TEAM DECISION PROBLEMS[1]

By R. Radner

University of California, Berkeley

1. Introduction. In a *team decision problem* there are two or more decision variables, and these different decisions can be made to depend upon different aspects of the environment, i.e., upon different information variables. For example, the "decision maker" may be a group of persons, each with access to different information (because of difficulty of communication, say), each deciding about something different, but participating in a common payoff as a result of their joint decision. As a second example, the decision maker might be an individual making different decisions in successive time periods, the payoff being a function of all the decisions made over the total time period. In such a case, if the decision maker does not "forget" anything from one time period to the next, then the problem is a typical sequential or dynamic programming problem. However, the keeping of records might be so costly that it would be worthwhile to forget some things, in which case new phenomena emerge.

Although from the most general point of view a team decision problem is a one-person problem, of a somewhat unconventional type, it is sometimes suggestive to talk about it in terms of the first, "many-person," example above, and that will be done in this paper. J. Marschak has called such a decision maker a *team* [19] to emphasize that there are no conflicts of interest between the members of the group. It should also be pointed out that differences of opinion (as embodied in different *a priori* distributions, for example) cannot be handled in this context, since these result formally in the same game-theoretic difficulties as do conflicts of interest.

The origin of the problems considered in this paper is Marschak's work on the theory of organization [19]. Marschak's approach is in the spirit of the theory of games and of decision theory. The immediate background of the present paper consisted of attempts by several workers to analyze some of the many-person aspects of organizations that are present even in the absence of many-person game complications, i.e., conflicts of interest and differences of opinion. These attempts took the form of study of some simplified examples of organizations, [4], [8], [18], [20], and [21]. Beckman [5] and McGuire [22] have also analyzed some real organizational problems from the point of view of the theory of teams.

Received November 23, 1958; revised February 26, 1962.

[1] Based on research done in part at the Cowles Commission for Research in Economics, under Contract Nonr-358(01), NR047-006, with the Office of Naval Research, and in part at the Center for Advanced Study in the Behavioral Sciences. Further work, and preparation of the final manuscript, was supported by the Office of Naval Research under contracts Nonr-222(43) and Nonr-222(77) with the University of California, Berkeley. Reproduction in whole or in part is permitted for any purpose of the United States Government.

Although this paper is not primarily concerned with conventional statistical methods and problems, some of the results suggest that under certain circumstances standard methods may not be appropriate for statistical inference in an organizational context. I shall give a primitive illustration at this point; for a more general discussion see Sections 6 to 8. Suppose that two people are each required to estimate a population mean m, from different but overlapping samples. Assume that person 1 observes ζ_1 and ζ_2 and that person 2 observes ζ_1 and ζ_3, where ζ_1, ζ_2, and ζ_3 are independent and identically distributed, with expected value m. Suppose further that the loss due to errors of estimation depends quadratically upon the errors of both persons, with an "interaction" between the errors, thus

$$(a_1 - m)^2 + (a_2 - m)^2 + 2q(a_1 - m)(a_2 - m),$$

where a_i is person i's estimate of m. It can be shown that the minimax estimates of m are

$$a_1 = [\zeta_1 + (1 + q)\zeta_2]/(2 + q), \qquad a_2 = [\zeta_1 + (1 + q)\zeta_3]/(2 + q)$$

(provided $q^2 < 1$). One also obtains these estimates by adopting the point of view of the Markoff theorem on least squares. Notice that these reduce to the standard estimates (sample means) only when the "interaction coefficient," q, is zero. Thus if errors in opposite directions are less costly than errors in the same direction ($q > 0$), then each person will give less weight to the common observation.

An important organizational problem is the determination of what statistical information shall be made available to the various decision makers in the organization. Implicit in the solution of such a problem is the determination of the best use that can be made of any given structure of information, i.e., the best decision functions. The results to be presented here are concerned with this latter problem. I have applied these results elsewhere [26] to the analysis of the evaluation of information in organizations.

In the present paper, certain team decision problems are analyzed from both the Bayes and the minimax points of view. Most of the paper is devoted to the case in which the payoff function is quadratic in the decision variables. The geometry of Hilbert space is helpful here in investigating the existence and uniqueness of Bayes decision functions, which can be interpreted as projections (Section 4).

If, in the quadratic payoff function, the coefficients of the quadratic terms are independent of the state of nature, then the situation is even more amenable to analysis (Section 5). It is shown that if the *a priori* distribution of the states of nature induces a normal distribution of all the information variables and of the coefficients in the linear terms of the payoff function, then the Bayes team decision function is linear in the information variables.

A team analogue of the Markoff problem of minimum variance unbiased estimation is solved (Sections 6 and 7), and the minimax properties of such solutions are investigated (Section 8).

Before going into the quadratic case, a general formulation of the team decision problem, and one general result on Bayes decision functions, are presented (Sections 2 and 3). A team decision function is called *person-by-person-optimal* if it cannot be improved by changing the decision function of any *one* person in the team. The above-mentioned result gives a sufficient condition for person-by-person-optimality to imply true optimality. (For a discussion of the case in which the payoff function is linear or piece-wise linear in the decision variables, and with linear constraints, see Radner [25], [24].)

From discussions with J. Marschak and with L. J. Savage, and from a study of Savage's recent book [27], I derived encouragement both to undertake and to continue this study, and help in arriving at the present formulation and in working out specific problems.

2. General formulation of the team decision problem. The team decision problem is concerned with a decision making unit (called here a *team*) that chooses a *decision* a from a set D and is thereupon rewarded according to its choice and the prevailing *state of the world* x. The decision variable a is actually an N-tuple $(a_1, \cdots, a_N)$ of *individual decision variables* a_i, with a_i in D_i, and the set D of possible values of the team decision variable is some subset of the Cartesian product of $D_1, \cdots, D_N$.

It will be assumed here that the reward is a real number $u(a, x)$ determined by a *payoff function* u. In a given situation the relevant state of the world may be the outcome of some random process, as in a prediction problem, or it may be a particular probability distribution of random events, as in an estimation problem, or it may be a combination of the two. Therefore let Z be a measurable space, whose class $\mathcal{Z}$ of measurable subsets is to be interpreted as the class of random events, and let P be a set of probability measures p on Z. The states of the world x will be represented by pairs (z, p), with z in Z and p in P.

Typically, the decision maker bases his decision upon some given information about the state of the world, according to some rule or decision function. This concept will be represented in this paper as follows. Assume that there are given a field $\mathcal{P}$ of subsets of P, and let $(X, \mathfrak{X})$ be the Cartesian product of the two measurable spaces, $(Z, \mathcal{Z})$ and $(P, \mathcal{P})$ (see Halmos [15], p. 140). For every $i = 1, \cdots, N$, let $\mathfrak{D}_i$ be a given field of subsets of D_i, and let $\mathcal{Y}_i$ be a given subfield of $\mathfrak{X}$; then any N-tuple $\alpha = (\alpha_1, \cdots, \alpha_N)$ such that, for each i, α_i is a $\mathcal{Y}_i$-measurable function from X to D_i will be called a *team decision function* based upon the N-tuple of *information subfields* $\mathcal{Y} = (\mathcal{Y}_1, \cdots, \mathcal{Y}_N)$. The functions α_i will be called *component decision functions.*

In some problems the decision functions available to the decision maker are, for external reasons, restricted to some subset of the set of functions measurable with respect to the information subfield, for example, by the requirement of linearity. In what follows, the set of decision functions available to the decision maker will be denoted by A.

A probably more familiar, but completely equivalent, way of representing information would be in terms of a transformation, that is, a measurable function from X to some other measurable space, say Y; component decision functions would then be measurable functions from Y_i to D_i. These two ways of representing information are equivalent, as has been said, and in this paper one or the other will be used, according to the nature of the particular problem being considered. For a discussion of the two sorts of representation, subfields and transformations, as well as the closely related concept of a "statistic," see Bahadur [1] and [2], and Bahadur and Lehmann [3].

For any given decision function α and any probability measure p in P, the expected payoff is

$$U(\alpha, p) = \int_Z u[\alpha(z, p), (z, p)]\, dp(z).$$

The problem for the team is to choose a decision function from the given set A that will make $U(\alpha, p)$ large in some sense. Two approaches, commonly known as *Bayes* and *minimax*, will be considered in this paper.

If G is a given *a priori* probability measure on $\mathcal{P}$, then the *Bayes payoff* for a decision function α is defined as

$$V(\alpha, G) = \int_P U(\alpha, p)\, dG(p), \qquad = \int_X u[\alpha(x), x]\, dF(x),$$

where $dF(x) = dp(z)\, dG(p)$; and α is a *Bayes decision function* if it maximizes the Bayes payoff on the set A.

Note that as soon as the Bayes approach is adopted, the distinction between Z and P ceases to have any real significance, for once a particular *a priori* measure G is given, the whole probability structure of the problem can be summarized in terms of a single measure F on X.

For any decision function α, and any p in P, the *risk* is defined as

$$\rho(\alpha, p) = \sup_{\alpha' \varepsilon A} U(\alpha', p) - U(\alpha, p).$$

Let $\mathcal{A}$ be a given field of subsets of A, and let A^* denote the set of all probability measures on $\mathcal{A}$; any such measure δ in A^* is called a *randomized decision function.* Extending the risk function to randomized decision functions in the obvious way, a randomized decision function $\hat{\delta}$ is called minimax in A^* if for all δ in A^*

$$\sup_{p \varepsilon P} \rho(\hat{\delta}, p) \leqq \sup_{p \varepsilon P} \rho(\delta, p).$$

It may be helpful to the reader to show how the usual one-person nonsequential statistical problem is represented within the framework just described. In this case, $N = 1$ and

(a) Z is the sample space;

(b) the single information subfield $\mathcal{Y}$ is the class of all sets $S \times P$, where S is in $\mathcal{Z}$, i.e., the true distribution of z is unknown;

(c) the payoff function depends upon x only through p; i.e., $u[a, (z, p)] = \tilde{u}[a, p]$. Because of (b), any decision function α can be represented by a measurable function $\tilde{\alpha}$ from Z to D. It is interesting that, if a function W is defined by

$$W(a, p) = \sup_{a' \varepsilon D} \tilde{u}(a', p) - \tilde{u}(a, p),$$

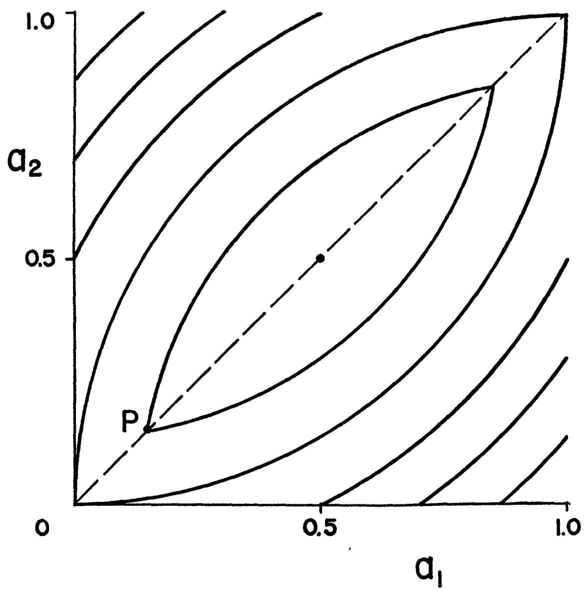
then the risk ρ is given by

$$\rho(\tilde{\alpha}, p) = \int_Z W(\tilde{\alpha}[z], p)\, dp(z),$$

as can easily be verified. In the statistical literature, the function W is often taken to be a datum of the decision problem (instead of being derived from a payoff function u, as was done here); in that case W is usually called the "loss" or "weight" function, and $\int_Z W(\tilde{\alpha}[z], p)\, dp(z)$ is called the "risk" function (see Wald [29], pp. 8, 10; Hodges and Lehmann [16], p. 182; Blackwell and Girshick [6], p. 82). Thus in this case the risk function ρ, as defined in this paper, does coincide with what is usually called risk. On the other hand, ρ is called the "loss" function by Savage [27], p. 163, and the "regret" function by Chernoff [9].

In this paper it will be assumed that for each i the set D_i of possible decisions for the ith team member is a Borel measurable subset of the real line. This assumption is not quite as special as it might at first seem, for a problem in which some D_i is a Borel measurable subset of M-dimensional Euclidean space could be recast in the present framework by replacing that person by M persons, all with the same information, each with a one-dimensional decision variable, and with their decision functions possibly subject to some joint constraint. Alternatively, such a D_i could be mapped onto a Borel set of the real line, the mapping being one-to-one and Borel-measurable both ways; this could be done even if the dimensions of D_i were infinite but denumerable (see Halmos [15], p. 159, exercise 7). Admittedly, although these devices achieve a certain technical generality for the present framework, in practice they might sometimes lead to unnecessary complexity and awkwardness. However, all of the results in this paper can also be reinterpreted directly in terms of vector decision variables.

3. Person-by-person maximization and stationarity in the Bayes problem. If $\hat{\alpha}$ is a Bayes team decision function relative to an *a priori* distribution, then the decision function $\hat{\alpha}_i$ for any one team member i must be best, given that every other member j uses the decision function $\hat{\alpha}_j$. Call a decision function α *person-by-person-maximal* if α cannot be improved by changing α_i for any one i alone. Thus any Bayes decision function is person-by-person maximal. The converse is not true, however, as the following example shows.



Consider a team of two members, whose payoff function is independent of x, with contour lines as in the accompanying figure; for example,

$$u(a_1, a_2) = \min\{-a_1^2 - (a_2 - 1)^2, -(a_1 - 1)^2 - a_2^2\}, \tag{3.1}$$

where each decision variable a_i is real. It is easily verified that any a for which $a_1 = a_2$ is person-by-person maximal, (e.g., point P in the figure) whereas the maximum of u is attained only at $a_1 = a_2 = \frac{1}{2}$. Note that u may be strictly concave in a, as is (3.1).

Suppose the decision functions of all but one of the team members are fixed; then the problem facing that one member becomes a one-person Bayesian problem, for the actions of the other members can then be considered as part of the "state of the world"; and he can therefore apply Bayes' rule. More precisely, let $\hat{\alpha}$ be person-by-person maximal, and suppose (with little, if any, real loss of generality) that for each i, conditional expectations given $\mathcal{Y}_i$ are *bona fide* expectations; then for every i and a.e. x, $\alpha_i(x)$ maximizes, with respect to a_i, the conditional expectation

$$\Phi_i(a_i, x) = E\{u[\hat{\alpha}_1(x), \cdots, a_i, \cdots, \hat{\alpha}_N(x), x] \mid \mathcal{Y}_i\}. \tag{3.2}$$

These N simultaneous conditions constitute the team analogue of the one-person Bayes rule, but as was just shown, they are, though necessary, not in general sufficient to determine the Bayes decision function, or functions.

If, in addition, $\Phi_i(a_i, x)$ is differentiable in a_i, then

$$(\partial/\partial a_i)\Phi_i(a_i, x) \mid_{a_i = \hat{\alpha}_i(x)} = 0.$$

Call a decision function α *stationary* if the Bayes payoff $V(\alpha) > -\infty$, and, with Φ_i defined as above,

$$(3.3) \qquad (\partial/\partial a_i)\Phi_i(a_i, x) \mid_{a_i=\alpha_i(x)} = 0 \text{ a.e.},$$

for every i. Thus, under suitable regularity conditions, a Bayes decision function is stationary. The following theorem gives one condition under which a stationary decision function is Bayes. This theorem will be applied in Sections 4 and 5.

The content of the theorem is essentially the following. Suppose that u is concave and differentiable in a for each x, that α is stationary, and that V satisfies certain conditions of finiteness. Then it is shown that for each δ such that $(\alpha + \delta)$ is not worthless, $V(\alpha + t\delta)$ is concave and differentiable in the real variable t, and its derivative vanishes at $t = 0$. Consequently, α maximizes V.

The Bayes payoff functional $V(\alpha) = V(\alpha, G)$ will be said to be *locally finite* at the decision function α if

1. $|(V\alpha)| < \infty$;

2. for any decision function δ such that $|V(\alpha + \delta)| < \infty$, there exist $k_1, \cdots, k_N$ all positive such that

$$|V(\alpha_1 + h_1\delta_1, \cdots, \alpha_N + h_N\delta_N)| < \infty$$

for all $h_1, \cdots, h_N$ for which $|h_1| \leqq k_1, \cdots, |h_N| \leqq k_N$.

In what follows it is assumed that the set A from which a decision function is to be chosen is actually the set of all $\alpha = (\alpha_1, \cdots, \alpha_N)$ such that α_i is $\mathcal{Y}_i$-measurable.

THEOREM 1. *If*

1. *$u(a, x)$ is concave (from below) and differentiable in a for* a.e. x,
2. $\sup_\beta V(\beta) < \infty$,
3. *V is locally finite at α,*
4. *α is stationary,*

then α is Bayes.

LEMMA. *If $f(c, x)$ is a concave function of the real variable c on the closed interval $[c', c'']$, for* a.e. x, *and $|Ef(c, x)| < \infty$ on $[c', c'']$, then*

$$(d^+/dc)Ef(c, x) \mid_{c=c'} = E(\partial^+/\partial c)f(c, x) \mid_{c=c'}$$

$$(d^-/dc)Ef(c, x) \mid_{c=c''} = E(\partial^-/\partial c)f(c, x) \mid_{c=c''},$$

where the superscripts $^+$ and $^-$ refer to differentiation from the right and left, respectively.

PROOF. The concavity of f implies that, for a.e. x, $(f[c, x] - f[c', x])/(c - c')$ is a nonincreasing function of c for $c' \leqq c \leqq c''$. Therefore, the first part of the lemma follows from the Lebesgue monotone convergence theorem (Halmos [15], Section 27, Theorem A); the second part follows by symmetry.

PROOF OF THE THEOREM. Suppose that $V(\alpha + \delta) > -\infty$. Let

$$f(k, x) \equiv u[\alpha_1(x) + k_1\delta_1(x), \cdots, \alpha_N(x) + k_N\delta_N(x), x],$$

and let $F(k) \equiv Ef(k, x)$. By the assumption of local finiteness F is finite in some neighborhood of 0, and, using the concavity of u, it is easy to verify that F is concave. Hence, for every i,

$$(\partial^+/\partial k_i)F(0) \quad \text{and} \quad (\partial^-/\partial k_i)F(0)$$

are finite. Hence, by the lemma, for every i,

$$(\partial^\pm/\partial k_i)F(0) = E(\partial^\pm/\partial k_i)f(0, x).$$

But it follows from assumption (1) that f is differentiable in k a.e., and hence

$$\begin{aligned}(\partial/\partial k_i)F(0) &= E(\partial/\partial k_i)f(0, x)\\ &= E\delta_i(x)(\partial/\partial a_i)u[\alpha(x), x],\end{aligned}$$

which last expectation is therefore finite.

On the other hand, by the lemma, applied to the condition of stationarity on α,

$$E\{(\partial/\partial a_i)u[\alpha(x), x] \mid \mathcal{Y}_i\} = 0; \tag{3.4}$$

hence $E(\partial/\partial a_i)u[\alpha(x), x]$ is finite, and therefore

$$\begin{aligned}E\delta_i(x)(\partial/\partial a_i)u[\alpha(x), x] &= E[E\{\delta_i(x)(\partial/\partial a_i)u[\alpha(x), x] \mid \mathcal{Y}_i\}]\\ &= E[\delta_i(x)E\{(\partial/\partial a_i)u[\alpha(x), x] \mid \mathcal{Y}_i\}]\\ &= 0.\end{aligned}$$

(See Doob [11], chapter I, Theorem 8.3.) Hence, for every i,

$$(\partial/\partial k_i)F(0) = 0.$$

Defining $k^0 = (1, \cdots, 1)$, it follows that $F(tk^0)$, as a function of the real variable t is differentiable at $t = 0$, and

$$(d/dt)F(tk^0)|_{t=0} = \sum_i (\partial/\partial k_i)F(k)|_{k=0} = 0 \tag{3.5}$$

(see Bonnesen and Fenchel [7], Section 13). Now $V(\alpha + t\delta) < \infty$ for all t (by Assumption 2), and is finite for $t = 0$ (Assumption 4) and for $t = 1$ (because it was assumed that $V(\alpha + \delta) > -\infty$.) Hence $V(\alpha + t\delta)$ is a finite, concave function of t for $0 \leqq t \leqq 1$. But $V(\alpha + t\delta) = F(tk^0)$, hence it follows from equation (3.5) that $V(\alpha + t\delta)$ has a minimum at $t = 0$, for $0 \leqq t \leqq 1$, which completes the proof of the theorem.

The following example shows that some condition like local finiteness is needed for the theorem just proved. Consider a team of two persons with the payoff function

$$u(a, x) = -\nu(x)(a_1 - a_2)^2 - 2a_1a_2,$$

where ν is any measurable function for which $\nu(x) > \frac{1}{2}$ (for concavity) and $E\nu(x) = \infty$; and suppose that only constant decision functions are allowed (i.e., $\mathcal{Y}_1 = \mathcal{Y}_2 = \{X\}$). Then the expected payoff is

$$V(a) = \begin{cases} -\infty, & \text{if} \quad a_1 \neq a_2 ; \\ -2a_1^2, & \text{if} \quad a_1 = a_2 . \end{cases}$$

Hence V is finite, but not locally finite at any a for which $a_1 = a_2$, and any such a is person-by-person maximal; but $a_1 = a_2 = 0$ is the single best decision.

4. The team with a quadratic payoff.

4.1. *Introduction.* This section will explore the consequence for the Bayes problem of assuming that for every state of the world x, the payoff is a quadratic function of the team decision, thus:

$$u(a, x) = \lambda(x) + 2a'\delta(x) - a'Q(x)a, \tag{4.1}$$

where a varies in N-dimensional Cartesian space R^N, and for every x, $\lambda(x)$ is real, $\delta(x)$ is in R^N and $Q(x)$ is an $N \times N$ symmetric matrix (λ, δ and Q all measurable). I want to consider only the situation in which, for a.e. x, $u(a, x)$ has a unique maximum in a; it will therefore be assumed that $Q(x)$ *is positive definite* for a.e. x.

It will be more convenient to speak in terms of loss (in a sense to be defined) rather than payoff. Maximizing (4.1), the one best team decision for any given typical x is

$$\gamma(x) \equiv Q^{-1}(x)\delta(x) \tag{4.2}$$

and the loss due to using any other decision a is

$$[a - \gamma(x)]'Q(x)[a - \gamma(x)]. \tag{4.3}$$

The expected loss, given the team decision function and the state of nature p, is

$$\sigma(\alpha, p) = E\{[\alpha(x) - \gamma(x)]'Q(x)[\alpha(x) - \gamma(x)] \mid p\}. \tag{4.4}$$

For any *a priori* distribution G, the Bayes expected loss, $E\sigma$, will be denoted by $\bar{\sigma}(\alpha, G)$, or sometimes just by $\bar{\sigma}(\alpha)$.

The expected loss function can be transformed to a certain extent without altering the problem. Roughly speaking, one can change the coordinate system in the decision space in any way that is compatible with the given structure of information. In particular, such a change could depend upon the state of the world, x. Let $T(x)$ be a measurable, $N \times N$-matrix-valued function of x that is nonsingular a.e., and such that for any decision function α, both $[T(x)\alpha(x)]_j$ and $[T^{-1}(x)\alpha(x)]_j$ are measurable-$\mathcal{Y}_j$; then the function β defined by $\beta(x) = T(x)\alpha(x)$ is a decision function if and only if α is, and the expected loss function

$$E[\beta(x) - T(x)\gamma(x)]'[T^{-1}(x)]'Q(x)T^{-1}(x)[\beta(x) - T(x)\gamma(x)]$$

defines a problem equivalent to that defined by (4.4).

For example, let the team members be divided into subgroups $I_1, \cdots, I_k$, and let $T(x)$ be a matrix with blocks $T_k(x)$ down the diagonal and zeros elsewhere, where $T_k(x)$ is a nonsingular matrix of order equal to the size of group

I_k, and measurable-$\bigcap_{k \epsilon I_i} \mathcal{Y}_k$. In particular, $T(x)$ can be taken to be any diagonal matrix such that, for every i, $t_{ii}(x)$ is measurable-$\mathcal{Y}_i$; typically, this will be the only type of T possible.

4.2. *Bayes decisions as projections—their existence and uniqueness.* A glance at equation (4.4) above shows that the team problem looks something like a prediction or estimation problem, with a quadratic form replacing the one-dimensional mean squared error. It is not surprising, therefore, that the Bayes solution can be described in terms of a projection in a suitable Hilbert space. Theorem 2 below describes the set of team decision functions that have finite Bayes expected loss. Theorem 3 proves that if any decision function has finite Bayes expected loss, then the best decision exists, is unique, and can be characterized as a projection.

In what follows it is to be understood that there is a given *a priori* probability measure on P, and that all expectations, probabilities, and references to "almost everywhere" are based on it. Any two functions on X that are equal a.e. will be considered equivalent, so that hereafter whenever any space of functions on X is introduced, it is to be understood that, strictly speaking, the object under discussion is really the corresponding quotient space modulo the space of functions that are zero a.e.

Let H be the space of all measurable functions α from X to R^N for which

$$E\alpha(x)'Q(x)\alpha(x) < \infty;$$

then, under the inner product

$$(\alpha, \beta) = E\alpha(x)'Q(x)\beta(x),$$

H is a (nontrivial) Hilbert space. This follows from the fact that the positive definite square root of $Q(x)$ induces a measurable isometry between H and the set of all measurable functions β from X to R^N for which $E\beta(x)'\beta(x) < \infty$; the details are routine and will be omitted.

In what follows let A be the set of all measurable α from X to R^N such that α_i is $\mathcal{Y}_i$-measurable. Also, denote the norm of an element β of H by $\|\beta\| = (\beta, \beta)^{\frac{1}{2}}$.

THEOREM 2. *For any measurable γ from X to R^N, the set F of α in A for which*

$$E[\alpha(x) - \gamma(x)]'Q(x)[\alpha(x) - \gamma(x)] < \infty$$

is either empty or it is the closed linear subvariety $A \cap (\gamma + H)$ of the complete linear variety $(\gamma + H)$ under the distance function

$$d(\alpha, \beta) = \|(\alpha - \gamma) - (\beta - \gamma)\|_H .$$

(For the special case, $N = 1$, of this theorem, see Girshick and Savage [12], Theorem 2.2.)

PROOF. Suppose that F is not empty. It follows that $F = A \cap (\gamma + H)$ and that $(\gamma + H)$ is complete under the given distance function; it remains to show that F is closed. Let α^0 be any element of F; then $(\gamma + H) = (\alpha^0 + H)$. The transformation that takes any β in $(\gamma + H)$ into $(\beta - \alpha^0)$ in H is an isometry from $(\gamma + H)$ onto H, and the image of F under this transformation is $A \cap H$. Therefore, F is closed if and only if $A \cap H$ is closed. Suppose that the sequence $\{\alpha^{(n)}\}$ in $A \cap H$ converges, in the norm of H, to α, which is therefore in H. By repeated application of Theorem 22.D (together with 25.A and 21.B) of Halmos [15], one can show that there is a subsequence $\{\alpha^{(k)}\}$ converging to α *coordinatewise* and pointwise a.e.; hence each α_i is $\mathcal{Y}_i$-measurable (or more exactly, has a $\mathcal{Y}_i$-measurable representation), which puts α in A.

An an example of a case in which F is empty, let x be uniformly distributed on the unit interval, let $N = 1$, $q(x) = 1/x^4$, $\gamma(x) = x$, and let A contain only constant functions. It is easy to verify that there is no real number a such that

$$\int_0^1 \frac{(a - x)^2}{x^4}\, dx < \infty.$$

The Bayes problem is one of finding an element of F that is closest to γ, in the sense of the distance d. Because of the isometry between $(\gamma + H)$ and H, this problem is equivalent to the problem of finding an element of $A \cap H$ that is closest to $(\gamma - \alpha^0)$ in the sense of the norm of H (where α^0 is as in the proof of Theorem 2). Therefore, if F is not empty, *it can be assumed without loss of generality that γ is in H*, and this will be done from now on. It then follows, of course, that $F = A \cap H$.

THEOREM 3. *If F is not empty, then there is a unique team decision function α that minimizes the Bayes expected loss $\bar{\sigma}(\alpha) = \|\alpha - \gamma\|$ on F, and α is the orthogonal projection of γ into F.*

PROOF. Immediate from Theorem 1 and the minimizing property of the orthogonal projection (see Halmos [14], Theorems 11.1 and 11.2).

4.3. *The condition of stationarity for the quadratic case.* In this section it is shown that, under a certain condition on the random matrix $Q(x)$, the hypothesis of Theorem 1 is satisfied, and therefore that a stationary decision function is Bayes.

Let $r(x)$ be the lower bound of $Q(x)$ with respect to the quadratic form $\sum_i q_{ii}(x)\, a_i^2$, i.e.,

$$r(x) = \min_a [a'Q(x)a]/[\sum_i q_{ii}(x) a_i^2];$$

and let r be the essential infimum of $r(x)$.

THEOREM 4. *If $r > 0$, and if α is stationary, then α is Bayes.* (Note that the hypothesis of this theorem is automatically satisfied if $Q(x)$ is a constant matrix, independent of x.)

PROOF. The present theorem will be proved if it can be shown that the hypothesis of Theorem 1 is satisfied. In interpreting Theorem 1, the reader should keep in mind that he is now concerned with expected loss. The only point not immediately obvious is that the expected loss $\bar{\sigma}$ is locally finite at α, which will now be demonstrated.

As already shown, there is no loss of generality in assuming that $\|\gamma\|^2 = E\gamma' Q\gamma < \infty$, and hence that the Bayes expected loss $\bar{\sigma}(\alpha) < \infty$ if and only if $\|\alpha\| < \infty$. Suppose, then, that $\|\alpha\| < \infty$ and $\|\alpha + \delta\| < \infty$; it follows by Theorem 2 that $\|\delta\| < \infty$. For a.e. x,

$$r(x)\sum_i q_{ii}(x)\delta_i^2(x) \leqq \delta(x)'Q(x)\delta(x).$$

Hence

$$E\sum_i q_{ii}(x)\delta_i^2(x) \leqq (1/r)\|\delta\|^2,$$

and thus for every i,

$$Eq_{ii}(x)\delta_i^2(x) < \infty.$$

Let $\delta^{(i)} = (0, \cdots, \delta_i, 0, \cdots, 0)$; then for every i, $\|\delta^{(i)}\| < \infty$, and hence by Theorem 2, $\|\alpha + \sum_i k_i\delta^{(i)}\| < \infty$, for all real $k_1, \cdots, k_N$, which concludes the proof.

In the quadratic case with $r > 0$ the condition for stationarity becomes (see equation (3.4))

$$E\{\sum_j q_{ij}(x)[\alpha_j(x) - \gamma_j(x)] \mid \mathcal{Y}_i\} = 0 \tag{4.5}$$

for every i and a.e. x, since Theorem 1 applies.

Under these conditions the equations for stationarity determine, of course, the orthogonal projection of γ into F. It can be shown that under these conditions F is a Cartesian product $\prod_i F_i$, where F_i is the space of all $\mathcal{Y}_i$-measurable functions β_i such that $Eq_{ii}(x)\beta_i^2(x)$ is finite. Note that the example at the end of Section 3 does not satisfy these conditions. There, the matrix $Q(x)$ is

$$\begin{bmatrix} q(x) & 1 - q(x) \\ 1 - q(x) & q(x) \end{bmatrix},$$

and $r(x) = 1/q(x)$, so that $r = $ r. On the other hand, Theorem 3 still applies, with F the set of all a such that $a_1 = a_2$.

5. The case of constant coefficients of the quadratic terms. It will now be assumed that the matrix Q is independent of x, i.e., is constant and known. A number of detailed results can be derived in this case, but the reader should keep in mind that this assumption represents an important loss of generality. As thus far discussed, the general quadratic payoff might be thought of as an approximation, for each x, to an arbitrary smooth payoff function in the neighborhood of the best team action $\gamma(x)$ corresponding to x. If Q is constant, the approximated payoff function in the neighborhood of $\gamma(x)$ is the same (or at least almost always the same) for each x, which is clearly a most special circumstance.

One who is familiar with the theory of minimum variance estimation and prediction might guess that the normal distribution would have a special place in the theory of the team with a quadratic payoff function. Such a guess would be correct, as will be shown in the present section. The main result here is that if the *a priori* distribution induces a normal distribution of all the information variables and the vector $\gamma(x)$, then the Bayes decision function is linear in the information variables. An explicit algorithm is given.

It will be more convenient to use the transformation, as opposed to the subfield, terminology in the next four sections.

For each $i = 1, \cdots, N$, let η_i be a measurable function from X to K_i dimensional Cartesian space R^{K_i}; $\eta_i(x)$ will represent the information that becomes available to the ith team member when x is the state of the world. In this section it will be assumed that, under the given *a priori* distribution, $\eta_1, \cdots, \eta_N$ and γ all have a joint normal distribution (see equations (4.1) and (4.2)). Let C_{ij} denote the matrix of covariances between the coordinates of η_i and the coordinates of η_j. There is no loss of generality in assuming that for every i, C_{ii} is the identity and η_i has mean zero. In summary,

$$\text{(5.1)} \qquad \text{Cov}(\eta_i, \eta_j) = C_{ij}, C_{ii} = I_{K_i}, E\eta_i = 0.$$

Let $\delta(x) = Q\gamma(x)$ (see equation (4.2)); the regression of δ on η_i is linear, i.e.,

$$\text{(5.2)} \qquad E[\delta_i \mid \eta_i(x) = y_i] = E\delta_i + d_i'y_i,$$

for some vector d_i.

THEOREM 5. *If, under the given a priori distribution, γ and the information functions $\eta_1, \cdots, \eta_N$ have a joint normal distribution, with parameters given by* (5.1) *and* (5.2), *then the components of the unique Bayes team decision function are linear,*

$$\text{(5.3)} \qquad \alpha_i[\eta_i(x)] = b_i'\eta_i(x) + c_i,$$

where the vectors b_i and the numbers c_i are determined by the systems of linear equations

$$\text{(5.4)} \qquad \sum_j q_{ij}C_{ij}b_j = d_i, \qquad i = 1, \cdots, N;$$

$$\text{(5.5)} \qquad \sum_j q_{ij}c_j = E\delta_i, \qquad i = 1, \cdots, N.$$

PROOF. The existence and uniqueness of the Bayes team decision function follows from Theorem 3; and by Theorem 4 it must satisfy the conditions of stationarity, i.e.,

$$\text{(5.6)} \qquad \sum_j q_{ij}E[\alpha_j \mid \eta_i(x) = y_i] = E[\delta_i \mid \eta_i(x) = y_i]$$

for all i, and all y_i in R^{K_i}. If the Bayes team decision function is to be linear, of the form (5.3), then (5.6), (5.1) and (5.2) imply that

$$\sum_j q_{ij}(b_j'C_{ji}y_i + c_j) = E\delta_i + d_i'y_i,$$

for all i and all y_i in R^{K_i}. From this it follows that if the Bayes team decision function is linear, then the b_i and c_i must satisfy (5.4) and (5.5). Hence the proof will be complete when it is shown that (5.4) and (5.5) have solutions. The solvability of (5.5) follows from the fact that Q is positive definite, and that of (5.4) from the following lemma.

LEMMA. *If C is a $K \times K$ symmetric nonnegative semi-definite matrix, partitioned symmetrically into blocks C_{ij}, such that C_{ii} is positive definite for every i; and if Q is an $N \times N$ symmetric positive definite matrix with elements q_{ij}; then the matrix H composed of blocks $q_{ij}C_{ij}$ is positive definite.*

(A special case of this lemma is the known result that the Hadamard product $((q_{ij}c_{ij}))$, of two positive definite matrices, $((q_{ij}))$ and $((c_{ij}))$, is positive definite. (See Halmos [13], Theorem 2, Section 85.) In this paper the matrix H of the lemma will be called the Hadamard product of Q and C (corresponding to the given partitioning of C).

PROOF OF LEMMA. Since C is nonnegative, it can be expressed as

$$C = \sum_p r(p)r'(p),$$

where for each p, $r(\text{p})$ is a vector in K-dimensional Cartesian space. For any K-vector v, let $\{v_i\}$ be a partitioning of v into subvectors, corresponding to the partitioning of C; then for every i and j

$$C_{ij} = \sum_p r_i(p)r_j'(p).$$

For any v

$$v'Hv = \sum_{ij} q_{ij}v_i'C_{ij}v_j = \sum_p \sum_{ij} q_{ij}v_i'r_i(p)r_j'(p)v_j = \sum_p \sum_{ij} q_{ij}w_i(p)w_j(p), \tag{5.7}$$

where $w_i(p) = v_i'r_i(p)$. Hence $v'Hv \geqq 0$ for all v. Now let $v \neq 0$; then for some i, $v_i \neq 0$. For that i, and for some p, $v_i'r_i(p) \neq 0$, because C_{ii} is positive definite. Hence from (5.7), $v'Hv > 0$ if $v \neq 0$, which completes the proof.

EXAMPLE 1. Let x_1 and x_2 be random variables with a joint normal distribution, with means zero, variances one, and correlation ρ. Suppose that the team has two members, that member i observes x_i, that member i's decision variable is a_i, and that the loss function is

$$a_1^2 + 2qa_1a_2 + a_2^2 - 2a_1x_1 - 2a_2x_2 + \text{a constant},$$

where $q^2 < 1$. In the notation of the theorem just presented, $\delta'(x) = (x_1, x_2)$, $Q = \begin{pmatrix} 1 & q \\ q & 1 \end{pmatrix}$, $C_{ii} = 1$, $C_{ij} = \rho$, $d_i = 1$, $E\delta_i = 0$. The Bayes decision function is

$$\alpha_i(x_i) = [x_i/(1 + q\rho)], \qquad i = 1, 2. \tag{5.8}$$

Note that

$$\gamma_i(x) = (x_i - qx_j)/(1 - q^2),$$

and

$$E(\gamma_i \mid x_i) = [(1 - q\rho)/(1 - q^2)]x_i,$$

so that $\alpha_i(x_i) = E(\gamma_i \mid x_i)$ only if $q = 0$ or $\rho = 1$.

One general way in which the conditions of the previous theorem can be fulfilled is as follows. Let $z_1, \cdots, z_N$ be N random vectors with a joint normal distribution, with known covariances, but with unknown means. However, letting z denote the combined vector $(z_1, \cdots, z_N)$ and K the dimension of z, suppose that the mean of z is known to lie in a linear subspace M of K-dimensional Cartesian space Z. In other words, the set P of probability measures p on Z is the set of all normal distributions on Z that have the given covariance structure and such that EZ is in M. The set P can therefore be represented by M, and X can be represented by the external direct sum $M + Z$. If the *a priori* distribution on M is itself normal, this induces a normal distribution on X.

EXAMPLE 2. Suppose that the team as a whole has available to it a random sample of k observations from an N-variate normal distribution with unknown mean μ and known covaraince σ_{ij}, but that for every i, member i knows only the sample values of the ith coordinate, and on the basis of those values needs to estimate the corresponding mean μ_i. Suppose further that the *a priori* distribution specifies that $\mu_1, \cdots, \mu_N$ are independent and normally distributed with means all zero and variances $\tau_1^2, \cdots, \tau_N^2$, respectively.

If z_i denotes the vector of k sample values of the ith coordinate, then in the notation of the (quadratic) team decision problem

$$x = (z_1, \cdots, z_N, \mu_1, \cdots, \mu_N), \ \eta_i(x) = z_i, \gamma_i(x) = \mu_i.$$

A straightforward calculation yields

$$\begin{aligned} Ez_i &= E\mu_i = 0 \\ E(\mu_i \mid z_i) &= e_i\bar{z}_i, \ E(\delta_i \mid z_i) = q_{ii}\bar{z}_i, \\ E(z_j \mid z_i) &= (\sigma_{ij}/\sigma_{ii})(z_i - e_i\bar{z}_i f) \end{aligned} \tag{5.9}$$

where

$$\begin{aligned} e_i &= [1 + (\sigma_{ii}/k\tau_i^2)]^{-1} \\ f' &= (1, 1, \cdots, 1) \ (k\text{-dimensional}) \\ \bar{z}_i &= (1/k)f'z_i \ (\text{the } i\text{th person's sample mean}). \end{aligned}$$

Rather than normalize the information variables so as to be able to apply Theorem 5 directly, it is easier to write down the stationarity condition for this example, knowing that the Bayes decision function is linear. If $\alpha_i(z_i) = b_i'z_i$, the stationarity condition (applying (5.9)) is

$$q_{ii}b_i'z_i + \sum_{j \neq i} q_{ij}b_j'(\sigma_{ij}/\sigma_{ii})(z_i - e_i\bar{z}_i f) = e_i\bar{z}_i, \qquad i = 1, \cdots. N. \tag{5.10}$$

It is easily verified that (5.10) is satisfied by

$$\begin{aligned} b_i &= (b_i^0/k)f, \\ \text{or} \quad b_i'z_i &= b_i^0\bar{z}_i, \end{aligned} \tag{5.11}$$

where the numbers b_i^0 are determined by the system

$$b_i^0 + [(1 - e_i)/q_{ii}\sigma_{ii}] \sum_{j \neq i} q_{ij}\sigma_{ij}b_j^0 = e_i, \qquad i = 1, \cdots, N. \tag{5.12}$$

In other words, the ith person's estimate of μ_i is proportional to his sample mean, but is not in general equal to it. It is interesting, however, that as the *a priori* information becomes more and more "vague," i.e., as $\tau_1^2, \cdots, \tau_N^2$ increase without limit, the numbers e_i approach 1 (see (5.9)), and therefore so do the coefficients b_i^0 (see (5.1)). The same effect is produced when the sample size k gets large.

It is *not* in general true that as the *a priori* distribution becomes more and more "vague," the Bayes team decision functions tend to a function independent of the matrix Q; this question will come up again in the next section.

6. Markoff decision functions. This section and the next deal with a problem that is the team analogue of the one-person problem of minimum-mean-square-error linear unbiased estimation, or the "Markoff problem." In the situation to be considered, each team member observes the value of a different random vector. The covariance structure in each vector space is known, as are the covariances between the vectors of different members, but the mean of the N-tuple of vectors is known only to be in a certain linear subspace of the direct sum of the N vector spaces. Each team member wants to estimate a given linear functional of the mean of his vector, the loss function for the team as a whole being a given quadratic form (determined by Q) in the errors of the estimates. Suppose further that the team wants to use only estimators for which the expected loss is bounded as a function of the mean, and finally, suppose that the team members want to keep their estimators simple and therefore restrict themselves to linear estimators. A Markoff estimator for the team is one that minimizes the expected loss for all possible values of the mean, subject to the two conditions of bounded expected loss and linearity. It is, of course, a very special circumstance that permits a single argument to minimize the expected loss for all values of a parameter at once. One result of this section is that the Markoff problem for the team is, in a certain sense, equivalent to an ordinary, one-person, Markoff problem, involving all the vectors together, in which the covariances between the vectors are weighted by the corresponding elements of the matrix Q; in particular, the Markoff problem for the team typically has a solution.

The requirement of bounded expected loss is close to the minimax principle in spirit, and, in fact, the next section shows that Markoff estimators are indeed minimax under certain conditions.

Actually, the interpretation of this problem as one of estimation is only suggestive. What *is* essential is that each team member's best decision under complete information, $\gamma_i(x)$, should be a linear function of the mean of his information vector $\eta_i(x)$.

Let the space Z of random events (see Section 2) be the direct sum of N real, finite dimensional vector spaces $Z_1, \cdots, Z_N$; let M be a given linear subspace of Z, and let P be a set of probability measures p on Z. Let $\zeta = (\zeta_1, \cdots, \zeta_N)$ denote the random vector in Z (with ζ_i in Z_i), and assume that:

1. For every p in P, $E(\zeta \mid p)$ is in M. For every m in M there is a p in P such that $E(\zeta \mid p) = m$.
2. All p's in P have the same matrix C of covariances between the coordinates of ζ (relative to some fixed coordinate systems in $Z_1, \cdots, Z_N$).

The matrix C can be partitioned into blocks C_{ij}, where C_{ij} is the matrix of covariances between the coordinates of ζ_i and the coordinates of ζ_j. There is no loss of generality in assuming that:

3. For every i, C_{ii} is nonsingular.

(If C_{ii} were singular, Z_i could be replaced by a linear subspace of itself, for which the corresponding covariance matrix would be nonsingular.)

Assume further that the team loss function is quadratic, of the form

$$[a - \gamma(x)]'Q[a - \gamma(x)], \tag{6.1}$$

and that the function γ is an N-tuple of linear functionals γ_i of $E(\zeta_i \mid p)$, i.e., for each i,

$$\gamma_i(x) = g_i'E(\zeta_i \mid p), \tag{6.2}$$

for some g_i in Z_i.

Finally, suppose that the set of allowable team decision functions A is the set of all α such that α_i is a (possibly nonhomogeneous) linear functional on Z_i, i.e., for every α in A,

$$\alpha_i(z_i) = b_{i0} + b_i'z_i, \tag{6.3}$$

for some number b_{i0} and some vector b_i in Z_i.

For any such team decision function α, the expected loss is

$$E \sum_{ij} q_{ij}[b_{i0} + b_i'\zeta_i - g_i'E\zeta_i][b_{j0} + b_j'\zeta_j - g_j'E\zeta_j] = \sum_{ij} q_{ij}b_i'C_{ij}b_j + \sum_{ij} q_{ij}[b_{i0} + (b_i - g_i)'E\zeta_i][b_{j0} + (b_j - g_j)'E\zeta_j].$$

The minimum expected loss is zero, for a given p in P. Hence expected loss is identical with *risk* as defined in Section 2.

The function α will be called a *bounded-risk decision function* if the above risk is bounded as $E\zeta$ varies in M. The Markoff problem is to choose, if possible, a single bounded-expected-loss decision function that minimizes the risk for all $E\zeta$ in M.

The idea of replacing the familiar constraint of unbiasedness with the equivalent (in this problem) but intuitively more reasonable constraint of bounded risk seems to be due to L. J. Savage.

For every i, let M_i be the set of all z_i in Z_i such that for some

$$m = (m_1, \cdots, m_N)$$

in M, $z_i = m_i$. It is clear that M_i is a linear subspace of Z_i. Let $M_i^{\perp}$ denote the orthogonal complement of M_i in Z_i; then α is a bounded-risk decision function if and only if, for every i,

$$(b_i - g_i) \quad \text{is in} \quad M_i^{\perp}, \tag{6.5}$$

as can be easily verified.

From this it follows that no linear bounded-risk decision function is admissible unless $b_{i0} = 0$ for all i; and whenever a bounded-risk decision function is referred to from now on it is to be understood that it is homogeneous. The risk for such a decision function is therefore

$$\sum_{ij} q_{ij} b_i' C_{ij} b_j \equiv b'Hb,$$

where H denotes the Hadamard product of Q and C (see Section 5), and $b' = (b_1', \cdots, b_N')$.

Let $\check{M}$ be the internal direct sum $\sum_i M_i$ of the subspaces M_i, let $\check{M}^{\perp}$ denote the orthogonal complement of $\check{M}$ in Z, and let $g' = (g_1', \cdots, g_N')$; *then* $\hat{\alpha}$ *is a solution of the Markoff problem if and only if* $\hat{\alpha}(z) = \hat{b}'z$, *where* $\hat{b}$ *minimizes* $b'Hb$ *subject to* $(b - g)$ *in* $\check{M}^{\perp}$.

It is clear that $\hat{\alpha}$ solves the ordinary (one-person) Markoff problem of finding a minimum variance linear unbiased estimator of $g'E(\zeta \mid p)$, *where* ζ *has the covariance matrix* H, and $E(\zeta \mid p)$ is known to lie in $\check{M}$. Let $\tilde{M}$ be the H-orthogonal complement of $\check{M}^{\perp}$ in Z. As is well known (see for example Scheffé [28], pp. 19–21) the solution of this last problem is given by

$$\hat{b} = Pg, \tag{6.6}$$

where P is the H-orthogonal projection onto $\tilde{M}$. Note that P is not affected by multiplying either Q or C by a positive constant.

EXAMPLE. Suppose there are only two team members. Let ζ_1, ζ_2, and ζ_3 be independent and identically distributed random variables with unknown mean m, and variance 1, and suppose that member one observes $\eta_1 = (\zeta_1, \zeta_2)$, and member two observes $\eta_2 = (\zeta_1, \zeta_3)$. The covariance matrix of η_1 and η_2 together is therefore

$$\left[\begin{array}{cc|cc} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ \hline 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right]$$

Suppose that $Q = \begin{pmatrix} 1 & q \\ q & 1 \end{pmatrix}$, and that each member needs to estimate m. The Markoff decision function $b = (b_{11}, b_{12}; b_{21}, b_{22})$ therefore minimizes

$$b'Hb = \sum_{ih} b_{ih}^2 + 2qb_{11}b_{21}$$

subject to the constraints: $b_{11} + b_{12} = b_{21} + b_{22} = 1$. The solution is easily found to be:

$$b_{i1} = 1/(2 + q)$$
$$i = 1, 2.$$
$$b_{i2} = (1 + q)/(2 + q)$$

Note that if $q = 0$, the best estimator for member i is his sample mean $\bar{\zeta}_i$. If $q > 0$, then the loss is greater the higher the (positive) correlation between the errors of the two members; hence less weight is given to the variable that is observed by both.

The case of M defined in terms of a parameter space. In many problems the set of possible values of the mean $E\zeta$ is defined as the image, under a linear transformation, of a "parameter space." Thus, let V be a vector space; for each i, let B_i be a linear transformation from V to Z_i, and suppose that for every possible value of $E\zeta$ there is a v in V such that, for every i,

$$E\zeta_i = B_i v.$$

Suppose, further, that each γ_i is a linear functional on V.

By examining the risk function, as before, one can easily show that there exists a linear bounded-risk decision function if and only if, for each i, there exists a vector g_i in Z_i such that for all v in V

$$g_i' B_i v = \gamma_i(v);$$

i.e.,

$$\gamma_i(v) = g_i' E(\zeta_i \mid v).$$

In other words, there exists a linear bounded-risk decision function if and only if the quantity to be "estimated" by the ith person depends upon the mean of the random vector ζ only through the mean of his own information vector.

Note that the set M of possible values of $E\zeta$ is a linear subspace of Z, and that $B_i(V)$ corresponds to the subspace M_i of the previous formulation.

7. Decomposable Markoff problems. In the (one-person) theory of minimum variance unbiased linear estimation, the Markoff estimator of a linear functional of the mean of the observed vector can be represented as that same linear functional of a projection of the observation into the linear subspace in which the mean is known to lie. This fact gives one meaning to the statement that this projection is the best (vector-valued) estimator of the mean itself.

In the team Markoff problem, this situation does not generally hold for the individual team members; when it does, the problem will be called *decomposable*. This section describes two situations in which decomposability occurs.

Formally, a team Markoff problem is called decomposable if, for every i, there exists a transformation S_i from Z_i into itself, such that, for all g in Z, the best decision function for member i (best "estimate" of $g_i' E[\zeta_i \mid p]$) is $g_i' S_i \zeta_i$. From

(6.6) it follows that a Markoff problem is decomposable if and only if, for every g in Z

$$(7.1) \qquad (Pg)' = (g_1'S_1, \cdots, g_N'S_N),$$

where P is the H-orthogonal projection onto $\tilde{M}$, i.e., if and only if P is a matrix with blocks S_i' down the diagonal, and zeros elsewhere. (Recall that $\tilde{M}$ is the H-orthogonal complement of $(\sum M_i)^{\perp}$.)

EXAMPLE 1. Suppose that in the matrix H, which is composed of blocks $q_{ij}C_{ij}$ (see the previous section), all the off-diagonal blocks ($i \neq j$) are zero. There is no loss of generality in assuming that each C_{ii} is in identity matrix. In this case, for every x and y in Z,

$$(7.2) \qquad x'Hy = \sum_i q_{ii}x_i'y_i .$$

It follows from (7.2) that x and y are H-orthogonal if and only if they are orthogonal ($x'y = 0$), and hence equation (7.1) is satisfied by taking S_i' to be the orthogonal projection, in Z_i, onto M_i.

In particular, the assumptions of this example will be satisfied if either the observations of different team members are uncorrelated, or the matrix Q is diagonal (no "interaction").

EXAMPLE 2. Suppose that all the spaces Z_i have the same dimension, K; suppose that for every i, $M_i = M_0$, where M_0 is a linear subspace of K dimensional Cartesian space; and suppose that for every i and j, $C_{ij} = \sigma_{ij}I_K$, where I_K is the identity matrix of order K. A special case of this example is suggested by example 2, Section 5, in which (1) the team as a whole has available to it a random sample of K observations from an N-variate normal distribution with unknown mean μ and known covariances σ_{ij}; (2) for every i, member i knows only the sample values of the ith coordinate; and (3) on the basis of those values needs to estimate the corresponding mean μ_i.

For any x and y in Z, by the assumptions of this example,

$$xHy = \sum_{ij} q_{ij}\sigma_{ij}x_i'y_j .$$

From this it quickly follows that (7.1) is satisfied by taking, for each i, S_i' to be the (ordinary) orthogonal projection, in K-space, onto the subspace M_i.

In particular, for the special case mentioned above, the Markoff estimate of μ_i is the sample mean of the observed values of the ith coordinate.

8. Minimax properties of Markoff decision functions. In the decision situation described in Section 6 it can be shown that the Markoff decision function is minimax in the set of *all* team decision functions, provided the set P of probability distributions is sufficiently rich in normal distributions (in a sense made precise below). More generally, suppose that the covariance matrix is known only to be one of some given class of such matrices; then, for certain such classes, there exists a Markoff decision function, relative to one of the covariance matrices in the class, that is minimax in the set of all team decision functions. Three cases of this kind will be considered in this section.

In the first case it is assumed that the covariance matrix is known to be a positive multiple of cC of some given matrix C, and that the factor c is known to be bounded. In this case the Markoff decision function relative to C is minimax.

In the second case it is assumed that the covariance between random variables observed by the same person are known, but that those between variables observed by different persons are unknown; here it is shown that there exists a Markoff decision function that is minimax, but the corresponding (least favorable) covariance matrix is not described explicitly.

In the third case, it is assumed that the covariance matrix is only known to be bounded relative to some given matrix C^0. In this case the Markoff decision function relative to C^0 is minimax. This might cause some uneasiness about the application of the minimax principle in this case. In such a context one typically would be willing to grant that the class of covariance matrices is bounded somehow, but would not have a very precise idea of the nature of the bound, whereas in this result the minimax decision function is quite sensitive to the form of the bound, as represented by C^0.

For related results in the one-person case see Wald [29], p. 142; Hodges and Lehmann [16], Theorem 6.5; Girshick and Savage [12]; and Radner [23].

As in the previous two sections, let Z be the direct sum of N real finite dimensional Cartesian spaces Z_i, and let M be a linear subspace of Z. Further, let $\mathcal{C}$ denote a set of symmetric nonnegative semi-definite matrices of order the dimension of Z; and for each C in $\mathcal{C}$, let $((C_{ij}))$ denote the partitioning into blocks corresponding to the direct summands Z_i.

Given $\mathcal{C}$, let P be a set of probability measures p on Z such that

1. For every p in P, the mean is in M and the covariance matrix is in $\mathcal{C}$.
2. For every m in M and C in $\mathcal{C}$ there is a *normal* probability measure p in P with mean m and covariance matrix C.

The assumption that P includes normal distributions is a natural one, since normality can rarely be ruled out as preposterous. (The form of the assumption given in (2) is, strictly speaking, slightly stronger than necessary for the purpose of this section, but has been put in that form for the sake of simplicity. See the remark after Lemma 2 below.)

Let the (quadratic) loss function be given by (4.3), with constant matrix Q, and let the set of allowable team decision functions be the set of *all* $\alpha = (\alpha_1, \cdots, \alpha_N)$ such that α_i is a real-valued measurable function on Z_i. The entire situation just described will be called here a *generalized Markoff situation.*

The main tool of this section is Lemma 2 below. Basic to this lemma is Lemma 1, stating that if the covariance matrix is known, then the corresponding Markoff decision function is minimax. It should be noted that, because of the convexity of the loss function, it is sufficient to consider only nonrandomized estimators (see [16], Theorem 3.2).

(Throughout the rest of this section, it is to be understood that all matrices are of order equal to the dimension of Z, unless explicit mention is made to the contrary.)

LEMMA 1. *If $\mathfrak{C}$ consists of a single matrix C such that C_{ii} is positive definite for every i, and if P contains only normal distributions, then the Markoff team decision function relative to C is minimax in the set of all team decision functions.*

PROOF. The proof consists in showing that the risk for the Markoff decision function, which is constant on P, is the limit of a sequence of Bayes risks, for some suitable sequence of *a priori* distributions on P. Adopting the "parameter space" approach (see the end of Section 6), let V be a Cartesian space of the same dimension as M, let B be the matrix of a nonsingular transformation from V to M, and set up a 1-1 correspondence v_p between V and P such that the mean corresponding to p is Bv_p. Consider a sequence of normal probability measures on V, each with mean zero, such that under the rth measure ($r = 1, 2$, etc. $\cdots$), the coordinates of v are independent, with variance r. Each of these *a priori* distributions on V induces a normal distribution on Z; therefore, by Theorem 5, Section 5, the corresponding Bayes team decision function is linear and homogeneous. Given r, the Bayes risk for the decision function $\alpha(z) = b'z$ is

$$\sum_{ij} q_{ij} b_i' C_{ij} b_j + r \sum_{ij} q_{ij}(b_i - g_i)' B_i B_j' (b_j - g_j), \tag{8.1}$$

where B_i are the blocks of B corresponding to the summands Z_i. As r increases without limit, the minimum of (8.1) with respect to b approaches the minimum of $\sum_{ij} q_{ij} b_i' C_{ij} b_j$ subject to the constraint

$$\sum_{ij} q_{ij}(b_i - g_i)' B_i B_j' (b_i - g_i) = 0. \tag{8.2}$$

This last condition is satisfied if and only if, for every i, $B_1'(b_i - g_i) = 0$. The reader can easily verify that this last is, in turn, equivalent to the condition,

$$(b_i - g_i) \quad \text{in} \quad M_i^{\perp}, \qquad i = 1, \cdots, N, \tag{8.3}$$

where $M_i^{\perp}$ is defined as in Section 6, completing the proof of the lemma.

For any (homogeneous) bounded-risk linear decision function, $\alpha(z) = b'z$ (and in particular for any Markoff decision function) the risk depends upon p in P only through the corresponding covariance matrix C in $\mathfrak{C}$, and is in fact equal to

$$b' H_c b \equiv \sum_{ij} q_{ij} b_i' C_{ij} b_j ,$$

where H_c denotes the Hadamard product of C and Q. Therefore the risk for such a decision function can simply be denoted by $\rho(b, C)$. Further, when reference is made to "the decision function b," it is to be understood as referring to the linear function, $\alpha(z) = b'z$.

The next lemma follows immediately from Lemma 1.

LEMMA 2. *In any generalized Markoff situation, if $\hat{b}$ is Markoff relative to $\hat{C}$ in $\mathfrak{C}$, and if $\rho(\hat{b}, C) \leqq \rho(\hat{b}, \hat{C})$ for all C in $\mathfrak{C}$, then $\hat{b}$ is minimax in the set of all team decision functions.*

According to Lemma 2, there is a Markoff decision function that is minimax if the risk function $\rho(b, C)$ has a saddle point (minimax) with $(b - g)$ in $\sum M_i^{\perp}$ and C in $\mathcal{C}$.

Note: For the proof of Lemma 2, condition (2) in the definition of a generalized Markoff situation could be replaced by the assumption that for every m in M there is a normal distribution in P with covariance matrix $\hat{C}$.

If $\hat{b}$ is minimax, and Markoff relative to $\hat{C}$, as in Lemma 2, then although there may be no least favorable *a priori* distribution, $\hat{C}$ is, in a sense, a least favorable covariance matrix. In the following three theorems three different classes $\mathcal{C}$ are considered; in each case there is some Markoff decision function, relative to a least favorable C, that is minimax.

The next theorem is a team analogue of Theorem 6.5 of Hodges and Lehmann [16]; its proof is immediate from Lemma 2, and is omitted.

THEOREM 6. *Let C be a fixed covariance matrix such that C_{ii} is positive definite for every i; let k be a fixed positive number; and let $\mathcal{C}$ be the set of all cC such that $0 \leqq c \leqq k$. Then the Markoff decision function $\hat{b}$ relative to kC is minimax.*

(Note that $\hat{b}$ does not depend upon k.)

The next theorem deals with the situation in which the covariance structure of the random vector observed by each team member is known, but the covariances between the vectors observed by different team members are unknown.

THEOREM 7. *Let $\mathcal{C}$ be the set of all C such that $C_{ii} = C_{ii}^0$, where C_{ii}^0 is, for each i, a fixed symmetric positive definite matrix; then there exists $\hat{C}$ in $\mathcal{C}$ such that the Markoff decision function $\hat{b}$ relative to $\hat{C}$ is minimax.*

PROOF. There is no loss of generality in assuming that C_{ii}^0 is an identity matrix, for each i. Let $\mathcal{H}$ be the set of all Hadamard products H_c of Q with a C in $\mathcal{C}$, and let $\mathcal{B}$ be the smallest closed convex set containing all b such that b is Markoff relative to C for some C in $\mathcal{C}$. By Lemma 2, it is sufficient to prove that $b'Hb$ has a saddle point (minimax) for b in $\mathcal{B}$ and H in $\mathcal{H}$. Consider $\mathcal{H}$ as imbedded in a vector space, in the usual way, and with the inner product of H and K defined as the trace of HK. According to a theorem of Kakutani [17], such a saddle point will exist if

1. $b'Hb$ is continuous on $\mathcal{B} \times \mathcal{H}$.
2. $\mathcal{B}$ and $\mathcal{H}$ are each convex and compact.
3. For every b in $\mathcal{B}$, the set of all H that maximize $b'Hb$ in $\mathcal{H}$ is convex.
4. For every H in $\mathcal{H}$, the set of all b that minimize $b'Hb$ in $\mathcal{B}$ is convex.

(Although the conditions given here are not explicitly written out in the paper cited, they are implicit in the proof of its Theorem 3. Kakutani's result is not the most general saddle-point theorem available, but is sufficient for this proof. For references to other literature on the subject see Debreu [10].)

It is easily verified that $\mathcal{H}$ is convex and closed; furthermore, $\mathcal{H}$ is bounded because for any H in $\mathcal{H}$ (recall that H is positive definite),

$$\text{tr}\,(H^2) \leqq [\text{tr}\,(H)]^2,$$

and the trace of H equals $t = \sum_i q_{ii} n_i$, where n_i is the dimension of Z_i. Hence $\mathcal{H}$ is compact.

To see that $\mathcal{B}$ is also bounded, let $\bar{b}$ be some fixed vector in $\mathcal{B}$, and for every H in $\mathcal{H}$ let $r(H)$ denote its smallest characteristic root; then for any H in $\mathcal{H}$ and for b Markoff relative to H,

$$b'br(H) \leqq b'Hb \leqq \bar{b}'H\bar{b} \leqq \bar{b}'\bar{b}\,\mathrm{tr}\,(H). \tag{8.4}$$

From (8.4),

$$b'b \leqq [(\bar{b}'\bar{b}\,\mathrm{tr}\,(H)/r(H)]. \tag{8.5}$$

On the one hand, trace $(H) = t$ (see above) for all H in $\mathcal{H}$. On the other hand, $\inf_{H \text{ in } \mathcal{H}} r(H) > 0$, which follows from the continuity and positivity of $r(H)$ on $\mathcal{H}$, and the compactness of $\mathcal{H}$. Hence $\mathcal{B}$ is compact.

For any fixed b in $\mathcal{B}$, $b'Hb$ is linear on $\mathcal{H}$, which is convex; hence condition (3) is satisfied. Condition (4) follows from the uniqueness of the Markoff decision function relative to any fixed H. Finally, $b'Hb$ is clearly continuous on $\mathcal{B} \times \mathcal{H}$, which completes the proof.

If nothing at all is known about C, that is, if $\mathcal{C}$ is taken to be the class of all possible covariance matrices on Z, then the risk for every decision function is unbounded on $\mathcal{C}$. To get a finite minimax value, the class $\mathcal{C}$ must be "bounded" in some sense. One such sense is provided by the concept of the norm of one quadratic form with respect to another. Let C^0 be a given symmetric positive definite matrix, and for any symmetric nonnegative semi-definite matrix C define the bound of C relative to C^0 by

$$\|C\| = \max_{b\varepsilon Z}\,(b'Cb/b'C^0b).$$

THEOREM 8. *Let $\mathcal{C}$ be the set of all C such that $\|C\| \leqq k$, where k is a given positive number and C^0 is a given positive definite matrix; then the Markoff decision function $\hat{b}$ relative to kC^0 is minimax. Note that $\hat{b}$ depends upon C^0 but not upon k.*

PROOF. Since Q is positive definite, there exist $q_i(n)$, $i, n=1, \cdots, N$, such that $q_{ij}=\sum_n q_i(n)q_j(n)$ for $i, j=1, \cdots, N$. If b is of bounded risk then for any C in $\mathcal{C}$,

$$\begin{aligned} \rho(b, C) &= \sum_{ij} q_{ij}b_i'C_{ij}b_j = \sum_n \sum_{ij} q_i(n)b_i'C_{ij}q_j(n)b_j \\ &\leqq \sum_n \|C\| \sum_{ij} q_i(n)b_i'C_{ij}^0q_j(n)b_j \\ &= \|C\|\rho(b, C^0) \leqq k\rho(b, C^0) = \rho(b, kC^0). \end{aligned} \tag{8.6}$$

In particular (8.6) holds for $\hat{b}$, so that Lemma 2 can be applied to complete the proof of the theorem.

REFERENCES

[1] BAHADUR, R. R. (1954). Sufficiency and statistical decision functions. *Ann. Math. Statist.* **25** 423–462.

[2] BAHADUR, R. R. (1955). Statistics and subfields. *Ann. Math. Statist.* **26** 490–497.

[3] BAHADUR, R. R. and LEHMANN, E. L. (1955). Two comments on "Sufficiency and statistical decision functions." *Ann. Math. Statist.* **26** 139–141.

[4] BECKMANN, M. (1952). On Marschak's model of an arbitrage firm. Cowles Commission Discussion Paper, Economics No. 2058, November 21, 1952 (unpublished).

[5] Beckmann, Martin J. (1958). Decision and team problems in airline reservations. *Econometrica* **26** 134–145.
[6] Blackwell, David and Girschick, Meyer A. (1954). *Theory of Games and Statistical Decisions*. Wiley, New York.
[7] Bonnesen, Tommy and Fenchel, Werner (1934). *Theorie der Konvexen Körper*. Springer, Berlin.
[8] Bratton, D. (1953). On the arbitrage problem. Cowles Commission Discussion Paper, Economics No. 2057, May 13, 1953 (unpublished).
[9] Chernoff, Herman (1954). Rational selection of decision functions. *Econometrica* **22** 422–443.
[10] Debreu, Gerard (1952). A social equilibrium existence theorem. *Proc. Nat. Acad. Science* **38** 886–892.
[11] Doob, Joseph L. (1953). *Stochastic Processes*. Wiley, New York.
[12] Girschick, M. A. and Savage, L. J. (1951). Bayes and minimax estimates for quadratic loss functions, 53–73. *Proc. Second Berkeley Symp. Math. Statist. Prob.* Univ. of California Press.
[13] Halmos, P. R. (1958). *Finite Dimensional Vector Spaces*. Van Nostrand, Princeton.
[14] Halmos, P. R. (1957). *Introduction to Hilbert Space and the Theory of Spectral Multiplicity*. Chelsea, New York.
[15] Halmos, P. R. (1950). *Measure Theory*. Van Nostrand, New York.
[16] Hodges, Jr., J. L. and Lehmann, E. L. (1950). Some problems in minimax point estimation. *Ann. Math. Statist.* **21** 182–197.
[17] Kakutani, Shizuo (1941). A generalization of Brouwer's fixed point theorem. *Duke Math. J.* **8** 457–459.
[18] Kiefer, J. and Orey, S. (1953). On the arbitrage problem. Cowles Commission Discussion Paper, Economics No. 2068, February 23, 1953 (unpublished).
[19] Marschak, J. (1955). Elements for a theory of teams. *Management Science* **1** 127–137.
[20] Marschak, J. and Radner, R. (1954). The firm as a team (abstract). *Econometrica* **22** 523.
[21] Marschak, J. and Waterman, D. (1952). On optimal communication rules for teams. Cowles Commission Discussion Paper, Economics No. 2058, November 21, 1952 (unpublished).
[22] McGuire, C. B. (1961). Some team models of a sales organization. *Management Science* **7** 101–130.
[23] Radner, R. (1958). Minimax estimation for linear regressions. *Ann. Math. Statist.* **29** 1224–1250.
[24] Radner, Roy (1959). The application of linear programming to team decision problems. *Management Science* **5** 143–150.
[25] Radner, R. (1955). The linear team: an example of linear programming under uncertainty, 381–396. *Proc. Second Symp. Linear Programming*. National Bureau of Standards, Washington.
[26] Radner, R. (1961). The evaluation of information in organizations, **1** 491–533. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* Univ. of California Press.
[27] Savage, L. J. (1954). *The Foundations of Statistics*. Wiley, New York.
[28] Scheffé, Henry (1959). *The Analysis of Variance*. Wiley, New York.
[29] Wald, A. (1950). *Statistical Decision Functions*. Wiley, New York.