Blackwell's Comparison of Experiments Theorem
for Information Systems with Finite Signal Sets
by Michael Pejic 8/05/09

Introduction The name of UC Berkeley Professor Emeritus David Blackwell is attached to many illustrious theorems in the field of statistics, one of which is his Comparison of Experiments theorem, proposed in 1951, on the equivalence of *sufficient* and *more informative* experiments. This theorem was recast in the economic language of information systems and *garbling* by J. Marschak and K. Miyasawa. Employing this language and notation, I will discuss an extension of Blackwell's result from systems with a finite number of states to those with uncountably many.

Definition of an Information System

Suppose a state $s \in S$ is drawn from the probability space $(S, \mathcal{S}, r)$, so that for a subset F of S that is in the σ-algebra $\mathcal{S}$,

$$\text{Prob}\left(s \text{ is in } F\right) = r\left(F\right) = \int_{s \in \ F} dr\left(s\right)$$

r is termed the *underlying probability measure*, and it determines an equivalence relation, $f \sim_r g$ for $\mathcal{S}$-measurable functions $f, g : S \to \mathbb{R}$ if f and g differ on a r-null set. The resulting equivalence classes, $[\cdot]_r$, form a real vector space under $c\left[f\right]_r = \left[cf\right]_r$ and $\left[f\right]_r + \left[g\right]_r = \left[f+g\right]_r$ for $c \in \mathbb{R}$.

Define a *information system* on $(S, \mathcal{S})$ to be a doublet of a finite set of possible signals, Y, together with a vector of nonnegative, $\mathcal{S}$-measurable functions, $\Lambda^Y = \left[\lambda_y\right]_{y \in Y}$, where the $\lambda_y : S \to \mathbb{R}$ satisfy

$$\sum_{y \in Y} \lambda_y\left(s\right) = 1 \text{ for all } s \in S \tag{1}$$

so for each $s \in S$, $\left(Y, 2^Y, \lambda\left(s\right)\right)$ is a probability space. Denote the space of all information systems $\left\langle Y, \Lambda^Y \right\rangle$ on $(S, \mathcal{S})$ by $\Upsilon\left(S, \mathcal{S}\right)$.

Various relevant measures and quantities can be determined in terms of r and Λ^Y. Define the vector of $\mathcal{S}$-measures on $\mathcal{S}$, $P^Y = \left[p_y\right]_{y \in Y}$, by $p_y = \lambda_y \cdot r$, so for any $F \in \mathcal{S}$,

$$p_y\left(F\right) = \int_{s \in F} \lambda_y\left(s\right) dr\left(s\right)$$

interpreted as the probability that s is in F and signal y has been received. Note that P^Y determines a $\mathcal{S} \times 2^Y$-probability measure, ρ, by

$$\rho\left(G\right) = \sum_{y \in Y} p_y\left(G_y\right)$$

for any G in the σ-algebra $\mathcal{S} \times 2^Y$, where $G_y = \{s \in S | \, (s, y) \in G\}$. Note that, using (1),

$$\sum_{y \in Y} p_y (F) = \sum_{y \in Y} \int_{s \in F} \lambda_y (s) \, dr (s) \tag{2}$$

$$= \int_{s \in F} \left(\sum_{y \in Y} \lambda_y (s) \right) dr (s) = \int_{s \in F} dr (s) = r (F)$$

Define the vector of numbers, $Q^Y = [q_y]_{y \in Y}$, by $q_y = p_y (S)$. q_y is interpreted as the probability that the signal y has been received, so $\left(Y, 2^Y, q\right)$ is a probability space. Note $q_y = 0$ is equivalent to $\lambda_y \sim_r 0$.

Lastly, define the vector of $\mathcal{S}$-probability measures, $\Pi^Y = [\pi_y]_{\substack{y \in Y \\ q_y \neq 0}}$, by

$$\pi_y = \frac{1}{q_y} p_y$$

For any $F \in \mathcal{S}$ and any $y \in Y$ such that $q_y \neq 0$, interpret $\pi_y (F)$ as the probability that s is in F given that signal y has been received, so for each $y \in Y$ such that $q_y \neq 0$, $(S, \mathcal{S}, \pi_y)$ is a probability space. Also, using (2)

$$\sum_{\substack{y \in Y \\ q_y \neq 0}} q_y \cdot \pi_y = \sum_{\substack{y \in Y \\ q_y \neq 0}} p_y = \sum_{y \in Y} p_y = r \tag{3}$$

since $p_y \leq q_y$, so $p_y = 0$ when $q_y = 0$. Hence, r is in the convex hull of those π_y for which $q_y \neq 0$.

Extreme Information Systems

Note information systems with the same set of signals form a convex set since for any $\xi \in [0, 1]$, $(1 - \xi) \Lambda^Y + \xi \Lambda'^Y$ has the required properties. Therefore, call a signal set $\left\langle Y, \Lambda^Y \right\rangle$ *extreme* if the λ_y only take on the values 0 or 1. Then, by (1), for every $s \in S$, there is a unique $\tilde{y} (s)$ such that $\lambda_{\tilde{y}(s)} (s) = 1$ and $\lambda_y (s) = 0$ for $y \neq \tilde{y} (s)$.

Product of information systems

For information systems on $(S, \mathcal{S})$, $\left\langle Y, \Lambda^Y \right\rangle$ and $\left\langle Y', \Lambda'^{Y'} \right\rangle$, denote by $\left\langle Y, \Lambda^Y \right\rangle \times \left\langle Y', \Lambda'^{Y'} \right\rangle$ the information system on $(S, \mathcal{S})$ with possible signals in $Y \times Y'$ and $\tilde{\Lambda}^{Y \times Y'} = \left[\tilde{\lambda}_{(y,y')} \right]_{(y,y') \in Y \times Y'}$ given by $\tilde{\lambda}_{(y,y')} = \lambda_y \cdot \lambda'_{y'}$. This is indeed an information system since the $\tilde{\lambda}_{(y,y')}$ are nonnegative, $\mathcal{S}$-measurable, and obey (1)

$$\sum_{(y,y') \in Y \times Y'} \tilde{\lambda}_{(y,y')} = \sum_{y \in Y} \sum_{y' \in Y'} \lambda_y \cdot \lambda'_{y'} = \sum_{y \in Y} \lambda_y \sum_{y' \in Y'} \lambda'_{y'} = 1 \cdot 1 = 1$$

Garbling

Given two information systems on $(S, \mathcal{S})$, $\langle Y, \Lambda^Y \rangle$ and $\langle Y', \Lambda'^{Y'} \rangle$, $\langle Y', \Lambda'^{Y'} \rangle$ is a *garbling* of $\langle Y, \Lambda^Y \rangle$, (or in Blackwell's terminology, $\langle Y, \Lambda^Y \rangle$ *is sufficient for* $\langle Y', \Lambda'^{Y'} \rangle$), denoted $\langle Y', \Lambda'^{Y'} \rangle \prec \langle Y, \Lambda^Y \rangle$, if there is a matrix of numbers, $B^{Y'Y} = [b_{y'y}]_{y' \in Y', y \in Y}$ such that:

i) $b_{y'y} \geq 0$ for all $y' \in Y'$ and $y \in Y$;
ii) $\sum_{y' \in Y'} b_{y'y} = 1$ for all $y \in Y$;
iii) $\sum_{y \in Y} b_{y'y} \lambda_y = \lambda'_{y'}$ for all $y' \in Y'$, or, in matrix form, $B^{Y'Y} \Lambda^Y = \Lambda'^{Y'}$.

$\langle Y', \Lambda'^{Y'} \rangle$ is a *r-garbling* of $\langle Y, \Lambda^Y \rangle$, , denoted $\langle Y', \Lambda'^{Y'} \rangle \prec_r \langle Y, \Lambda^Y \rangle$, if there is a matrix of numbers, $B^{Y'Y} = [b_{y'y}]_{y' \in Y', y \in Y}$ satisfying **(i)** and **(ii)** but with **(iii)** replaced by

iv) $\sum_{y \in Y} b_{y'y} [\lambda_y]_r = \left[\lambda'_{y'}\right]_r$ for all $y' \in Y'$, or, in matrix form, $B^{Y'Y} \left[\Lambda^Y\right]_r = \left[\Lambda'^{Y'}\right]_r$.

Note that for any $\mathcal{S}$-measure r, if $B^{Y'Y} \Lambda^Y = \Lambda'^{Y'}$, then

$$B^{Y'Y} \left[\Lambda^Y\right]_r = \left[B^{Y'Y} \Lambda^Y\right]_r = \left[\Lambda'^{Y'}\right]_r$$

so

$$\left\langle Y', \Lambda'^{Y'} \right\rangle \prec \langle Y, \Lambda^Y \rangle \Rightarrow \left\langle Y', \Lambda'^{Y'} \right\rangle \prec_r \langle Y, \Lambda^Y \rangle$$

Also, for any $\mathcal{S}$-measurable function f and $\mathcal{S}$-measures $r, \tilde{r}$ with $\tilde{r} \ll r$, then $[f]_{\tilde{r}} \supset [f]_r$ since all r-null sets are $\tilde{r}$-null by the definition of absolute continuity. Hence, if $B^{Y'Y} \left[\Lambda^Y\right]_r = \left[\Lambda'^{Y'}\right]_r$, then

$$B^{Y'Y} \left[\Lambda^Y\right]_{\tilde{r}} = \left[B^{Y'Y} \left[\Lambda^Y\right]_r\right]_{\tilde{r}} = \left[\Lambda'^{Y'}\right]_{\tilde{r}}$$

so

$$\left\langle Y', \Lambda'^{Y'} \right\rangle \prec_r \langle Y, \Lambda^Y \rangle \Rightarrow \left\langle Y', \Lambda'^{Y'} \right\rangle \prec_{\tilde{r}} \langle Y, \Lambda^Y \rangle$$

There is a converse to the above, with an added technical condition,

Proposition 1 If $\langle Y', \Lambda'^{Y'} \rangle$ is a r-garbling of $\langle Y, \Lambda^Y \rangle$ for every $\mathcal{S}$-measure r, and the $\lambda_y \in \Lambda^Y$ are linearly independent, then $\langle Y', \Lambda'^{Y'} \rangle$ is a garbling of $\langle Y, \Lambda^Y \rangle$.

Proof Assume the $\lambda_y \in \Lambda^Y$ are linearly independent and suppose that $\langle Y', \Lambda'^{Y'} \rangle$ were not a garbling of $\langle Y, \Lambda^Y \rangle$. Then there are two cases: either there is some $v_1 \in S$ for which there is no matrix $B^{Y'Y}$ obeying **(i)** and

(ii) such that $\Lambda'^{Y'}(v_1) = B^{Y'Y}\Lambda^Y(v_1)$, or for every $s_1 \in S$ there is some $B^{Y'Y}(s_1)$ obeying **(i)** and **(ii)** such that $\Lambda'^{Y'}(s_1) = B^{Y'Y}(s_1)\Lambda^Y(s_1)$. In the first case, let $V = \{v_1\}$. The second has two subcases: either there are some $v_1, v_2 \in S$ for which there is no matrix $B^{Y'Y}$ obeying **(i)** and **(ii)** such that

$$\left[\begin{array}{cc} \Lambda'^{Y'}(v_1) & \Lambda'^{Y'}(v_2) \end{array}\right] = B^{Y'Y}\left[\begin{array}{cc} \Lambda^Y(v_1) & \Lambda^Y(v_2) \end{array}\right]$$

in which case, let $V = \{v_1, v_2\}$; or for every $s_1, s_2 \in S$ there is some $B^{Y'Y}(s_1, s_2)$ obeying **(i)** and **(ii)** such that

$$\left[\begin{array}{cc} \Lambda'^{Y'}(s_1) & \Lambda'^{Y'}(s_2) \end{array}\right] = B^{Y'Y}(s_1, s_2)\left[\begin{array}{cc} \Lambda^Y(s_1) & \Lambda^Y(s_2) \end{array}\right]$$

in which case one continues with two additonal subcases. This process can be continued until one reaches the subcase where for every $s_1, \ldots, s_{|Y|} \in S$ there is some $B^{Y'Y}\left(s_1, \ldots, s_{|Y|}\right)$ obeying **(i)** and **(ii)** such that

$$\left[\begin{array}{ccc} \Lambda'^{Y'}(s_1) & \cdots & \Lambda'^{Y'}\left(s_{|Y|}\right) \end{array}\right] = B^{Y'Y}\left(s_1, \ldots, s_{|Y|}\right)\left[\begin{array}{ccc} \Lambda^Y(s_1) & \cdots & \Lambda^Y\left(s_{|Y|}\right) \end{array}\right]$$

Since $\lambda_y \in \Lambda^Y$ are linearly independent, there are $v_1, \ldots, v_{|Y|} \in S$ such that

$$\left[\begin{array}{ccc} \Lambda^Y(v_1) & \cdots & \Lambda^Y\left(v_{|Y|}\right) \end{array}\right]$$

is invertible, so $B^{Y'Y}\left(v_1, \ldots, v_{|Y|}\right)$ is unique. By assumption, $\left\langle Y', \Lambda'^{Y'}\right\rangle$ is not a garbling of $\left\langle Y, \Lambda^Y\right\rangle$, so there is some $v_{|Y|+1}$ such that

$$\Lambda'^{Y'}\left(v_{|Y|+1}\right) \neq B^{Y'Y}\left(v_1, \ldots, v_{|Y|}\right)\Lambda^Y\left(v_{|Y|+1}\right)$$

For this subcase, let $V = \left\{v_1, \ldots, v_{|Y|+1}\right\}$.

Now define the $\mathcal{S}$-probablity measure r by, for each $F \in \mathcal{S}$,

$$r(F) = \frac{|F \cap V|}{|V|}$$

By the construction of V, $\left\langle Y', \Lambda'^{Y'}\right\rangle$ is not a r-garbling of $\left\langle Y, \Lambda^Y\right\rangle$, establishing the proposition.

Equivalent condition to r-garbling

Condition **(iv)** implies that

$$\textstyle\sum_{y \in Y} b_{y'y}p_y = p'_{y'} \text{ for all } y' \in Y' (\text{ or } B^{Y'Y}P^Y = P'^{Y'})$$

$$\textstyle\sum_{y \in Y} b_{y'y}q_y = q'_{y'} \text{ for all } y' \in Y' (\text{ or } B^{Y'Y}Q^Y = Q'^{Y'})$$

which in turn imply

$$\frac{1}{q'_{y'}} \sum_{\substack{y \in Y \\ q_y \neq 0}} b_{y'y} q_y \pi_y = \pi'_{y'} \text{ for all } y' \in Y' \text{ with } q'_{y'} \neq 0$$

Note for all $y' \in Y'$ with $q'_{y'} \neq 0$,

$$\frac{1}{q'_{y'}} \sum_{\substack{y \in Y \\ q_y \neq 0}} b_{y'y} q_y = \frac{1}{q'_{y'}} \sum_{y \in Y} b_{y'y} q_y = \frac{1}{q'_{y'}} q'_{y'} = 1$$

so for such a y', $\pi'_{y'}$ is in the convex hull of the $\pi_y \in \Pi^Y$.

There is a converse with an added technical condition.

Proposition 2 If the $\pi_y \in \Pi^Y$ are linearly independent (which is equivalent to $[\lambda_y]_r$ being linearly independent for those $y \in Y$ for which $\lambda_y \nsim_r 0$), and for each $y' \in Y'$ with $q'_{y'} \neq 0$, $\pi'_{y'}$ is in the convex hull of the $\pi_y \in \Pi^Y$, then $\langle Y', \Lambda'^{Y'} \rangle$ is a r-garbling of $\langle Y, \Lambda^Y \rangle$.

Proof Assuming the conditions in the proposition, then there are nonnegative numbers $c_{y'y}$ such that

$$\sum_{\substack{y \in Y \\ q_y \neq 0}} c_{y'y} = 1$$

and

$$\sum_{\substack{y \in Y \\ q_y \neq 0}} c_{y'y} \pi_y = \pi'_{y'}$$

for each $y' \in Y'$ with $q'_{y'} \neq 0$. Then, multiplying each side by $q'_{y'}$ and summing,

$$\sum_{\substack{y' \in Y' \\ q'_{y'} \neq 0}} \sum_{\substack{y \in Y \\ q_y \neq 0}} q'_{y'} c_{y'y} \pi_y = \sum_{\substack{y' \in Y' \\ q'_{y'} \neq 0}} q'_{y'} \pi'_{y'} = r$$

using (3). Since the π_y are linearly independent, then by (3), this implies

$$\sum_{\substack{y' \in Y' \\ q'_{y'} \neq 0,}} q'_{y'} c_{y'y} = q_y$$

for all $y \in Y$. Then, by defining $b_{y'y}$ by

$$b_{y'y} = \begin{cases} \dfrac{q'_{y'} c_{y'y}}{q_y} & \text{if } q'_{y'} \neq 0, q_y \neq 0 \\ 0 & \text{if } q'_{y'} = 0 \text{ or } q_y = 0 \end{cases}$$

properties **(i)** and **(ii)** clearly hold. Property **(iv)** also holds since for all $y' \in Y'$ with $q'_{y'} \neq 0$ (equivalent to $\lambda'_{y'} \nsim_r 0$)

$$\sum_{\substack{y \in Y \\ q_y \neq 0}} c_{y'y} \pi_y = \pi'_{y'}$$

$$\Rightarrow \sum_{\substack{y \in Y \\ q_y \neq 0}} c_{y'y} \frac{\lambda_y \cdot r}{q_y} = \frac{\lambda'_{y'} \cdot r}{q'_{y'}}$$

$$\Rightarrow \sum_{\substack{y \in Y \\ q_y \neq 0}} \frac{q'_{y'} c_{y'y}}{q_y} \lambda_y \cdot r = \lambda'_{y'} \cdot r$$

$$\Rightarrow \sum_{y \in Y} b_{y'y} \left[\lambda_y\right]_r = \left[\lambda'_{y'}\right]_r$$

whereas for those $y' \in Y'$ with $q'_{y'} = 0$,

$$0 = \sum_{y \in Y} b_{y'y} \left[\lambda_y\right]_r = \left[\lambda'_{y'}\right]_r = 0$$

so $\left\langle Y', \Lambda'^{Y'} \right\rangle$ is a r-garbling of $\left\langle Y, \Lambda^Y \right\rangle$.

An interpretation of the garbling matrix

Consider the information system $\left\langle Y', \Lambda'^{Y'} \right\rangle$ given by first receiving signal y from a information system $\left\langle Y, \Lambda^Y \right\rangle$ and then passing that signal through a stochastic process, $g : \Omega \times Y \to Y'$, to get a signal in Y'. Then the probability of getting signal y' out of the stochastic process given signal y was sent in is precisely $b_{y'y}$ and $\left\langle Y', \Lambda'^{Y'} \right\rangle$ is a garbling of $\left\langle Y, \Lambda^Y \right\rangle$.

While the above process is one possible model, there is a simple interpretation in another case. For $\left\langle Y', \Lambda'^{Y'} \right\rangle$ a garbling of an extreme information system $\left\langle Y, \Lambda^Y \right\rangle$, consider the product information system, $\left\langle Y, \Lambda^Y \right\rangle \times \left\langle Y', \Lambda'^{Y'} \right\rangle$. For $(y, y') \in Y \times Y'$ with $q_y \neq 0$, the probability of receiving signal y' given reception of signal y is

$$\text{Prob}\left(\text{receive signal } (y, y') \middle| \text{receive signal } (y, \star)\right)$$

$$= \frac{\text{Prob}\left(\text{receive signal } \left(y, y'\right)\right)}{\text{Prob}\left(\text{receive signal } \left(y, \star\right)\right)} = \frac{\tilde{q}_{(y,y')}}{q_y}$$

$$= \frac{1}{q_y} \int_{s \in S} \tilde{\lambda}_{(y,y')}\left(s\right) dr\left(s\right) = \frac{1}{q_y} \int_{s \in S} \lambda_y\left(s\right) \lambda'_{y'}\left(s\right) dr\left(s\right)$$

Since $\left\langle Y', \Lambda'^{Y'}\right\rangle$ is a garbling of $\left\langle Y, \Lambda^Y\right\rangle$,

$$\sum_{\tilde{y} \in Y} b_{y'\tilde{y}} \lambda_{\tilde{y}} = \lambda'_{y'}$$

Using this

$$\lambda_y \cdot \lambda'_{y'} = \lambda_y \cdot \left(\sum_{\tilde{y} \in Y} b_{y'\tilde{y}} \lambda_{\tilde{y}}\right) = \sum_{\tilde{y} \in Y} b_{y'\tilde{y}} \lambda_y \cdot \lambda_{\tilde{y}} = b_{y'y} \lambda_y$$

since $\left\langle Y, \Lambda^Y\right\rangle$ is an extreme information system. Then

$$\text{Prob}\left(\text{receive signal } \left(y, y'\right) \middle| \text{receive signal } \left(y, \star\right)\right)$$

$$= \frac{1}{q_y} \int_{s \in S} b_{y'y} \lambda_y\left(s\right) dr\left(s\right) = \frac{1}{q_y} b_{y'y} q_y = b_{y'y}$$

Garbling providing a directed set

Note the notion of garbling induces the structure of a directed set onto the collection of information systems, $\Upsilon\left(S, \mathcal{S}\right)$. This is reflexive since by taking $B^{Y'Y}$ to be the identity (1's on the diagonal, 0's elsewhere) every information system is a garbling of itself. Transitivity can be shown to hold. If $\left\langle Y', \Lambda'^{Y'}\right\rangle \prec \left\langle Y, \Lambda^Y\right\rangle$ and $\left\langle Y'', \Lambda''^{Y''}\right\rangle \prec \left\langle Y', \Lambda'^{Y'}\right\rangle$, let the corresponding garbling matrices be written $B^{Y'Y} = \left[b_{y'y}\right]_{y' \in Y', y \in Y}$ and $\tilde{B}^{Y''Y'} = \left[\tilde{b}_{y''y'}\right]_{y'' \in Y'', y' \in Y'}$. Consider the matrix $\breve{B}^{Y''Y} = \left[\breve{b}_{y''y}\right]_{y'' \in Y'', y \in Y}$ given by

$$\breve{b}_{y''y} = \sum_{y' \in Y'} \tilde{b}_{y''y'} b_{y'y} \text{ for all } y'' \in Y'', y \in Y$$

Clearly $\breve{b}_{y''y}$ is nonnegative. Property **(ii)** holds since for any $y \in Y$,

$$\sum_{y'' \in Y''} \breve{b}_{y''y} = \sum_{y'' \in Y''} \sum_{y' \in Y'} \tilde{b}_{y''y'} b_{y'y} = \sum_{y' \in Y'} \left(\sum_{y'' \in Y''} \tilde{b}_{y''y'}\right) b_{y'y} = \sum_{y' \in Y'} b_{y'y} = 1$$

using **(ii)** twice. From its definition, **(iii)** holds. Since such a $\breve{B}^{Y''Y}$ exists, $\left\langle Y'', \Lambda''^{Y''}\right\rangle \prec \left\langle Y, \Lambda^Y\right\rangle$. Lastly, given any two information systmens, $\left\langle Y, \Lambda^Y\right\rangle$,

$\left\langle Y', \Lambda'^{Y'} \right\rangle \in \Upsilon(S, \mathcal{S})$, there is another information system, $\left\langle Y, \Lambda^{Y} \right\rangle \times \left\langle Y', \Lambda'^{Y'} \right\rangle \in \Upsilon(S, \mathcal{S})$ such that both $\left\langle Y, \Lambda^{Y} \right\rangle \prec \left\langle Y, \Lambda^{Y} \right\rangle \times \left\langle Y', \Lambda'^{Y'} \right\rangle$ and $\left\langle Y', \Lambda'^{Y'} \right\rangle \prec \left\langle Y, \Lambda^{Y} \right\rangle \times \left\langle Y', \Lambda'^{Y'} \right\rangle$. For $B^{YY \times Y'}$, take entries $b_{y(\tilde{y}, y')}$ to be 1 if $y = \tilde{y}$ and 0 otherwise, and for $\breve{B}^{Y'Y \times Y'}$, take entries $\breve{b}_{y'(y, \tilde{y})}$ to be 1 if $y' = \tilde{y}$ and 0 otherwise. These are readily seen to satisfy the required properties **(i)**, **(ii)**, and **(iii)**.

Similarly, r-garbling also induces a directed set structure on $\Upsilon(S, \mathcal{S})$.

Aside–Garbling providing a partial ordering

Note garbling does not in general provide a partial odering on $\Upsilon(S, \mathcal{S})$ since if $\left\langle Y, \Lambda^{Y} \right\rangle$ is a garbling of $\left\langle Y', \Lambda'^{Y'} \right\rangle$ and $\left\langle Y', \Lambda'^{Y'} \right\rangle$ is a garbling of $\left\langle Y, \Lambda^{Y} \right\rangle$, it cannot be concluded that $\left\langle Y, \Lambda^{Y} \right\rangle = \left\langle Y', \Lambda'^{Y'} \right\rangle$, even after identifying information systems that differ by a permutation of signals. For example, let

$$\left\langle Y, \Lambda^{Y} \right\rangle = \left\langle Y = \{1, 2\}, \Lambda^{Y} = \begin{bmatrix} \lambda_1(s) \\ \lambda_2(s) \end{bmatrix} = \begin{bmatrix} \xi \\ 1 - \xi \end{bmatrix} \right\rangle$$

for some $\xi \in]0, 1[$, and let

$$\left\langle Y', \Lambda^{Y'} \right\rangle = \left\langle Y = \{1'\}, \Lambda^{Y} = \begin{bmatrix} \lambda'_{1'}(s) \end{bmatrix} = \begin{bmatrix} 1 \end{bmatrix} \right\rangle$$

Then $\left\langle Y, \Lambda^{Y} \right\rangle$ is a garbling of $\left\langle Y', \Lambda'^{Y'} \right\rangle$ with

$$B^{YY'} = \begin{bmatrix} \xi \\ 1 - \xi \end{bmatrix}$$

and $\left\langle Y', \Lambda'^{Y'} \right\rangle$ is a garbling of $\left\langle Y, \Lambda^{Y} \right\rangle$ with

$$B^{Y'Y} = \begin{bmatrix} 1 & 1 \end{bmatrix}$$

however, $\left\langle Y, \Lambda^{Y} \right\rangle \neq \left\langle Y', \Lambda'^{Y'} \right\rangle$.

On the other hand, in some cases having $\left\langle Y, \Lambda^{Y} \right\rangle$ be a garbling of $\left\langle Y', \Lambda'^{Y'} \right\rangle$ and $\left\langle Y', \Lambda'^{Y'} \right\rangle$ be a garbling of $\left\langle Y, \Lambda^{Y} \right\rangle$ does imply that $\left\langle Y, \Lambda^{Y} \right\rangle = \left\langle Y', \Lambda'^{Y'} \right\rangle$, up to a permutation of signals. Without loss of generality, take $|Y'| \leq |Y|$. Then, if there are $|Y|$ states, $s_1, \ldots, s_{|Y|} \in S$ such that the vectors $\left\{ \Lambda^{Y}(s_1), \ldots, \Lambda^{Y}\left(s_{|Y|}\right) \right\}$ form a linearly independent set,

$$\begin{aligned} \begin{bmatrix} \Lambda^{Y}(s_1) & \cdots & \Lambda^{Y}\left(s_{|Y|}\right) \end{bmatrix} &= B^{YY'} \begin{bmatrix} \Lambda'^{Y'}(s_1) & \cdots & \Lambda'^{Y'}\left(s_{|Y|}\right) \end{bmatrix} \\ &= B^{YY'} \tilde{B}^{Y'Y} \begin{bmatrix} \Lambda^{Y}(s_1) & \cdots & \Lambda^{Y}\left(s_{|Y|}\right) \end{bmatrix} \end{aligned}$$

implies that

$$B^{YY'}\tilde{B}^{Y'Y} = 1_{|Y|\times|Y|}$$

However, that is impossible unless $|Y'| \geq |Y|$, so it must be that $|Y'| = |Y|$. Then $B^{YY'}$ and $\tilde{B}^{Y'Y}$ are stochastic matrices that are inverses, but that is only possible if they are permutation matrices.

Constructing a $\Upsilon(S,\mathcal{S})$-net

Let $\mathcal{M}(S,\mathcal{S})$ be the collection of all $\mathcal{S}$-probability measures on S. Note $\mathcal{M}(S,\mathcal{S})$ is convex. Given a function $f : \mathcal{M}(S,\mathcal{S}) \to \mathbb{R}$ and an underlying probability measure r, there is a map $E_Y[f] : \Upsilon(S,\mathcal{S}) \to \mathbb{R}$ given by

$$E_Y[f] = \sum_{\substack{y \in Y \\ q_y \neq 0}} q_y f(\pi_y)$$

which gives a $\Upsilon(S,\mathcal{S})$-net on $\mathbb{R}$, with $\mathbb{R}$ ordered by the standard ordering.

Proposition 3 If $f : \mathcal{M}(S,\mathcal{S}) \to \mathbb{R}$ is convex, the map $E_Y[f]$ is isotone.

Proof Since, for $\langle Y, \Lambda^Y \rangle$, $\langle Y', \Lambda'^{Y'} \rangle \in \Upsilon(S,\mathcal{S})$ with $\langle Y', \Lambda'^{Y'} \rangle \prec_r \langle Y, \Lambda^Y \rangle$, $\pi'_{y'}$ is in the convex hull of the $\pi_y \in \Pi^Y$,

$$\begin{aligned}
E_{Y'}[f] = \sum_{\substack{y' \in Y' \\ q'_{y'} \neq 0}} q_{y'} f\left(\pi'_{y'}\right) &= \sum_{\substack{y' \in Y' \\ q'_{y'} \neq 0}} q_{y'} f\left(\frac{1}{q'_{y'}} \sum_{\substack{y \in Y \\ q_y \neq 0}} b_{y'y} q_y \pi_y\right) \\
&\leq \sum_{\substack{y' \in Y' \\ q'_{y'} \neq 0}} q_{y'} \frac{1}{q'_{y'}} \sum_{\substack{y \in Y \\ q_y \neq 0}} b_{y'y} q_y f(\pi_y) \\
&= \sum_{\substack{y \in Y \\ q_y \neq 0}} \left(\sum_{\substack{y' \in Y' \\ q'_{y'} \neq 0}} b_{y'y} \right) q_y f(\pi_y)
\end{aligned}$$

However, for $q'_{y'} = 0 \Leftrightarrow \lambda'_{y'} \sim_r 0$,

$$[0]_r = \sum_{y \in Y} b_{y'y} [\lambda_y]_r$$

which implies $b_{y'y} = 0$ for all y and y' with $q'_{y'} = 0$ and $q_y \neq 0$. Therefore, for $y \in Y$ such that $q_y \neq 0$

$$\sum_{\substack{y' \in Y' \\ q'_{y'} \neq 0}} b_{y'y} = \sum_{y' \in Y'} b_{y'y} = 1$$

so

$$E_{Y'}[f] \leq \sum_{\substack{y \in Y \\ q_y \neq}} q_y f(\pi_y) = E_Y[f]$$

so $\left\langle Y', \Lambda'^{Y'} \right\rangle \prec_r \left\langle Y, \Lambda^Y \right\rangle \Longrightarrow E_{Y'}[f] \leq E_Y[f]$.

Similarly, if f is concave, $\left\langle Y', \Lambda'^{Y'} \right\rangle \prec_r \left\langle Y, \Lambda^Y \right\rangle \Longrightarrow E_{Y'}[f] \geq E_Y[f]$.

Payoff function

Let A be the set of possible actions resulting from reception of the signals. Let $u : A \times S \to \mathbb{R}$ be the payoff function, with $u(a, \cdot) : S \to \mathbb{R}$ a $\mathcal{S}$-measurable function for all actions $a \in A$. Then define $\breve{u} : A \times \mathcal{M}(S, \mathcal{S}) \to \mathbb{R}$ by

$$\breve{u}(a, \mu) = (u(a, \cdot)\mu)(S) = \int_{s \in S} u(a, s)\, d\mu(s)$$

Note $\breve{u}$ is linear in the second entry, so for any $a \in A$, $\mu, \nu \in \mathcal{M}(S, \mathcal{S})$, and $\xi \in [0, 1]$

$$\breve{u}(a, (1-\xi)\mu + \xi\nu) = (1-\xi)\breve{u}(a, \mu) + \xi\breve{u}(a, \nu)$$

Let $\hat{u} : \mathcal{M}(S, \mathcal{S}) \to \mathbb{R}$ be defined by

$$\hat{u}(\mu) = \sup_{a \in A} \breve{u}(a, \mu)$$

Note $\hat{u}$ is a convex function since for any $\mu, \nu \in \mathcal{M}(S, \mathcal{S})$, and $\xi \in [0, 1]$,

$$\hat{u}((1-\xi)\mu + \xi\nu) = \sup_{a \in A} \breve{u}(a, (1-\xi)\mu + \xi\nu)$$

$$= \sup_{a \in A} ((1-\xi)\breve{u}(a, \mu) + \xi\breve{u}(a, \nu))$$

by the linearity of $\breve{u}$ in the second entry. Since both $1 - \xi$ and ξ are non-negative, by the property of the supremum, this is in turn less than or equal to

$$(1-\xi)\sup_{a \in A} \breve{u}(a, \mu) + \xi \sup_{a \in A} \breve{u}(a, \nu) = (1-\xi)\hat{u}(\mu) + \xi\hat{u}(\nu)$$

More informative

Following Blackwell's terminology, if for two information systems, $\left\langle Y, \Lambda^Y \right\rangle$, $\left\langle Y', \Lambda'^{Y'} \right\rangle \in \Upsilon(S, \mathcal{S})$,

$$E_{Y'}[\hat{u}] \leq E_Y[\hat{u}]$$

for all payoff functions u and underlying $\mathcal{S}$-probability measures r, $\left\langle Y, \Lambda^Y \right\rangle$ is said to be *more informative* than $\left\langle Y', \Lambda'^{Y'} \right\rangle$, denoted $\left\langle Y', \Lambda'^{Y'} \right\rangle \subset \left\langle Y, \Lambda^Y \right\rangle$. Being more informative is clearly reflexive and transitive.

If, given a fixed $\mathcal{S}$-measure r,

$$E_{Y'}[\hat{u}] \leq E_Y[\hat{u}]$$

for all payoff functions u and underlying $\mathcal{S}$-probability measures $\tilde{r}$ with $\tilde{r} \ll r$, $\langle Y, \Lambda^Y \rangle$ is said to be *more r-informative* than $\langle Y', \Lambda'^{Y'} \rangle$, denoted $\langle Y', \Lambda'^{Y'} \rangle \subset_r \langle Y, \Lambda^Y \rangle$. Being more r-informative is clearly reflexive and transitive.

From the definitions, one has for any underlying $\mathcal{S}$-probability measure r,

$$\left\langle Y', \Lambda'^{Y'} \right\rangle \subset \langle Y, \Lambda^Y \rangle \Rightarrow \left\langle Y', \Lambda'^{Y'} \right\rangle \subset_r \langle Y, \Lambda^Y \rangle$$

and for any underlying $\mathcal{S}$-probability measures $r, \tilde{r}$ with $\tilde{r} \ll r$,

$$\left\langle Y', \Lambda'^{Y'} \right\rangle \subset_r \langle Y, \Lambda^Y \rangle \Rightarrow \left\langle Y', \Lambda'^{Y'} \right\rangle \subset_{\tilde{r}} \langle Y, \Lambda^Y \rangle$$

Blackwell's Comparison of Experiments Theorem

Given two signal sets, $\langle Y, \Lambda^Y \rangle, \langle Y', \Lambda'^{Y'} \rangle \in \Upsilon(S, \mathcal{S})$, $\langle Y, \Lambda^Y \rangle$ is more informative than $\langle Y', \Lambda'^{Y'} \rangle$ if $\langle Y', \Lambda'^{Y'} \rangle$ is a garbling of $\langle Y, \Lambda^Y \rangle$, or in symbols,

$$\left\langle Y', \Lambda'^{Y'} \right\rangle \prec \langle Y, \Lambda^Y \rangle \Longrightarrow \left\langle Y', \Lambda'^{Y'} \right\rangle \subset \langle Y, \Lambda^Y \rangle$$

Also, given a fixed $\mathcal{S}$-measure r, $\langle Y, \Lambda^Y \rangle$ is more r-informative than $\langle Y', \Lambda'^{Y'} \rangle$ if $\langle Y', \Lambda'^{Y'} \rangle$ is a r-garbling of $\langle Y, \Lambda^Y \rangle$, or in symbols,

$$\left\langle Y', \Lambda'^{Y'} \right\rangle \prec_r \langle Y, \Lambda^Y \rangle \Longrightarrow \left\langle Y', \Lambda'^{Y'} \right\rangle \subset_r \langle Y, \Lambda^Y \rangle$$

The converse of the latter holds with the added technical condition that with r as the underlying $\mathcal{S}$-probability measure, $\pi_y \in \Pi^Y$ are linearly independent. The converse of the former holds with the added technical condition that $\lambda_y \in \Lambda^Y$ are linearly independent.

Proof of theorem ($\Rightarrow$)

Since for any payoff function u, $\hat{u}$ is convex, then for any underlying $\mathcal{S}$-probability measure r on S,

$$\left\langle Y', \Lambda'^{Y'} \right\rangle \prec \langle Y, \Lambda^Y \rangle \Longrightarrow \left\langle Y', \Lambda'^{Y'} \right\rangle \prec_r \langle Y, \Lambda^Y \rangle \Longrightarrow E_{Y'}[\hat{u}] \geq E_Y[\hat{u}]$$

using **Proposition 3**, so $\langle Y, \Lambda^Y \rangle$ is more informative than $\langle Y', \Lambda'^{Y'} \rangle$. Also, given a fixed $\mathcal{S}$-probability measure r, then for any underlying $\mathcal{S}$-probability measure $\tilde{r}$ with $\tilde{r} \ll r$,

$$\left\langle Y', \Lambda'^{Y'} \right\rangle \prec_r \langle Y, \Lambda^Y \rangle \Longrightarrow \left\langle Y', \Lambda'^{Y'} \right\rangle \prec_{\tilde{r}} \langle Y, \Lambda^Y \rangle \Longrightarrow E_{Y'}[\hat{u}] \geq E_Y[\hat{u}]$$

using **Proposition 3**, so $\langle Y, \Lambda^Y \rangle$ is more r-informative than $\langle Y', \Lambda'^{Y'} \rangle$.

Proof of theorem ($\Leftarrow$)

With the added condition, if $\langle Y', \Lambda'^{Y'} \rangle$ is not a r-garbling of $\langle Y, \Lambda^Y \rangle$, then by **Proposition 2** there must be some $y' \in Y'$ with $q'_{y'} \neq 0$ such that $\pi'_{y'}$ is not in the convex hull of the $\pi_y \in \Pi^Y$. Let $\mathcal{B}$ be the Banach space of r-equivalence classes of $\mathcal{S}$-measurable functions equipped with the essential supremum norm. Define $T : \mathcal{B} \to \mathbb{R}^{|\{y \in Y | q_y \neq 0\}|+1}$ by

$$T(f) = \begin{bmatrix} E[f|y] \\ E[f|y'] \end{bmatrix}_{\substack{y \in Y \\ q_y \neq 0}}$$

where

$$E[f|y] = \int_{s \in S} f \, d\pi_y$$

$$E[f|y'] = \int_{s \in S} f \, d\pi'_{y'}$$

T is continuous since there are only finitely many components and each component has operator norm 1, hence T is bounded. Therefore, the kernel of T is a closed subspace of codimension at most $|Y|+1$, so $B/\ker T$ is a finite dimensional vector space. Each $\pi_y \in \Pi^Y$ and $\pi'_{y'}$ then correspond to elements, x_y for $y \in Y$ with $q_y \neq 0$ and $x'_{y'}$, in the dual space, $(B/\ker T)^*$, itself finite-dimensional, hence locally convex. Then, since $\pi'_{y'}$ is not in the convex hull of the $\pi_y \in \Pi^Y$, $x'_{y'}$ is not in the convex hull of $\{x_y\}_{y \in Y, q_y \neq 0}$ because if there were some linear relation among the x_y for $y \in Y$ with $q_y \neq 0$ and $x'_{y'}$ not present among the $\pi_y \in \Pi^Y$ and $\pi'_{y'}$, then that linear relation among the $\pi_y \in \Pi^Y$ and $\pi'_{y'}$ would be a nonzero $\mathcal{S}$-measure, absolutely continuous with respect to r, that would give 0 when integrated over S against any function in B, an impossibility. Hence, by the separation theorem, there is a continuous linear functional in $(B/\ker T)^{**}$, call it Φ, and a $c \in \mathbb{R}$ such that

$$\Phi\left(x'_{y'}\right) > c > \sup\left\{ \Phi(x) \,\middle|\, x \in \text{convex hull}\{x_y\}_{\substack{y \in Y \\ q_y \neq 0}} \right\} \tag{4}$$

$$= \max\{\Phi(x_y)\}_{\substack{y \in Y \\ q_y \neq 0}}$$

since linear functionals are convex. Since $B/\ker T$ is finite dimensional, it is reflexive, so

$$\Phi(x) = \langle x, f + \ker T \rangle$$

for some $f \in B$. Then

$$\Phi(x_y) = \langle x_y, f + \ker T \rangle = E[f|y]$$

so (4) becomes

$$E[f|y'] > c > \max\{E[f|y]\}_{\substack{y \in Y \\ q_y \neq 0}} \tag{5}$$

Now consider the payoff function $u : \{0,1\} \times S \to \mathbb{R}$ given by

$$\begin{cases} u(0,s) = c \\ u(1,s) = f(s) \end{cases}$$

Then $\breve{u} : \{0,1\} \times \mathcal{M}(S, \mathcal{S}) \to \mathbb{R}$ is given by

$$\begin{cases} \breve{u}(0,\mu) = c \\ \breve{u}(1,\mu) = E_\mu[f] = \int_{s \in S} f \, d\mu \end{cases}$$

and $\hat{u} : \mathcal{M}(S, \mathcal{S}) \to \mathbb{R}$ is

$$\hat{u} = \max(c, E_\mu[f])$$

Using (5), then

$$E_Y[\hat{u}] = \sum_{\substack{y \in Y \\ q_y \neq 0}} q_y \hat{u}(\pi_y) = \sum_{\substack{y \in Y \\ q_y \neq 0}} q_y \max(c, E[f|y]) = c \sum_{\substack{y \in Y \\ q_y \neq 0}} q_y = c$$

whereas

$$E_{Y'}[\hat{u}] = \sum_{\substack{\tilde{y}' \in Y' \\ q'_{\tilde{y}'} \neq 0}} q'_{\tilde{y}'} \hat{u}\left(\pi'_{\tilde{y}'}\right) = \sum_{\substack{\tilde{y}' \in Y \\ q'_{\tilde{y}'} \neq 0}} q'_{\tilde{y}'} \max(c, E[f|\tilde{y}'])$$

$$= q'_{y'} E[f|y'] + \sum_{\substack{\tilde{y}' \in Y \setminus \{y'\} \\ q'_{\tilde{y}'} \neq 0}} q'_{\tilde{y}'} \max(c, E[f|\tilde{y}'])$$

$$\geq q'_{y'}(E[f|y'] - c) + c$$

Since there is a payoff function and an underlying $\mathcal{S}$-probability measure absolutely continuous with respect to r, namely r itself, for which $E_{Y'}[\hat{u}] \geq E_Y[\hat{u}]$, $\left\langle Y', \Lambda'^{Y'} \right\rangle$ is not more r-informative than $\left\langle Y, \Lambda^Y \right\rangle$. Therefore, if $\left\langle Y', \Lambda'^{Y'} \right\rangle$ is more r-informative than $\left\langle Y, \Lambda^Y \right\rangle$, $\left\langle Y', \Lambda'^{Y'} \right\rangle$ must be a r-garbling of $\left\langle Y, \Lambda^Y \right\rangle$.

Now suppose $\langle Y, \Lambda^Y \rangle$ is more informative than $\langle Y', \Lambda'^{Y'} \rangle$ and that $\lambda_y \in \Lambda^Y$ are linearly independent. Then, given any underlying $\mathcal{S}$-probability measure r, there is a $\mathcal{S}$-probability measure $\tilde{r}$ such that $r \ll \tilde{r}$ and, with $\tilde{r}$ as the underlying $\mathcal{S}$-probability measure, $\pi_y \in \Pi^Y$ are linearly independent. By the preceding, then $\langle Y', \Lambda'^{Y'} \rangle$ is a $\tilde{r}$-garbling of $\langle Y, \Lambda^Y \rangle$, hence a r-garbling of $\langle Y, \Lambda^Y \rangle$. Since r was arbitrary, by **Proposition 1** $\langle Y', \Lambda'^{Y'} \rangle$ is a garbling of $\langle Y, \Lambda^Y \rangle$.