DISCUSSION PAPER NO. 416

AN INTRODUCTION TO INFORMATION STRUCTURES */

by

David A. Malueg

March, 1980

Northwestern University

Department of Economics

*/ This paper has benefitted from my discussions with Paul Milgrom and from his comments on an earlier version of this paper. Financial support from the National Science Foundation (Grant SOC 77-08896, D.T. Mortensen, Principal Investigator and Grant SOC 77-15793, S.Reiter, Principal Investigator is acknowledged.)

Comments are welcome.

AN INTRODUCTION TO INFORMATION STRUCTURES

I. Introduction

This paper discusses information structures, which can be viewed as mechanisms through which an agent may learn about his environment. Consequently, they may also be useful in modeling the beliefs of agents. We are interested in the comparisons of such structures.

The purposes herein are several. First, as an introduction to this topic, this paper gathers from various articles results on information structures and their comparisons. In part II we examine the original work by Blackwell. In a statistical setting he gives necessary and sufficient conditions for one information structure to be as "good" as another, in an appropriate sense. We work through these papers, citing Blackwell for proofs of the theorems stated. Economists will recognize in Blackwell's criterion of "informativeness" the antecedent of Rothschild-Stiglitz "riskiness."

Secondly, in part III we develop these comparisons in an economic setting as presented by Marschak and Miyasawa. They introduce two additional relations between information structures, viz. fineness and garbling. We show that while these comparisons are not generally the same, in the special case of noiseless signals they are equivalent.

In situations involving more than two signals or more than two agents the particular relation between signals becomes important and characterizations sharper than informativeness are desirable. This is also discussed in section III. However, general results in this direction will be hard to come by. We show this in part IV through an example in which a "better" signal may or may not be preferred.

Lastly, we conclude in section V with speculation about questions for future research.

II. The Blackwell Papers

Introduction, Informativeness, and the Standard Experiment

If a statistician wishes to infer from experiment which of n states (or hypotheses) is the true state of the world, what experimental procedure should he use? Blackwell (1951,1953) has responded to this question by characterizing those procedures which are dominated regardless of the experimenter's prior over states. Formally, an n-tuple, $\alpha = (q_1,\ldots,q_n)$, of probability measures on a Borel field B of subsets of a bounded outcome space X is an experiment; $x \in X$ is distributed according to q_i if the i^{th} state obtains. Upon observing an outcome $x \in X$ the statistician chooses an action d from a specified set D, incurring loss L(i,d) if q_i is the true distribution. Thus with each $d \in D$ is associated a loss vector $w(d) = (L(1,d),\ldots,L(n,d))$. Denote by A the range of w, $A = w(D) \subset R^n$. We will assume D to have been restricted so as to yield compact, convex A.[1] We may now view the statistician as choosing an action $a \in A$, $a = (a_1,\ldots,a_n)$, where a_i is the loss incurred if x is distributed according to q_i. (α,A) will denote this decision problem. Before observing an outcome of his experiment the statistician constructs a decision function f, a B-measurable function from X to A, which specifies the action a to be taken when x is observed. With every decision procedure $f(x) = (a_1(x),\ldots,a_n(x))$ is associated a conditional risk vector $v(f) = (\int a_1(x)dq_1,\ldots,\int a_n(x)dq_n)$ where $\int a_i(x)dq_i$ is the risk with procedure f, conditional on q_i being the distribution of x. $R(\alpha,A)$ will denote the range of v as f varies over decision procedures in the problem (α,A). Given our assumptions on A, $R(\alpha,A)$ is a compact, convex subset of R^n. $R(\alpha,A)$ is the set of conditional risk vectors which may be attained in the decision problem (α, A).

Since we naturally expect "better" information should be more "useful," we are led to the following definition.

Definition 1: Let α, β be two experiments. We say α is <u>as informative as</u> β, denoted $\alpha \supset \beta$, if for every A it is true that $R(\alpha,A) \supset R(\beta,A)$.

Thus, α is as informative as β if every conditional risk vector attainable in (β,A) is attainable in (α,A).

Theorem 2:[2] The following are equivalent to $\alpha \supset \beta$.

(i) For every A and every $v \in R(\beta,A)$, there is a $v^* \in R(\alpha,A)$ with $v_i^* \leq v_i$ for all i.

(ii) For every A and every choice of $c_i \geq 0$, $\Sigma c_i = 1$,

$$\min_{v \in R(\alpha,A)} \Sigma c_i v_i \leq \min_{v \in R(\beta,A)} \Sigma c_i v_i \quad .$$

(iii) For every A, $\min_{v \in R(\alpha,A)} \Sigma v_i \leq \min_{v \in R(\beta,A)} \Sigma v_i$.

(iv) For every A, $\min_{v \in R(\alpha,A)} (\max_i v_i) \leq \min_{v \in R(\beta,A)} (\max_i v_i)$.

(iv) says that an agent playing a minimax strategy in a game against nature can do at least as well by basing his strategy on α rather than on β, irrespective of the set of actions available to him. (ii) states that regardless of the statistician's prior over the n states or hypotheses, his Bayes risk ($\min_{v \in R(\cdot,A)} \Sigma c_i v_i$) is no greater with α than with β. It is through consideration of (iii), the case in which a uniform prior is held over hypotheses, that we reduce α to Blackwell's standard experiment.

Before he has collected any data, an experimenter might think all hypotheses equally likely to be true. For any α we can construct a new measure

$nq_o = q_1 + \ldots + q_n$. q_o is the statistician's prior over X when he has a univorm prior over states. It follows from the Radon-Nikodym Theorem that there exists an essentially unique non-negative B-measurable function $p_i(x)$, $i = 1,\ldots,n$, such that for any $S \epsilon B$, $q_i(S) = \int_S np_i(x)dq_o$. Further, for any

$$S \epsilon B,\ q_o(S) = \frac{1}{n}\Sigma\, q_i(S) = \int_S \Sigma\, p_i(x)dq_o\ ,\ \text{which implies } \Sigma\, p_i = 1 \text{ a.s.},$$

and we may redefine p_i so this latter condition holds identically. Let $P = \{ (p_1,\ldots,p_n) \mid p_i \geq 0,\ i = 1,\ldots,n, \text{ and } \Sigma\, p_i = 1 \}$ and define, for any Borel subset A of P, $M_i(A) = q_i\ \{x \epsilon X \mid p(x) \epsilon A\}$, where $p(x) = (p_1(x),\ldots,p_n(x))$. Since $p(x)$ tells us as much about i as does x, we should expect experiment α to be equivalent to $\alpha^* = (M_1,\ldots,M_n)$. This is the import of

Theorem 3:[3] For every A, $R(\alpha,A) = R(\alpha^*,A)$.

α^* is the <u>standard experiment</u> associated with α. With $nM_o = M_1 + \ldots + M_n$, we have $M_i(S) = n\int_S p_i dM_o$ and since $1 = M_i(P) = n\int_P p_i dM_o$, it follows that

$E(p) = (1/n,\ldots,1/n)$; on the other hand, every probability measure M_o on P having mean $(1/n,,\ldots,1/n)$ defined a standard experiment. Hence, "<u>the class of standard experiments is essentially equivalent to the class of probability measures over P with mean $(1/n,\ldots,1/n)$</u>,"[4] and so informativeness provides a partial ordering of this class. The measure M_o is called the <u>standard measure</u> of α; if M, m are standard measures of experiments α, β, respectively, we will write $M \supset m$ to mean $\alpha \supset \beta$.

Since the outcome spaces of two experiments need bear no similarity to one another, the reduction to the standard experiment highlights the fact that an experiment is useful only as it allows a statistician to update his prior--and this only involves the conditional distributions $M_1,\ldots,M_n$.

Employing the concepts developed above, Blackwell proves[5]

Theorem 4: For two standard measures M,m, $M \supset m$ if and only if for every continuous, convex $g(p)$, $\int g(p)dM \geq \int g(p)dm$.

Theorem 5: If $n = 2$, $M \supset m$ if and only if $\int_0^y F_M(t)dt \geq \int_0^y F_m(t)dt$ for all y, where $F_M(t) = M\{p_1 \leq t\}$ and $F_m(t) = m\{p_1 \leq t\}$

Sufficiency

The concept of sufficiency provides another tool for the comparison of experiments. "If B, C are Borel fields of subsets of X, Y respectively, a stochastic transformation T is a function $Q(x,E)$ defined for all $x \in X$ and $E \in C$ which for fixed E is a B-measurable function of x and for fixed x is a probability measure on C. For any probability measure m on B, the function $M(E) = \int Q(x,E)dm(x)$ is a probability measure on C, denoted by Tm."[6] Further, T is called mean-preserving if $\int y dQ(x,y) = x$.

Definition 6: $\alpha = (M_1,\ldots,M_n)$ is said to be sufficient for $\beta = (m_1,\ldots,m_n)$, denoted $\alpha \succcurlyeq \beta$, if there exists a stochastic transformation T such that $m_i = TM_i$, $i = 1,\ldots,n$, where T is independent of i.

Grossman et al. have suggested the following interpretation of sufficiency.[7] A statistician who observes an outcome of experiment α is able to generate an experiment which is as informative as β; if x is observed in experiment α the statistician "draws" an observation x' from an urn in which x' is distributed according to the density $Q(x,x')$. Since someone who observes the drawn x' will not have known which x had been observed, integration is over X with respect to M_i. Further, if Q depended on i, the statistician would not know from which urn to draw x'; therefore we require Q to be independent of i.

This interpretation of sufficiency makes it clear that information will be "lost" through the stochastic transformation; an optimal decision based on β is not likely to be as good as that based on α. Consequently, we expect that $\alpha \succcurlyeq \beta$ implies $\alpha \supset \beta$. The more remarkable result is that if $\alpha \supset \beta$, then $\alpha \succcurlyeq \beta$. Blackwell initially (1951) shows the equivalence of sufficiency and informativeness when $n = 2$, later (1953) extending the result for any positive integer. In terms of standard measures, sufficiency admits the following representation.

Theorem 7:[8] $M \succcurlyeq m$ if and only if there is a mean-preserving stochastic transformation T with $M = Tm$.

If $\alpha = (M_1,\ldots,M_n)$ is an experiment with outcome space X; $\beta = (m_1,\ldots,m_n)$ is an experiment with outcome space Y; X and Y have N and N' elements respectively; and $\alpha \succcurlyeq \beta$, then the stochastic transformation takes the form $m_i(y_j) = \sum_{k=1}^{N} b_{kj} M_i(x_k)$ where $b_{kj} \geq 0$ and $\sum_{j=1}^{N'} b_{kj} = 1$; $i = 1,\ldots,n$ $j = 1,\ldots,N'$. This establishes the following:

Theorem 8:[9] Let $P = (M_i(x_k))$, $i = 1,\ldots,n$, $k = 1,\ldots,N$ and $Q = (m_i(y_j))$, $i = 1,\ldots,n$, $j = 1,\ldots,N'$ be the $n \times N$ and $n \times N'$ Markov matrices associated with experiments α and β respectively. Then $\alpha \succcurlyeq \beta$ if and only if there exists a $N \times N'$ Markov matrix T such that $Q = PT$.

We have given two characterizations of sufficiency which are seemingly contradictory in appearance. If $\alpha = (M_1,\ldots,M_n)$ and $\beta = (m_1,\ldots,m_n)$ and M and m are their respective standard measures, then $\alpha \succcurlyeq \beta$

(i) if and only if there exists a stochastic transformation T such that $m_i = TM_i$, $i = 1,\ldots,n$;

(ii) if and only if there exists a mean-preserving stochastic transformation T^* such that $M = T^*m$.

In (i) we condition on state i being "true" and so would like the dispersion of observations from this distribution to be small. In the extreme case the mass for x would be concentrated at a single point according to M_i, though we could still have a non-degenerate distribution m_i for x'. Alternatively, when we take i to be the true state, we want our observation to accept or reject this hypothesis as confidently as possible; hence we want the "variance" of this distribution to be low.[10]

On the other hand, in (ii) the statistician assigns a uniform prior to states. Now the less preferred experiment is the one that yields a smaller "variety" of observations; here in the extreme case m could be a point mass and M still a non-degenerate distribution. The standard experiment with lower dispersion is less able to allow one to distinguish confidently between states.

Because of the equivalence of $\succcurlyeq$ and $\supset$, Theorem 8 furnishes a useful representation of informativeness when X and Y are finite; the more general stochastic transformation is useful when X and Y are infinite. Grossman et al. use this latter characterization of informativeness in analyzing the demand for a drug which has an uncertain effect on health. Specifically, they show that if one's level of health in period t, z_t, is represented as a random variable, $\tilde{z}_t = \alpha + \beta y_t + \tilde{e}_t$, where y_t is the drug dosage in period t, $\{\tilde{e}_t\}$ is a series of serially uncorrelated standard normal random variables, α and y_t are known ex ante, then larger drug doses make the observation of an individual's health a more informative signal of the drug's effectiveness, β. It is shown that if $0 < y' < y$, then the conditional density of $\tilde{z}'$ can be written as $h(z' \mid \alpha + \beta y') = \int_{-\infty}^{+\infty} v^*(z' \mid z) h(z \mid \alpha + \beta y) dz$, where $v^*(z' \mid z)$ is the density of z' given z. $v^*(\cdot \mid \cdot)$ is a stochastic transformation. Since a Bayesian updates his beliefs in accord with his experience, information gathered now influences future decisions. In order to take advantage of this "learning by doing," a Bayesian consequently chooses a larger drug dosage than a non-Bayesian holding the same prior over the drug's effectiveness.

Combined Experiments

The last topic we shall consider in Blackwell's papers is the combination of experiments. We shall only state here a definition and two results, returning to this issue when we consider information structures from an economic perspective. Let us suppose n, X, and Y are finite and P and Q are experiments as in Theorem 8. Further, suppose M_i, m_i are the marginal distributions of m_i^*. Then the combined experiment, denoted (P,Q), is represented by the $n \times NN'$ matrix R with (i, (N-1)j + k) entry $m_i^*(x_j, y_k)$. For example, if

$$P = \begin{pmatrix} M_1(x_1) & M_1(x_2) \\ M_2(x_1) & M_2(x_2) \end{pmatrix} \quad \text{and} \quad Q = \begin{pmatrix} m_1(y_1) & m_1(y_2) \\ m_2(y_1) & m_2(y_2) \end{pmatrix}$$

then

$$R = \begin{pmatrix} m_1^*(x_1,y_1) & m_1^*(x_1,y_2) & m_1^*(x_2,y_1) & m_1^*(x_2,y_2) \\ m_2^*(x_1,y_1) & m_2^*(x_1,y_2) & m_2^*(x_2,y_1) & m_2^*(x_2,y_2) \end{pmatrix} .$$

We have the following intuitive results.[11]

Theorem 9: If P and Q are $n \times N$ and $n \times N'$ experiments and R is an $n \times N''$ experiment which is independent of both, then $P \supset Q$ implies $(P,R) \supset (Q,R)$.

Theorem 10: If $P^{(1)} \supset Q^{(1)}$ and $P^{(2)} \supset Q^{(2)}$, $P^{(1)}$ and $P^{(2)}$ are independent and $Q^{(1)}$ and $Q^{(2)}$ are independent, then $(P^{(1)},P^{(2)}) \supset (Q^{(1)},Q^{(2)})$.

One implication of Theorem 10 is that if one of two experiments available to a statistician is more informative than the other, then the statistician should take all (independent) observations from the more informative experiment, regardless of how beliefs change as he gathers information.

III. An Economic Interpretation

The statistical literature discussed above has an immediate economic interpretation, which we shall now develop.[12]

Let $(\mathscr{S}, S, \pi)$ be a probability space; $s \in S$ represents a state of nature, S is the Borel σ-field of subsets of $\mathscr{S}$, and π is a probability measure on $(\mathscr{S}, S)$. Throughout, π will denote a generic probability measure, its domain being indicated by its arguments. A partition $\hat{\Omega}$ of $\mathscr{S}$ is a collection of mutually disjoint sets, $\{\omega\} \subset S$, such that $\bigcup \omega = \mathscr{S}$. A signal χ is a measurable mapping from $\mathscr{S}$ to a measurable space $(\mathscr{Z}, Z)$. χ^{-1} induces a partition X of $\mathscr{S}$; thus, for any $x \in X$, s and $s' \in x$, it is true that $\chi(s) = \chi(s')$. The partition X together with the induced conditional probability measure $\pi(\cdot|E)$, $E \in S$, on X will be termed an information structure generated by χ. In this way X represents those subsets of $\mathscr{S}$ within which an agent observing χ cannot distinguish states, while π establishes the relationship between states and the observed signal. It should be noted that we are no longer concerned explicitly with the outcome space of a signal, as we were in section II; while these approaches are equivalent for comparisons by informativeness, the use of partitions will allow a more natural definition of "fineness," to be given below.

Each economic agent maximizes his expected value of a real-valued measurable state-dependent payoff function which is defined on $\mathscr{S} \times \Delta$, where Δ is the class of decisions available to the agent. Letting $\psi = \{u_i\}_{i \in N}$ be a collection of payoff functions, we call $\hat{\Omega}_\psi$ a payoff-<u>adequate</u> partition of $\mathscr{S}$ with respect to ψ if for every $\omega \in \hat{\Omega}_\psi$, s and $s' \in \omega$, $u \in \psi$, and $d \in \Delta$ (d is a constant decision), it is true that $u(s,d) = u(s',d)$; the coarsest payoff-adequate partition, Ω_ψ, is termed the payoff-<u>relevant</u>

partition of $\mathscr{S}$ with respect to ψ. This captures the notion that a group of agents might view with indifference the distinctions between some states. For example, in consumer theory indifference curves represent a payoff-relevant partition of commodity space for an individual. It is clear that in decision problems involving agents represented by ψ it is suggicient to consider Ω_ψ instead of $\mathscr{S}$. Characterization of each agent is complete when we associate with each agent a probability measure π which represents his beliefs over states.

For simplicity we shall take X and Ω to be finite.[13] An agent for whom Ω is payoff-adequate initially faces the following problem:

$$(1) \qquad \max_{d \in \Delta} \sum_{\omega \in \Omega} u(\omega, d)\, \pi(\omega) \;.$$

(Allowing ω as an argument of u is a convenient but harmless abuse of notation.) However, the agent who observes a signal $\chi(\omega)$ will use this information in forming a decision function $\delta(x)$ to

$$(2a) \qquad \max_{\delta \in \Delta} \sum_{\Omega} u(\omega, \delta(\chi(\omega)))\, \pi(\omega)$$

or equivalently,

$$(2b) \qquad \max_{\delta \in \Delta} \sum_{\Omega} u(\omega, \delta(x))\, \pi(\omega|x), \quad \forall\, x \in X.$$

We define the value in (2a) as $V(X;u,\pi)$, which needs be as great as the value in (1).

The analogy with part II is clear. Each $\omega \in \Omega$ represents a hypothesis. After observing a signal (experiment) a decision-maker revises his beliefs about the state of the world and chooses an action which is optimal relative to these new beliefs. Now we ask, "When will one information structure be as valuable as another, regardless of an agent's prior over states?" Value is the analogue of risk.

Definition 11: With respect to (u,π), the signal represented by X is as valuable as that represented by X' if and only if $V(X;u,\pi) \geq V(X';u,\pi)$.

Definition 12: X is as fine as X', denoted $X \varphi X'$, if and only if $\forall x \in X$, $\exists x' \in X'$ such that $x \subset x'$.

If X is as fine as X', then X is obviously as valuable as X' for any (u,π), since by observing the signal χ one knows exactly the outcome of χ'. Furthermore, if we do not fix the class ψ with respect to which we are comparing two signals, then the (partial) ordering of information structures by fineness is equivalent to the ordering by value.

Theorem 13:[14] For any two information structures X and X', X is as valuable as X' for all (u,π) if and only if $X \varphi X'$.

While sufficiency is clear, necessity is less obvious. If X and X' are not ordered by fineness, each will distinguish states in $\mathscr{S}$ which the other does not. We can then construct payoff functions for which these distinctions are crucial. Consequently, a different information structure will be preferred with each payoff function.

Refinement provides only a relatively rough comparison of information structures. Henceforth we shall take Ω as fixed and consider only those signals χ for which $\pi(\cdot|\omega)$ on X is the same for all priors on Ω. Such signals are said to provide statistical information;[15] comparisons sharper than refinement can be made in this restricted class. We have seen in section II that Blackwell has provided necessary and sufficient conditions for one (statistical) information structure to be as valuable as another for any (u,π) for which Ω is payoff-adequate.

If we require agreement not only on $\pi(x|\omega)$ and $\pi(x'|\omega)$ but also on $\pi(x,x'|\omega)$, then a further comparison is available.

Definition 14: X' is a garbling of X if and only if one of the following equivalent conditions holds:[16]

(i) $\pi(x'|x,\omega) = \pi(x'|x)$,

(ii) $\pi(\omega|x,x') = \pi(\omega|x)$,

(iii) $\pi(x,x'|\omega) = \pi(x'|x)\ \pi(x|\omega)$.

Thus, when X' garbles X we see x is a sufficient statistic for (x,x'); further, since it is always the case that $\pi(x'|\omega) = \sum_{x \in X} \pi(x'|x,\omega)\ \pi(x|\omega)$, (iii) shows the stochastic transformation relating $\pi(x'|\cdot)$ and $\pi(x|\cdot)$ has a special form. When $X \supset X'$, the stochastic transformation involved is arbitrary. Therefore, we should expect fewer signals to be ordered by the criterion of garbling than that of informativeness, though in return we get a stronger and possibly more useful characterization of the relation between two so ordered signals.

Our criteria for comparing information structures are related as follows (a proof is given in Appendix A).

Lemma 15: (a) $X \varphi X' \Rightarrow$ X' is a garbling of X.

(b) X' is a garbling of X $\Rightarrow X \supset X'$.

The foregoing discussion suggests the converse of Lemma 15 is false. We establish this with two examples.

Example 1

$\pi(y,x'|\omega_1)$

	y_1	y_2
x'_1	10/45	12/45
x'_2	20/45	3/45

$\pi(y,x'|\omega_2)$

	y_1	y_2
x'_1	5/45	24/45
x'_2	10/45	6/45

Example 2

$\pi(y,x''|\omega_1)$

	y_1	y_2
x''_1	50/225	41/225
x''_2	100/225	34/225

$\pi(y,x''|\omega_2)$

	y_1	y_2
x''_1	15/225	62/225
x''_2	60/225	88/225

In both examples,

$$L_o \equiv \begin{pmatrix} \pi(y_1|\omega_1) & \pi(y_2|\omega_1) \\ \pi(y_1|\omega_2) & \pi(y_2|\omega_2) \end{pmatrix} = \begin{pmatrix} 2/3 & 1/3 \\ 1/3 & 2/3 \end{pmatrix}$$

while in Example 1

$$L_1 \equiv \begin{pmatrix} \pi(x_1'|\omega_1) & \pi(x_2'|\omega_1) \\ \pi(x_1'|\omega_2) & \pi(x_2'|\omega_2) \end{pmatrix} = \begin{pmatrix} 22/45 & 23/45 \\ 29/45 & 16/45 \end{pmatrix} .$$

and in Example 2

$$L_2 \equiv \begin{pmatrix} \pi(x_1''|\omega_1) & \pi(x_2''|\omega_1) \\ \pi(x_1''|\omega_2) & \pi(x_2''|\omega_2) \end{pmatrix} = \begin{pmatrix} 91/225 & 134/225 \\ 77/225 & 148/225 \end{pmatrix} .$$

In Example 2

$$L_2 = L_o \cdot \begin{pmatrix} 33/75 & 40/75 \\ 21/75 & 54/75 \end{pmatrix} ,$$

and hence $L_o \supset L_2$; further it is easy to check that X" is not a garbling of Y. In Example 1, L_1 garbles L_o, with

$$\pi(x_1'|y_1) = 1/3 \quad \pi(x_2'|y_1) = 2/3$$
$$\pi(x_1'|y_2) = 4/5 \quad \pi(x_2'|y_2) = 1/5 \quad ;$$

clearly Y is not finer than X'. Lemma 15 and the Examples establish the following, which is written with obvious notation.

Theorem 16: $\quad$ fineness $\underset{\not\Longleftarrow}{\Longrightarrow}$ garbling $\underset{\not\Longleftarrow}{\Longrightarrow}$ informativeness

For a single decision-maker who is able to choose one of several signals, all of which are ordered by informativeness, the above distinctions will be irrelevant, for the most valuable signal is the most informative. However, in other situations the particular relation between information structures is important. For example, it has been shown[17] that in a Vickery auction for an object of unknown quality, if two bidders have the same private valuation of the object when quality is known, if they hold the same priors over the quality, and if each is allowed to observe a private signal (about the quality) before submitting his bid, then if the second bidder's information is a garbling of the first, the second cannot expect a positive payoff from the auction. In this case the second bidder will prefer to observe an even less informative signal, but one which is conditionally independent of the first bidder's. (It is not hard to show that the <u>distributions</u> for such signals exist.)

In another context, agents may pool their information or they might be allowed to choose to observe two or more of several signals ordered by informativeness. In addition to Theorems 9 and 10, the following is useful with respect to such combined experiments.

Theorem 17:

(a) if $X \varphi X'$, then and only then $X \varphi (X,X')$.

(b) if X' garbles X, then and only then does (X,X') garble X.

(c) if $X \supset X'$, then $(X,X') \supset X$, where the ordering may or may not be strict.

(d) if $X \not\supset X'$ and $X' \not\supset X$, then $(X,X') \supset X$ and X', where the ordering <u>is</u> strict.

Proof: (a) is clear. (b) and (d) are shown in Appendix B. From (b) it follows that if X' garbles X, then $X \supset (X,X')$ and the ordering in (c) is weak. To finish the proof of (c) we have from Example 2 above

$$R \equiv \begin{pmatrix} \pi(y_1,x''_1|\omega_1) & \pi(y_1,x''_2|\omega_1) & \pi(y_2,x''_1|\omega_1) & \pi(y_2,x''_2|\omega_1) \\ \pi(y_1,x''_1|\omega_2) & \pi(y_1,x''_2|\omega_2) & \pi(y_2,x''_1|\omega_2) & \pi(y_2,x''_2|\omega_2) \end{pmatrix}$$

$$= \begin{pmatrix} 50/225 & 100/225 & 41/225 & 34/225 \\ 15/225 & 60/225 & 62/225 & 88/225 \end{pmatrix}$$

and

$$L_o \equiv \begin{pmatrix} 2/3 & 1/3 \\ 1/3 & 2/3 \end{pmatrix} \quad \text{as before.}$$

By refinement, $R \supset L_o$. In order to see R is strictly more informative than L_o we solve for the 2x4 matrix M such that $R = L_o M$:

$$M = \begin{pmatrix} 85/225 & 140/225 & 20/225 & -20/225 \\ -20/225 & 20/225 & 83/225 & 142/225 \end{pmatrix};$$

since M is unique but not Markov, $L_o \not\supset R$.

(b) and (d) give us sufficient conditions for one to know whether it is worthwhile to pool information. Furthermore, in Examples 1 and 2, $Y \supset X' \supset X''$ and X' is a garbling of Y, while X' is not (though it is dependent on Y). Consequently, we have constructed another example where a less informative signal may be preferred. Furthermore, we conjecture that a necessary and sufficient condition for the combination of two signals to be strictly more informative than either alone is that neither be a garbling of the other.

<u>Noiseless Information</u>

Definition 18: X is <u>noiseless</u> with respect to Ω if and only if $\Omega \varphi X$.[18]

Noiseless signals are those which tell us with certainty in which subset of Ω lies the "true" state. However, since by coarsening its partition of $\mathscr{S}$ every signal can be made noiseless--and hence by refinement the original partition is more informative than the noiseless one--noiselessness is <u>not</u> an optimality property. On the other hand, since $\Omega \varphi \Omega$, there always <u>exists</u> an undominated noiseless partition of Ω. We shall show, nonetheless, when attention is restricted to noiseless signals, fineness, garbling, and informativeness provide equivalent comparisons. With this in mind, one is assured of a matrix representation of fineness (cf. Theorem 8); we show this takes a special form.

Lemma 19: Suppose X is noiseless with respect to Ω; $P = (\pi(x_j|\omega_i))$ and $Q = (\pi(x'_k|\omega_i))$, $i = 1,\dots,n$, $j = 1,\dots,N$, $k = 1,\dots,N'$ are the $n \times N$ and $n \times N'$ matrices representing X and X' respectively. Then $X \varphi X'$ if and only if there exists a $N \times N'$ Markov matrix $B = (\beta_{kj})$ such that $\beta_{kj} = 0$ or 1, $k = 1,\dots,N$, $j = 1,\dots,N'$ and $Q = PB$.

Proof:

($\Rightarrow$) $X \varphi X'$ implies $\forall\, x \in X$, $x' \in X'$ we have $\pi(x'|x) = 0$ or 1. Further, $\pi(x'|\omega) = \sum_{x \in X} \pi(x,x'|\omega) = \sum_{x \in X} \pi(x'|x)\, \pi(x|\omega)$ since fineness implies garbling. Let $\beta_{kj} = \pi(x'_j|x_k)$, which is either 0 or 1. Then $Q = PB$, and B is Markov. (In fact this part of the proof does not use noiselessness of X and so furnishes a characterization of refinement in general.)

($\Leftarrow$) Fix $x' \in X'$ and $\omega \in \Omega$; we have $\pi(x'|\omega) = \sum_{x \in X'} \beta_{xx'}\, \pi(x|\omega)$, where $\beta_{xx'} = 0$ or 1. Since X is noiseless, $\exists!\ x^* \in X$ such that $\omega \subset x^*$. Consequently, $\pi(x'|\omega) = \beta_{x^*x'} = \pi(x'|x^*,\omega)$, and so $\beta_{x^*x'}\, \pi(x^*,\omega) = \pi(x^*,x',\omega)$. Summing over Ω, $\pi(x^*)\, \beta_{x^*x'} = \pi(x^*,x')$ and hence

$$\pi(x^*,x') = \begin{cases} \pi(x^*) & \text{if } \beta_{x^*x'} = 1 \\ 0 & \text{if } \beta_{x^*x'} = 0\,, \end{cases}$$

implying either $x^* \subset x'$ or $x^* \cap x' = \emptyset$. (Actually, we only have $\pi(x'|x^*) = 0$ or 1. Only null sets would cause difficulty with the inclusion relations; in this case we could find another partition which is <u>equally</u> as informative as X' for which the inclusions hold. Alternatively, $\pi(x'|x^*) = 0$ or 1 $\forall\, x' \in X'$, $x^* \in X$ could be taken as the definition of refinement.) To complete the proof we must show for arbitrary $\hat{x} \in X$ and $x' \in X'$ that either $\hat{x} \subset x'$ or $x' \cap \hat{x} = \emptyset$. But since X is noiseless, $\exists\ \hat{\omega} \in \Omega$ such that $\hat{\omega} \subset \hat{x}$. Repeating the above procedure for x' and $\hat{\omega}$ finishes the proof.

Theorem 20: In the consideration of noiseless information structures

$$\text{fineness} \Longleftrightarrow \text{garbling} \Longleftrightarrow \text{informativeness}.$$

Proof: Given Theorem 16 it suffices to show informativeness $\Rightarrow$ fineness. Suppose $X \supset X'$; then there is a Markov matrix B such that $Q = PB$, where Q and P are as in Lemma 19. This can be rewritten as $\pi(x'|\omega) = \Sigma\, \beta_{xx'}\, \pi(x|\omega)$, $\forall\, x' \in X'$, $\omega \in \Omega$. Now fix $x' \in X'$, $\omega \in \Omega$. Since X and X' are noiseless, $\pi(x'|\omega) = 0$ or 1, and $\exists!\, x^* \in X$ such that $\pi(x^*|\omega) = 1$. Therefore, $\beta_{x^*x'} = \pi(x'|\omega) = 0$ or 1. Since this holds for any $x' \in X'$, $\omega \in \Omega$, B consists only of zeros and ones, which with Lemma 19 implies $X \varphi X'$.

Theorem 20 stands in marked contrast with Theorem 16. For noiseless information structures X and X' to be more valuable in combination than either alone it is necessary and sufficient that neither be a refinement of the other. Also, the criterion of refinement is easier to check than informativeness, as it only involves an inclusion relation rather than a matrix condition.

IV. Limitations on Comparisons: an example

As noted earlier, when a single agent chooses to observe one among several signals, he is only interested in relative informativeness. However, when combined experiments are considered, a sharper characterization of signals is useful. This is also the case in the bidding example. Thus, it appears that in situations involving two or more information structures the additional characterization of garbling and finess may be useful and interesting. This section shows comparisons in these larger situations will be difficult.

We consider the following non-cooperative game between two agents. There are two states of the world, (ω_1, ω_2), and there corresponds to each state a different payoff matrix.

ω_1:

	t_1		t_2	
s_1	1/2	0	0	1/2
s_2	0	1	1	0

ω_2:

	t_1		t_2	
s_1	0	3/4	3/4	0
s_2	1	0	0	1

The top player is allowed pure strategies (t_1,t_2); the side player, (s_1,s_2). The state of the world is unknown before the players take an action. However, each is allowed to observe a private signal and may base his strategy on that signal. The side player observes the signal Y while the top player is allowed a choice between X and X' (here we associate partitions with signals as before). His choice will be known to the side player. Further, we assume initially each player believes $\pi(\omega_1) = 1/2$. The information structures are represented by the following distributions.

$\pi(y,x|\omega_1)$

	y_1	y_2
x_1	10/45	12/45
x_2	20/45	3/45

$\pi(y,x|\omega_2)$

	y_1	y_2
x_1	5/45	24/45
x_2	10/45	6/45

$\pi(y,x'|\omega_1)$

	y_1	y_2
x_1'	182/675	91/675
x_2'	268/675	134/675

$\pi(y,x'|\omega_2)$

	y_1	y_2
x_1'	77/675	154/675
x_2'	148/675	296/675

It can be shown that Y is strictly more informative than X, which is strictly more informative than X'. Further X is a garbling of Y but X' is not.

We now seek the expected payoff to the top player as he observes X or X'. There is an equilibrium in mixed strategies. Let $\sigma(y_i)$ denote the probability with which the side player plays s_1, given he has observed y_i, $i = 1,2$, and the top player observes signal X (though the realization is not revealed to the side player); analogously define $\hat{\sigma}(y_i)$ when the top player observes X'.

We find that $\sigma(y_1) = 4/5 \qquad \sigma(y_2) = 1/2$

$$\hat{\sigma}(y_1) = 16/21 \qquad \hat{\sigma}(y_2) = 10/21.$$

Further in equilibrium, the expected payoff to the top player is 3/8 if he observes X, 8/21 if he observes X'. Hence in this case, the less informative (but ungarbled) of two signals is preferred.

To see that this is not a general result in this context we have the following example. Let the signals and priors be as before; consider the following payoff matrices:

ω_1:

	t_1	t_2
s_1	2 0	0 2
s_2	0 1	1 0

ω_2:

	t_1	t_2
s_1	0 1	1 0
s_2	1 0	0 1

We now find $\sigma(y_1)=1/4 \qquad \sigma(y_2) = 1$

$$\hat{\sigma}(y_1) = 1/6 \qquad \hat{\sigma}(y_2) = 2/3,$$

and the top player's expected payoff when he observes X is 5/8; when he observes X', 7/12. Now the garbled, but more informative signal is preferred.

V. Conclusion

Information structures present one mode by which beliefs or the production of information might be modeled. For comparisons of such structures in the context of a single agent choosing a single signal we have the theorems of Blackwell. Further areas for research include the comparisons of combined signals as suggested earlier, and also the consequences of differential information among several agents. Information structures seem a fruitful way to model such differentials. By examining the game implicit in the bidding model, to isolate the difficulties raised in part IV, general results on comparisons of information in multi-person settings might be obtained. Interesting results might also arise by considering information in the context of repeated games. Lastly, another avenue for research would be to uncover situations in which different actions _generated_ different information structures, as in the example of Grossman _et al_.

Notes

[1]Convexity would have followed from consideration of randomized decisions; compactness is assumed for simplicity. Initially, Blackwell (1951) only assumes A to be bounded, though later (1953) he assumes compactness and convexity.

[2]Blackwell (1951), Theorem 2.

[3]*Ibid.*, Theorem 3.

[4]*Ibid.*, p. 95.

[5]*Ibid.*, Theorem 4 and Theorem 5.

[6]Blackwell (1953), p. 266.

[7]Grossman et al., pp. 538-9.

[8]Blackwell (1951), p. 97.

[9]Blackwell (1953), Theorem 5.

[10]This and the following paragraph are meant only to make plausible the claim that informativeness implies sufficiency. For example, dispersion would be irrelevant if the supports of the n different distributions were disjoint, for then the outcome of an experiment immediately identifies the state. This point will become clearer in section III. The remarks in the text are best interpreted, bearing in mind the example in which M_i, m_i are normal distributions Also, we have the caveat that the relations between relative informativeness and variance are not the same. For an example, see Rothschild-Stiglitz.

[11] Blackwell and Girschick, Theorems 12.3.1 and 12.3.2.

[12] Also see Marschak and Miyasawa, Marschak and Radner.

[13] We shall now drop the subscript on Ω_Ψ, understanding Ω to be payoff-relevant for the problem at hand.

[14] Marschak and Radner, pp. 55-7.

[15] Milgrom and Stokey.

[16]
$$\frac{\pi(x'|x,\omega)}{\pi(x'|x)} \equiv \frac{\pi(\omega|x,x')}{\pi(\omega|x)} \equiv \frac{\pi(x,x'|\omega)}{\pi(x|\omega)\ \pi(x'|x)}$$

These identities hold since each ratio equals $\dfrac{\pi(\omega,x,x')\ \pi(x)}{\pi(\omega,x)\ \pi(x,x')}$.

Garbling requires these ratios equal 1. See Marschak and Miyasawa, p. 149.)

[17] Milgrom.

[18] An alternative definition of <u>noiselessness</u> is to say X is noiseless with respect to Ω if $\pi(\omega|x) = 0$ or 1 for any $\omega \in \Omega$, $x \in X$. But here X is "virtually" perfect information since Ω is payoff adequate for each agent, and so seems to offer fewer interesting possibilities (cf. Marschak and Miyasawa, footnote 8).

Appendix A

Proof of Lemma 15:

(a) Since $X \varphi X'$, $\forall\, x \in X$, $x' \in X'$ either

(Ai) $x \subset x'$ so $\pi(x,x'|\omega) = \pi(x|\omega)$

or

(Aii) $x \cap x' = \emptyset$ so $\pi(x,x'|\omega) = 0.$

In (Ai) $\pi(x'|x) = 1$, in (Aii) $\pi(x'|x) = 0$. Hence, condition (iii) in the definition of garbling is met.

(b) Rewrite Theorem 8 as follows: $X \supset X'$ if and only if $\forall\, x' \in X'$, $x \in X$, $\omega \in \Omega$, $\exists$ non-negative numbers $\gamma_{xx'}$ such that $\sum_{x' \in X'} \gamma_{xx'} = 1$, $\forall\, x$ and $\pi(x'|\omega) = \sum_{x \in X} \gamma_{xx'}\, \pi(x|\omega)$. Since X' garbles X, $\pi(x,x'|\omega) = \pi(x|\omega)\, \pi(x'|x)$.

Therefore, $\pi(x'|\omega) = \sum_{x \in X} \pi(x,x'|\omega) = \sum_{x \in X} \pi(x'|x)\, \pi(x|\omega)$

Let $\gamma_{xx'} \equiv \pi(x'|x)$; application of our rewritten theorem completes the proof.
This proof of (b) is found in Marschak and Radner, pp. 65-6.

Appendix B

Proof of Theorem 17:

(b) Suppose X' garbles X. Then $\forall\ x_j,\ x_k \in X,\ x' \in X'$, and $\omega \in \Omega$

$$\pi(x_j,x'|x_k,\omega) = \begin{cases} 0 = \pi(x_j,x'|x_k) & j \neq k \\ \pi(x'|x_k,\omega) = \pi(x'|x_k) = \pi(x_k,x'|x_k) & j = k. \end{cases}$$

So our first condition for garbling is satisfied.

Now suppose (X,X') garbles X. Then $\forall\ x_j,\ x_k \in X,\ x' \in X'$ and $\omega \in \Omega$, $\pi(x_j,x'|x_k,\omega) = \pi(x_j,x'|x_k)$. Summing over j we get $\pi(x'|x_k,\omega) = \pi(x'|x_k)$, so X' garbles X, finishing the proof of (b).

(d) We have $(X,X') \supset X'$ by refinement. Now suppose $X \supset (X,X')$. Then $X \supset (X,X') \supset X'$, which contradicts the assumption that $X \not\supset X'$. Hence $(X,X') \supset X$ and the comparison is strong. Likewise, (X,X') is strictly more informative than X'.

References

Blackwell, D. (1951). "Comparison of Experiments," Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, pp. 93-102. University of California Press, California.

__________ (1953). "Equivalent Comparisons of Experiments," The Annals of Mathematical Statistics, 24, pp. 265-72.

__________, and Girshick, M.A. (1954). Theory of Games and Statistical Decisions. John Wiley and Sons, Inc., New York.

DeGroot, M.H. (1970). Optimal Statistical Decisions. McGraw-Hill, Inc., New York.

Green, J., and Stokey, N. (1978). "Two Representations of Information Structures and Their Comparisons," Economic Series, Technical Report No. 271, Institute for Mathematical Studies in the Social Sciences, Stanford University.

Grossman, S.J., Kihlstrom, R.E., and Mirman, L.J. (1977). "A Bayesian Approach to the Production of Information and Learning by Doing," The Review of Economic Studies, 44:3, pp. 533-47.

Marschak, J., and Miyasawa, K. (1968). "Economic Comparability of Information Systems," International Economic Review, 9, pp. 137-74.

__________, and Radner, R. (1972) Economic Theory of Teams (Cowles Foundation Monograph 22). Murray Printing Co., Massachusetts.

Milgrom, P. (1978). "A Bidding Model of Price Formation Under Uncertainty," Discussion Paper No. 364, the Center for Mathematical Studies in Economics and Management Science, Northwestern University.

__________, and Stokey, N. (1979). "Information, Trade, and Common Knowledge," Discussion Paper No. 377, the Center for Mathematical Studies in Economics and Management Science, Northwestern University.

Rothschild, M. and Stiglitz, J.E. (1970). "Increasing Risk: I. A Definition," Journal of Economic Theory, 2, pp. 225-43.

Royden, H.L. (1968). Real Analysis, Second Edition. The Macmillan Co., U.S.A.