# EQUIVALENT COMPARISONS OF EXPERIMENTS[1]

By David Blackwell

*Stanford University and Howard University*

**1. Summary.** Sherman [8] and Stein [9] have shown that a method given by the author [1] for comparing two experiments is equivalent, for experiments with a finite number of outcomes, to the original method introduced by Bohnenblust, Shapley, and Sherman [4]. A new proof of this result is given, and the restriction to experiments with a finite number of outcomes is removed. A class of weaker comparisons—comparison in $k$-decision problems—is introduced, in three equivalent forms. For dichotomies, all methods are equivalent, and can be described in terms of errors of the first and second kinds.

**2. Introduction.** An ordered collection $\alpha = (m_1, \cdots, m_n)$ of probability measures on a Borel field $\mathcal{B}$ of subsets of a space $X$ will be called an *experiment.* Any pair $(\alpha, A)$, where $A$ is a closed bounded convex subset of $n$-space corresponds to a decision problem as follows. A point $x \varepsilon X$ is selected according to one of the distributions $m_i$; the statistician observes $x$ and then chooses an action $d$ from a given set $D$, incurring a loss $L(i, d)$. If we associate with $d$ the vector $w(d) = (L(1, d), \cdots, L(n, d))$, the range of $w(d)$ as $d$ varies over $D$ is the set $A$ associated with the problem. Thus we may replace $D$ by $A$, and suppose that the statistician chooses a point $a = (a_1, \cdots, a_n) \varepsilon A$, incurring loss $a_i$ when $m_i$ is the distribution of $x$. By using randomized decision procedures we increase $A$ to its convex hull, and for simplicity we suppose $A$ closed and bounded as well as convex.

A decision function in the problem $(\alpha, A)$ is a $\mathcal{B}$-measurable function $f$ from $X$ into $A$, specifying for each $x$ the action $a = f(x)$ to be taken when $x$ is observed. When $m_i$ is the distribution of $x$, the expected loss from $f$ is $v_i(f) = \int a_i(x)\, dm_i(x)$; the vector $v(f) = (v_1(f), \cdots, v_n(f))$ is called the loss vector of $f$, and the range of $v(f)$ as $f$ varies over all decision functions in the problem $(\alpha, A)$ will be denoted by $B(\alpha, A)$. The set $B(\alpha, A)$ will be a closed, bounded, convex subset of $n$-space [2].

For two experiments $\alpha$, $\beta$ with the same $n$, following Bohnenblust, Shapley, and Sherman, we say that $\alpha$ is *more informative than* $\beta$, written $\alpha \supset \beta$, if for every $A$ we have $B(\alpha, A) \supset B(\beta, A)$, that is if every loss vector attainable in problem $(\beta, A)$ is also attainable in $(\alpha, A)$. For any experiment $\alpha = (m_1, \cdots, m_n)$, let $p_i(x)$ be the density of $m_i$ with respect to $\sum_1^n m_i$, let $p(x) = [p_1(x), \cdots, p_n(x)]$, and let $m_\alpha$ denote the distribution of $p(x)$ when $x$ has distribution $\sum_1^n m_i/n$. Then $m_\alpha$ is a probability measure defined on the set $P$ of all vectors

Received 11/17/52.

[1] This paper was prepared under an Office of Naval Research Contract.

$p = (p_1, \cdots, p_n)$ with $p_i \geqq 0$ and $\sum_1^n p = 1$, and

$$\int p_i dm_\alpha = 1/n; \tag{1}$$

the center of gravity of $m_\alpha$ is the point $(1/n, \cdots, 1/n)$. The measure $m_\alpha$ is called by Bohnenblust, Shapley, and Sherman the standard measure associated with the experiment $\alpha$. Their basic results connecting $m_\alpha$ and $\supset$ are summarized as Theorem 1 below (for a proof see [1]).

THEOREM 1. *Every probability measure on $P$ with property* (1) *is the standard measure of some experiment; two experiments $\alpha$ and $\beta$ have the same standard measure if and only if $B(\alpha, A) = B(\beta, A)$ for all $A$; $\alpha \supset \beta$ if and only if for every continuous convex function $\phi(p)$ on $P$, $\int \phi \, dm_\alpha \geqq \int \phi \, dm_\beta$.*

An alternative method of comparing two experiments $\alpha$, $\beta$, introduced by the author [1], can best be described in terms of the concept of *stochastic transformation.* If $\mathcal{B}$, $\mathcal{C}$ 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 $E \in \mathcal{C}$ which for fixed $E$ is a $\mathcal{B}$-measurable function of $x$ and for fixed $x$ is a probability measure on $\mathcal{C}$. For any probability measure $m$ on $\mathcal{B}$, the function $M(E) = \int Q(x, E) \, dm(x)$ is a probability measure on $\mathcal{C}$, denoted by $Tm$. If $X$, $Y$ are Borel sets in $n$-space and $\mathcal{B}$, $\mathcal{C}$ are the Borel subsets of $X$, $Y$, $T$ is called *mean-preserving* if $\int y dQ(x, y) = x$ for all $x$.

If $\alpha = (m_1, \cdots, m_n)$ and $\beta = (M_1, \cdots, M_n)$ are two experiments, with $m_i$, $M_i$ defined on Borel fields $\mathcal{B}$, $\mathcal{C}$ of $X$, $Y$ respectively, we shall say that $\alpha$ is *sufficient for* $\beta$, written $\alpha \succ \beta$, if there is a stochastic transformation $T$ with $Tm_i = M_i$ for $i = 1, \cdots, n$. Thus $\alpha \succ \beta$ means that the statistician, observing the result $x$ of $\alpha$, can, by selecting $y$ according to $Q(x, E)$, obtain a result equivalent to the result of observing $\beta$.

The concept $\succ$ also has a description in terms of standard measures, summarized in

THEOREM 2. [1]. *$\alpha \succ \beta$ if and only if there is a mean-preserving stochastic transformation $T$ with $Tm_\beta = m_\alpha$.*

If $\alpha \succ \beta$ and $\phi$ is any continuous convex function on $P$,

$$\begin{aligned} \int \phi \, dm_\alpha &= \int \left( \int \phi(p) \, dQ(q, p) \right) dm_\beta(q) \\ &\geqq \int \phi \left( \int p dQ(q, p) \right) dm_\beta(q) \\ &= \int \phi \, dm_\beta, \end{aligned}$$

so that, from Theorem 1 we obtain

THEOREM 3. *$\alpha \succ \beta$ implies $\alpha \supset \beta$.*

The converse of Theorem 3 has been proved, for experiments with a finite number of outcomes, by Sherman [8] and Stein [9]. In Section 3 we give a new proof of the Sherman-Stein theorem, and in Section 4 extend the theorem to arbitrary experiments.

**3. The Sherman-Stein Theorem.** If the space $X$ of outcomes of the experiment $\alpha$ is finite, consisting say of $x_1, \cdots, x_N$, then $\alpha$ is characterized by the $n \times N$ Markov matrix $P = \| p_{ij} \|$, where $p_{ij} = m_i(x_j)$, and conversely every Markov matrix can be interpreted as an experiment. For two Markov matrices $P$, $Q$ with the same $n$, we write $P \supset Q$, $P \succ Q$ if the corresponding experiments are related by $\supset$, $\succ$ respectively.

THEOREM 4. *If $P$, $Q$ are $n \times N_1$, $n \times N_2$ Markov matrices with $P \supset Q$, then for every $N_2 \times n$ matrix $D$ there is an $N_1 \times N_2$ Markov matrix $M$ with*

$$\text{Trace } (PMD) \leqq \text{Trace } (QD).$$

PROOF. Let $A$ be the convex hull of the rows of $D$. The decision function $f$ in problem $(Q, A)$ selecting the $j$th row of $D$ when $j$ is observed has $v_i(f) = \sum_j q_{ij} d_{ji}$, the $i$th diagonal element of $QD$.

Since $P \supset Q$, there is a decision function $g$ in problem $(P, A)$, selecting say $a_j \varepsilon A$ when $j$ is observed, with $v_i(g) = \sum_j p_{ij} a_{ji} = v_i(f)$ for all $i$. Since $a_j \varepsilon A$, there are nonnegative numbers $m_{jk}$ with $\sum_k m_{jk} = 1$ such that $a_{ji} = \sum_k m_{jk} d_{ki}$ for all $i$. Thus $v_i(g) = \sum_{j,k} p_{ij} m_{jk} d_{ki}$, which is the $i$th diagonal element of $PMD$. It follows that $M$ has not only the property asserted in the theorem but the stronger property that $PMD$ and $QD$ have identical diagonal elements.

THEOREM 5. *$P \succ Q$ if and only if there is a Markov matrix $M$ with $PM = Q$.*

This is simply a restatement of the definition of $\succ$ for the special case $X = (1, \cdots, N_1)$, $Y = (1, \cdots, N_2)$, since a stochastic transformation becomes simply an $N_1 \times N_2$ Markov matrix.

THEOREM 6. (Sherman-Stein theorem). *$P \supset Q$ implies $P \succ Q$.*

PROOF. Consider the function $h(D, M) = \text{Trace } (Q - PM)D$, as $M$ varies over all $N_1 \times N_2$ Markov matrices and $D$ varies over all $N_2 \times n$ matrices with $0 \leqq d_{ki} \leqq 1$ for all $k$, $i$. Since $h$ is bilinear and the ranges of $D$, $M$ are closed, bounded, and convex, $h$ has a saddle point [3], that is there exist $D_0$, $M_0$ with $h(D_0, M) \geqq h(D_0, M_0) \geqq h(D, M_0)$ for all $D$, $M$. From Theorem 4, there is an $M$ with $h(D_0, M) \leqq 0$, so that $h(D, M_0) \leqq 0$ for all $D$. Writing $U = Q - PM_0$, we have

$$\text{Trace } (UD) \leqq 0 \text{ for all } D,$$

so that $u_{ik} \leqq 0$ for all $i$, $k$. Since $U$ is the difference of two Markov matrices, $\sum_{i,k} u_{ik} = 0$, so that $u_{ik} = 0$ for all $i$, $k$ and $PM_0 = Q$. Thus by Theorem 5, $P \succ Q$.

An alternative form of the Sherman-Stein theorem is

THEOREM 7. *If $m_1$ and $m_2$ are any two probability measures on a finite subset $X$ of n-space such that for every continuous convex $\phi$ defined on the convex hull of $X$,*

$\int \phi\, dm_1 \geqq \int \phi\, dm_2$, *then there is a mean-preserving stochastic transformation* $T$ *with* $Tm_2 = m_1$.

From Theorems 1 and 2, Theorem 7 implies Theorem 6. Theorem 7 was proved for $n = 1$ by Hardy, Littlewood, and Polya [6], for $n = 2$ without the restriction that $X$ be finite by the author, and in the form given here by Sherman [8] and Stein [9].

PROOF OF THEOREM 7. From Theorems 1 and 2, Theorem 6 implies Theorem 7 if $X \subset P$ and the common center of gravity of $m_1$, $m_2$ is $(1/n, \cdots, 1/n)$, since in this case $m_1$, $m_2$ are the standard measures of experiments. Imbedding $X$ in $n + 1$ space and performing an appropriate linear transformation reduces the general case in $n$-space to that of standard measures in $n + 1$ space and completes the proof.

A direct proof of Theorem 7, using the methods of Theorem 6 and not appealing to Theorems 1 and 2 can be given.

**4. Equivalence of $\supset$ and $>$.** In this section we extend Theorem 7, replacing the requirement that $X$ be finite by the weaker requirement that $X$ be bounded. For any two probability measures $m$, $M$ on a bounded subset $X$ of $n$-space, we write $M \supset m$ if for every continuous convex $\phi$ on the convex hull of $X$ $\int \phi\, dM \geqq \int \phi\, dm$ and $M > m$ if there is a mean-preserving stochastic transformation (abbreviated m.p.s.t.) $T$ with $Tm = M$. We shall prove

THEOREM 8. *If $M \supset m$, then $M > m$.*

The method of proof consists of approximating $m$, $M$ by measures concentrated on finite sets and using Doob's martingale convergence theorems. We first prove

A. *There exist sequences of measures $m_n$, $M_n$ each concentrated on a finite set, with $m_N < m_{N+1} \subset m \subset M < M_{N+1} < M_N$ for all $N$, and for every open set $O$*

$$m_N(O) \to m(O), \qquad M_N(O) \to M(O) \qquad \text{as } N \to \infty.$$

PROOF. For any $n$-vector $a = (a_1, \cdots, a_n)$ with integral coordinates, let $C(N, a)$ denote the cube consisting of all $t = (t_1, \cdots, t_n)$ with $2^{-N}a_i \leqq t_i < 2^{-N}(a_i + 1)$, let $Z(N, a)$ be the center of gravity of $m$ on $C(N, a)$ and let $m_N$ assign to $Z(N, a)$ measure $m(C(N, a))$. It is easily verified that $m_N$ has the required properties.

To define $M_N$, let $Q_N(t, E)$ for $t \,\varepsilon\, C(N, a)$ concentrate on the $2^n$ vertices of $C(N, a)$ assigning to vertex $2^{-N}(a_1 + \epsilon_1, \cdots, a_n + \epsilon_n)$, where $\epsilon_i = 0$ or 1, measure $b_1 b_2 \cdots b_n$, where $b_i = 2^{-N}A_i + 1 - t_i$ if $\epsilon_i = 0$ and $b_i = t_i - 2^{-N}a_i$ if $\epsilon_i = 1$. The function $Q_N(t, E)$ is a m.p.s.t. $U_N$, and if we define $M_N = U_N M$, we have also $M_N = U_N M_{N+1}$, so that $M_N$ has the required properties.

B. *There exist sequences $T_N$, $V_N$, $W_N$ of* m.p.s.t. *each from a finite set of n-space to a finite set of n-space with*

$$\text{(a)}\quad m_{N+1} = T_N m_N, \qquad \text{(b)}\quad M_{N-1} = V_N M_N, \qquad \text{(c)}\quad M_N = W_N m_N,$$

and

$$\text{(d)} \quad W_N = V_{N+1} W_{N+1} T_N .$$

PROOF. From $A$ there exist sequences $T_N$ and $V_N$ with properties (a) and (b). Also from $A$, $m_N \subset M_N$, so that, from Theorem 7, there is a m.p.s.t. $Y_N$ from a finite set to a finite set with $M_N = Y_N m_N$. For $D > N$, write

$$Y_{ND} = V_{N+1} \cdots V_D Y_D T_{D-1} \cdots T_N ,$$

so that

$$Y_{ND} = V_{N+1} Y_{N+1,D} T_N \qquad \text{for } D > N + 1,$$

and

$$M_N = Y_{ND} m_N \qquad \text{for } D > N.$$

Let $D \to \infty$ through a subsequence for which $Y_{ND}$ converges for all $N$, say to $W_N$. Then $W_N$ satisfies (c) and (d).

PROOF OF THEOREM 8. We specify the joint distribution of two sequences $x_1, x_2, \cdots, y_1, y_2, \cdots$, of $n$-dimensional chance variables by

C. For any $N$, the variables $x_1, \cdots, x_N, y_N, \cdots, y_1$ form a Markov chain in the order written. The distribution of $x_1$ is $m_1$ and the conditional distributions of $x_{i+1}$ given $x_1$, $y_N$ given $x_N$, and $y_{i-1}$ given $y_i$, are specified by $T_i$, $W_N$, and $V_i$ respectively.

Part (d) of B guarantees that the requirements C are consistent, and Kolmogorov's extension theorem [7] then asserts the existence of $x_1, x_2, \cdots, y_1, y_2, \cdots$, with property C. Parts (a), (b), (c) of B imply that $x_N$, $y_N$ have distributions $m_N$, $M_N$ respectively. Also the sequence

$$x_1, x_2, \cdots, \cdots, y_2, y_1$$

forms a martingale [5] in the order written; by Doob's martingale theorem [5], $x_N \to x^*$, $y_N \to y^*$ as $N \to \infty$, and $E(y^* \mid x^*) = x^*$. From $A$, $x^*$ and $y^*$ have distributions $m$, $M$ respectively, so that $Q(x, E) = \text{Prob}\,\{y^* \varepsilon E \mid x^* = x\}$ is a m.p.s.t. $T$ with $Tm = M$. This completes the proof.

**5. $k$-decision problems.** In this section we introduce a comparison somewhat weaker than $\succ$. The following lemma will be useful.

LEMMA. *For any experiment $\alpha$ and any closed, bounded set $C$ with convex hull $A$, $B(\alpha, A) =$ convex hull of $B(\alpha, C)$.*

PROOF. Since both $B(\alpha, A)$ and $B(\alpha, C)$ are closed and $B(\alpha, A)$ is convex [2], it suffices to show that every $v(f) \varepsilon B(\alpha, A)$ can be approximated by points in the convex hull of $B(\alpha, C)$. We may suppose that $f$ assumes only a finite number of values $a_1, \cdots, a_N$, since every $f$ can be approximated by $f$'s of this kind. Say

$$S_j = \{f(x) = a_j\}, \qquad a_j = \sum_{i=1}^{r} \lambda_{ji} c_i, \lambda_{ji} \geqq 0, \qquad \sum_i \lambda_{ji} = 1.$$

For any $h = (h_1, \cdots, h_N)$, $1 \leqq h_i \leqq r$, define

$$f(h) = c_{h_j} \text{ for } x \in S_j, \lambda(h) = \prod_{j=1}^{r} \lambda_{jh_j}.$$

Then $v(f(h)) \in B(\alpha, C)$, and $\sum_h \lambda(h) v[f(h)]$ has for its *sth* coordinate

$$\begin{aligned} \sum_h \lambda(h) \sum_j \int_{S_j} c_{h_j s} \, dm_s &= \sum_{h,j} m_s(S_j(c_{h_j s} \lambda(h))) \\ &= \sum_j m_s(S_j) \left(\sum_i c_{is} \left(\sum_{h:h_j = i} \lambda(h)\right)\right) \\ &= \sum_j m_s(S_j) \left(\sum_i \lambda_{ij} c_{is}\right) = \sum_j a_{js} m_s(S_j) \\ &= s^{th} \text{ coordinate of } v(f). \end{aligned}$$

This completes the proof.

APPLICATION 1. *Let $\alpha$ be any experiment, let $\mathcal{S} = (S, \cdots, S_k)$ be any partition of $X$ into $k$ disjoint $\mathcal{B}$-measurable sets, let $P(\mathcal{S})$ be the $n \times k$ Markov matrix with $p_{ij} = m_i(S_j)$, let $\mathcal{P}^*_{\alpha k}$ be the range of $P(\mathcal{S})$, and let $\mathcal{P}_{\alpha k}$ be the set of all $n \times k$ Markov matrices $P$ which have the property $\alpha \succ P$. Then $\mathcal{P}_{\alpha k}$ is the convex hull of $\mathcal{P}^*_{\alpha k}$.*

This is the special case of the lemma applied to the experiment $\alpha'$ consisting of $nk$ measures $M_{ij}$ with $M_{ij} = m_i$ for $j = 1, \cdots, k$ and $C$ consisting of the $kn \times k$ Markov matrices $P_1, \cdots, P_k$, where $P_j$ has the *jth* column identically 1 and the remaining columns identically zero.

APPLICATION 2. *For any experiment $\alpha$ and any closed bounded convex set $A$ which is the convex hull of the set of $k$ points $d_1, \cdots, d_k$, $B(\alpha, A)$ is the range of* diag *$PD$ as $P$ varies over $\mathcal{P}_{\alpha k}$, where* diag *$U$ for any $n \times n$ matrix $U = \| u_{ij} \|$ denotes the $n$-vector $(u_{11}, u_{22}, \cdots, u_{nn})$ and $D$ is the $k \times n$ matrix whose rows are $d_1, \cdots, d_k$.*

If $C$ consists of $d_1, \cdots, d_k$, and $f$ is any decision function in $(\alpha, C)$, say $S_j = \{f = d_j\}$. Then the $s$th coordinate of $v(f)$ is

$$\sum_j m_s(S_j) d_{js},$$

so that

$$v(f) = \text{diag } P(\mathcal{S}) D.$$

Thus $B(\alpha, C) =$ range of diag $PD$ as $P$ varies over $\mathcal{P}^*_{\alpha k}$. From the lemma, the convex hull of $B(\alpha, C)$ is $B(\alpha, A)$, and from Application (1) the convex hull of the range of diag $PD$ as $P$ varies over $\mathcal{P}^*_{\alpha k}$ is the range of $PD$ as $P$ varies over $\mathcal{P}_{\alpha k}$.

THEOREM 9. *Let $\alpha$, $\beta$ be two experiments with the same $n$. The following conditions are equivalent:*

(1) $$\mathcal{P}_{\alpha k} \supset \mathcal{P}_{\beta k}$$

(2) *For every $A$ which is the convex hull of a set of $k$ points, $B(\alpha, A) \supset B(\beta, A)$.*

(3) *For every convex function $\phi$ on $n$-space which is the maximum of $k$ linear functions, $\int \phi \, dm_\alpha \geqq \int \phi \, dm_\beta$.*

PROOF. Suppose (1) and let $v \in B(\beta, A)$, where $A$ is the convex hull of $d_1, \cdots, d_k$. Then $v =$ diag $PD$ for some $P \in \mathcal{P}_{\beta k}$.

Since $\mathcal{P}_{\beta k} \subset \mathcal{P}_{\alpha k}$, $v = \text{diag } PD$ for some $P \varepsilon \mathcal{P}_{\alpha k}$ and $v \varepsilon B(\alpha, A)$. Thus (1) implies (2).

Now suppose (2) and let $P \varepsilon \mathcal{P}_{\beta k}$. Then for any closed bounded convex set $R$, let $v \varepsilon B(P, R)$, say $v = v(f)$, where $f(j) = r_j \varepsilon R, j = 1, \cdots, k$. Then $v \varepsilon B(P, R^*)$, where $R^*$ is the convex hull of $r_1, \cdots, r_k$. Since $B(P, R^*) \subset B(\beta, R^*) \subset B(\alpha, R^*)$, $v \varepsilon B(\alpha, R^*)$ and consequently $v \varepsilon B(\alpha, R)$. Thus $\alpha \supset P$ for any $P \varepsilon \mathcal{P}_{\beta k}$ and, by Theorem 8, $\alpha > P$. Since $\mathcal{P}_{\alpha k}$ contains all $n \times k$ Markov matrices $P$ with $\alpha > P$, $P \varepsilon \mathcal{P}_{\alpha k}$ and $\mathcal{P}_{\beta k} \subset \mathcal{P}_{\alpha k}$. Thus (2) implies (1).

In considering (3), we use the fact that the standard measure $m_P$ of an $n \times k$ Markov matrix $P$ is concentrated on $k$ points, which follows immediately from the definition. Suppose (3), let $\phi$ be the maximum of any finite set $\mathcal{L}$ of linear functions, and let $P \varepsilon \mathcal{P}_{\beta k}$. There is a $\psi$, the maximum of $k$ functions in $\mathcal{L}$, which agrees with $\psi$ on the $k$ points on which $m_p$ is concentrated. Then $\int \phi \, dm_\alpha \geqq \int \psi \, dm_\alpha \geqq \int \psi \, dm_\beta \geqq \int \psi \, dm_P = \int \phi \, dm_P$, so that from Theorems 1 and 8, $\alpha > P$. Thus $P \varepsilon \mathcal{P}_{\alpha k}$, $\mathcal{P}_{\beta k} \subset \mathcal{P}_{\alpha k}$ and (3) implies (1).

Finally, suppose (1) and let $\phi = \max (L_1, \cdots, L_k)$; say

$$U_j = \{L_j(p) = \phi(p), L_i(p) < \phi(p) \text{ for } i < j\}.$$

If $S_j = \{p(x) \; U_j\}$, $\mathcal{S} = (S_1, \cdots, S_k)$ is a partition of $X$ and the experiment $P = P(\mathcal{S})$ associated with $\beta$ and $\mathcal{S}$ (see Application 1) has a standard measure $m_P$ with

$$m_P(U_j) = m_\beta(U_j),$$

so that

$$\int \phi \, dm_\beta = \int \phi \, dm_P.$$

Since $P \varepsilon \mathcal{P}_{\beta k}$, (1) implies $P \varepsilon \mathcal{P}_{\alpha k}$, so that $\int \phi \, dm_\alpha \geqq \int \phi \, dm_P = \int \phi \, dm_\beta$. This completes the proof.

If two experiments $\alpha, \beta$ with the same $n$ satisfy any of the three equivalent conditions of Theorem 9, we shall say that $\alpha$ is more *informative than* $\beta$ *for k-decision problems*, written $\alpha >_k \beta$. Condition (2) is the direct analogue of $\supset$, and condition (1) is analogous to $>$, since it requires that every experiment with $k$ outcomes producible from $\beta$ is also producible from $\alpha$. Clearly $>_{k+1}$ implies $>_k$, and if $\alpha >_k \beta$ for all $k$, then $\alpha > \beta$, since $\alpha >_k \beta$ for all $k$ implies $\int \phi \, dm_\alpha \geqq \int \phi \, dm_\beta$ for every $\phi$ which is the maximum of a finite number of linear functions and hence, by approximation, for every continuous convex $\phi$. An alternative statement is: if every experiment with a finite number of outcomes which is producible from $\beta$ is also producible from $\alpha$, then $\beta$ is itself producible from $\alpha$.

Stein (unpublished paper) has shown that in general $>_{k+1}$ is actually stronger than $>_k$. For $n = 2$, however, all $>_k$ for $k \geqq 2$ are equivalent.

THEOREM 10. *If* $\alpha$ *and* $\beta$ *are two experiments with* $n = 2$, *then* $\alpha >_2 \beta$ *implies* $\alpha > \beta$.

PROOF. For $n = 2$, the standard measures $m_\alpha$ and $m_\beta$ are defined on the line segment $p_i \geqq 0, p_1 + p_2 = 1$. On this line segment, every function $\phi$ which is the maximum of a finite number of linear functions is representable as $\sum a_i\phi_i$, where $a_i > 0$ and each $\phi_i$ is a maximum of two linear functions. Consequently $\alpha \succ_2 \beta$ implies $\alpha \succ_k \beta$ for all $k$ and hence $\alpha \succ \beta$.

COROLLARY. *Let $A$ be the line segment joining* (0, 1) *and* (1, 0). *If* $B(\alpha, A) \supset B(\beta, A)$, *then* $\alpha \succ \beta$.

PROOF. For any line segment $A'$ in the plane, there is a transformation

$$L: \begin{aligned} x' &= ax + b \\ y' &= cx + d \end{aligned}$$

with $LA = A'$. Since $LB(\alpha, A) = B(\alpha, LA)$ and similarly for $\beta$, we have $B(\alpha, A') \supset B(\beta, A')$, so that $\alpha \succ_2 \beta$ and consequently $\alpha \succ \beta$.

For the $A$ of the corollary, the boundary of the set $B(\alpha, A)$ consists of two curves, joining (0, 1) and (1, 0), one of which is the reflection of the other about (1/2, 1/2). Denote by $f_\alpha(t)$ the minimum of $u$ for which $(t, u) \varepsilon B(\alpha, A)$. Then $\alpha \succ \beta$ if and only if $f_\alpha(t) \leqq f_\beta(t)$ for all $t$, $0 \leqq t \leqq 1$. The function $f_\alpha(t)$ is a nonincreasing convex function of $t$, representing the minimum attainable error of the second kind when the error of the first kind is fixed at $t$. Thus an alternative statement of the corollary is:

*$\alpha$ is more informative than $\beta$ if and only if at every level $t$ the error of the second kind with $\alpha$ is less than or equal to the corresponding error with $\beta$.*

Since if $\alpha \succ \beta$, then an experiment with $n$ independent observations with $\alpha$ is more informative than the corresponding experiment with $\beta$ [1] we obtain

THEOREM 11. *If for a sample of size* 1 *at every level $t$ the probability of an error of the second kind with $\alpha$ does not exceed the corresponding probability for $\beta$, then the same is true for every sample size.*

## REFERENCES

[1] DAVID BLACKWELL, "Comparison of experiments," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1951, pp. 93–102.

[2] DAVID BLACKWELL, "The range of certain vector integrals," *Proc. Amer. Math. Soc.*, Vol. 2 (1951), pp. 390–395.

[3] H. F. BOHNENBLUST, S. KARLIN, AND L. S. SHAPLEY, "Games with continuous convex payoff," *Contributions to the Theory of Games*, Princeton University Press, 1950, pp. 181–192.

[4] H. F. BOHNENBLUST, L. S. SHAPLEY, AND S. SHERMAN, Unpublished paper.

[5] J. L. DOOB, "Continuous parameter martingales," *Proceedings of the Secondary Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 1951, pp. 269–277.

[6] G. H. HARDY, J. E. LITTLEWOOD, AND G. POLYA, *Inequalities*, Cambridge University Press, 1934.

[7] A. N. KOLMOGOROV, "Grundbegriffe der wahrscheinlichkeitsrechnung," *Ergebnisse der Mathematik*, No. 3, 1933.

[8] S. SHERMAN, "On a theorem of Hardy, Littlewood, Polya, and Blackwell," *Proc. Nat. Acad. Sci.*, U.S.A., Vol. 37 (1951), pp. 826–831.

[9] CHARLES STEIN, "Notes on the Comparison of Experiments," (mimeographed), University of Chicago, 1951.