JOSEPH D. SNEED

ENTROPY, INFORMATION, AND DECISION*

It is acceptable common usage to speak of some sentences or propositions as being more informative than others, or perhaps as conveying more information. We speak of a sentence or proposition as being more informative to one person than to another, and of its being more informative to a person today than it was yesterday. It is also not unusual to speak of the information conveyed by a sentence or proposition (perhaps to a person) as well as the information it conveys (to a person) about a certain subject matter. In all these usages of the words 'informativeness' and 'information' denote a property of sentences or propositions. Whether or not they denote the same property in all such usages may not be clear, but it is nevertheless clear that there is a well-established and intuitively acceptable usage of 'informativeness' to denote some property or properties of sentences or propositions. Whether one speaks of sentences or propositions here is not crucial to the discussion, and subsequently I will speak only of propositions.

It is reasonably clear that statistical communication theory (or information theory) is not directly concerned with any concept of information involved in these usages. In most authoritative expositions of statistical communication theory some care is taken to explain that the engineering problems of efficient communication are treated independently of the informativeness, in any sense involved in the usage exemplified above, of the messages transmitted [12]. Despite these disclaimers, it appears that the distinction between the concept of information employed in statistical communication theory and the concept of information as a property of propositions is not always strictly observed and some have seen a need to reiterate and elucidate this distinction [2]. I shall not review this work here but only point out that I am concerned with one aspect of informativeness – a property of propositions.

In considering the informativeness of propositions it is customary to distinguish two kinds of informativeness – semantic and pragmatic [3]. A semantic property of a proposition is, roughly speaking, one which a

Synthese **17** (1967) 392–407. © *D. Reidel Publishing Co., Dordrecht-Holland*

proposition possesses in virtue of what must be the case for it to be true – its truth conditions. Thus one should expect to be able to determine whether a proposition is semantically informative, or semantically more informative than another, by examining only the truth conditions of the propositions. In particular, one should not have to consider whether anyone knows or believes the proposition to be true, or whether believing or knowing that it is true might be useful to anyone. In contrast, a property is pragmatic if a proposition has it in virtue of certain properties possessed by people who have certain attitudes, such as belief and desire, toward the proposition. An example is a property which a proposition has in virtue of the usefulness to some individual of knowing or believing that the proposition is true. Thus, in determining whether a proposition is pragmatically informative one should expect to consider more than just the truth conditions of the proposition (though these might be relevant too). Things like the beliefs, desires, and capabilities of people – their attitudes toward other propositions – as well as their attitudes toward the proposition in question might be relevant.

There is a significant body of literature, beginning with the work of Carnap and Bar-Hillel, which undertakes to distinguish and explicate various senses of semantic information and/or semantic informativeness [3, 4, 6]. In this work a mathematical formalism very similar to the formalism employed in statistical communication theory plays a leading role. Even though the concept of information being explicated is explicitly recognized as being distinct from the communication engineer's concept, one is tempted by the similarity of formalism to say that the same concepts are applicable in explicating both concepts of information. In particular, one might be tempted to say that the concept of entropy is crucial in understanding both concepts of information.

So far as I know, there has been no systematic effort directed toward developing a comprehensive theory of pragmatic informativeness. However, much of the work done by statisticians concerned with the design of experiments can be viewed as attempting to provide a partial account of the notion of pragmatic informativeness. They are concerned with expounding principles for ranking various questions one might ask in accordance with their pragmatic informativeness [11]. In this effort attempts have also been made to apply the mathematical formalism of statistical communication theory, most notably by Lindley [9]. Recently Adams has undertaken to provide what one might view as an account of the pragmatic informativeness of scales of measurement, employing the concepts of statistical communication theory [1].

In this paper an outline of a theory of pragmatic informativeness within the framework of Bayesian decision theory will be suggested. The relevance of the concepts of statistical communication theory to this theory will be investigated. In particular, it will be shown that, in certain well-defined circumstances, these concepts are undoubtedly useful, but that attempts to apply such concepts outside these circumstances are not, in any obvious way, fruitful. The work of Lindley and Adams will be examined in the light of these results.

In explicating the concept of pragmatic informativeness, I propose to focus on one feature of this concept. Roughly speaking, the feature is this. The more informative a proposition is to an individual, the more he ought to be willing to pay to find out whether the proposition is true. A paradigm case of informativeness in this sense might be a situation in which a government is paying a free-lance spy on the basis of the informativeness of the facts he transmits.

We can get a firmer hold on this sense of 'informativeness' by considering how one might fit it into a theory of subjective probability and utility. The theory I have in mind is the one recently proposed by Jeffrey [7].[1] In this theory two related measures, P (probability) and D (desirability or utility), are defined on B, a Boolean algebra of propositions. One might ask whether it is possible to define a third measure I (informativeness) on this Boolean algebra, related to both P and D, such that $I(x)$ is intuitively identified as the value to the agent of discovering whether or not x is true.

One of the first difficulties that occurs to one considering this question is this. Roughly speaking, how much it is worth to the agent to discover the truth value of x will depend on what courses of action the agent believes are open to him. The amount a government is willing to pay for the spy's facts will depend on what its various policy alternatives are. The amount a man is willing to pay for knowing whether or not a used car needs a valve job depends on whether or not the option of buying the car is one that he believes open to him.

This suggests that it might be appropriate to consider a concept of 'informativeness relative to a given decision problem'. The decision problem would be characterized by a set of propositions believed by the agent to be mutually exclusive and jointly exhaustive and also believed by him to be propositions he can make true at will. If we did this, then the informativeness measure on the agent's propositions would not be unique for the agent. But if we like we can remove this part of the relativeness by considering informativeness relative to the most comprehensive decision problem the agent can face at a given time. This is the problem in which the alternatives are the longest conjunctions of propositions, each one of which is an alternative in some decision problem. If we did this the informativeness measure would still depend on which propositions the agent believed he could make true as well as on other facts about his P and D measures.

One also encounters a difficulty in deciding exactly what entities are to serve as the domain of the informativeness measure. At first glance, the set of elements B of the Boolean algebra of propositions seems to be the obvious choice. But it seems clear that finding out the truth value of x should be worth exactly as much as finding out the truth value of $\bar{x}$. This suggests that the appropriate domain for the informativeness measure might be the *set* of sets of mutually exclusive and jointly exhaustive propositions. However, somewhat more generality and intuitive plausibility can be obtained by taking the domain to be the *set* of all sets of propositions

$$C = (c_1, c_2, \ldots, c_n)$$

that the agent believes to be mutually exclusive and jointly exhaustive in the sense that

$$P(c_1 \vee c_2 \vee \ldots \vee c_n) = 1$$
$$P(c_i \wedge c_j) = 0 \quad \text{if } i \neq j.$$

Let χ be the set of all sets of propositions that the agent believes to be mutually exclusive and jointly exhaustive. The informativeness measure on χ would be a measure of how much it is worth to the agent to discover which one of the propositions in the set is true, in the context of a given decision problem.

With this intuitive understanding of the sense of 'informativeness' we are considering, our task is one of defining, in terms of the agent's P and D functions, the informativeness value, relative to a given decision problem, of a set of propositions that the agent believes to be mutually exclusive and jointly exhaustive.

The most natural way to do this appears to be to relate the informativeness value of the set of propositions

$$C = (c_1, c_2, \ldots, c_m)$$

to the value of the strategies the agent has available to him if he knows he is going to be told which member of C is true. More precisely, we might say that the informativeness value of C, relative to the decision problem characterized by the set of courses of action

$$A = (a_1, a_2, \ldots, a_n)$$

is the expected value of an experiment designed to discover which member of C is true.

If we let S_C^A be the set of strategies that could be based on C, the typical member of S_C^A will look like:

$$S^{\{j_i\}} = \bigwedge_{i=1}^{m} A_{j_i}/C_i; \quad j_i \in N = (1, 2, \ldots, n) \tag{1}$$

where 'x/y' means 'y causes x' (For a discussion of the role of the causal operation '/' in expressing strategies see [13].) Then clearly the informativeness value, relative to A, of being told which member of C is true is

$$I(C, A) = \max_{\{j_i\}} D(S^{\{j_i\}}) - \max_{a_i \in A} D(a_i). \tag{2}$$

In most cases of interest

$$D(S^{\{j_i\}}) = \sum_{i=1}^{m} D(a_{j_i} \wedge c_i)\, P(c_i), \tag{3}$$

and

$$\max_{S^{\{j_i\}} \in S} D(S^{\{j_i\}}) = \sum_{i=1}^{m} \max_{j_i \in N} D(a_{j_i} \wedge c_i)\, P(c_i). \tag{4}$$

The informativeness of C relative to the decision problem A is then just the difference between the desirability of the most desirable strategy based on C and the desirability of the most desirable 'pure' course of action in A. The essentials of this account of the value of an experiment are due to Savage [11].

It is not apparent how the concepts of statistical communication theory are relevant to the sense of informativeness explicated by (2). Indeed, so long as we confine our attention to the most general characterization of a decision problem, it is not apparent that they are relevant in any way whatsoever in the general case. There simply seems to be no way these concepts could play a role in evaluating (2). I am not prepared to offer any argument for this claim beyond reporting an unsuccessful effort to find an application for these concepts. However, if we consider a special type of decision problem then we find that these concepts can play a significant role in evaluating the informativeness of C, relative to a decision problem of this type.

To characterize this special class of decision problems consider a set of propositions,

$$X_0 = \{x_1, x_2, \ldots, x_N\}$$

which the agent believes to be mutually exclusive and jointly exhaustive. Suppose that the agent is interested in discovering which member of X is true and that he has the following means at his disposal. He may partition X_0 into $n \leqslant N$ disjoint, non-void subsets

$$X_1^1, X_2^1, \ldots, X_n^1$$

of his own choosing and pay a fee r to be told reliably which member of the partition contains the true member of X. Say he is told X_1^1. If

$$N(X_1^1) \geqslant n$$

he may then partition X_1^1 into n disjoint, non-void sub-sets

$$X_{i1}^2, X_{i2}^2, \ldots, X_{in}^2$$

and pay r to be reliably told which member of this partition contains the true member of X. If

$$1 < N(x_i^1) < n$$

he may pay r to be told which member of X_i^1 is true. If $N(\chi_i^1) = 1$ the agent has obviously achieved his aim and stops the procedure.

There are, for a given n and N, a finite number v different questioning procedures of this sort that the agent could employ in attempting to discover which member of X is true. Call these *'n-ary questioning procedures for X at constant rate r'*.

Let

$$Q_X = (q_1, q_2, \ldots, q_v)$$

be a set of mutually exclusive and jointly exhaustive propositions describing the employment of these different n-ary questioning procedures to discover which member of X_0 is true. Suppose now that the decision problem faced by the agent is one in which the set of courses of action he believes open to him is Q_X. The agent is thus concerned with computing

$D(q_i)$ = the expected value of discovering which member of X is true using the n-ary questioning procedure q_i.
$= K - C(q_i)$.

where K = the value of finding which member of X is true

$C(q_i)$ = the expected cost of the n-ary questioning procedure described by q_i.

Let us now consider the informativeness value of the set of propositions C, relative to the decision problem characterized by Q_X. It is in considering informativeness, relative to this special sort of decision problem, that the concepts of statistical communication theory seem to naturally apply. To see this we shall need to express these concepts in a notation amenable to the formulation of decision theory we are using.

Let B be the set of all propositions in the agent's propositional algebra and χ the set of all sets of propositions the agent believes to be mutually exclusive and jointly exhaustive. Define the operation $\otimes$ on χ in the following way:

(D-1) If

$$A = (a_1, a_2, \ldots, a_n)$$
$$B = (b_1, b_2, \ldots, b_m)$$

are included in χ, then

$$A \otimes B \equiv (a_i \wedge b_j)_{i=1}^{n} {}_{j=1}^{m}.$$

It is easy to show that $A \otimes B \in X$. One may also define a natural equivalence relation '$\sim$' on X such that

$$A \otimes B \sim B \otimes A$$
$$(A \otimes B) \otimes C \sim A \otimes (B \otimes C)$$
$$A \otimes A \sim A.$$

Define the functions H_n from χ into the real numbers by:

(D-2) For all $X = \{x_1, x_2, \ldots, x_n\} \in \chi$

(5) $$H_n(X) = -\sum_{i=1}^{N} P(x_i) \lg_n P(x_i)$$

and the functions H_n from $\chi \times B$ into the real numbers by:

(D-3) For all $X = \{x_1, x_2, \ldots, x_N\} \in \chi$ and $y \in B$

(6) $$H_n(X, y) = -\sum_{i=1}^{N} P(x_i, y) \lg_n P(x_i, y).$$

We shall call $H_n(X)$ the *n-ary entropy of X* and $H_n(X, y)$, the *n-ary entropy of X, given y* [8].

The following results are relevant to the subsequent discussion. All follow straightforwardly from the preceding definitions [8].

(T-1) For all $A = \{a_1, a_2, \ldots, a_N\} \in \chi$
$B = \{b_1, b_2, \ldots, b_M\} \in \chi$

(7) $$H_n(A \otimes B) = H_n(A) + \sum_{i=1}^{N} P(a_i) H_n(B, a_i),$$

and in the case that A and B are mutually independent, i.e.,

$$P(a_i \wedge b_j) = P(a_i) P(b_j)$$

for all $a_i \in A$ and $b_j \in B$, then

(8) $$H_n(A \otimes B) = H_n(A) + H_n(B).$$

Also

(9) $$\sum_{i=1}^{N} P(a_i) H_n(B, a_i) \leqslant H_n(B),$$

(10) $$H_n(A \otimes B) \leqslant H_n(A) + H_n(B),$$

and

(11) $H_n(A, a_i) = 0.$

The entropy functions we have defined may be related to the decision problem we are considering by way of the following results from statistical communication theory.

(T-2) If

$$X = \{x_1, x_2, \ldots, x_N\} \in \chi \quad \text{and}$$
$$Q_X = \{q_1, q_2, \ldots, q_v\}$$

is a set of propositions describing the n-ary question procedures at constant rate r for X, then

(i) For all $q_i \in Q_X$

(12) $rH_n(X) \leqslant C(q_i)$

(ii) There exists a $q_j \in Q_X$ such that

(13) $C(q_j) < r(H_n(X) + 1).$

This theorem can be regarded as setting bounds on the *minimal* expected cost of discovering which member of X is true using the n-ary questioning procedures described by members of Q_X, i.e.

(14) $$rH_n(X) \leqslant \min_{q_i \in Q_X} C(q_i) < r(H_n(X) + 1).$$

We shall also want to consider the expected cost of q_i, given that y is true, i.e. the expected cost of $q_i \wedge y$. A theorem analogous to T-2 yields the following result:

(15) $$rH_n(X, y) \leqslant \min_{q_i \in Q_X} C(q_i \wedge y) < r(H_n(X, y) + 1).$$

These results are essentially consequences of the so-called "noiseless coding theorem" [10]. This can be seen if one recognizes that there is a one-one correspondence between what I have called "an n-ary questioning procedure at constant rate r for the set of propositions X", and what communication theorists call "a separable n-ary code for the message ensemble X". This correspondence is employed by Cox [5] in a similar context. The quantity $C(q_i)/r$ is to be identified with the average length of the encoded message.

Consider now the informativeness value, relative to Q_X, of being told which member of C is true.

An elementary calculation shows that (2), the definition of $D(q_i)$, and (15) lead to

$$(16) \qquad r(H_n(X) - \sum_{i=1}^{N} P(C_i)\, H(X, C_i) - 1) < I(C, Q_X)$$
$$\leqslant r(H_n(X) - \sum_{i=1}^{N} P(C_i)\, H_n(X, C_i) + 1).$$

We can thus put upper and lower bounds on $I(C, Q_X)$ which are expressed as values of entropy functions. What is the intuitive significance of this fact?

Consider first the special case in which

$$(17) \qquad \min_{q_i \in Q_X} C(q_i) = rH_n(X)$$
$$\min_{q_{j_i} \in Q_X} C(q_{j_i} \wedge C_i) = rH_n(X, c_i).$$

Note that these equalities will always hold approximately for large values of $H_n(X)$ and $H_n(X, C_i)$ and may hold in other cases also. In this special case the minimum expected cost of discovering which member of X is true by an n-ary questioning procedure at rate r is exactly r times the n-ary entropy of X. In this case

$$(18) \qquad I(C, Q_X) = r(H_n(X) - \sum_{i=1}^{N} P(c_i)\, H_n(X, c_i)).$$

Note that in this special case

$$I(C, Q_X) \geqslant 0$$

with equality holding if and only if C and X are mutually independent. This is because

$$\sum_{i=1}^{N} P(C_i)\, H_n(X, c_i) \leqslant H_n(X).$$

Also note, that as a consequence of (11),

$$I(X, Q_X) = rH_n(X).$$

We can look at this result intuitively in the following way. The best questioning procedure is the one with the lowest expected cost. Clearly which questioning procedure this is will depend on the agent's probability distribution on X. The value of the agent's entropy function for X can be used to bracket this minimum expected cost. In this special case we assume the minimum expected cost actually achieves its lower bound. When the agent discovers which member of C is true his probability distribution on X will change (unless C and X are mutually independent). In general this will cause a change in the expected costs of the various questioning procedures and perhaps a different one will have the smallest expected cost. For each $c_i \in C$, the minimum expected cost after c_i has been discovered to be true will be, in our special case, equal to $rH_n(X, c_i)$. Of course it might be that the agent's probability distribution on X changes in such a way that the minimum expected cost of discovering which member of X is true is actually greater after he discovers which member of C is true than it was before. In our special case, this will happen if the agent finds out that c_i is true and

$$H_n(X) < H_n(X, c_i).$$

But if we consider the agent's *expected value* of the minimum expected cost of discovering which member of X is true after discovering which member of C is true, in our special case

$$\sum_{i=1}^{N} P(c_i)\, H_n(X, c_i),$$

we find that it will never be greater than the minimum expected cost before finding out which member of C is true. This is to say, it is never rational for the agent to pay someone not to tell him which member of C is true.

It is clear now that the informativeness value of finding out which member of C is true, relative to the decision problem of choosing the best n-ary questioning procedure at rate r for discovering which member of X is true simply the difference between the minimum expected cost of such a question procedure before discovering which member of C is true and the expected value of this minimum expected cost after discovering which member of C is true. In our special case where the minimal expected costs are equal to the entropy we get expression (18) for the informativeness value. We can thus say that the informativeness value of C, relative to Q_X, is proportional to the expected value of the change in the agent's entropy of X when he discovers which member of C is true. Roughly, more informative C's are those expected to produce a greater change in the agent's entropy. If one is faced with the decision problem Q_X and has a choice of several 'experiments' – several C's$\in\chi$ which one might choose to learn which member of C is true – it is clear, at least in this special case, that one will want to choose the experiment which is expected to produce the greatest change in the entropy of X.

In the general case where we do not have the equalities (17), (i.e. where the minimum expected costs do not actually achieve their lower bound) we can not identify the informativeness value of C, relative to Q_X, with the agent's expected change in his entropy of X when he finds out which member of C is true. But it is still the case that the expected change in the entropy of X is what determines the bounds on the informativeness value. $I(C, Q_X)$ is confined to an interval of width $2r$ around the quantity

$$r(H_n(X) - \sum_{i=1}^{N} P(c_i)\, H_n(X, c_i)).$$

It is still the case here that, when given a choice, one should choose the experiment which is expected to produce the greatest change in the entropy of X, provided you are faced with the decision problem Q_X.

In view of these results one might be led to say that the n-ary entropy of X for the agent, $H_n(X)$, is a measure of the agent's uncertainty about which member of X is true, and likewise that $-H_n(X)$ is a measure of the agent's information about which member of X is true. This is a perfectly legitimate intuitive way of describing the role of the entropy function in the preceding discussion. Its appeal is enhanced by the fact that it appears to correspond to the usage of $H_2(X)$ in the theory of semantic information. In this theory, $-H_2(X)$ is identified as the expected value of the amount of semantic information conveyed by the members of X, and the quantity

$$(19) \qquad -[H_2(X) - \sum_{i=1}^{N} P(c_i)\, H_2(X, c_i)]$$

is identified as the expected value of the change in this expected value due to discovering which member of C is true [4]. It is important to keep in mind that the probabilities appearing in (19) are so-called 'logical probabilities', so that some argument at least is needed to show that there is any more than a coincidental similarity of formalism linking semantic information theory and our expression (16) where the probabilities are subjective probabilities.

Despite its intuitive appeal and analogy with the theory of semantic information, it appears that this way of describing the role of $H_n(X)$ in a theory of pragmatic informativeness can lead to misunderstandings of the following sort. One might be led to suggest that we pursue analogy with the theory of semantic information and identify the informativeness of finding out which member of C is true, for someone who is in some way 'interested' in which member of X is true, as the amount by which finding out which member of C is true can be expected to reduce his uncertainty about which member of X is true, i.e.

$$\mathscr{I}(X, C) = H_n(X) - \sum_{i=1}^{N} P(c_i)\, H_n(X, c_i). \tag{20}$$

Thus, if one had to choose between various 'experiments', i.e., various C's, one would choose the one which maximized $\mathscr{I}(X, C)$, regardless of the specific nature of his interest in X.

Our discussion has shown that if the agent's interest in X is a very special one – finding which member of X is true using an n-ary questioning procedure at constant rate – then indeed he should choose the C which maximizes $\mathscr{I}(X, C)$. But it does not support the further conclusion that, whatever his interest in X is, he should choose C to maximize $\mathscr{I}(X, C)$. Counter-examples are easily given. Consider

$$X = (x_1, x_2, \ldots, x_5)$$
$$C = (c_1, c_2)$$
$$C' = (c'_1, c'_2)$$

$$P(x_i) = 0.2 \text{ for all } i$$

$P(a, b)$

$a \backslash b$	c_1	c_2	c'_1	c'_2
x_1	0.4	0.3	0.5	0.2
x_2	0.3	0.4	0.2	0.5
x_3	0.1	0.1	0.1	0.1
x_4	0.1	0.1	0.1	0.1
x_5	0.1	0.1	0.1	0.1

Suppose the agent's interest in X is that he is to pick one member of X and receive 100 dollars if the member he picks is true and nothing otherwise. Clearly he will pick an X_i, which has a maximum probability, and his expected return will be $P(x_i) \times 100$ dollars. It is also clear that C' is a more valuable experiment than C since, whatever its outcome, the agent's maximum probability for a member of X, conditional on that outcome, is 0.5 while for C it is only 0.4. However, C provides a greater value for $\mathscr{I}(X, C)$ than C',

$$\mathscr{I}(X, C) = 0.29$$
$$\mathscr{I}(X, C') = 0.25.$$

With the interest that the agent has in X in this situation it is clear that he should choose C' over C although C maximizes $\mathscr{I}(X, C)$.

In the light of this sketch of a theory of pragmatic informativeness and the role of the entropy function it is interesting to consider briefly some other efforts aimed at applying the entropy concept to questions of pragmatic informativeness.

It has been suggested by Lindley [9] that $-H_2(X)$ is a measure of the agent's information about the state of nature with respect to X, quite apart from considerations of the relevance of this information to any specific decision problem facing the agent. If the purpose of the agent's investing in experiments is to increase his store of this sort of pure disinterested information about X, and $-H_2(X)$ is a measure of this information, then clearly one should, when faced with a choice of experiments choose one which maximizes $\mathscr{I}(X, C)$. Adams [1] seems to be applying this suggestion to a special case when he identifies the informativeness of scales of measurement (his example is the Mohs hardness scale) as $\mathscr{I}(X, C)$ where C is a set of propositions describing the possible results of a measurement on this scale (in his example of the possible results of scratch tests) and X is a set of propositions describing the possible results of some experiment (an experiment to determine the scratch behavior of two minerals).

It seems reasonably clear that these authors are proposing or appealing to something like a theory of pragmatic informativeness. They expound criteria for choosing more or less informative questions when the purpose in asking questions is to increase the store of disinterested information about something. Our discussion has shown that their criteria cannot be generally applicable. There are interests – aside from accumulating disinterested information – which are not best served by choosing questions in the way they suggest. This does not, of course, show that their suggestions are not of value in situations where one is concerned with obtaining more pure, disinterested information.

One might, however, raise a question about the status of the notion of pure, disinterested information. It is not readily obvious that anyone can be correctly described as being concerned with doing experiments which increase his store of this sort of information. One conceivable answer to this query might be to identify pure, disinterested information as semantic information. If one accepts the available accounts of semantic information *and* assumes that the applicability of their results extends beyond the simple languages for which they have been developed, then chosing experiments to maximize $\mathscr{I}(X, C)$ will maximize the expected increase in the expected value of the amount of semantic information conveyed by the members of X. Of course, it still remains to be demonstrated that anyone can be correctly described as being concerned with increasing his store of semantic information.

Another conceivable answer to this query is to explicate the concept of pure disinterested information about X in terms of the expected cost of discovering which member of X is true by an n-ary questioning procedure at constant rate. It does not seem implausible to identify the agent's quantity of disinterested information about X with minus one times the least he should expect to pay to find out which number of X is true by some n-ary questioning procedure. He has less information the more he expects to pay, and it is disinterested in the sense that whatever value he attaches to finding out which member of X is true is not relevant to the analysis.

If we explicate the concept of pure, disinterested information in this way then our previous discussion provides a justification for taking $-H_n(X)$ as a measure this sense of information which is independent of any theory of semantic information.

BIBLIOGRAPHY

[1] Adams, E. W., 'On the Nature and Purpose of Measurement', *Synthese* **16** (1966) 125–168.

[2] Bar-Hillel, Y., 'An Examination of Information Theory', *Philosophy of Science* **22** (1955) 86–105.

[3] Bar-Hillel, Y., 'Semantic Information and its Measures' in Y. Bar-Hillel, *Language and Information*, Reading, Mass., 1964.
[4] Bar-Hillel, Y. and Carnap, R., 'An Outline of a Theory of Semantic Information', in Y. Bar-Hillel, *Language and Information*, Reading, Mass., 1964.
[5] Cox, R. T., *The Algebra of Probable Inference*, Baltimore 1961.
[6] Hintikka, J., 'On Semantic Information', in *Proceedings of the International Colloquium on Logic, Physical Reality and History at the University of Denver*, New York (forthcoming).
[7] Jeffrey, R. C., *The Logic of Decision*, New York 1965.
[8] Khinchin, A. I., *Mathematical Foundations of Information Theory*, New York 1957.
[9] Lindley, D. V., 'On a Measure of the Information Provided by an Experiment', *Annals of Mathematical Statistics* **29** (1956) 986–1005.
[10] Reza, F. M., *An Introduction to Information Theory*, New York 1961.
[11] Savage, L. J., *Foundations of Statistics*, New York 1954.
[12] Shannon, C. and Weaver, W., *The Mathematical Theory of Communication*, Urbana, Ill., 1949.
[13] Sneed, J. D., 'Strategy and the Logic of Decision', *Synthese* **16** (1966) 270–283.

REFERENCES

* The research reported here was supported by NSF grant GS-1263.

[1] I doubt that anything I have to say depends essentially on this theory, but the ontology of the Jeffrey theory seems more convenient than others for formulating my claims.