{3})\ (0) + (\tfrac{1}{3})\ (\tfrac{6+2}{2}) = 2\ \tfrac{2}{3}$$

points per share.

We can also conceive of a third informant B*, who exerts himself to give even more precise messages than B: he breaks up the whole range of possible price-changes into a very large odd number of intervals of which the middle one is centered around zero. The amount of information in his set of messages will exceed by far that of A's and B's messages, making B* a very expensive information instrument indeed. But the value of his information to you, while exceeding the value of B's information, will still fall short of the value of A's information (although it will approach it): because B* fails and A does not fail to make the useful distinction between rises and declines of stock when they are within the middle interval. As so often, the brand least costly to make is the best for the buyer.

5. The Payoff Function. In our example, you have used A's or B's information on stock prices in order to buy or sell. Your actual gain (or loss) from each action was determined by both the action and the actual price change, according to the following "payoff function": if the price changes by +x points, and you buy (sell), you gain (lose) x points per share; if the price changes by -x points, your gains from buying or selling are, respectively, -x and +x points. If we denote the three-valued action variable by a, with $a = +1$ (buying), $a = -1$ (selling), and $a = 0$ (do nothing) then the gain u is equal to xa. We shall use Greek ω as a symbol for a payoff function, and the corresponding Latin letter u, for the value of this function. In our case

(1) $u = \omega(x,a) = xa$.

It is important to make explicit an assumption tacitly made so far: the decision-maker maximizes the expected payoff. In our example, the payoff was measured in money. It is more general to define as the payoff that function of the action and the state of the world, that is being maximized by the decision-maker. That such a function (defined on all possible outcomes of actions, whether these outcomes be money amounts or not) exists is an assumption about the decision-maker's behavior. It was shown by Von Neumann and Morgenstern [1948] that it follows from certain simple maxims of consistent behavior which it is reasonable to advise a person to use. The "payoff function" and the "criterion function" are the same thing.* (Footnote on p. 9a)

Under conditions of our example the payoff function (1) resulted in your preferring informant A to informant B. But we can change the payoff function so as to reverse your preferences. Instead of assuming transaction costs neglible let us assume you have to pay (in brokerage fees, taxes, etc.) 2 points per share on each purchase or sale. With this new payoff function, ω^*

$$u = \omega^*(x,a) = \begin{matrix} xa - 2, & a \neq 0 \\ 0 \quad, & a = 0, \end{matrix}$$

it becomes advantageous for you to abstain from transaction whenever the price is predicted to change by less than 2 points. B enables you to apply such a rule. But A does not. As a result, A will help you to an expected profit of $3 - 2 = 1$ point per share. But B will help you to more, viz., to

$$\left(\frac{1}{3}\right)\left(\frac{6+2}{2}\right) - 2) + \left(\frac{1}{3}\right)(0) + \left(\frac{1}{3}\right)\left(\frac{6+2}{2}\right) - 2) = 1\frac{1}{3}$$

points per share. The value of information from B has now become larger than that from A.

6. Information Structure. What is the explanation for this switching of the comparative values of the two information instruments? Let X denote the whole set of possible price-changes, i.e., the whole interval between, and including, -6 and $+6$. Thus $X = [-6,+6]$. Consider its following subsets:

$$X_1 = [-6, -2) \; ; \; X_2 = [-2, 0) \; ; \; X_3 = [0,+2) \; ; \; X_4 = [+2, +6].$$

Informant A, in fact, partitions X into two subsets, one message corresponding to each subset: one subset consists of X_1 and X_2, and the other

---

* Here is an example of non-monetary payoffs: To compare the information values of a slow and a fast military communication system Thorton Page [1957] assumed that the lack of information - due to communication delay - on each of the relevant external variables "degrades the decision" in a given way. In our terminology, the best decision possible on the basis of delayed information results in a diminished payoff. The information values computed by Page are, in effect, the expected payoffs under each of the compared communication systems.

subset consists of X_3 and X_4. Informant B partitions X into three subsets: X_1; X_4; and "X_2 and X_3". When the payoff function is ω of equation (1) (the case of zero transaction costs), the client's choice of action is influenced by whether the predicted price change belongs to "X_1 and X_2" or to "X_3 and X_4"; but it is not important for him to distinguish between X_1 and X_2, as both will dictate the same appropriate action ("sell"); similarly, the distinction between X_3 and X_4 is unimportant. Hence the greater value of A when payoff function is ω. But when it is the function ω^* of equation (2) (i.e., transaction costs = 2), the distinction between X_1 and X_2, and also between X_3 and X_4, becomes important enough to make B a more valuable information instrument than A.

Thus each information instrument is characterized by the way it partitions the set (X) of all possible states of the environment. We call each way of partitioning (or, briefly, each partition) of X an "information structure." We have seen that, whether a particular information structure yields a greater expected payoff than another partition depends in general on the payoff function, as our example has shown. Thus, in general, the ranking of information structures (and instruments) according to their value is a "subjective" matter inasmuch as it depends on the usefulness of information for a given user.

The question arises naturally whether there are pairs of partitions such that the ranking of their values is not influenced by the payoff function. It is easily seen that such ("objective") ranking is possible <u>if and only if one partition is a sub-partition of the other</u> in the sense that each of the subsets in the former partition is contained in some subset of the latter. For example, suppose informant C uses four messages, corresponding to the partition of X into the 4 subsets X_1, X_2, X_3, X_4 given above. This partition is a sub-partition of the two-set partition used by A as well as of the three-set partition used by B. And it is clear that any client of C has all the knowledge that a client of A or B has. Hence, for any client, whatever his payoff function, the information value of A or B can never exceed that of C (with the particular payoff function ω used above, information value of C is strictly larger than that of B and equals that of A; under ω^*, the positions of A and B are interchanged). Thus if one information structure is a subpartition of another, their values are ranked independently of the payoff function. The converse is also true. For (as pointed out by Roy Radner), if neither partition is a sub-partition of the other then there will be three states x_1, x_2, x_3 such that, under one partition, x_2 belongs into the same subset with x_1 but not with x_3; while under the other partition, x_2 belongs into the same subset with x_3 but not with x_1. The first or the second partition will have greater information value depending on whether it is more important to distinguish x_2 from x_3 or from x_1. Thus let, in our example, $x_1 = -3$, $x_2 = -1$, $x_3 = +1$. A's information structure is not a sub-partition of B's, nor conversely; and this shows up in the fact that A permits to distinguish x_2 from x_3 but not from x_1; while B permits to distinguish x_2 from x_1 but not from x_3. When the payoff function is ω (transaction cost nil) it is the former distribution that matters; when the payoff function is ω^*, it is the latter.

We can also regard each partition, or information structure, as a function (operator) η, that translates each state of environment - i.e., each element x of X - into a message $y = \eta(x)$ (y is the value of the function η). For a particular y, all those states $x_1, x_2, \ldots$ for which $\eta(x_1) = \eta(x_2) = \ldots = y$, form a particular one of the subsets generated by the information structure η.

Let P be the probability distribution on X . (When X is finite we can write p_x for the probability that a particular state x obtains.) It is clear that the distribution P , together with the information structure η, will determine the probability distribution over the set of messages y . And since the amount of information is a property of the probability distribution of messages, it is a property of the pair (P, η). In our example, the (uniform) distribution (P) of price-changes, and the information structures — say, η^A, η^B,...— characterizing the informants A, B,... — were such as to generate the distribution of messages: $(\frac{1}{2}, \frac{1}{2})$ in the case A ; $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ in the case B,... Even more: P and η taken together determine completely the joint distribution of the messages and of the states of the world; and, hence, also the conditional probability that the world is in a given state if the message received is a given one. This remark will become somewhat less trivial later when we drop our assumption that the informants were faultless.

We can now summarize our concepts formally: Write X = set of states of the world; x = a particular state; P = probability distribution on X [when X is finite, p_x is the probability of a given x]; a = action (= decision); u = payoff, ω = payoff function, η = information structure, y = message, α = rule of action, i.e., a function associating a message y with an action a .

Then y = $\eta(x)$; a = $\alpha(y)$; and u = $\omega(x, a)$. Hence u = $\omega(x, \alpha(y))$ = $\omega(x, \alpha(\eta(x)))$: the payoff depends on the state of the world and will vary with the three functions ω, α, and η. The expected payoff is $U = E_x u = E_x \omega(x, \alpha(\eta(x)))$ [in the case of finite X , $U = \sum_x p_x \omega(x, \alpha(\eta(x)))$]. Thus the expected payoff depends on the functions ω, α, η, and P . We can write $U = U(\alpha, \eta; \omega, P)$, separating by a semicolon the functions the decision-maker can choose (α and η) from those he cannot: the payoff function and the distribution function of the states of the world.

If the rule of action is a good one, $\alpha = \alpha^*$, say, U is maximized over the set of all possible action rules. That is, $U(\alpha^*, \eta; \omega, P) = \max_{\alpha} U(\alpha, \eta; \omega, P) = V(\eta; \omega, P)$, say, is the maximum expected payoff achievable with the information structure η, given ω and P. We call $V(\eta; \omega, P)$ the <u>information value</u> of η and want to emphasize that, in general, it depends on ω. On the other hand, the probability distribution of messages, or (more comprehensively) the joint probability distribution of messages and states of the world, depends on η, P but is independent of ω.

If I am in possession of an information instrument characterized by structure η^0 of information, I shall be willing to offer the <u>demand price</u> $V(\eta^1; \omega, P) - V(\eta^0; \omega, P)$ for the instrument with information structure η^1 if the difference is positive. In particular η^0 may mean "no information": a degenerate partition of X into a single subset (itself).

7. <u>Non-Equiprobable Messages</u>. In our examples so far, a faultless informant sent equiprobable messages. But our concepts apply to more general cases: it remains true that the value of a given information structure, and the ranking of the values for different information structures, depend on the payoff function.

On the other hand, the amount of information is independent of the payoff function. In particular, the information is highest, regardless of the payoffs, when the messages are equiprobable, and hence remove the "highest degree of uncertainty." Consider the following type of payoff functions, ω, and the following probabilities p_x of two alternative states of nature:

	State of nature $x =$	1	2
Action $a = 1$		r_1	0
2		0	r_2
Probability	$p_x =$	p_1	p_2 ,

where $p_1 + p_2 = 1$ and the rewards r_i are positive. How much will you pay to the informant who tells you without fault the actual state of nature? Call this information structure, η^1. With its help you will always choose the appropriate action, $a = x$, and earn, on the average, $V_1 = (\eta^1; \omega, P) = p_1 r_1 + p_2 r_2$. If you act without the knowledge of the actual state (we shall then say you possess information structure η^0), your expected reward is $p_1 r_1$ or $p_2 r_2$ depending on the action you choose once for all (or between these two amounts if you randomize your actions). In this case, the highest expected reward is $V_0 = \max(p_1 r_1, p_2 r_2)$. Hence, assuming (without loss of generality) $p_2 r_2 \geq p_1 r_1$, the demand price for η^1 is

$$V_1 - V_0 = (p_1 r_1 + p_2 r_2) - p_1 r_1 = p_2 r_2 .$$

This quantity reaches its maximum, not when "the degree of uncertainty" is highest ($p_1 = p_2 = 1/2$) but when $p_2 r_2 = p_1 r_1$, $p_1 = r_2/(r_1 + r_2)$; this is $\neq 1/2$ unless $r_1 = r_2$.

Similarly, while, for a three-message set, the information amount is highest when $p_i = 1/3$ ($i = 1, 2, 3$), in the following example the demand price is highest at different probability distributions of messages. Consider two payoff functions, ω' and ω''; and assume a probability distribution P (with $p_2 = 1/3$) :

ω'

	x = 1	2	3
a = 1	r_1	0	0
2	0	r_2	r_2
p_x =	p_1	1/3	p_3

ω''

	x = 1	2	3
a = 1	r_1	r_1	0
2	0	0	r_2
p_x :	p_1	1/3	p_3

Thus, if the payoff function is ω' it is unimportant to distinguish between the states 2 and 3 ; the demand price is the same as in the two-states example just used. Accordingly, it reaches its maximum when $p_1 = r_2 / (r_1 + r_2)$. Similarly, when the payoff function is ω'' , maximum demand price is reached when $p_3 = r_2 / (r_1 + r_2)$. And in the particular case when $r_1 = r_2$, the probability distribution resulting in the highest demand price is, in the case of each of the two payoff functions: (1/2, 1/3, 1/6) and (1/6, 1/3, 1/2), respectively, not (1/3, 1/3, 1/3)! The "highest degree of uncertainty" does not correspond to the highest demand price for information dispelling it.

There are of course many parameters of the probability distribution P on the (finite) set $X = (x_1, \ldots, x_n)$ of states of nature, that have the property of reaching the maximum when $p_1 = p_2 = \ldots = p_n$. Of such parameters, Claude Shannon's entropy measure, $-\sum p_i \log p_i$, has the further property that it increases with the minimum number of symbols needed to distinguish messages about long sequences of states of nature. To see this, compute the number of distinct sequences of length T . Each sequence in which $x_1, \ldots, x_n$ occur $t_1, \ldots, t_n$ times respectively $(\sum t_i = T)$ has the probability $p_1^{t_1} p_2^{t_2} \ldots p_n^{t_n}$; and since, with T large, it is practically certain that $t_i = Tp_i$, each such sequence has the probability $(p_1^{p_1} p_2^{p_2} \ldots p_n^{p_n})^T$ Its reciprocal, $(p_1^{p_1} p_2^{p_2} \ldots p_n^{p_n})^{-T}$ is therefore the number of distinct sequences; and its logarithm, $-T \sum p_i \log p_i$ is proportional to the number of digits needed to distinguish between messages about such sequences. As in our earlier exposition (of the special case when the states of nature are equiprobable and T = 1, and therefore the messages are equiprobable), the number of digits can be taken to be roughly proportional to the cost of transmitting information about T successive states. Hence the average cost per transmitted occurrence is proportional to $H = -\sum p_i \log p_i$, the entropy. This quantity has its maximum when all p_i are equal (= 1/n).

Changing our terminology we can also regard the set $X = (x_1, \ldots, x_n)$ as a vocabulary, and each state of nature as a sequence of words describing it. For each word, the probability of its occurrence will, in general, depend on its predecessors in the sequence; but we omit this complication for the present discussion, for our main conclusion remains valid: the cost of information is increasing roughly with the number of symbols needed and this is measured by the amount of information defined as the entropy, a statistical parameter characterizing the set of potential messages and independent of the payoff function. But the value of information, and its demand price, do depend on the payoff function.

8. Special Classes of Payoff Functions. The binary relation "η' is a sub-partition of η''" induces a partial ordering on all partitions of the set X. We have seen that

(3) $V(\eta'; \omega, P) \geq V(\eta''; \omega, P)$ for all P, ω if and only if η' is a sub-partition of η''. If two information structures are not related by sub-partitioning, it is possible to reverse the comparative ranking of their information values by changing the payoff function. However, we may consider restricting the payoff function to some special class Ω such that, for all payoff functions, the ranking of information values depends on η and P only. For example, we may define some numerical function $K(\eta, P)$

such that

(4) $V(\eta' ; \omega, P) \geq V(\eta'' ; \omega, P)$ for all ω in Ω if and only if $K(\eta', P) \geq K(\eta'', P)$. In particular, since the probability distribution of message is determined by η and P, the number $K(\eta, P)$ may be chosen to be a parameter of that distribution - for example, the entropy characterizing it. Is there such a class Ω?

9. Faulty Information. A particularly important class Ω of payoff functions is defined by the concept of faulty information, or a "noisy channel." In this case, we regard the variable, state of nature x as a pair (x^e, x^i) where x^e denotes the external or environmental state, and x^i, the internal state (i.e., the state of the information instrument). Internal state is a variable with the following property: it influences the message but not the payoff. * The message is $y = \eta(x^e, x^i) = \eta(x)$ as before; but the payoff function depends now only on x^e and the action variable a. That is, there exists a function $\bar{\omega}$ such that the payoff

(5) $u = \omega(x,a) = \omega(x^e, x^i, a) = \bar{\omega}(x^e, a)$.

A simple question which, on the face of it, seems to put our previous results in doubt, will illustrate these concepts. Suppose an expert offers to predict which of two things will happen (e.g., whether a stock will rise or fall). Won't you be willing to pay him more (certainly, not less!) if you

* From another point of view, the faulty information is the general case and the non-faulty the special one, with x^i constant.

know him to be right 90% of the time than if you know him to be right only 60% of the time? If so, the information value is larger (or at least not smaller) in the former than in the latter case, regardless of how important the event predicted is for you, i.e., regardless of how it (together with any given action of yours) affects its profit. No change in the payoff function will alter the ranking of information values. Note moreover that if the probability of the expert's being right is, say, 40%, he is as good as if he were right 60% of the time since you can always replace his message by its opposite. Thus the information value is smallest when the probability of error is = 50%, i.e., when the "degree of uncertainty" and the entropy is maximal: an ominous coincidence!

To check the intuition let x^i = state of the instrument = +1 or -1 according to whether the informant is right or wrong; and denote the two values of x^e, the external variable in question, by +1 and -1. With this notation the message y is the product $x^e x^i$:

(6) $y = x^e x^i = \eta(x^e, x^i)$

The information structure η thus defined is the partition of the set of X of pairs (x^e, x^i) into two subsets, with $x^e x^i$ = +1 and -1, respectively. By the conditions of the problem, there are two possible actions, a = +1 and -1 (say) such that a = +1 is the appropriate (i.e., the most profitable) action when the external state x^e = +1; and a = -1 is appropriate when x^e = -1. That is, if we denote the four possible payoffs $u = \bar{\omega}(x^e, a)$ by

(7) $\bar{\omega}(+1, +1) = r_1$; $\bar{\omega}(-1, -1) = r_2$; $\bar{\omega}(-1, +1) = s_1$, $\bar{\omega}(+1, -1) = s_2$,

then

(8) $\min(r_1, r_2) \geq \max(s_1, s_2)$,

i.e., r_1 , r_2 are (relative) gains, and s_1, s_2 are (relative) losses. This property of $\bar{\omega}$, a special case of the independence of ω on x^i, defines the class Ω of admitted payoff functions. We shall confirm that, if the information structure η is defined as in (6), then the ranking of the information values $V(\eta; \omega, P)$ for varying P is the same for all payoff functions in Ω. Let $Pr(x^e = +1) = p^e = 1 - q^e$; let $Pr(x^i = +1) = p^i = 1 - q^i$ = "probability of the informant's being right." The variable x^i was, in effect, assumed to be independent of x^e so that the four joint probabilities of (x^e, x^i) are : $p^e p^i$, $p^e q^i$, $q^e p^i$, $q^e q^i$. If $p^i \geq \frac{1}{2}$ we can verify that a good rule is to obey the informant : $a^* = \alpha^*(y) = y$; if $p^i < \frac{1}{2}$ a good rule is $\alpha^*(y) = -y$. Assume $p^i \geq \frac{1}{2}$, without loss of generality. Then by (6), $u = \bar{\omega}(x^e, y) = \bar{\omega}(x^e, x^e x^i)$. Therefore, for the two payoffs r_1, r_2 that are maximal in the sense of (8), the probabilities are:

$$Pr(u = r_1) = Pr(x^e = +1, x^e x^i = +1) = Pr(x^e = +1, x^i = +1) = p^e p^i$$

$$Pr(u = r_2) = Pr(x^e = -1, x^e x^i = -1) = Pr(x^e = -1, x^i = +1) = q^e p^i ,$$

adding up p^i. Hence the larger p^i the larger the expected payoff. Thus for any fixed payoff function in Ω, if experts A and B are characterized by $p^i_A > p^i_B \geq 1/2$ (or, more generally, if $\left| p^i_A - 1/2 \right| > \left| p^i_B - 1/2 \right|$) expert A has greater information value.

For a two-signal channel whose probability of not conveying a falsehood is p^i, the capacity (maximum rate of transmission) is, under conditions of our problem

$$(9) \quad C = p^i \log p^i + q^i \log q^i + \log 2 .$$

This quantity reaches its maximum at $|p^i = 1/2|$. It increases with $|p - 1/2|$, just as does the information value. However the (monotone) function that relates information value (and also the demand price) to C is not a simple one, and this function does depend on the parameters r_1, r_2, s_1, s_2 that distinguish one element of the class Ω from another. The information value (assuming $p^i \geq 1/2$) is

$$V_1 = p^i(r_1 p^e + r_2 q^e) + q^i(s_2 p^e + s_1 q^e) ;$$

without using information the expected payoff would have been:

$$V_o = \max(r_1 p^e + s_1 q^e , s_2 p^e + r_2 q^e).$$

Assuming, without loss of generality, that the second term in parentheses is maximal, we obtain as the demand price

$$V_1 - V_o = p^i p^e(r_1 - s_2) - q^i q^e (r_2 - s_1) .$$

Since $q^i = 1 - p^i \geq 1/2$ and by (8), both V_1 and $V_1 - V_o$ are linearly increasing in p^i. If the case $p^i < 1/2$ is added (and accordingly, p^i and q^i interchanged) one sees that the value of information as well as the demand price increases linearly with $|p^i - 1/2|$ (and hence are monotone increasing functions of the channel capacity C), with coefficients depending on the particular payoff function assumed.

In a "fair bet" with a fixed stake s we would have, in particular:

$$s_1 = s_2 = -s ; r_1/s = q^e/p^e ; r_2/s = p^e/q^e ; V_o = 0 ; V_1 = 2s|p^i - 1/2|.$$

This last result will be compared with the one of the next Section.

10. <u>A Case when Information Value coincides with Channel Capacity.</u> This interesting case was constructed by J. R. Kelly, Jr. [1956] of Bell Telephone Laboratories. In presenting it, we shall find it logically instructive to continue the use of our formal notation even though it is made to look incongruous by the intuitive simplicity of the problem. The problem retains all the characteristics of our very last example (fair bet) except that the set of actions and the payoff functions are defined differently. The stake s varies from bet to bet, the gambler having decided in advance that he will bet a fixed proportion ℓ of the capital available after the previous bet. Thus

the action a is now described by a pair of variables:

$$a = (k, \ell)\ ;\ k = +1, -1\ ;\ 0 \leq \ell \leq 1\ .$$

For simplicity, assume $p^e = 1/2$ (however, Kelly obtains his result also for an arbitrary p^e). Denoting by c the capital before the bet, the money losses and gains become:

$$s_1 = s_2 = -c\ell\ ;\ r_1 = r_2 = c\ell.$$

However, the payoff u, i.e., the random variable whose expected value is maximized, is assumed to be, not the money gain or loss, but the exponential rate of growth of capital, i.e., the logarithm of the ratio between the capital after the before the bet:

$$(10) \qquad u = \begin{cases} \log(1+\ell) & \text{if gambler wins} \\ \log(1-\ell) & \text{if gambler loses.} \end{cases}$$

With $p^i \geq 1/2$ (no loss of generality), a good rule requires again that the informant's advice be followed. This makes the expected payoff* equal to

$$Eu = p^i \log(1+\ell) + q^i \log(1-\ell),$$

---

* Kelly justifies the result in a different way, trying to avoid the concept of a criterion (payoff) function whose <u>expectation</u> is being maximized. He uses the time-sequence involved in his problem: the sequence of capital amounts, c_t after the t-th bet. He computes the <u>probability limit</u> of the average

$$\frac{1}{T}\sum_{t=1}^{T} \log(c_t / c_{t-1})$$

where $c_t = c_o (1+\ell)^{W_t}(1+\ell)^{L_t}$, and W_t = number of wins, L_t = number of losses in the first t bets, and $W_t + L_t = t$. This does, in effect, amount to assuming the criterion function (10); although the proof seems formally to duplicate the reasoning estiablishing the entropy formula for the transmission of a long sequence of messages about varying states of nature, Section 7.

which has maximum when $\ell = 2p^1 - 1$. Hence the information value

(11) $$V_1 = \max_{\ell} \ Eu = p^1 \log p^1 + q^1 \log q^1 + \log 2$$

is by (9) equal to C, the channel's capacity. This is also equal to the demand price since the maximum expected payoff obtained without information is

$$V_0 = \max_{\ell} \left(\frac{1}{2} \log (1 + \ell) + \frac{1}{2} \log (1 - \ell) \right) = 0 .$$

At the risk of being pedantic, it will be helpful to relate the results to our concept. We have as in the preceding section

$$x = (x^e, x^i) ; \ y = \eta(x) = x^e x^i .$$

Moreover: $a = \alpha(y) = (k, \ell) = [\kappa(y), \ell]$, say;

$$u = \omega(x,a) = \bar{\omega}(x^e,a) = \bar{\omega}(x^e, k, \ell) = \log (1 + k\ell x^e) ;$$

$$V_1 = \max_{\alpha} \ E_x u = \max_{\ell} \max_{\kappa} \ E_x \log (1 + \kappa(y)\ell x) = \max_{\ell} \ E_x \log (1 + y\ell x^e)$$
$$\sum_{x^i} \sum_{x^e} p(x^i,x^e) \log (1 + (x^e)^2 \ell x^i) ;$$

Since $(x^e)^2 = 1$, we have, denoting by $p(x^i)$ ($= p^i$ or q^i) the (marginal) probability of x^i

$$V_1 = \max_{\ell} \sum_{x^i} p(x^i) \log (1 + \ell x^i) = \max_{\ell} [p^i \log (1 + \ell) + q^i \log (1 - \ell)] = C .$$

To use Kelly's summary (p. 926): "If a gambler bets on the input symbol to a communication channel and bets his money in the same proportion each time a particular symbol is received his capital will grow (or shrink) exponentially. If the odds are consistent with the probabilities of occurrence of the transmitted symbols (i.e., equal to their reciprocals*), the maximum value of this exponential rate of growth will be equal to the rate of transmission of information."

---

* This is the case of fair odds, with p^e arbitrary; but Kelly treated also the case when odds are not fair.

In constructing his example, Kelly was motivated (p. 917-918) by the desire to attach "a value measure to a communication system" without using a function "defined on pairs of symbols which tells how bad it is to receive a certain symbol when a specified symbol is transmitted." Kelly calls this function "a cost function": in our notation, it is the function $\bar{\omega}$ in the definition of the payoff: $u = \bar{\omega}(x^e, \alpha^*(y))$ where $\bar{\omega}$ is the payoff function (independent of the state x^i of the information instrument), and where $\alpha^*(y)$ is the action appropriate to the received symbol; that is, the action rule α^* maximizes the expectation Eu:

$$\max_{\alpha} E\bar{\omega}(x^e, \alpha(y)) = E\bar{\omega}(x^e, \alpha^*(y)) .$$

This interpretation of Kelly's term "cost function" *) is confirmed by his further reference to the "utility theory of Von Neumann" and to the property of the "cost function" that "it must be such that...a system must be preferable to another if its average cost is less." Kelly believes that the "cost function approach" "is too general to shed any light on the specific problems of communication theory. The distinguishing feature of a communication system is that the ultimate receiver (thought of here as a person) is in a position to profit from any knowledge of the input symbols or even from a better estimate of their probabilities. A cost function, if it is supposed to apply to a communication system, must somehow reflect this feature."

---

*) The word "cost" stands presumably for the negative difference between two payoffs of which the smaller one is due to poorer information. This should not be confused with our use of the word cost, as the cost to the seller of information.

In this weak form, Kelly's requirement is, in fact, satisfied by <u>any</u> payoff function. For, by the definition of maximum, if a^* maximizes $\omega(x,a)$ and b^* maximizes $\omega(y,b)$ then $\omega(x,a^*) \geq \omega(x,b^*)$: the profit is improved if the action is chosen in full knowledge of the true state x. And similarly with ω replaced by its expectation, x by its distribution P, y by some estimate of P, and the action variable a by α, the rule of action.

It seems however that Kelly has in mind one or both of the following two stronger requirements. One of them is: given any payoff function, the maximum expected payoffs associated with different channels should be <u>ordered</u> according to the capacities of these channels. A still stronger requirement would be, that the maximum expected payoff be equal or proportional to, or an increasing linear function (or some other fixed increasing function) of the channel capacity independent of the payoff function; so that channel capacity is a <u>measure</u> (up to a fixed increasing transformation) of the expected payoff. We may call these two requirements the ordering and the measuring requirements, respectively.

We have seen in Section 9 on faulty information, that the <u>ordering</u> requirement is satisfied by any payoff function, for the case when the channel transmits two alternative symbols (messages) only. The case of any number of alternative messages has been studied by Beckmann [ ] and McGuire [ ]. They tell me that the following results were previously obtained by Blackwell, Sherman and Stein [ ]: in general the ordering requirement is satisfied if and only if the conditional probabilities of external states x^e given the messages y' transmitted by one of the two compared channels are fixed linear functions of the conditional probabilities of the x^e given the messages y'' transmitted by the second channel. (Beckmann also showed that in the case of 2-message channels this condition is always satisfied.)

It follows *a fortiori* that the stronger, or measurement, requirement is not, in general, satisfied; we have seen in Section 9 that it is not satisfied even for 2-message channels.

Suppose, for example, the gambler would have to pay a progressive income tax on his gains. The exponential rate of growth of capital would then be, not $\log(1 + \ell x^i)$ but $\log f(1 + \ell x^i)$, with f a concave function. And the result (11) would not obtain.

This agrees with Kelly's own result. Although he declared the intention (p. 918) "to take some real-life situation which seems to possess the essential features of a communication problem, and to analyze it without the introduction of an arbitrary cost function", he concludes that the particular criterion followed by his gambler - and resulting in the information value being equal to channel capacity - is related to the particular assumptions made: the results would be different, writes Kelly if we "for example, suppose the gambler's wife allowed him to bet one dollar each week but not to reinvest his winnings."

The re-investment feature of Kelly's example has attracted the attention of R. Bellman and R. Kalaba [1957 a and b] who regarded it as a case in dynamic programming, a good program being characterised by an optimal constant ℓ^*. These writers are quite clear that the coincidence in this case, of information value and channel capacity, is due to the logarithmic nature of the payoff function chosen,

11. <u>Incentive Fee to Forecaster</u>. In an article "Measures of the Value of Information", John McCarthy [1956] attaches to this term still a different meaning. Generalizing a concept due to I. J. Good [1952, p. 113] he defines what we might call an <u>efficient incentive fee function</u>. Good speaks of a "fair fee" and McCarthy, - more appropriately, we think - speaks of "a payoff rule to keep the forecaster honest." The payoff in question is <u>not</u> the value of information in our sense. It is not the worth of information to its user.

The question raised by both Good and McCarthy is extremely interesting and, in fact, opens up a new field of problems in the economics of information. Yet an incentive fee to the expert is not the same thing as the value of his service to the client. If I am in a hurry and am aware that exactly 25 cents added to the taxi driver's tip will suffice to make him skillfully avoid the traffic lights, I shall conditionally promise him exactly an extra quarter; though by arriving in time for my appointment I shall gain $1,000. If on this basis, the expected values of his gain and of my gain are computed, two quite different numbers will result. The same applies to the services of a forecaster.

Good's problem presented by him rather casually in a single paragraph can be stated as follows. Denote by $p = (p_1, \ldots, p_n)$ the vector of the probabilities of alternative events $k = 1, \ldots, n$. The client does not know p but has an <u>a priori</u> expectation $Ep = (Ep_1, \ldots, Ep_n) = (\bar{p}_1, \ldots, \bar{p}_n)$. The expert will tell the client an estimate y of p, $y = (y_1, \ldots, y_n)$, and receive a fee $f(y_k)$ if the event k happens. We shall call the function f an efficient incentive function if f has the following properties:

b) the expected fee $F = F(y) = \sum p_k f(y_k)$ is largest when $y = p$ (i.e., when the expert's estimates are perfect):

(12) $$F(y) = \sum p_k f(y_k) \leq \sum p_k f(p_k) = \max_y F(y) = F^*, \text{ say.}$$

b) the expected fee is nil if the expert does not know more than the client:

(13) $$\sum \bar{p}_k f(\bar{p}_k) = 0 .$$

Good states that both requirements are satisfied by

(14) $$f(y_k) = A \log y_k + B , A > 0 ; \quad k = 1,\ldots,n.$$

To see that (12) is satisfied when f has this form*, we can maximize $\sum p_k \log y_k$ with respect to the y_k subject to $\sum y_k = 1$; and then check that the maximizing vector y is non-negative. A little more insight is gained if we remember that $\log y_k$ is a concave function of y_k ; if $z_k > 0$ for all k and $\sum z_k = s$ then

$$\log(\sum p_k \cdot z_k/s) \geq \sum p_k \log(z_k/s),$$

$$\log \sum p_k z_k \geq \sum p_k \log z_k \text{ ; let } z_k = y_k/p_k \text{ ;}$$

$$0 \geq \sum p_k \log y_k - \sum p_k \log p_k \text{ ; hence if } A > 0 \text{ then}$$

$$\sum p_k (A \log p_k + B) \geq \sum p_k (A \log y_k + B) \text{ for any } B .$$

---

* McCarthy refers to an unpublished proof by Gleason, of the uniqueness of the logarithmic solution (14). Beckmann constructed the following counterexample, with $n = 2$:

(14*) $$f(y_k) = \int_{1/2}^{y_k} \frac{g(|t - 1/2|)}{t} dt , \; k = 1,2.$$

where g is an arbitrary positive-valued function. The expression

$$p_1 f(y_1) + (1 - p_1) f(1 - y_1)$$

is maximized when $y_1 = p_1$.

Hence, the logarithmic incentive function satisfies (12). It yields the maximum expected fee

$$(15) \quad F^* = \max_y F(y) = A \sum p_k \cdot (\log p_k + B),$$

a linearly increasing function of the entropy parameter that characterizes the true probability distribution of the external events. By the very nature of the problem, the comparative importance of the events and the client's actions for the client's welfare, and hence the payoff function to the client, does not enter the function f, nor the maximum expected fee F^*. This latter quantity is not, as McCarthy seems to assert, " a good measure of what it is worth to be given these probabilities." It does measure, not the worth to the client, but the (expected) cost to the client if, to stimulate the expert's efforts to give a good forecast, the client has agreed to pay him according to an efficient incentive fee of a certain form.

On the other hand, as Good remarks, his result is affected by the money utility function of the expert (not the client): if the expert maximizes, not the expectation of his money fee but, say, of its logarithm, the efficient fee schedule ceases to be related to entropy as in (15). Thus, the payoff function to the expert (the function whose value the expert tries to maximize) does affect the expected cost (not the expected payoff) to the client.

The result is also influenced by the client's "beliefs" (not by his uses or tastes, as expressed in his payoff function): for it is easily seen that in order for the incentive function (14) to satisfy the second efficiency property, (13), the number B must be

$$B = - \sum \bar{p}_k \log \bar{p}_k .$$

As to the parameter A : it is an item for bargaining between the client and the expert, as McCarthy has correctly remarked in a more general context.

In the particular case when, a priori, all events are equally probable, all $\bar{p}_k = 1/n$, $B = \log n$, $f(y_k) = A \log (ny_k)$. And if $n = 2$, $f(y_k) = A \log (2y_k)$, $F^* = A(p_1 \log p_1 + p_2 \log p_2 + \log 2)$.

Thus Kelly's result in (11) is formally a special case of Good's result. But the content is quite different. Kelly's maximand is the expected payoff to the client; the variable with respect to which it is maximised is the decision of the client, viz., the choice of the optimal fraction ℓ to be re-invested; and p_1, p_2 are characteristics of the expert (the "channel"), viz., the probabilities of his being right or wrong; p_1, p_2 are known to the client. Good's maximand, on the other hand, is the expected payoff to the expert; it is maximized with respect to the expert's decision, viz., the effort to make a better estimate of the p's; and the p's are characteristics (unknown to the client) of the external environment (the "source"), not of the expert. Kelly's quantity is related to the demand price, Good's to the supply price, of information.

McCarthy has generalized Good's problem as follows: instead of looking for a single efficient incentive function f, the same for all y_k, and depending on one variable only, he looks for n functions $f_1,\ldots,f_n$, each depending on the vector $y = (y_1,\ldots,y_n)$ of the expert's estimates. If the event k happens, the expert receives $f_k(y)$. The set $(f_1,\ldots,f_n)$ of functions is an efficient incentive function (or, in McCarthy's words, a payoff rule that keeps the forecaster honest) if, regardless of the value of $p = (p_1,\ldots,p_n)$ the expected fee $\sum p_k f_k(y)$ is maximized if and only if $y = p$, i.e., if $y_k = p_k$ for each k. McCarthy then states, without proof, that the set $(f_1,\ldots,f_n)$ is efficient if and only if $f_k(y) = (\partial / \partial y_k) \phi (y)$, where ϕ is a convex function homogeneous of the first degree. *) (Footnote on p. 29a)

---

* However, Good's case seems to provide a counter-example showing that the homogeneity property does not seem to be necessary. We know that if, for every h, $f_h(y) = A \log y_h + B$ then F is an efficient incentive function. Find $\phi(y)$ such that, for every h,

$$f_h(y) = \partial\phi / \partial y_h = A \log y_h + B ;$$

then $\partial^2\phi / \partial y_h \, \partial y_j = 0$, and hence $\phi(y)$ is a sum:

$$\phi(y) = \sum \int f_h(y) dy_h = A \sum y_h \; (\log y_h - 1) + B \sum y_h + C = A \sum y_h \log y_h + K,$$

where $K = B + C - A$.

It is easily seen that $\phi(y)$ is indeed convex. But it is not homogeneous of first degree: for $\phi(ry) - r\phi(y) = K(1-r) + Ar \log r$;

this cannot vanish for all r and fixed K, A.

Granting at least the sufficiency part of this theorem of McCarthy it would follow that the logarithmic incentive function is not the only efficient one. Beckmann's function (14*) in a previous footnote seems to confirm this for Good's special case. Consequently, the maximum expected fee does not have to be related to the entropy formula.

We may question moreover whether the client is really interested in the efficient incentive function defined as the one that "keeps the expert honest", i.e., encourages him to bring every y_k as close to p_k as possible. For different values of k, the (signed) error $y_k - p_k$ may be of different importance to the client. This brings us back to the discussion in Section 9, with each signal y interpreted now more generally, as an estimate of the probability distributions $p = (p_1, \ldots, p_n)$ of the external state of the world, x^e (in Section 9, each p_k was 1 or 0). McCarthy does undertake this analysis. Using his words (but our notation): "Suppose that on the basis of the forecaster's prediction the client chooses the action a out of the actions open to him and that his payoff if the event k occurs is $\omega(x^e_k, a)$. His expectation will be $g(p) = \max_a \sum_k p_k \omega(x^e, a)$ if a is chosen optimally." To this we have to remark that for a to be chosen optimally on the basis of the client's prediction y, the value $a = \alpha^*(y)$ (say) has to satisfy

$$\sum_k y_k \omega(x^e_k, \alpha^*(y)) = \max_a \sum_k y_k \omega(x^e_k, a) = g(y)$$

This value of a will, in general, <u>not</u> maximize

$$\sum_k p_k \omega(x^e_k, a) .$$

The expectation of client's payoff is therefore

$$(16) \quad \sum_k p_k \omega(x^e_k, \alpha^*(y)) \leq \max_a \sum_k p_k \omega(x^e_k, a) = g(p);$$

to be sure, the inequality sign drops if $y = p$, a situation that is approached but, in general, not achieved by the use of the efficient incentive function which was described. Depending on the payoff function ω it is possible that the difference between the two averages in (16) is diminished more easily by encouraging the expert to make very precise estimates of some p_k while treating others with less care.

Some tantalizingly short remarks of McCarthy give the promise of a fruitful analysis of the non-zero sum game between the expert and the client, and of the expert's effort. In this approach, the entropy formula loses its significance; and the payoff functions of the expert and the client must gain in significance.

References

Beckmann, M.

Bellman, R. and Kalaba, R. [a], On the role of dynamic programming in statistical communication theory, IRE Transactions of the Professional Group on Information Theory, 1957, IT-3, 197-203.

Bellman R. and Kalaba, R. [b], Dynamic programming and statistical communication theory. Proc. Nat. Acad. Sci., 1957, 43, 749-751.

Blackwell, D., Sherman, and Stein,

Feinstein, A., A new basic theorem of information theory, IRE Transactions of Prof. Group, 1954, 2-22.

Gilbert, E. N., An outline of information theory. The American Statistician, 1958, 12, 13-19.

Good, I. J., Rational decisions, J. Royal Stat. Soc., B., 1952, 14, No. 1, pp. 107-115, especially p. 112.

Hitch, Ch., Economics and military operations research, Rev. Ec. Stat., 1958, 50, 199-209.

Kelly, J., A new interpretation of information rate. Bell Syst. Tech. J., 1956, 35, 917-926.

Marschak, J., Towards an economic theory of information and organization, Decision Processes, Coombs, Thrall and Davis, eds. New York 1954, 187-220.

Marschak, J. and Radner, R., Economic Theory of Teams, Chapters 2 and 3 (mimeo.) Cowles Foundation Discussion Paper No. 59, 1958.

McCarthy, J., Measures of the value of information. Proc. Nat. Acad. Sci., 1956, 42, 654-655.

McGuire, C. B.,

Page, Thornton, The value of information in decision-making. Proceedings of the First International Conference on Operations Research (Oxford 1957) Operations Research Society of American, 1957, 306-314.

Savage, L. J., The Foundations of Statistics, New York 1954.

Shannon, C.

Von Neumann, J. and O. Morgenstern. Theory of Games and Economic Behavior, 2nd ed., 1948.