# Two Representations of Information Structures and Their Comparisons*

Jerry R. Green** and Nancy L. Stokey***

September 1978

IMSSS Technical Report No. 271
Stanford University

Consider a decision maker concerned with the true value of an unknown parameter $\theta \in \Theta$. Let $\Theta$ be a finite set, $\Theta = \{\theta_1, ..., \theta_m\}$ , and let the prior probability distribution be fixed and denote it by $r = (r_1, ..., r_m)$ .

The decision maker's problem is to maximize the expected value of his utility by the choice of an action from a set A. Utility is written as $u$: $\Theta \times A \to \mathbb{R}$. The selection of $a \in A$ is made after receiving some information about $\theta$.

Let us give two alternative formalizations of the idea of an *information structure.*

I. A set $Y$, and a collection of $m$ probability distributions $\pi_1, ..., \pi_m$ on $Y$. This will be denoted $(Y, \pi)$.

---

[0]* This work was supported by National Science Foundation Grant SOC at the Institute for Mathematical Studies in the Social Sciences, Stanford University and National Science Foundation Grants SOC71-03803 and APR77-06999, Harvard University. Correspondence with D. Blackwell and C.B. McGuire is gratefully acknowledged.

** Department of Economics, Harvard University.

*** Graduate School of Management, Northwestern University.

II. A set $X$, a measure $\mu$ on $\Theta \times X$ such that the marginal distribution of $\theta$ is $r$, and a partition $\mathcal{S}$ of $X$. This will be denoted $(X, \mu, \mathcal{S})$.

The interpretation of (I) is that a point $y \in Y$ is observed according to the distribution $\pi_i$ if $\theta_i$ is the true state. Then the action $a \in A$ is chosen to maximize

$$\sum_{i=1}^{m} \lambda\left(\theta_i \mid y\right) u\left(\theta_i, a\right),$$

where $\lambda$ is the posterior distribution of $\theta$ given $y$:

$$\lambda\left(\theta_i \mid y\right) = \frac{\pi_i(y) r_i}{\sum_k \pi_k(y) r_k}.$$

The interpretation of (II) is that an event $S \in \mathcal{S}$ is observed and the conditional distribution of $\theta$ is computed to be

$$\frac{\mu\left(\{\theta\} \times S\right)}{\mu\left(\Theta \times S\right)}.$$

Obviously any information structure can be expressed in either form: for given $(Y, \pi_1, ..., \pi_m)$ we can let $X \equiv Y$, $\mu\left(\theta_i, x\right) \equiv r_i \pi_i\left(x\right)$, and $\mathcal{S} \equiv \left(\{x\}\right)_{x \in X}$. Similarly, $(Y, \pi_1, ..., \pi_m)$ can be constructed immediately given $(X, \mu, \mathcal{S})$.

Therefore, it would seem that description (I) is generally to be preferred because of its greater simplicity—there is really no need to describe a set $X$ in detail since the relevant objects are only the elements of $\mathcal{S}$, and the joint distribution on $\Theta \times \mathcal{S}$.

Description (II) is useful, however, when one has to compare information structures. Information structures can be (partially) ordered according to the maximized value of the objective function. One is said to be *more informative* than the other if the objective function can be made at least as great for every possible utility function $u$.[1] Let $(Y, \pi)$ and $(Y', \pi')$ be two information structures. We assume further that $Y$ and $Y'$ are finite sets: $Y = \{y_1, ..., y_n\}, Y' = \left\{y_1', ..., y_{n'}'\right\}$. Blackwell [1951] and others have shown that $(Y, \pi)$ is more informative than $(Y', \pi')$ if and only if there exists a Markov matrix $B$ satisfying

$$\Pi' = \Pi B, \tag{1}$$

where $\Pi_{m\times n} = [\pi_i(y_j)]$, $\Pi'_{m\times n'} = [\pi'_i(y'_{j'})]$.[2] Equation (1) can be used to check, constructively, whether two information structures are related by this criterion. This amounts to showing that a system of inequalities has a solution because the required Markovian character of $B$ entails $b_{jj'} \geq 0$ for all $j, j'$, as well as $\sum_{j'} b_{jj'} = 1$ for all $j$.

If the two information structures $(Y, \pi)$ and $(Y', \pi')$ were expressed in form (II) as $(X, \mu, \mathcal{S})$ and $(X', \mu', \mathcal{S}')$ the required computations would be precisely the same. However, if they could be expressed so that the same set $X$ and the same joint distribution $\mu$ were used, and so that $\mathcal{S}$ is a refinement of $\mathcal{S}'$, then the conclusion that $(Y, \pi)$ is more informative than $(Y', \pi')$ would be immediate upon inspection. We would like to know whether such a choice of $(X, \mu)$, $\mathcal{S}$ and $\mathcal{S}'$ generally exists.

Let us view this problem the other way around. If $(X, \mu)$ is fixed, then the set of all partitions of $X$ corresponds to a set of information structures which are naturally (partially) ordered by the criterion of refinement. Of course, these information structures may be only a small subset of all possible structures. Moreover, even if $(Y, \pi)$ is more informative than $(Y', \pi')$ and if $(Y, \pi)$ and $(Y', \pi')$ are equivalent to $(X, \mu, \mathcal{S})$ and $(X, \mu, \mathcal{S}')$ respectively, for two partitions $\mathcal{S}$ and $\mathcal{S}'$, it may not be true that $\mathcal{S}$ refines $\mathcal{S}'$. That is to say that refinement implies more informativeness, but not vice versa–for example, $\mathcal{S}'$ may have more elements than $\mathcal{S}$.

We know that $(Y, \pi)$ can be converted directly into $(X, \mu, \mathcal{S})$ by identifying points in $Y$ with members of S. However there are other descriptions of form (II), refinements of the n-member partition, with identical informational characteristics.

Definition: We will say that $(X, \mu, \mathcal{S})$ *represents* $(Y, \pi)$ if there is a mapping: $f : \mathcal{S} \to Y$ such that

i) for each $y \in Y$ and each $S \in f^{-1}(y)$,

$$\frac{\mu(\{\theta_i\} \times S)}{\mu(\Theta \times S)} = \lambda(\theta_i \mid y), \qquad i = 1, ..., n$$

ii) for each $y \in Y$

$$\sum_{S \in f^{-1}(y)} \mu(\Theta \times S) = \sum_{i=1}^{m} \pi(y \mid \theta_i) r_i.$$

That is, $(X, \mu, \mathcal{S})$ represents $(Y, \pi)$ if there is a class of subcollections of $\mathcal{S}$ with the properties that the induced posterior on $\Theta$ is insensitive to the element $S$ within any subcollection, being identically to equal to $\lambda(\cdot \mid y)$ for some $y \in Y$, and that the probability of occurrence of each such subcollection is the same as that for the $y$ with which it is associated. Thus, if $T$ is the collection $(f^{-1}(y))_{y \in Y}$ generated in this way, then $(X, \mu, \mathcal{S})$ and $(X, \mu, \mathcal{T})$ are really equivalent information structures, both being representations of $(Y, \pi)$. The partition $\mathcal{S}$ merely refines $T$ in a way that imparts no extra information about $\theta$.

Definition: Two information structures $(Y, \pi)$ and $(Y', \pi')$ satisfying Blackwell's condition (1) are said to be *imbeddable* if there exists $(X, \mu, \mathcal{S}, \mathcal{S}')$ such that

$(X, \mu, \mathcal{S})$ represents $(Y, \pi)$

$(X, \mu, \mathcal{S}')$ represents $(Y', \pi')$

$\mathcal{S}$ is a refinement of $\mathcal{S}'$.

Theorem 1: Any two information structures $(Y, \pi)$, $(Y', \pi')$ satisfying (1) are imbeddable.

Proof: Let $X = Y \times Y'$,

$$\mu(\{(\theta_i, x)\}) = \mu\left(\left\{\left(\theta_i, y_j, y'_{j'}\right)\right\}\right) = r_i \pi_i(y_j) b_{jj'},$$

where $B_{n \times n'} = (b_{jj'})$ is the Markov matrix in (1) relating $\Pi$ and $\Pi'$. Further, take

$$\mathcal{S} = \left\{\left(y_j, y'_{j'}\right) \mid y_j \in Y,\ y'_{j'} \in Y'\right\}$$

and

$$\mathcal{S}' = \left\{Y \times \left\{y'_{j'}\right\} \mid y'_{j'} \in Y'\right\}.$$

Clearly $\mathcal{S}$ refines $\mathcal{S}'$ and $\mu$ is a probability distribution on $\Theta \times X$.

To show that $(X, \mu, \mathcal{S})$ represents $(Y, \pi)$ take

$$f(S) \equiv f\left(\left\{\left(y_j, y'_{j'}\right)\right\}\right) = y_j,$$

so that

$$f^{-1}(\overline{y}_j) = \left\{\left\{\left(\overline{y}_j, y'_{j'}\right)\right\} \mid y'_{j'} \in Y'\right\}.$$

Then

$$\frac{\mu\left(\left\{\left(\theta_i, y_j, y'_{j'}\right)\right\}\right)}{\mu\left(\Theta \times \left\{\left(y_j, y'_{j'}\right)\right\}\right)} = \frac{r_i \pi_i\left(y_j\right) b_{jj'}}{\sum_{k=1}^{m} r_k \pi_k\left(y_j\right) b_{jj'}} = \lambda\left(\theta_i \mid y_j\right),$$

verifying part i) of the definition of representation and

$$\begin{aligned}
\sum_{S \in f^{-1}(y_j)} \mu\left(\Theta \times S\right) &= \sum_{j'=1}^{n'} \mu\left(\Theta \times \left\{\left(y_j, y'_{j'}\right)\right\}\right) \\
&= \sum_{j'=1}^{n'} \sum_{i=1}^{m} r_i \pi_i\left(y_j\right) b_{jj'} \\
&= \sum_{i=1}^{m} r_i \pi_i\left(y_j\right),
\end{aligned}$$

since $B$ is Markov matrix, verifying ii).

To show that $(X, \mu, \mathcal{S}')$ represents $(Y', \pi')$ take

$$f(S') = f\left(Y \times \left\{\overline{y}'_j\right\}\right) = \overline{y}'_j$$

so that

$$f^{-1}\left(\overline{y}'_j\right) = Y \times \left\{\overline{y}'_j\right\}.$$

(That is, $f^{-1}$ has only one member in $\mathcal{S}'$). Thus using (1),

$$\begin{aligned}
\frac{\mu\left(\left\{\theta_i\right\} \times Y \times \left\{y'_{j'}\right\}\right)}{\mu\left(\Theta \times Y \times \left\{y'_{j'}\right\}\right)} &= \frac{\mu\left(\left\{\theta_i\right\} \times S'\right)}{\mu\left(\Theta \times S'\right)} \\
&= \frac{\sum_{j=1}^{n} r_i \pi_i\left(y_j\right) b_{jj'}}{\sum_{k=1}^{m} \sum_{j=1}^{n} r_k \pi_k\left(y_j\right) b_{jj'}} \\
&= \frac{r_i \pi'_i\left(y'_{j'}\right)}{\sum_{k=1}^{m} r_k \pi_k\left(y'_{j'}\right)} \\
&= \lambda\left(\theta_i \mid y'_{j'}\right)
\end{aligned}$$

verifying i). Property ii) follows, as above, by applying Blackwell's theorem to $\mu(\Theta \times \mathcal{S}')$. Q.E.D.

Bart McGuire has suggested the following matrix interpretation of Theorem 1. Write the $n \times n'$ Markov matrix $B$ as the product of two Markov matrices, $E$ and $C$ where $E$ is $n \times (nn')$ and $C$ is $(nn') \times n'$, as follows:

$$E = \begin{pmatrix} b_{11}...b_{1n'} & 0\ldots0 & & 0\ldots0 \\ 0\ldots0 & b_{21}...b_{2n'} & & \vdots \\ & & \ddots & \vdots \\ 0\ldots0 & 0\ldots0 & & 0\ldots0 \\ & & \ddots & \vdots \\ & & & 0\ldots0 \\ & & & b_{n1}...b_{nn'} \end{pmatrix}$$

has the (row) vector $b_{k\cdot}$ as shown in the $k^{\text{th}}$ row and zeros elsewhere, and

$$C = \begin{pmatrix} I_{n'} \\ I_{n'} \\ \vdots \\ I_{n'} \end{pmatrix}$$

stacks $n$ identity matrices, each of which is $n' \times n'$.

By definition $\Pi B = (\Pi E)\, C = \Pi'$. Therefore the information structure given by $(Y'', \Pi E)$ (where $Y''$ has $n \times n'$ elements)[3] is more informative than $\Pi'$, and they are related by refinement as indicated by the matrix $C$. That is, all observations $Y_k'' = 1, ..., n \times n'$ in $Y''$ are associated to some $y_j',\in Y'$ according to $j' = k \mod n'$.

On the other hand, by choosing an $(n \times n') \times n$ Markov matrix $D$ given by

$$D = \begin{pmatrix} \underline{1}_{n'} & \underline{0}_{n'} & & & & \underline{0}_{n'} & \underline{0}_{n'} \\ \underline{0}_{n'} & \underline{1}_{n'} & & & & \underline{0}_{n'} & \underline{0}_{n'} \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ \underline{0}_{n'} & \underline{0}_{n'} & & & & \underline{1}_{n'} & \underline{0}_{n'} \\ \underline{0}_{n'} & \underline{0}_{n'} & & & & \underline{0}_{n'} & \underline{1}_{n'} \end{pmatrix},$$

$n$ blocks, each consisting of column vectors (of length $n'$) of zeros and ones, we have that $(\Pi E)\, D = \Pi$, so that $(Y'', \Pi E)$. Therefore $(Y, \Pi)$ and $(Y'', \Pi E)$ are equivalent information structures.

These results establish the existence of an information structure $(Y'', \Pi E)$ equivalent to $(Y, \pi)$ which can be viewed as a refinement of $(Y', \pi')$.

Theorem 1 is useful in the following way: Suppose a proposition is valid for any pair of partitions $\mathcal{S}, \mathcal{S}'$ of $X$, where $(X, \mu)$ is arbitrary and $\mathcal{S}$ refines $\mathcal{S}'$. Is it valid to conclude that the same proposition can be asserted for any two information structures $(Y, \pi)$ , $(Y', \pi')$ such that $(Y, \pi)$ is more informative than $(Y', \pi')$? This theorem allows one to give a positive answer by simply constructing $(X, \mu, \mathcal{S}, \mathcal{S}')$ so as to imbed $(Y, \pi)$, $(Y', \pi')$. Because it is often easier to work with refinements of partitions that with systems of inequalities such as (1), this theorem may provide a convenient analytical tool.

Theorem 1 shows how to construct $(X, \mu)$ to imbed any two information structure that are ordered by Blackwell's criterion. The following theorem, on the other hand, treats a fixed $(X, \mu)$ and characterizes alternative families of partitions of $X$ that all represent the same information structures $(Y, \pi)$ and $(Y', \pi')$.

Theorem 2: Given any two information structures $(Y, \pi)$ and $(Y', \pi')$, and any set $X$ for which there exist $(X, \mu, S)$ and $(X, \mu, S')$ representing $(Y, \pi)$ and $(Y', \pi')$ respectively:

1) $(Y,\pi)$ is more informative than $(Y',\pi')$ if and only if the finest partition of $X$ that represents $(Y,\pi)$ is a refinement of every partition of $X$ that represents $(Y',\pi')$;

2) $(Y,\pi)$ is more informative than $(Y',\pi')$ if and only if for every partition $\mathcal{S}$ of $X$ that represents $(Y,\pi)$ and every partition of $\mathcal{S}'$ of $X$ that represents $(Y',\pi')$ the least common refinement of $\mathcal{S}$ and $\mathcal{S}'$ represents $(Y,\pi)$.

The proof follows the lines of Theorem 1.

**Reference**

Blackwell, D. [1951], "Comparison of Experiments" in Second Berkeley Symposium on Mathematical Statistics and Probability, J. Neyman (Ed.), University of California Press.

**Notes**

[1]Actually, Blackwell [1951] requires that the dominance apply for all utilities and all prior probabilities. However his theorem shows that the partial ordering obtained for any fixed, positive, prior coincides with that for any other such prior.

[2]This theorem generalizes to infinite sets $\Theta$ and $Y$. The corresponding condition is stated in terms of a Markov kernel.

[3]The construction above, as in Theorem 1, requires the use of a space with $nn'$ points. One can observe that there is not, in general, an information structure $(Y'',\Pi'')$ equivalent to $(Y,\Pi)$ and refining $(Y',\Pi')$ (i.e. $\Pi''D'' = \Pi'$ for $D''$ a matrix of zeros and ones), when $Y''$ has fewer than $nn'$ points. For example consider $n = n' = 2$ and

$$\Pi = \begin{pmatrix} .6 & .4 \\ .3 & .7 \end{pmatrix}, \qquad \Pi' = \begin{pmatrix} .48 & .52 \\ .69 & .31 \end{pmatrix},$$

so that

$$B = \begin{pmatrix} .2 & .8 \\ .9 & .1 \end{pmatrix}$$

uniquely satisfies $\Pi B = \Pi'$.

Let us try to find $(Y'', \Pi'')$ as required with $Y''$ consisting of three points. Without loss of generality we can take

$$C'' = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Thus

$$E'' = \begin{pmatrix} .10 & .10 & .80 \\ .45 & .45 & .10 \end{pmatrix}.$$

If $\Pi E''$ is equivalent to $\Pi$, then we must be able to find a Markov matrix $D''$ so that $(\Pi E'')\, D'' = \Pi$. (The other direction, "$\Pi$ more information the $\Pi E$," is trivially verified by $E''$ itself.) Solving for $D$ in this numerical example one derives directly that $d''_{31} > 1$, so that $D''$ cannot be non-negative as required.

Similar examples can be given for any $n, n'$. In some special cases, where certain degeneracies occur, this construction can be done for less than $n \times n'$ points. But these are not really relevant, and $n \times n'$ is the general lower bound.