COWLES FOUNDATION DISCUSSION PAPER NO. 71

Comparisons of Information Structures*

by

C. B. McGuire**

April 21, 1959

---

* Research undertaken by the Cowles Commission for Research in Economics under Contract Nonr-358(01), NR 047-066 with the Office of Naval Research.

** For numerous helpful discussions I wish to express my thanks to Martin Beckmann and Jacob Marschak, whose interests in the matters dealt with here led to my own.

Comparisons of Information Structures

1. Introduction: The Conflict between Dollars and Bits as Measures of Information

The following simple example, due to Marschak, serves to introduce the questions to be discussed. Consider a speculator who must choose between two actions:

1) to buy today and sell tomorrow one unit of a commodity,

2) to sell today and buy tomorrow one unit of the same commodity.

Everything the speculator knows about the change of price from today to tomorrow will be regarded as summed up in his personal probability distribution over the set of price changes. For simplicity let us suppose that he regards all price changes in the interval $[-1, 1]$ as equally likely. Without forecasting help, clearly, the best he can do will result in an expected profit not greater than zero. In order to consider some of the alternative types* of price forecasts

---

* The legitimacy of restricting attention here to those apparently very special kinds of forecasts will be discussed below.

---

which may be bought by the speculator, let us denote by x the true price change and by $\eta(x) = y$ the message resulting from forecast η. In these terms, consider the following alternative forecasts:

$$\text{Forecast } \eta_2 : \quad \eta_2(x) = \begin{cases} \text{"up"} & \text{if } -1 \leq x \leq 0 \\ \text{"down"} & \phantom{\text{if }} 0 < x \leq 1 \end{cases}$$

$$\text{Forecast } \eta_3 : \eta_3(x) = \begin{cases} \text{"up"} & -1 \le x \le -1/3 \\ \text{"small change"} \quad \text{if} & -1/3 < x \le 1/3 \\ \text{"down"} & 1/3 < x \le 1 \end{cases}$$

$$\text{Forecast } \eta_4 : \eta_4(x) = \begin{cases} \text{"up a lot"} & -1 \le x \le -1/2 \\ \text{"up a little"} & -1/2 < x \le 0 \\ \text{"down a little"} & 0 < x \le 1/2 \\ \text{"down a lot"} & 1/2 < x \le 1 \end{cases} \quad \text{if}$$

(Notice that these forecasts are, by definition, always correct.) How much is each of these forecasts worth to the speculator? The results of η_2 clearly enable him to act in as perfect a fashion as would a perfect forecast $(\eta(x) = x)$; his expected gross profit is 1 , so it can be said that the value of η_2 to the speculator is 1 . Forecast η_3 is of less value; when either of the signals "up" or "down" occurs he can take a perfect action, but when "small change" is the signal he is uncertain whether to buy or sell. Gross profit in this case--and hence the value of Forecast η_3 --is 2/3 . Forecast η_4 has value equal to that of η_2 , but no more, because no use can be made of the added detail in the prediction. In general, then, as Marschak has pointed out, the forecasts which subdivide the x interval into an <u>even</u> number of equal intervals have value 1 ; those which subdivide into an <u>odd</u> number of equal intervals have value less than 1 , no matter how fine the subdivision.

For any finite set $E = \{E_1, \ldots, E_n\}$ of mutually exclusive and exhaustive events with probabilities $p(E) = \{p(E_1), \ldots, p(E_n)\}$ Shannon [1] has proposed a measure

$$(1) \qquad H[p(E)] = -\sum_i p(E_i) \log p(E_i)$$

of the corresponding state of uncertainty.* The quantity (1) can be regarded

---

* For any i for which $p(E_i) = 0$, $p(E_i) \log p(E_i)$ is defined to be equal to zero.

---

as the "entropy" of the situation, or alternatively as the amount of "information" gained when one of E_i is revealed as the true value. Applying Shannon's measure to the forecasting systems available to the speculator in the fashion

$$H(\eta_3) \equiv H(1/3, 1/3, 1/3) = \log 3, \text{ etc.},$$

we find that $H(\eta_2) < H(\eta_3) < H(\eta_4) < \ldots$, an ordering of "information systems" clearly in conflict with the value ordering. That the two measures do not agree is not surprising, of course, for Shannon's gives as much credit for information that the speculator cannot use (e.g., $x \epsilon [0, 1/2]$, as distinct from $x \epsilon (1/2, 1])$ as for information that he can use.

At this point the Shannon information theorist would object with the claim that the H measure can be adapted to situations where some events are of interest and others not. Consider two finite systems of mutually exclusive and exhaustive events $E = \{E_i\}$ and $F = \{F_j\}$ with associated joint probabilities $p(EF) = \{p(E_iF_j)\}$. The most important property of the H measure is that

(2) $$H[p(EF)] = H[p(F)] + \mathcal{E}_i H[p(F|E_i)] .$$

Property (2), in fact, together with continuity and a maximum at $p = (\frac{1}{n}, \ldots, \frac{1}{n})$ determine the function H uniquely (up to a multiplicative constant).* In words, (2) states that knowledge of the outcome of E reduces

---

* A proof is given in Khinchin [2], p. 9-13.

---

the uncertainty--or entropy--of the compound system EF by the amount of uncertainty of the E system. Since whatever uncertainty remains is entirely in the F system, it can be said that knowledge of E has produced an amount of information about F equal to

(3) $$R[p(EF)] \equiv H[p(F)] - \mathcal{E}_i H[p(F|E_i)]$$

or equivalently (with obvious notation)

(3') $$R(EF) = H_F \mathcal{E}_E [p(F|E)] - \mathcal{E}_{EF} H[p(F|E)]$$

or

(3") $$R(EF) = H(E) + H(F) - H(EF) .$$

R , called by engineers--for reasons not relevant here--the "rate of transmission," can be shown to be always nonnegative. Expression (3") justifies our use of symmetric notation. If E and F are independent, then $R(EF) = 0$; if they are identical $R(EF) = H(F) = H(E)$ --both characteristics in accord with intuitive demands we would make on an information measure.

Before applying the measure R in the example of the speculator we must decide what variables play the roles of E and F in (3). For a given forecast η_n , the set $Y_n = \{y_1, \ldots, y_n\}$ of possible outcomes clearly corresponds to E in (3)--that is, Y_n is the system whose "relevant" information content we wish to measure, "relevance" being determined by our choice of the F-set. At least two choices seem appropriate enough to warrant investigation:

(i) the set of possible directions of price change;

$$\text{i.e.,}\quad X = \{+, -\}$$

(ii) the whole set of possible price changes;

$$\text{i.e.,}\quad X = \{x : x \epsilon [-1, 1]\}.$$

Let us examine the first one. Using (3") for $R(X, Y_n)$, we observe that $H(X) = \log 2$, $H(Y_n) = \log n$, and when n is even

$$H(X, Y_n) = H(Y_n) = \log n$$

and when n is odd

$$H(X, Y_n) = \frac{n-1}{n} \log n + \frac{2}{2n} \log 2n$$

$$= \log n + \frac{1}{n} \log 2 .$$

For the "relevant" information content of η_n therefore we have

$$R(X, Y_n) = \begin{cases} \log 2 & \text{when } n \text{ is even} \\ \frac{n-1}{n} \log 2 & \text{when } n \text{ is odd} \end{cases}$$

an ordering of the η_n which agrees perfectly with the "value" ordering established earlier. That all is not well, however, becomes apparent as soon as other forecasting systems are examined. Consider the system η defined by:

$$(4) \qquad \eta(x) = \begin{cases} y_1 & \text{if } -4/5 \leq x < -2/5 \\ y_2 & \text{if } -2/5 \leq x < 0 \\ y_3 & \text{if } 0 \leq x < 2/5 \\ y_4 & \text{if } 2/5 \leq x \leq 4/5 \\ y_5 & \text{if } 4/5 < |x| \leq 1 . \end{cases}$$

It is obvious that $R(X,\eta) = R(X,Y_5) > R(X,Y_3)$. The value of η to the speculator is easily calculated: When y_5 is observed he is indifferent between buying or selling and his conditional expected profit is zero; when y_1, y_2, y_3, y_4 are observed the respective appropriate actions are clear and the conditional expected profits are $3/5$, $1/5$, $1/5$, $3/5$. Expected profit--equal in this example to the value of η --is therefore $8/25$. But the value of η_3 is $4/9 > 8/25$. Hence either the R-measure is not in full agreement with the value measure, or our procedure for determining "relevance" was improper.

Calculation of R for case (ii), where all price changes are distinguished, calls for definitions of R and H for infinite sets. Instead of extending these definitions we shall simply define X to be the set of mutually exclusive and exhaustive half-open intervals of length ε between -1 and 1. $R(X,Y_n)$ is then approximately equal to log n, which is the H-measure we started with. Case (ii) is seen therefore to lead again to an ordering of the η in conflict with the value ordering.

The source of conflict in Case (ii) has been described: information not relevant to the choice between actions is weighted equally with information that <u>is</u> relevant to this choice. In Case (i), only information relevant in this sense is counted, but no allowance is made for the difference in profits at stake when the price change is high in absolute value and when it is low.

Following Marschak we have demonstrated, by means of an example, the difficulties of resolving the conflict between the ordering of information systems based on value and the ordering established, in various ways, by the Shannon information measure. At this point we shall dismiss the speculator and, from a more general point of view, discuss the aims of ordering information systems and the various methods that have been proposed for doing so.

2. The Basic Comparison and the Relation "$\lesssim_M$".

Our formulation of an abstract decision problem--and most of the notation as well--will follow that of [3]. Briefly, $x \in X$ denotes the state of nature exhaustively described, $\pi(x)$ a probability distribution over X which we suppose finite. The decision maker chooses an action $a \in A$ (also finite) and, depending on the true state x, receives a utility payoff $\omega(a,x)$. The information about x available to the decision maker prior to making his choice is the distribution π and a "signal" or observation $y \epsilon Y$ determined by a function η on X called the "information structure"; thus $\eta(x) = y$. A "decision function" α from Y to A describes the response to each signal. For given η the decision maker will choose that α which maximizes

$$(5) \qquad \Omega(\alpha,\eta,\omega,\pi) \equiv \mathcal{E}\omega[\alpha(\eta(x)),x] .$$

Define

$$(6) \qquad \hat{\Omega}(\eta,\omega,\pi) \equiv \Omega(\hat{\alpha},\eta,\omega,\pi) \equiv \max_{\alpha} \Omega(\alpha,\eta,\omega,\pi) .$$

For given ω and π, the value of η to the decision maker may be defined (see [3], p. 14) as

$$(7) \qquad V(\eta,\omega,\pi) \equiv \hat{\Omega}(\eta,\omega,\pi) - \max_{a} \mathcal{E}\omega(a,x) .$$

This is the quantity that must be compared with whatever costs are connected with the use of the structure η. In this paper we shall not be concerned with these costs.

Suppose N is the set of information structures any one--but only one--of which is available for use by the decision maker. Let P be the set of all possible probability distributions π on X, and let U be the set of all possible independent payoff functions ω on $A \times X$. For any point $(\omega,\pi) \in U \times P$, (7) establishes a complete ordering of the $\eta \epsilon N$. Let us denote this ordering by "$\geq_{(\omega,\pi)}$", and call it the "Basic Comparison." Now consider a subset $M \subset U \times P$ and the ordering "$\geq_M$" defined by

(8) $\eta' \geq_M \eta''$ if and only if $V(\eta',\omega,\pi) \geq V(\eta'',\omega,\pi)$ for all $(\omega,\pi)\epsilon M$.

The relation "$\geq_M$" will be of interest especially when (i) M is empirically interesting and (ii) "$\geq_M$" can be calculated without brute-force repetitive recourse to (7). It is clear that in general we cannot expect "$\geq_M$" to yield a <u>complete</u> ordering over N except when M is very small (e.g., $M = \{(\omega,\pi)\}$) or when N is small.

In the following sections we describe several methods that have been proposed for comparing the $\eta \epsilon N$; we add one method which, so far as we know, is new. In our view the different methods are not competitors, despite occasional claims to the contrary (e.g., [7], p.997), but rather each is equivalent to the relation "$\geq_M$" for some M or class of M's . To be sure, some M's are more interesting than others, but this is a different matter. It must also be emphasized that no invidious comparison of different proposals is intended; not all of the proponents were addressing themselves to the problem that concerns us here.

3. The Shannon Comparison I*

Let $A = 2^X$ (i.e., the set of subsets of X), so that choice of an act $a \epsilon A$ is equivalent to selecting a subset of X . For given π define a payoff function by

$$(9) \qquad \omega(a,x) \equiv \log \pi(x \mid a) \; .$$

For simplicity let us denote the subset of X which is the inverse image of a signal $y \epsilon Y$ for a certain information structure η by y ; that is $\eta^{-1}(y) = y$. It is easy to show that

$$(10) \qquad \underset{a \epsilon A}{\text{Max}} \; \mathcal{E}[\omega(a,x) \mid x \epsilon y] = \mathcal{E}[\omega(y,x) \mid x \epsilon y] = -H[\pi(x|y)]$$

so that $\hat{\alpha}(y) = y$ and

$$(11) \qquad \mathcal{E}[\omega(\hat{\alpha}(\eta(x)),x)] = - \underset{yx}{\mathcal{E}} H[\pi(x|y)] \; .$$

One can see this by considering the problem

$$(12) \qquad \underset{z_x}{\text{Max}} \; \sum_{x \epsilon y} \pi(x|y) \log z_x$$

$$\text{subject to} \quad z_x \geq 0 \quad \text{and} \quad \sum_x z_x = 1 \; ,$$

a sufficient condition for the solution of which is that $\pi(x|y)/z_x = \text{const.}$ for all $x \epsilon y$; with $z_x = \pi(x|y)$, we have (10).

---

* This terminology, as well as the application, is ours.

By (10) and (11) we see that the value of an information structure equals the "rate of transmission," that is,

$$V(\eta,\omega,\pi) = H[\pi(x)] - \underset{yx}{\mathcal{E}}H\,[\pi(x|y)] \equiv R(\eta,\pi) \; . \tag{13}$$

Thus the Shannon Comparison I amounts to a relation "$\geq_M$" which completely orders N , where M consists of the single point (ω,π) with ω satisfying (9). Since there are many such M's (one for each π) there are many such "$\geq_M$" 's. We denote this class of M's by $\mathcal{M}_1$. One might argue that the Shannon Comparison I tells us nothing new, since every M it can deal with can also be dealt with by the Basic Comparison. But this is always true if we are willing to do enough computing. The main shortcoming of the Shannon's Comparison I is that its application is limited to very special payoff functions.

4. The Shannon Comparison II*

Once more, let π be given, and $A = 2^X$. Let $\sigma(x) = s$ be some given partition of X ; we shall write $\pi(s)$ for $\sum_{\sigma(x)=s} \pi(x)$. Define the payoff function by

$$\omega(a,x) \equiv \log \pi[\sigma(x)|a] \equiv \omega(a,s) \; . \tag{14}$$

---

* Again, the terminology and application are ours.

Consider the problem

$$(15) \qquad \max_{z_{\sigma(x)}} \sum_x \pi(x|y) \log z_{\sigma(x)} = \max_{z_s} \sum_s \left[\sum_{x:\sigma(x)=s} \pi(x|y) \right] \log z_s$$

$$\text{subj to} \quad z_s \geq 0 \quad \text{and} \quad \sum_s z_s = 1 \ .$$

A sufficient condition for the solution is now that $\pi(s|y)/z_s$ = const. for all s, so $z_s = \pi(s|y)$ is a solution. Consequently $\alpha(y) = y$ as before. Instead of (10), however, we now have

$$(16) \qquad \max_{a \in A} \mathcal{E}[\omega(a,x)|x\epsilon y] = -H[\pi(s|y)]$$

and in place of (11)

$$(17) \qquad \mathcal{E}[\omega(\hat{\alpha}(\eta(x)),x)] = -\mathcal{E}_y H_s [\pi(s|y)] \ .$$

Value of structure η can still be written in the "rate of transmission" form,

$$(18) \qquad V(\eta,\omega,\pi) = H[\pi(s)] - \mathcal{E}_y H_s [\pi(s|y)] \equiv R_\sigma(\eta,\pi)$$

but the definition of "relevance" here is different from that of (13) in the same way that the definition differed in Cases (i) and (ii) in the example of the speculator.

For the Shannon Comparison II therefore "$\geq_M$" again completely orders N . For each π and each σ there is an M for which Comparison II is appropriate, namely the set consisting of the single point (ω,π) with ω defined by (14). Call this set of M's $\mathcal{M}_2$. Notice that for the identity function $\sigma(x) = x$, Comparison II is equivalent to Comparison I; II therefore is more general. Another way of putting this is: $\mathcal{M}_1 \subset \mathcal{M}_2$.

5. The Marschak-Radner Comparison*

---

* The version presented here is a slight generalization of that in [3]; errors in the present exposition should not be attributed to Marschak and Radner.

In the comparisons described so far M has been very small--one point, in fact. Marschak and Radner [3] have proposed a comparison based on very large M's . Let there be specified any set $X' \subset X$. Let $P' \subset P$ be any set possessing the following two properties

(19) If $\pi \epsilon P'$ and $x \notin X'$ then $\pi(x) = 0$

(20) For any $x_1, x_2 \in X'$ there is a $\pi \epsilon P'$ with $\pi(x_1)\pi(x_2) > 0$.

Then "$\geq_M$" for $M = U \times P'$ is a Marschak-Radner comparison. They assert that if M is of this type, that is to say, if $M \in \mathcal{M}_3$, then

(21) $\eta \geq_M \eta'$ if and only if η is as fine a partition of X' as η' .

The "if" part of (21) is obvious; the "only if" part can be argued as follows. Suppose η is not as fine a partition of X' as η' . Then, denoting respective inverse images of η and η' by y and z , there exists a $y \cap X'$ not wholly contained in any $z \cap X'$. Hence there exist points x_1

and x_2 in $y \cap X'$ such that $\eta'(x_1) \neq \eta'(x_2)$. By (20) there is a $\pi \epsilon P'$ with $\pi(x_1)\pi(x_2) > 0$. Define $\omega(a,x)$ by the entries in the following table:

(22)

	x_1	x_2	.	.	.
a_1	$\pi(x_2)$	$-\pi(x_1)$		0	
a_2	$-\pi(x_2)$	$\pi(x_1)$			
.	0			0	
.					
.					

Then, $\hat{\Omega}(\eta,\omega,\pi) = 0 < 2\pi(x_1)\pi(x_2) = \hat{\Omega}(\eta',\omega,\pi)$ contradicting $\eta \geq_M \eta'$.

Thus for $M \epsilon \mathcal{M}_3$, (21) provides an easy means of comparison.

The class $\mathcal{M}_3$ of M's which define Marschak-Radner comparisons is quite large; among others it includes $M = U \times P$, i.e., the set of <u>all</u> (ω,π) pairs, and it includes the set $M = U \times \pi$ for any $\pi \epsilon P$. Notice however that not all sets $U \times P'$ are in $\mathcal{M}_4$; there does not always exist an X' related to P' by (19) and (20). As an example, consider $X = \{x_1,x_2,x_3\}$ and $P' = \{\pi_1,\pi_2\}$ with

$$\pi'(x_1) > 0 \qquad \pi''(x_1) = 0$$

$$\pi'(x_2) > 0 \qquad \pi''(x_2) > 0$$

$$\pi'(x_3) = 0 \qquad \pi''(x_3) > 0 .$$

In general it is true that if $M_\epsilon \subset M_1 \subset U \times P$ then $\eta \gtrsim_{M_1} \eta'$ implies $\eta \gtrsim_{M_2} \eta'$. Let $M_1 = U \times P_1$ be a Marschak-Radner M and let $M_2 = (\omega, \pi)$, with $\pi \in P_1$ and ω a "Shannon payoff function" of type (9)or (14). It follows that

(23) $\eta \gtrsim_{M_1} \eta'$ implies $R_\sigma(\eta, \pi) \geq R_\sigma(\eta', \pi)$ for $\pi \in P_1$ and arbitrary σ .

Since the $M \in \mathcal{M}_3$ are very large, it is to be expected that the resulting "$\lesssim_M$" relations order N very incompletely; this is another way of saying that "$\lesssim_M$" for $M \in \mathcal{M}_3$ is a very strong condition. We have already seen in Section 1 an occasion where the Marschak-Radner criterion fails to yield a comparison, namely between η_n and η_{n+1} .

6. The "Clear Action" Comparison

In this section we attempt to devise a comparison somewhat weaker than that of Marschak and Radner by restricting the set of payoff functions to which the comparison applies. Suppose there is given an arbitrary function τ on X ; as before, we will denote the subset $\{x \epsilon X : \tau(x) = t\}$ by t . Consider the set $U_\tau \subset U$ of payoff functions ω for which

(24) $$\text{If } \tau(x_1) = \tau(x_2) \text{ then } \underset{a}{\text{maximizer}}\ \omega(a,x_1) = \underset{a}{\text{maximizer}}\ \omega(a,x_2) .$$

Thus a decision maker whose payoff function is U_τ is not interested in knowing the value of x , he only cares about t . And once t is known the proper action to take is clear. In the speculator example the payoff function was a member of U_τ for $\tau(x) \equiv$ sign of x .

As in the last section, we now pick an X' and make use of (19) and (20) to define a P' . Next we set $M = U_\tau \times P'$. $\mathcal{M}_4$ is the class of M's so constructed. Paralleling the Marschak-Radner development then, we assert that for any $M \epsilon \mathcal{M}_4$

(25) $\eta \geq_M \eta'$ if and only if for any $y \epsilon Y$ either $y \cap X' \subset x \cap X'$ for some $z \epsilon Z$ or $y \cap X' \subset t \cap X'$ for some $t \epsilon T$.

(Again y and z are inverse images respectively of the functions η and η' ; Y and Z have the same number of elements; A and T have the same number of elements.)

<u>Proof</u>: Let α denote a decision function based on information y, and β on z. Pick an arbitrary $(\omega,\pi) \in M$. Construct a function α^* as follows. If $y \cap X' \subset z \cap X'$ for some z, set $\alpha^*(y) = \hat{\beta}(z)$. For those $y \cap X'$ not contained in a $z \cap X'$, $y \cap X' \subset t \cap X'$ for some t; set $\alpha^*(y) = \text{Maximizer } \omega(a,x)$ for any x for which $\tau(x) = t$. It follows that $\omega[\hat{\beta}(z),x] \leq \omega[\alpha^*(y),x] \leq \omega[\hat{\alpha}(y),x]$ for all $x \in X'$. Hence $\eta \geq_M \eta'$.

As before, the converse is somewhat less obvious. Suppose that some $y \cap X'$ is contained neither in a $t \cap X'$ nor a $z \cap X'$. Then there exist points $x_1, x_2 \in y \cap X'$ such that $\eta'(x_1) \neq \eta'(x_2)$. Two possibilities arise. If $\tau(x_1) \neq \tau(x_2)$, then define $\omega(a,x)$ by (22), and the argument following (22) applies. If $\tau(x_1) = \tau(x_2)$, definition of ω by (22) would put ω outside U_τ. But since $y \cap X'$ is not contained in any $t \cap X'$, there exists a point $x_3 \in y \cap X'$ such that $\tau(x_1) = \tau(x_2) \neq \tau(x_3)$. Since $\eta'(x_1) \neq \eta'(x_2)$, either $\eta'(x_3) \neq \eta'(x_1)$ or $\eta'(x_3) \neq \eta'(x_2)$. In either event (20) assures the existence of a $\pi \in P'$ which makes an argument analogous to (22) again applicable.

Thus in (25) we have a computable means of carrying out the comparison "$\geq_M$" for $M \in \mathcal{M}_4$.

It will be recalled in Section 4 that the identity function $\sigma(x) = x$ yielded a subset of $\mathcal{M}_2$ which was exactly $\mathcal{M}_1$, thus showing the Shannon Comparison II to be a generalization of the Comparison I. Similarly here, when $\tau(x) = x$ (25) and (21) are equivalent. This is easily seen: if $y \cap X'$ is not in a $z \cap X'$, then it is in an $t \cap X'$. But if $\tau(x) = x$, then every set t consists of a single point, so $y = s$. Since the z's cover X, $y \subset z$. Hence η is as fine a partition of X' as η'.

Paralleling (23) we have, obviously,

(26) $\eta \geq_M \eta'$ for $M = U_\tau \times P' \in \mathfrak{M}_4$

implies

$$R_\tau(\eta,\pi) \geq R_\tau(\eta',\pi) \quad \text{for} \quad \pi \in P' .$$

Notice also that the "Clear Action" comparison ranks all the η_n of Section 1.

7. The Bohnenblust-Shapley-Sherman Comparison*

Once again, let there be given a partition σ of X, and let U' be the set of payoff functions u with the property that $\sigma(x') = \sigma(x'')$ implies $u(a,x') = u(a,x'')$ for all $a \in A$.** For a given probability distribution π define P' to be the set of probability distributions π' related to π in the following way:

(27) If $\sigma(x') = \sigma(x'')$ and $\pi(x') > 0$ then

$$\frac{\pi'(x'')}{\pi'(x')} = \frac{\pi(x'')}{\pi(x')} \quad ;$$

if $\pi(x') = 0$ then $\pi'(x') = 0$.

---

* Reported in Blackwell [4], pp. 93-96, and Savage [8], pp. 148-153.

** This restriction of u yields what Savage has tentatively called a "partition problem." He argues that "Modern statistics has no name for this type of problem, because it recognizes no other type." [8], pp. 120-121.

$\mathcal{M}_5$ is the collection of sets $M = U' \times P'$ with U' and P' so defined.

Since the conditional probability of x given s is the same for any $\pi \in P'$, we can define $p(y|s)$ to be the conditional probability that $\eta(x) = y$ given that $\sigma(x) = s$. Let n be the number of elements in S. Define the likelihood ratio for y as

$$q(y) = \left(\frac{p(y)s_1}{\sum_{i=1}^{n} p(y|s_i)} , \ldots, \frac{p(y|s_n)}{\sum_{i=1}^{n} p(y|s_i)} \right) . \tag{28}$$

Let Q be the set of n-tuples $q = (q_1, \ldots, q_n)$ with $\Sigma q_i = 1$ and $q_i \geq 0$. Each s_i row yields a distribution $m_i(q)$, say, over Q. That is to say, $m_i(q)$ is the probability that likelihood ratio q occurs when the true s is s_i.

For given $M = U' \times P'$, Bohnenblust, Shapley, and Sherman show that the n distributions $m_1, \ldots, m_n$ completely characterize the information structure η. If, instead of observing y distributed according to the $p(y|s_i)$, the decision maker observes q distributed according to the m_i, an equivalent problem results: whatever $(\pi,\omega) \in M$ obtains, the new problem results in the same expected payoff as the old. Furthermore, since $\frac{m_i}{\sum_1^n m_i} = q_i$, it is clear that $\sum_1^n m_i$ completely characterizes the m_i.

Define $m_o = \frac{1}{n} \sum_1^n m_i$; m_o is called the "standard measure" for η. Suppose we are given a distribution m_o over Q ; the question arises whether m_o is or is not a "standard measure" for some information structure η . Construct the m_i as follows: $m_i(q) = nq_i m_o(q)$. Then $\sum_q m_i(q) = 1$ if $\sum_q q_i m_o(q) = \frac{1}{n}$. Therefore m_o is a standard measure for some η if the mean of m_o is $\left(\frac{1}{n},\frac{1}{n}, \ldots, \frac{1}{n}\right)$.

If two structures η,η' have the same standard measure m_o then $\eta =_M \eta'$.* Thus the equivalence class of η's is equivalent to the class of

---

* This relationship Savage calls "virtual equivalence."

---

distributions m over Q with mean $\left(\frac{1}{n},\frac{1}{n}, \ldots, \frac{1}{n}\right)$.

As a computational tool the authors prove the following theorem: Let m be the standard measure for η and m' for η' ; then

(29) $\eta \gtrsim_M \eta'$ if and only if for every continuous convex function g on Q

$$\sum_q g(q)\ m\ (q) \geq \sum_q g(q)\ m'(q)\ .$$

Without commenting on the usefulness of (29) we shall pass on the Blackwell comparison which covers the same (π,ω)-territory.

8. The Blackwell Comparison*

* Blackwell [4] and [5], and Blackwell and Girshick [6], Chapter 12.

Restricting attention to an $M \in \mathcal{M}_5$, we can use $p(y|s)$ and $p(z|s)$ respectively to describe information structures η and η' . Blackwell calls η "sufficient" for η' relative to the partition σ if there exist numbers b_{yz} such that

$$(30) \qquad p(z|s) = \sum_y p\,(y|s) b_{yz}$$

$$\sum_z b_{yz} = 1$$

$$\text{and } b_{yz} \geq 0 \,.$$

Notice that this concept of sufficiency is not the same as the customary Neymann-Fisher one; η need not be a contraction of η' as the latter definition would require.

The usefulness of Blackwell's concept of sufficiency is summed up in the following theorem*, due in its present form to S. Sherman and Charles

* A concise proof is given in Blackwell [5], p. 267. Blackwell himself has shown that the theorem holds in much more general (i.e., infinite) contexts than those dealt with in this paper.

Stein and also to Martin Beckmann, who achieved the same result independently by essentially the same method of proof:

(31) For $M \in \mathcal{M}_5$, $\eta \geq_M \eta'$ if and only if η is sufficient for η' .

In interpreting this theorem, the reader is urged to consult the next section where certain matters concerning the definition of [illegible]$_5$ are discussed.

A graphical representation of sufficiency will be found useful. Suppose η is sufficient for η'. Using the b_{yz} of (30), define

$c_{yz} = \frac{\pi(y)}{\pi(z)} \cdot b_{yz}$, where $\pi \in P'$ and $\pi(s_i) = \frac{1}{n}$ for $i = 1, .., n$.
It follows that

$$\text{(32)} \qquad \pi(s|z) = \sum_y \pi(s|y)c_{yz}$$

$$\sum_y c_{yz} = 1$$

$$c_{yz} \geq 0 .$$

The results of Bohnenblust, Shapley, and Sherman concerning a standard measure ensure that the selection of π entails no loss of generality. For each y, $\pi(s|y)$ is a point in n-space. The statement of (32) is that for each z, the point $\pi(s|z)$ must lie within the convex hull of the set of $\pi(s|y)$ points, if η is sufficient for η'.* The large triangle in

---

* Sometimes it is more convenient to start from the c_{yz} instead of from the b_{yz} as we have done. If things were perfectly symmetric, no comment would be needed, but they are not. A statement equivalent to "η is sufficient for η'" is: there exist c_{yz} satisfying (32) <u>and</u> $\pi(y) = \sum_z c_{yz}\pi(z)$ for the "standard" π. We assume this last condition satisfied in our discussion of the graphical representation; a careful check might show it is violated somewhere in Figure 1.

Figure 1 represents that part of the $\sum_s \pi(s|\ ) = 1$ surface in the positive orthant for the case $n = 3$. Each of the polygons within this triangle represents a different η. Obviously $\eta \geq_M \eta'$, $\eta'' \geq_M \eta'$, but neither $\eta \geq_M \eta''$ nor $\eta'' \geq_M \eta$. From the graph it is clear that an (k-1)-signal system can never be sufficient for an k-signal system $(k \leq n)$ whose $\pi(s|\ )$ vectors are linearly independent.

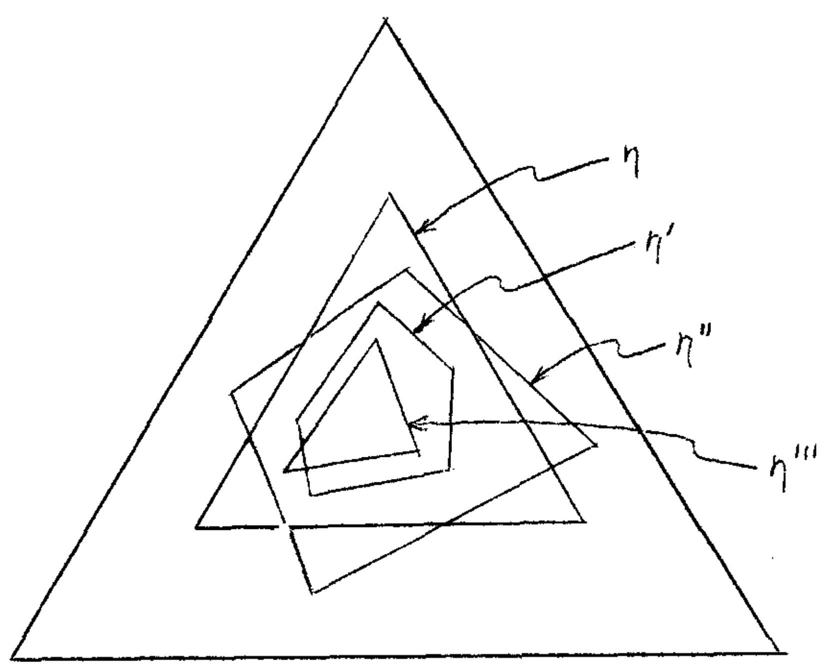

Figure 1

Since each $M \in \mathcal{M}_5$ contains a (π,ω) with ω a Shannon payoff function of type (14), it is true that

(33) $\eta \geq_M \eta'$ for $M \in \mathcal{M}_5$ implies $R_\sigma(\eta,\pi) \geq R_\sigma(\eta',\pi)$ for every $\pi \in P'$.

When the partition σ is the identity partition $\sigma(x) = x$, then $M = U \times P \in \mathcal{M}_5$. In this case (31) becomes equivalent to the Marschak-Radner requirement (21). The argument is as follows. Pick an arbitrary x . Since $p(y|x) = p(z|x) = 0$ for $y \nsupseteq \eta(x)$ and $z \nsupseteq \eta'(x)$, $b_{\eta(x),\eta'(x)} = 1$ by (30), and $b_{yz} = 0$ for disjoint y and z . Now take any other point x' such that $\eta(x') = \eta(x)$. By (30) $p(\eta'(x)|x') = p(\eta(x)|x') = 1$, so $x' \in \eta'(x)$. Therefore every y is contained in some z ; hence η is an extension of η' as the Marschak-Radner condition requires.

Notice however that as $\mathcal{M}_5$ has been defined it is not true that $\mathcal{M}_5 \supset \mathcal{M}_3$. It is an interesting question whether, without essentially damaging (30) and (31), $\mathcal{M}_5$ can be enlarged by admitting to membership sets $U'' \times P''$ where U'' is the set of all "stepwise" payoff functions over a <u>proper</u> subset X'' of X and where $P'' \subset P'$ is only large enough to ensure that for every pair s,s' there is a $\pi \in P''$ with $\pi(s)\pi(s') > 0$. If this extension can be carried out successfully the enlarged $\mathcal{M}_5$ would contain $\mathcal{M}_3$.

9. <u>The Blackwell k-Comparison</u>

Up to now we have been assuming implicitly that the set A of actions available to the decision maker is "large." If, to take a trivial case, A consists of a single element, then all η in N are of equal value for any (π,ω) pair. Intuition suggests that as A increases in size the number of structures η' inferior to a given structure η decreases. To see this, consider the two structures η' and η'' depicted in Figure 1. When A is very large, the Blackwell "sufficiency" comparison tells us that (for the M on which Figure is based) neither $\eta' \gtrsim_M \eta''$ nor $\eta'' \gtrsim_M \eta'$. Although the η' polygon <u>almost</u> contains the entire η'' triangle, the Sherman-Stein-Beckmann theorem* asserts that there is a $\omega \in U'$ which

* The proof of the theorem <u>does</u> require an assumption that A is large, although in Blackwell's presentation the entrance of this assumption is not explicit.

sufficiently exploits the tiny advantage of η'' over η' to make η'' the superior structure <u>for that</u> ω. One would expect the number of such payoff functions to be small according as the piece of the η'' triangle outside the η' polygon is small. The assured existence of a ω able to capture arbitrarily small obtrusions of this kind implies the existence of very fine gradations of payoff functions within U'. The size of A is an important (but of course not sole) determinant of this degree of fineness. In Figure 1 it is not difficult to believe that if only two actions are available $\eta' \geq_M \eta'''$ because U does not contain a payoff function quite able to exploit the tiny triangle.

Although we have no argument to support the representation of Figure 2, the reader may find it suggestive. For a given $M \in \mathcal{M}_5$ and a given structure η, we may ask for what η' is $\eta \geq_M \eta'$ and how does this set of inferior structures change with changes in A ? It may be possible to state the

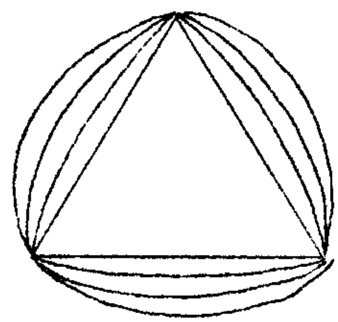

Figure 2

answer geometrically in something like the following terms. Let the triangle in Figure 2 be the η polygon as defined earlier. If A is large, say 12 elements, the η' polygon must lie entirely within the central triangle of Figure ϵ. If A has 8 elements, then η' will be inferior if its polygon lies entirely within the first ring beyond the triangle. If A has 6 elements the η' polygon need only be contained in the second ring from the center; etc. The question of whether these rings (if such they be) have some simple mathematical characterization is of course crucial for application.

Define the class $\mathcal{M}_6(k)$ as follows. If $M \in \mathcal{M}_5$, let $M' \subset M$ contain those payoff functions whose definition requires an A of not more than k elements. Let $\mathcal{M}_6(k)$ contain any M' so constructed. Clearly $\mathcal{M}_6(\infty) = \mathcal{M}_5$, and for any $M \in \mathcal{M}_6(k+1)$ there is an $M' \in \mathcal{M}_6(k)$ with $M' \subset M$. It is not true of course that $\mathcal{M}_6(k) \subset \mathcal{M}_6(k+1)$.

For $M \in \mathcal{M}_6(k)$ Blackwell has examined* the relation "$\leq_M$", which he

---

* [5], pp. 269-27

---

calls a "k-comparison." We shall relate only one easily stated result from among the several he has achieved. Some of his results, it should be said, are of more immediate practical importance than this one.

(34) For any $M \in \mathcal{M}_5$ and corresponding $M' \in \mathcal{M}_6(k)$ (i.e., $M' \subset M$):

$\eta \geq_{M'} \eta'$ if and only if

$\eta' \geq_M \eta^*$ implies $\eta \geq_M \eta^*$ for any k-signal structure η^*.

In very rough graphical terms: if in Figure 3 $\eta \geq_{M'} \eta'$, then any triangle η^* with non-vacuous region A must also have a non-vacuous region B (recall the footnote on p. [illegible]).

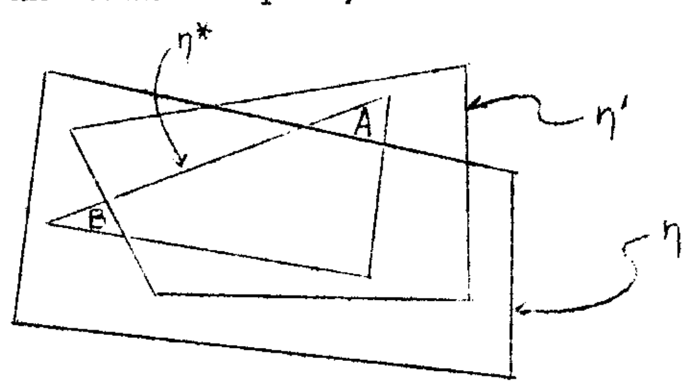

Figure 3

10. The Lindley Comparison

With respect to a given $M \in \mathfrak{M}_5$, Lindley [7] suggests that η be called "as informative" as η' if

(35) $R_\sigma(\eta,\pi) \geq R_\sigma(\eta',\pi)$ for all $\pi \in P'$.

In our terms, (35) could be restated as follows. Let the set $M' \subset M$ be the collection of (ω,π) pairs such that $\pi \in P'$ and ω is the Shannon payoff function of type (14) corresponding to π. (Thus M' has one (ω,π) pair for each $\pi \in P'$.) Let $\mathfrak{M}_7$ be the class of such sets M', one for each $M \in \mathfrak{M}_5$. Then clearly, $\eta \geq_M \eta'$ for $M \in \mathfrak{M}_5$ implies $\eta \geq_M \eta'$ for $M \in \mathfrak{M}_7$ which is the same as saying that η is as informative, in Lindley's sense, as η'.

Definition of such a class $\mathfrak{M}_7$ serves little purpose however: no computationally useful theorem accompanies the relation "$\geq_M$". And in view of the very special nature of the payoff functions to which $\mathfrak{M}_7$ relates, one wonders what purpose Lindley could have had in mind.

The puzzle is not diminished by his discussion of his comparison relative to Blackwell's. He observes that since his comparison is strictly weaker it leads to a more nearly complete ordering of the set N of all information structures. "The smallness of the [part of N not superior nor inferior to a given η] is a satisfactory feature of the comparison..., for ideally all [structures] would be comparable."* But then, of course, the Basic Comparison is better still -- in fact <u>perfect</u>.

---

* [7], pp. 997-998. In Lindley's defense it must be said that he proposes his comparison for use in a non-decision-problem context if in fact -- and he admits he is not sure -- there is such a thing. Our critical remarks seem justified in any case.

11. Conclusion: A Warning and a Suggestion

The main purpose of this paper has been to suggest for use in discussing the different comparisons a vocabulary that makes somewhat clearer the applicability of each. Since little has been said about the reasons for wanting such comparisons it may be well to point to three -- one obvious, one perhaps debatable, and a third which is clearly specious and not a valid reason at all.

The first reason is simply computational. A decision maker with given ω and π wishes to determine the value to him of each of several different information structures. There is no question of how to do this in principle, but the mechanics of the Basic Comparison are quite often difficult to carry out. We need methods of comparison which fasten on the easily recognizable gross characteristics of information structures. The comparisons discussed do just this. They will be found useful even if no further progress is made along these lines -- an unlikely prospect, as will be argued below.

The second reason for such comparisons is suggested by economic applications. Suppose a decision maker or organization must make a long-term commitment to one of several available information structures, the chosen structure to be used repeatedly for a variety of problems. The commitment might take the form of a purchase of durable machinery such as a telescope, or a telephone or radio system. Or it might be institutional: a subscription to a news service, or the promulgation of a certain set of rules for communication. A straightforward decision-theory approach to the problem would be to view the choice of the information structure and the corresponding choice of actions for every one of the variety of problems as part of one grand problem. The practical difficulty of following this counsel suggests as an alternative that that information structure be chosen which is good against any other for each one of the "little" decision problems to be faced. If this latter set of problems happens to fall in an M for which we have an easy "$\geq_M$" comparison, the choice among information structures may be greatly facilitated. Only by experience will we learn the value of this particular application of the theory.

The third "reason," which is no reason at all, is inserted here only as a warning of how the theory must <u>not</u>, at least in general, be used. One might suppose, naively, that if η is superior to η' for all (ω,π) pairs in a certain M, then η is superior to η' for the decision maker who is uncertain which (ω,π) problem faces him but who knows that $(\omega,\pi) \in M$. It is easy to show however that such an application of the theory presupposes that nothing is changed by restriction of the whole theory to convex M's, which is false. To see this, consider the two "Clear Action" payoff functions:

ω_1:

	x_1	x_2	x_3	x_4
a_1	9	9	-9	-9
a_2	-9	-9	9	9
a_3	8	0	8	0
a_4	0	8	0	8

and

ω_2:

	x_1	x_2	x_3	x_4
a_1	-9	-9	9	9
a_2	9	9	-9	-9
a_3	8	0	8	0
a_4	0	8	0	8

and the two information structures:

η_1: $\eta_1(x_1) = \eta_1(x_2) \neq \eta_1(x_3) = \eta_1(x_4)$

η_2: $\eta_2(x_1) = \eta_2(x_3) \neq \eta_2(x_2) = \eta_2(x_4)$.

Let π be the same for ω_1 and ω_2. Then clearly $\eta_1 > \eta_2$ for (ω_1, π) and (ω_2, π) but $\eta_2 > \eta_1$ for $(\frac{1}{2}\omega_1 + \frac{1}{2}\omega_2, \pi)$.

Finally we wish to suggest two directions in which the theory might be extended.

First, unfinished investigations suggest that there is almost certain to be a single generalization of which the "Clear Action" comparison and the Blackwell Comparison are both special cases. The value of such a generalization would be, of course, that it would contain more than the sum of these two parts.

Second, there remain easily recognizable gross characteristics of information structures which are relevant to value comparisons but which, as yet, are not exploited formally by any of the "$\leq_M$" relations we have described.

In Figure 4 the outside triangle again represents, as in Figure 1, the positive piece of the $\sum_s p(s| \ ) = 1$ plane. Two information structures are depicted.

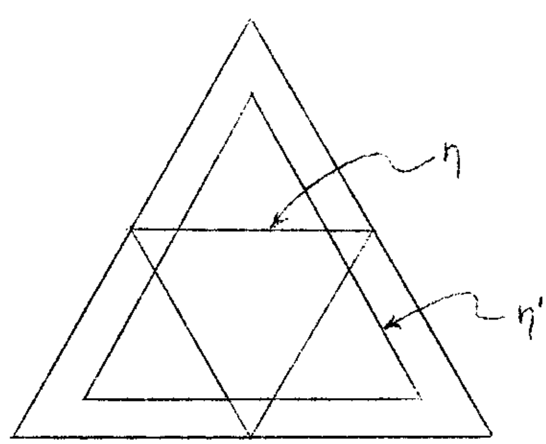

Figure 4

Consider the payoff function

	x_1	x_2	x_3
a_1	1	1	-1
a_2	1	-1	1
a_3	-1	1	1

and let $\pi(x_i) > 0$ for each i .

Clearly, in this situation, η is superior to η' no matter how close η' is to the outside triangle. Certainly this superiority holds true for a large class of payoff functions similar to the one defined. Yet we are not helped to this easy conclusion by any of the formal comparisons known to us.

References

[1] Shannon, C. E., and Warren Weaver, The Mathematical Theory of Communication, University of Illinois Press, Urbana (1949).

[2] Khinchin, A. I., Mathematical Properties of Information Theory, Dover Publications, New York, (1957). (Translation of "The Entropy Concept In Probability Theory" and "On the Fundamental Theorems of Information Theory".)

[3] Marschak, J. and R. Radner, "Economic Theory of Teams," Cowles Foundation Discussion Paper No. 59, October 29, 1958.

[4] Blackwell, David, "Comparison of Experiments," Second Berkeley Symposium on Mathematical Statistics and Probability, J. Neymann (ed), University of California Press, 1951, pp. 93-102.

[5] ______________ "Equivalent Comparisons of Experiments," Annals of Mathematical Statistics, 24:265-272 (1952).

[6] ______________ and M. A. Girshick, Theory of Games and Statistical Decisions, John Wiley and Sons (1954), especially Chapter 12, "Comparison of Experiments".

[7] Lindley, D. V., "On a Measure of the Information Provided by an Experiment," Annals of Mathematical Statistics, 27:986 (1956).

[8] Savage, Leonard J., The Foundations of Statistics, John Wiley and Sons, New York, 1954.