Rational Selection of Decision Functions

Herman Chernoff
Stanford University

April 10, 1953
Rational Selection of Decision Functions

Herman Chernoff

1. Summary.

A set of postulates for the rational selection of a decision function is discussed. It is shown that for the class of problems involving n states of nature these postulates imply that the choice among the available decision functions should be made as though each state of nature were equally likely.

2. Introduction.

Abraham Wald [1] introduced a generalized formulation of problems of Statistical Inference. This formulation is that of the Theory of Decision Functions. In this theory, there is a risk function \( r(s, d) \) which measures the risk or expected loss of the statistician if he applies strategy \( d \) when \( s \) is the state of nature. The statistician is given that the state of nature is an unknown element of a set \( S \) of possible states of nature. The strategy \( d \) to be selected is one of a set \( D \) of strategies available. Each strategy represents a plan of how to react to information which becomes available and is called a decision function. (The action or decision taken is a function of the data observed.) The statistician's object is to select an element \( d \) of \( D \) for which \( r(s, d) \) is as small as possible. Since there is seldom a uniformly best strategy, i.e., a \( d \) for which

\[
r(s, d) \leq r(s, d')
\]

for all \( s \) and \( d' \), there is a question as to what constitutes a good strategy.

This formulation seems to apply not only to statistical problems but to a great variety of problems of scientific method. It has often been conjectured that there is a rational approach to the solution of problems of scientific method but satisfactory approaches have not come forth. It would seem that if a rational approach does exist, it would involve the application
of a formulation which completely specifies the problems to be treated and of a set of postulates representing a conception of rational and consistent behaviour. It is natural to ask whether the decision function formulation does adequately specify the problems it was designed to treat. If the answer to this question is yes and if a rational approach exists, then it should be possible to construct a set of postulates of rational behaviour associated with the decision function formulation and to demonstrate that these postulates indicate a satisfactory criterion for selecting a "good" decision function.

The fundamental purpose of this paper is to see whether the theory of decision functions shows promise of being applicable in "real" problems and not necessarily to specify how to apply the theory. Hence we shall make certain idealizations or simplifications in the class of problems to be treated. Unless these simplifications are so poorly selected as to extract some of the essence from our problem, the treatment of the simplified class of problems should provide a more or less definite answer to our problem. In other words, if the decision function formulation cannot be used to rationally treat the simplified class of problems, it would seem reasonable to believe that one cannot hope to use this formulation to treat the more complicated case of "real" problems. On the other hand an affirmative result in the simplified case would indicate hope, but not certainty, for the affirmative result in the "real" case.

The simplifications that we shall make include the following. First we shall assume that there are only a finite number of possible states of nature in a given problem, i.e., \( S \) is always a finite set. We shall not indicate what constitutes a state of nature except through the postulates. We shall assume that the result of applying strategy \( d \) when state \( s \) is the
true state may be measured in terms of a measurable utility $u(s,d)$, which we shall call the payoff function. We shall assume that the statistician knows the payoff function even though this assumption implies that the statistician has a tremendous computing ability. In connection with this we shall further assume that there is no cost attached to any computational difficulties which may be associated with a particular criterion. It is the author's belief that the effects of these last two assumptions should be carefully examined before an affirmative result is obtained on the existence of a rational approach for the unsimplified class of "real" problems.

In our discussion of rational and consistent behaviour we do not imply that there are human beings who act in accord with our postulates. We are interested in determining whether an idealized intelligent individual, with tastes indicated by a utility function, can act rationally in certain situations of ignorance. Our notion of what constitutes rationality is specified by our postulates.

We shall have occasion to refer to the fact that, in the decision function formulation, experience does not have any role in selecting the strategy or decision function. Instead, experience, together with the decision function selected, determines the particular action to be taken at any time. We shall say more about the role of experience in connection with one of the postulates.

Section 3 will be devoted to discussing and indicating objections to some criteria which have been proposed for selecting a "good" strategy. In section 4, we shall present our definitions and postulates and then we shall proceed to obtain our main results in section 5. Finally various possible interpretations of these results will be indicated in section 6.

In this section we shall present some criteria which have previously been considered and we shall indicate some objections to these criteria.

Wald [1] tentatively suggested that a possible criterion would be that of minimizing the maximum risk. We treat risk as negative utility, i.e., \( r(s,d) = -u(s,d) \). This criterion selects that strategy \( d \) for which

\[
\max_{s \in S} r(s,d)
\]

is a minimum. (We use \( S \) and \( S' \) to denote is an element of and is not an element of respectively.) We shall abbreviate this criterion as min max risk.

It is interesting that such a criterion has been shown to be optimal in the problem of zero sum two person games, which problem has a strong formal resemblance to the decision function problem. In fact if one could assume that nature is an opponent whose gain is the statistician's loss and that \( s \), the unknown state of nature, is a strategy selected by nature, then the min max risk criterion would be a rational one. However, these assumptions seem unreasonable. Hence the analogy with two person games may not be used to justify the use of min max risk. On the other hand various people have objected to certain consequences of applying this criterion. For example, consider the risk function represented by the matrix in Table I.

Table I

<table>
<thead>
<tr>
<th></th>
<th>( d_1 )</th>
<th>( d_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( s_1 )</td>
<td>1</td>
<td>99</td>
</tr>
<tr>
<td>( s_2 )</td>
<td>100</td>
<td>99</td>
</tr>
<tr>
<td>( \max_{s \in S} r(s,d) )</td>
<td>100</td>
<td>99</td>
</tr>
</tbody>
</table>
This example illustrates the fact that the min max risk criterion sometimes allows very slight differences in a disastrous state to completely determine the choice of strategy. This property of min max risk leads many individuals to feel uncomfortable about the min max criterion.

An alternative was suggested by Savage [4] who pointed out that in all the examples Wald treated, the risk matrix (function) had zero as a minimum in each row, i.e., \( \min_{d \in D} r(s, d) = 0 \). Savage suggested that what Wald had in mind was to minimize the maximum regret where regret is defined by

\[
\rho(s, d) = r(s, d) - \min_{d \in D} r(s, d).
\]

In other words, one may consider the risk to be composed of the sum of an inevitable risk due to \( s \) and of a regret or risk due to the selection of a strategy in ignorance of \( s \). In Table II, the risk matrix of Table I is represented as such a sum.

<table>
<thead>
<tr>
<th>( r(s, d) )</th>
<th>( d_1 )</th>
<th>( d_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( s_1 )</td>
<td>1</td>
<td>99</td>
</tr>
<tr>
<td>( s_2 )</td>
<td>100</td>
<td>99</td>
</tr>
</tbody>
</table>

Table II

<table>
<thead>
<tr>
<th>( \rho(s, d) )</th>
<th>( d_1 )</th>
<th>( d_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( s_1 )</td>
<td>0</td>
<td>98</td>
</tr>
<tr>
<td>( s_2 )</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Since nothing can be done about the inevitable risk due to \( s \), it may be argued that the statistician should confine his attention to the regret matrix. If one applies min max regret in the above problem, the strategy selected is \( d_1 \). This choice seems to be more in accord with the decision that most people faced with this problem would make. Unfortunately the min max regret criterion has several drawbacks. Firstly, it has never been clearly demonstrated that differences in utility do in fact measure what one
may call regret. In other words it is not clear that the "regret" of going from a state of utility 5 to a state of utility 3 is equivalent in some sense to that of going from a state of utility 11 to one of utility 9. Secondly, one may construct examples where slight differences in one state of nature almost completely determine the strategy to be selected. Such examples tend to produce the same feelings of unease which led many to object to min max risk. One may consider the risk matrix in Table III as an example.

Table III

\[ r(s,d) = \rho(s,d) \]

<table>
<thead>
<tr>
<th></th>
<th>( d_1 )</th>
<th>( d_2 )</th>
<th>( d_3 )</th>
<th>( d_4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( s_1 )</td>
<td>1</td>
<td>99</td>
<td>0</td>
<td>10,000</td>
</tr>
<tr>
<td>( s_2 )</td>
<td>100</td>
<td>99</td>
<td>10,000</td>
<td>0</td>
</tr>
<tr>
<td>( s_3 )</td>
<td>0</td>
<td>0</td>
<td>10,000</td>
<td>10,000</td>
</tr>
</tbody>
</table>

\[ \max_{s \in S} \rho(s,d) \]

<table>
<thead>
<tr>
<th>( s )</th>
<th>( d_1 )</th>
<th>( d_2 )</th>
<th>( d_3 )</th>
<th>( d_4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( s_1 )</td>
<td>100</td>
<td>99</td>
<td>10,000</td>
<td>10,000</td>
</tr>
</tbody>
</table>

In this example each row has a minimum risk of zero and hence the risk and regret matrices are identical. We now find ourselves selecting \( d_2 \) in circumstances which remarkably resemble those of Table I. A third objection which the author considers very serious is the following. In some examples the min max regret criterion may select a strategy \( d_3 \) among the available strategies \( d_1, d_2, d_3, \) and \( d_4 \). On the other hand if for some reason \( d_4 \) is made unavailable, the min max regret criterion will select \( d_2 \) among \( d_1, d_2, \) and \( d_3 \). The author feels that for a reasonable criterion the presence of an undesirable strategy \( d_4 \) should not have an influence on the choice among the remaining strategies.** An example is given in Table IV.
Table IV

\[ r(s,d) = \rho(s,d) \quad [d_4 \text{ available}] \]

<table>
<thead>
<tr>
<th></th>
<th>(d_1)</th>
<th>(d_2)</th>
<th>(d_3)</th>
<th>(d_4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(s_1)</td>
<td>200</td>
<td>168</td>
<td>120</td>
<td>0</td>
</tr>
<tr>
<td>(s_2)</td>
<td>0</td>
<td>48</td>
<td>120</td>
<td>300</td>
</tr>
</tbody>
</table>

\[ \max_{s \in S} \rho(s,d) = 200 \quad 168 \quad 120 \quad 300 \]

\[ r(s,d) \quad \rho(s,d) \quad [d_4 \text{ unavailable}] \]

<table>
<thead>
<tr>
<th></th>
<th>(d_1)</th>
<th>(d_2)</th>
<th>(d_3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(s_1)</td>
<td>200</td>
<td>168</td>
<td>120</td>
</tr>
<tr>
<td>(s_2)</td>
<td>0</td>
<td>48</td>
<td>120</td>
</tr>
</tbody>
</table>

\[ \max_{s \in S} \rho(s,d) = 80 \quad 48 \quad 120 \]

Another approach that has been applied has been to show that certain postulates are satisfied by any criterion which selects a strategy by assuming an a priori probability distribution on the possible states of nature \([6,7,8]\). This approach, sometimes called the subjective approach, depends on postulates to which this writer has objections which will be indicated later. It should be emphasized that this consequence of this postulate system leads to a class of criteria instead of to a unique criterion. From an objective point of view such a result might be regarded as a denial of the existence of a rational approach via the decision function formulation.

4. Postulates of Rationality.

In this section we shall present the postulates which we consider descriptive of a rational approach to the problem of selecting a strategy.
Before proceeding with the postulates let us first consider the possible effects of introducing postulates which should not be present, of not introducing postulates which should be present and of doing both. Introducing postulates which should not be present may result in either eliminating from consideration all possible "rational" criteria or in a tendency to cut down the number of "rational" criteria to be considered. In the first case one may obtain negative results for our major question on the existence of rational criteria. In the second case one may find oneself in a deceptively satisfying position of recognizing the existence of only one "rational" criterion. While the postulates that are presented will have to satisfy the reader and cannot sneak in unobserved, it is possible that the author and the readers may forget to specify a few postulates which should be introduced. The effect of such omission can be only to increase the number of "rational" criteria to be considered. Hence it would be wise to thoroughly inspect each criterion which succeeds in satisfying the postulates presented. If some postulates are unfairly introduced and others omitted, it is quite conceivable that a rational criterion would be eliminated from consideration and replaced by an unsatisfactory criterion. In view of these remarks we shall be careful to specify those postulates concerning which there may be some controversy and to discuss in detail any of these which are found to play a critical role.

The author's conception of a rational criterion can be precisely described only in terms of the postulates. On the other hand an intuitive notion leading to the selection of these postulates is the following. In a given problem facing a statistician, he should first eliminate those strategies which are obviously bad. He should then dispose of some of the
remaining which, while not as obviously bad, still fail to make the grade. After a certain amount of eliminating, the remaining strategies will be considered adequate. The statistician will have no reason to prefer any of these strategies to the others. The set of these adequate strategies will be called the solution of the problem. It is not implied that the statistician necessarily considers that any two elements of the solution are equivalent. For example if \( d_1 \) and \( d_2 \) are elements of the solution and \( d_2 \) is replaced by a strategy \( d_3 \) which is definitely better than \( d_2 \), it is still possible that \( d_1 \) will be an element of the solution. To give a more concrete example, we assume that a rational individual who can not decide on the relative merits of trips to Florida and California may not be obligated to go to Florida if the train fare there is reduced by one dollar. It may be remarked here that investigations carried out by the "subjective" approach [6, 7, 8] do not apply this point of view. Instead, it is assumed that there is a complete ordering of strategies. In other words for any two strategies one is either to be preferred to the other or they are equivalent in the sense that a slight improvement in one of the strategies will make it preferred to the other. The author feels that such an assumption is too much to make if it is desired to determine whether a rational approach exists via the decision function formulation.

In view of the Decision Function formulation we introduce the following definition.

**Definition 1.** A problem \( Q = (S, D, u) \) consists of two non null sets \( S \) and \( D \) and a bounded real valued utility function \( u \) defined for \( s \in S \) and \( d \in D \). The value of the function \( u \) is called the payoff.

We may consider the problem \( Q \) to be represented by a matrix where each row represents the payoff corresponding to an element \( s \in S \) and each column the payoffs corresponding to an element \( d \in D \).
As was pointed out in the criticism of the principle of min max regret, it is desirable to have some sort of consistency between the "optimal" strategies for related problems. Hence we should deal with a criterion for selecting a good strategy for each of a class of problems, and the criterion should be "consistent" for problems of this class. We shall later discuss reasons for limiting this class.

Definition 2. A non null set of problems \( \{Q\} \) is called a general problem \( G \).

Definition 3. A general problem \( G \) is rationally solvable by \( C(Q) \) if postulates 1-9 are satisfied.

Postulate 1. \( C(Q) \) is a non null subset of \( \mathcal{U}(Q) \) for every \( Q \) in \( G \).

Definition 4. If \( u(s, d_1) = u(s, d_2) \) for every \( s \in S \), we write \( d_1 \sim d_2 \) and read \( d_1 \) is equivalent to \( d_2 \). If \( u(s, d_1) \geq u(s, d_2) \) for every \( s \in S \) and \( u(s, d_1) > u(s, d_2) \) for some \( s \in S \), we write \( d_1 \succ d_2 \) and read \( d_1 \) dominates \( d_2 \).

The relationship \( d_1 \succ d_2 \) means \( d_1 \succ d_2 \) or \( d_1 \sim d_2 \).

Note that \( \succ \) represents a partial ordering defined on \( \mathcal{D}(\cdot) \) for each problem \( Q \) and has meaning only in reference to the problem \( Q \). Our second postulate deals with this ordering in an obvious fashion.

a: If for some \( Q \in G \), \( d_1 \in C(Q) \) and \( d_1 \sim d_2 \), then \( d_1 \in C(Q) \).

b: If for some \( Q \in G \), \( d_1 \succ d_2 \), then \( d_2 \in C(Q) \).

Our third postulate will deal with the principle that relabeling the states of nature and strategies should not affect our solution so long as the payoffs are not changed.
Definition 5. The problems $Q_1 = (S_1, D_1, u_1)$ and $Q_2 = (S_2, D_2, u_2)$ are called isomorphic if there is a one to one correspondence $s_2 = f(s_1)$ between $S_1$ and $S_2$ and a one to one correspondence $d_2 = g(d_1)$ between $D_1$ and $D_2$ such that

$$u_2[f(s_1), g(d_1)] = u_1[s_1, d_1] \text{ for } s_1 \in S_1, d_1 \in D_1.$$

Postulate 3.

If $Q_1 = (S_1, D_1, u_1)$ and $Q_2 = (S_2, D_2, u_2)$ are isomorphic and $Q_1$ and $Q_2$ are elements of $G$, then $C(Q_2) = f[C(Q_1)]$.

It should be remarked that this postulate is not as innocuous as it may seem. In fact, the non-acceptance of this postulate is essential in the "subjective" approach. The main objection which may be raised to it is the following. It specifically prohibits the possibility of using previous information or prejudice to give one state of nature more weight than another. One may argue that in any real problem facing an individual, there is usually a considerable accumulation of data which may serve as a clue to indicate which is the true state of nature. A possible reply to this argument is the following. Our task is to investigate whether the theory of decision functions leads to an approach to scientific problems which is something more rational than a dependence on instinct. Then we may assume that at some original point in his history our rational statistician was ignorant (had no information) and was faced with a problem $Q$. Granting that abysmal ignorance must be associated with lack of ability to understand the problem, we may concede that whatever actions the statistician took up to now may not have been very wise ones. At this point however, the statistician is wise enough to be able to formulate the problem that faced him originally and to
notice that certain strategies which were available to him then were in-
consistent with the actions he took up to now. His problem now is to
"solve" a modification of his original problem, namely that problem where
the class of available strategies is diminished by omitting those strategies
inconsistent with his actions up to now. In this new problem, the
experience he has had since his origin will be applied, like his future data,
by substitution into the strategy or decision function and not by the giving
of different weights to the various states of nature.

We have several times considered related problems where one is obtained
from another by deleting some of the available strategies. In this connection
let us introduce

Definition 6. We say that \( C_1 = (S_1, D_1, u_1) \) is strategically contained in
\( C_2 = (S_2, D_2, u_2) \) and write \( C_1 \subseteq C_2 \) if \( D_1 \subseteq D_2 \) and
\[ u_2(s, d) = u_1(s, d) \text{ for } s \in S, d \in D_1. \]

One of our objections to the min max regret criterion leads us to
formulate the following postulate.

Postulate 4. If \( C_1 = (S_1, D_1, u_1) \) and \( C_2 = (S_2, D_2, u_2) \) are elements of \( G \)
and \( C_1 \subseteq C_2 \), then
\[ C(Q_2) \subseteq C(Q_1) \cup (D_2 - D_1). \]

This last relation may also be stated as follows:
\[ C(Q_2) \cap D_1 \subseteq C(Q_1). \]

We emphasize here that the statement of postulate 4 is worded so that
the rational statistician who is not "decided" on his choice between trips
to Florida and California may remain undecided when the train fare to Florida
is reduced by $1. More precisely, the first four postulates do not prevent
the possibility that \( C(Q_2) \) consists of two strategies \((d_1, d_2)\) where \( d_1 \in D_1 \) and \( d_2 \in D_2 - D_1 \) while \( C(Q_1) \) consists of two strategies \((d_1, d_3)\) where \( d_3 \in D_1 \) and \( d_2 \not\subset D_3 \).

In postulate 4 we have not given the situation where the elements of \( D_2 - D_1 \) are all bad the special treatment it deserves. In that case our intuitive notion of a solution should lead us to postulate that eliminating \( D_2 - D_1 \) from consideration does not affect our solution. Hence we should supplement postulate 4 by postulate 5 or possibly by the stronger postulate 5*.

Postulate 5. If \( Q_1 = (S, D_1, u_1) \) and \( Q_2 = (S, D_2, u_2) \) are elements of \( G \), \( Q_1 \subsetneq Q_2 \), and every element of \( D_2 - D_1 \) is dominated by some element of \( D_1 \), then \( C(Q_2) = C(Q_1) \).

Postulate 5*. If \( Q_1 = (S, D_1, u_1) \) and \( Q_2 = (S, D_2, u_2) \) are elements of \( G \), \( Q_1 \subsetneq Q_2 \), and \( C(Q_2) \subset D_1 \), then \( C(Q_1) = C(Q_2) \).

While postulate 5* is not imposed on our definition of a rational solution, we shall be careful to check whether it is satisfied by the criteria which satisfy the other postulates.

Postulate 4 should be further supplemented by the additional requirement that our problem is not affected by making available strategies which are extraneous in the sense that they are equivalent to strategies already available.

Postulate 6. If \( Q_1 = (S, D_1, u_1) \) and \( Q_2 = (S, D_2, u_2) \) are elements of \( G \), \( Q_1 \subsetneq Q_2 \), and every element of \( D_2 \) is equivalent to some element of \( D_1 \), then \( C(Q_2) \cap D_1 = C(Q_1) \).
With assistance from postulate 2, postulate 6 implies that $C(Q_2)$ can be obtained from $C(Q_1)$ by adding to $C(Q_1)$ those elements of $D_2 - D_1$ equivalent to elements of $C(Q_1)$.

Let us now consider the role of mixed or randomized strategies. A typical example of a mixed strategy would be that of selecting $d_1$ if a coin falls heads and $d_2$ if the coin falls tails. Statisticians have frequently applied randomized strategies in problems of design of experiments, in problems of testing hypotheses when for one reason or another it has been desired to use a specified level of significance while dealing with discrete densities, and, in essence, in the very act of taking random samples. We should not fail to take into account the possibility of using mixed strategies. Indeed the fact that a statistician can not be prevented from going into a corner and tossing a coin will influence us later to confine our attention only to problems where all mixed strategies are available. To help discuss certain situations involving mixed strategies we shall first make the simplification of dealing with problems involving only a finite number of strategies and with problems derived from these by permitting randomization.

Definition 7.

A mixture of the elements of a finite set $D$ is a real valued function $\xi$ defined on $D$ with the properties

\begin{align*}
(1) & \quad 0 \leq \xi(d) \leq 1 \quad \text{for } d \in D \\
(2) & \quad \sum_{d \in D} \xi(d) = 1
\end{align*}

The mixture $\xi = p \xi_1 + (1-p)\xi_2$ is said to be a proper mixture of $\xi_1$ and $\xi_2$ if $\xi_1$ and $\xi_2$ are mixtures and $0 < p < 1$. A mixture $\xi$ is called pure if $\xi(d) = 1$ for some $d \in D$. 
Definition 8.
\[ \zeta^* = (S, D^*, u^*) \text{ is the randomized (or mixed) version of } Q = (S, D, u) \]
if \( D^* \) is the class of mixtures of the elements of the finite set \( D \) and
\[ u^* (s, \zeta) = \int_{d \in D} \zeta(d) u(s, d). \]

Our concept of a solution immediately suggests the following

Postulate 7. If \( Q \in \mathcal{Q} \) is a randomized version of a problem, any mixture of
two elements of \( C(Q) \) is an element of \( C(Q) \).

We shall now discuss a postulate suggested to the author by H. Rubin.
While this postulate has important consequences it seems extremely reasonable.
This postulate is suggested by the following situation. Suppose that a
rational statistician would select \( C(Q) \) for his solution of the problem \( Q \).
Suppose that his problem is modified so that if a coin falls heads, he will
obtain \( u_0 (s) \) independent of \( d \) and if the coin falls tails, he will obtain
\( u(s, d) \). It would seem reasonable to expect the rational statistician's
choice in the modified problem to be the same as in the original problem.
Of course we must assume that the probability \( p \) of a tail does not depend
on \( s \) for otherwise the event tail would provide the statistician with infor-
mation concerning \( s \) which might affect his choice.

Postulate 8. If \( Q_1 = (S, D, u_1) \) and \( Q_2 = (S, D, u_2) \) are elements of \( \mathcal{G} \) and
\[ u_2 (s, d) = p u_1 (s, d) + (1-p) u_0 (s) \quad 0 < p < 1 \]
where \( u_0 (s) \) is a utility function, then \( C(Q_2) = C(Q_1) \).

Let us suppose that in a problem \( Q \), there are two states of nature \( s_1 \)
and \( s_2 \) such that \( u(s_1, d) = u(s_2, d) \) for all \( d \in D \). Then it does not seem
unreasonable to propose considering combining the two states into a single
state "\( s_1 \) or \( s_2 \)" and postulating that this combination does not affect the
solution. The justification for such a postulate must be involved with the conception of what constitutes an unknown state of nature. The author has been deliberately vague on this point and hence cannot justify the use or non-use of such a postulate at this point. In fact this postulate is incorporated but will be avoided by our restricting ourselves to considering only general problems where all problems have the same number of states of nature. Some more discussion concerning this postulate will follow later.

Postulate 9. If

(i) \( Q_1 = (S_1, D, u_1) \) and \( Q_2 = (S_2, D, u_2) \) are elements of \( G \),

(ii) \( S_1 \subseteq S_2 \),

(iii) \( u_1(s, d) = u_2(s, d) \) for \( s \in S_1 \), \( d \in D \), and

(iv) for every \( s_2 \in S_2 \), there is an \( s_1 \in S_1 \) such that

\[
\begin{align*}
    u_1(s_1, d) &= u_2(s_2, d) \\
    &\text{for all } d \in D,
\end{align*}
\]

then

\[
C(Q_1) = C(Q_2).
\]

In terms of the matrices representing the problems, we may have phrased this postulate as follows \( C(Q_1) = C(Q_2) \) if \( Q_2 \) is derived from \( Q_1 \) by adding rows equivalent to rows of \( Q_0 \).

Finally we shall present a postulate which indicates a property that one may like a rational approach to have. On the other hand, it seems to the author to be a condition which should be put on a provisional basis. When \( Q_1 \) and \( Q_2 \) are two problems differing only in the strategies available, we shall let \( Q_1 + Q_2 \) represent the problem where the set of strategies available is \( D_1 \cup D_2 \). Then postulate 10 essentially states that if a strategy
is satisfactory for $Q_1$ and for $Q_2$ then it should be satisfactory for $Q_1 + Q_2$.

**Definition 2.** If $Q_1 = (S, D_1, u_1)$ and $Q_2 = (S, D_2, u_2)$ where $u_1(s, d) = u_2(s, d)$ if $d \in D_1 \cap D_2$, then $Q_1 + Q_2 = (S, D_1 \cup D_2, u)$ where $u = u_1(s, d_1)$ for $d \in D_1$, $i = 1, 2$.

**Postulate 10.** If $Q_1$, $Q_2$, and $Q_1 + Q_2$ are elements of $G$, then $C(Q_1) \cap C(Q_2) \subset C(Q_1 + Q_2)$.

We present a list of the above definitions and postulates with cryptic (and non-precise) descriptive comments.

**Definitions**

1. problem
2. general problem
3. rational solvability by $C(Q)$
4. partial ordering
   equivalence
   dominance
5. isomorphism
6. strategic inclusion
7. mixed strategies
   proper mixture
   pure strategy
8. mixed problem
9. strategic addition

$Q = (S, D, u)$

$G = \{Q\}$

postulates 1-9

$u_2[f(s_1), g(d_1)] = u_1[s_1, d_1]$

$Q_1 \sqcup Q_2$

$0 \leq \xi(d) \leq 1$, $\xi(d) = 1$

$p \xi_1 + (1-p) \xi_2$, $0 < p < 1$

$\xi(d) = 1$ for some $d$

$u^*(s, \xi) = \sum \xi(d_1) u(s, d_1)$
Postulates

1. \( C(Q) \) non null

2. a) \( d_1 \in C(Q), d_2 \sim d_1 \Rightarrow d_2 \in C(Q) \)
   b) \( d_1 \cong d_2 \Rightarrow d_2 \in C(Q) \)

3. \( C(Q) \) invariant under isomorphism.

4. \( Q_1 \cap Q_2 \Rightarrow C(Q_2) \cap D_1 \subseteq C(Q_1) \).

5. Elimination of dominated strategies.

5\* \( Q_1 \supseteq Q_2, C(Q_2) \supseteq D_1 \Rightarrow C(Q_2) = C(Q_1) \)


7. Convexity: \( C(Q) \) contains all mixtures based on \( C(Q) \).

8. Rubin's Postulate: \( C(Q_1) = C(Q_2) \) if \( u_2 = p u_1 + (1-p) u_0 \) (a)

9. Combining equivalent states of nature.

10. \( C(Q_1) \cap C(Q_2) \subseteq C(Q_1 + Q_2) \)

5. Main Results.

In this section several main results will be presented. First we shall apply Rubin's postulate to show that (1) regret matrices are relevant and that (2) mixed strategies are not especially necessary. By the latter clause we mean that if a mixture of some strategies is in the solution then these strategies are also in the solution. Of course the relevance of regret matrices does not imply that the criterion min max regret has merit. Finally we shall show that for \( Q_n \), the class of all mixed problems involving \( n \) states of nature and a finite number of pure strategies, the unique criterion satisfying the postulates is equivalent to assuming that each state of nature has an a priori probability of \( 1/n \).
Unless otherwise stated, each theorem will assume that the general
game is rationally solvable by $C(Q)$. There will be added to the statement of
each theorem a list of those postulates which are applied in the proof of the
theorem.

Theorem 1. (Converse of Convexity).

If (i) $Q = (S, D^*, u) \in \mathcal{O}$ is a mixed problem,

(ii) $Q_1$ isomorphic to $Q_2$ and $Q_2 \supseteq Q$ imply that

$Q_1$ and $Q_2$ are elements of $\mathcal{O}$, and

(iii) $\xi_0 \in C(Q)$ is a proper mixture of $\xi_1$ and $\xi_2$.

then

$\xi_1$ and $\xi_2$ are elements of $C(Q)$, $(P_3, \gamma_4, \gamma_8^o)$

Proof: We shall first show that $\xi_1 \in C(Q)$. There is a number $p$, $0 < p < 1$,
such that $\xi_0 = p \xi_1 + (1-p) \xi_2$ and

$u(s, \xi_0) = p u(s, \xi_1) + (1-p) u(s, \xi_2)$ for $s \in S$.

Let $Q_1 = (S, D^*, u_1)$ where

$u_1(s, \xi) = p u(s, \xi_1) + (1-p) u(s, \xi_2)$ for $s \in S$, $\xi \in D^*$.

Rubin's postulate is applicable if we let $u(s, \xi_2)$ play the role of $u_0(s)$
and if $Q_1 \in \mathcal{O}$. Hence to prove $\xi_1 \in C(Q)$ it suffices to show that $Q_1 \in \mathcal{O}$
and $\xi_1 \in C(Q)$. To show these we shall construct an isomorphism between $Q_1$
and $Q_2 \supseteq Q$ where $\xi_1$ is transformed into $\xi_0$ and $\xi_2$ into $\xi_2^o$. Once this
isomorphism is constructed it follows that $Q_1$ and $Q_2$ are elements of $\mathcal{O}$ and
by postulate 1 that $\xi_0 \in C(Q_2)$. Finally, by postulate 3, $\xi_1 \in C(Q_1)$. A
similar argument may be applied to $\xi_2$. 
All that remains is the construction of the isomorphism. Since \( Q \) is a mixed problem, for every element \( \xi^* \in D^* \) there is an element
\[
\xi = g(\xi^*) = p \xi^* + (1-p) \xi_2 \in D^*.
\]
The function \( g \) is a 1-1 correspondence such that
\[
g(\xi_1) = \xi_0 \quad \text{and} \quad g(\xi_2) = \xi_2.
\]
Let
\[
D_2 = \{ \xi : \xi = g(\xi^*) , \xi^* \in D^* \}
\]
and
\[
u_2(s, \xi) = u(s, \xi) \quad \text{for} \quad s \in S, \xi \in D_2.
\]
Then it is obvious that
\[
Q_2 = (S_2, D_2, u_2) \supseteq Q
\]
Furthermore, since
\[
u_2(s, g(d^*)) = \nu_1(s, d^*) \quad \text{for} \quad s \in S, d^* \in D,
\]
\( Q_2 \) is isomorphic to \( Q_1 \).

Theorem 2. (Relevance of Regret)

If \( Q_1 = (S, D, u_1) \) and \( Q_2 = (S, D, u_2) \) are elements of \( G \) and
\[
u_2(s, d) = \nu_1(s, d) + u_0(s) \quad \text{for} \quad s \in S, d \in D
\]
and if there is a \( p, 0 < p < 1 \) such that \( Q_3 = (S, D, u_3 = p \nu_2) \) is an element of \( G \) and \( \frac{p}{1-p} \nu_0(s) \) is a utility function, then \( C(Q_2) = C(Q_1) \). (P_6)

Proof: Let \( \nu_0^*(s) = 0 \). Then
\[
\nu_3(s, d) = p \nu_2(s, d) + (1-p) \nu_0^*(s) = p \nu_1(s, d) + (1-p) \frac{p \nu_0(s)}{1-p}
\]
for \( s \in S, d \in D \). Applying Rubin's postulate twice we have \( C(Q_3) = C(Q_2) \) and \( C(Q_3) = C(Q_1) \).
Theorem 2 implies that one may subtract a constant from each row [a different constant may be used for each row] without affecting the solution providing that \( G \) is not too small. Hence the regret matrix represents a canonical form into which each problem may be transformed.

We shall now treat the general problem \( Q_n \) which is defined as follows.

**Definition 10.** \( Q_n \) is the class which contains (i) all mixed problems \( Q = (S, D^*, u) \) where \( S \) has \( n \) elements and \( D^* \) is the class of mixtures of a finite set of strategies and (ii) all problems isomorphic to such mixed problems.

Our main result will be to prove that \( Q_n \) is rationally solvable by one and only one criterion \( C(Q) \) and that this criterion is equivalent to assuming that each state of nature is equally likely. To help simplify the proof of this result we shall introduce the (geometric) representation of a problem.

**Definition 11.** If the elements of \( S \) are relabeled \( s_1, s_2, \ldots, s_n \), then the set

\[
R = \{ x(d) = [u(s_1, d), u(s_2, d), \ldots, u(s_n, d)]: d \in D \}
\]

is a representation of the problem \( Q = (S, D, u) \). The point \( x(d) \) is the representative of \( d \) relative to the relabeling above.

We note that several elements of \( D \) may have the same representative, many problems may have the same representation, and there may be several representations of a particular problem depending on the relabeling of the elements of \( S \). The lack of a 1-1 correspondence between representations and problems causes no real difficulty, once some simple properties are noted. For example, if \( Q_1 \) and \( Q_2 \) are isomorphic, they have the same representations. Equivalent strategies have the same representative relative to any relabeling.
The representation of a problem of $Q_n$ is a closed bounded convex set in $n$ dimensional space generated by a finite number of points whose coordinates are possible utilities and conversely. In fact, for any such convex set $R$ we may construct a canonical problem $Q_R$, with representation $R$ by letting the axes be the states of nature and the vertices correspond to pure strategies. We now introduce the notion of the solution of a representation.

Definition 12. The solution of a representation $R$ of $Q$ obtained by a relabeling $\lambda$ of the elements of $S$ is given by $C_{Q, \lambda}(R)$, the set of representatives of elements of $C(Q)$ relative to $\lambda$.

Lemma 1. If $Q \in Q_n$ has representation $R$, then $C_{Q, \lambda}(R)$ depends only on $R$ and may be denoted by $C(R)$. ($P_2$, $P_3$, $P_6$).

Proof. It is obvious that there is a problem $Q_1 \in Q_n$ such that $Q_1 \equiv Q$ and $Q_1$ is isomorphic to $Q$, the canonical problem of $R$. By postulate 3,

$C_{Q_1, \lambda}(R) = C_{Q, \lambda}(R)$ where $C_{Q, \lambda}(R)$ is the solution of $R$ of $Q_R$ relative to the trivial relabeling and depends only on $R$. There remains to be shown only that $C_{Q_1, \lambda}(R) = C_{Q, \lambda}(R)$. But this is implied by postulate 6 together with 2.

Lemma 1 relates the solution of problems with the same representation.

Now let us relate the solutions of different representations of the same problem.

Lemma 2. If $R_1$ and $R_2$ are representations of problems of $Q_n$ and $R_2$ may be obtained from $R_1$ by a relabeling $\lambda$ of the axes of the $n$ dimensional space, then $C(R_2)$ may be obtained from $C(R_1)$ by the same relabeling $\lambda$. ($L_1$).
Proof: It is easily seen that there is a problem \( R_1 \) for which \( R_1 \) and \( R_2 \) are two representations. Therefore \( C(R_1) \) and \( C(R_2) \) are the sets of representatives of \( C(R_1) \) corresponding to the trivial relabeling and to \( \lambda \).

The lemma follows.

The presence of these two lemmas eliminates the difficulties which may have arisen from the lack of a 1-1 correspondence between problems and representations. Now one may translate the postulates and theorems into the language of representations. For example postulate 7 implies that \( C(R) \) is convex for the representation of a problem of \( G_n \). Theorem 1 implies that if \( x \in C(R) \) is an interior point of a line segment in \( R \), this line segment is in \( C(R) \). Theorem 2 implies that if \( R_2 \) may be obtained from \( R_1 \) by a translation, then \( C(R_2) \) may be obtained from \( C(R_1) \) by the same translation.

We are now prepared to prove four lemmas which immediately imply the main theorem.

Lemma 3. If \( R = \{ x = (x_1, x_2, \ldots, x_n): m \leq x_i \leq M \text{ for } 1 \leq i \leq n, \sum_{i=1}^{n} x_i = c \} \)

where \( m \) and \( M \) are possible utilities, then \( C(R) = R_0 \) \( (P_1, P_2, T_1, L_1, L_2) \).

Proof: If \( m \) and \( M \) are possible utilities, \( R \) is the representation of a problem of \( G_n \). The case \( m = M \) is trivial since \( R \) has only one point then. We assume that \( m < c < M \) and hence no point of \( R \) has all components equal to \( M \) nor all components equal to \( m \). Let \( r(x) \) be the number of components of \( x \) equal to \( m \) and \( s(x) \) the number of components of \( x \) equal to \( M \).

Then \( 0 \leq r(x) + s(x) \leq n \), \( 0 \leq r(x) < n \), and \( 0 \leq s(x) < n \). It suffices to show that there is an element \( x \in C(R) \) for which \( r(x) = s(x) = 0 \) for then,
every point of \( R \) is on a line segment in \( R \) of which \( x \) is an interior point and theorem 1 would then yield our result. \( C(R) \) is non empty (postulate 1). Let \( x \in C(R) \). Suppose \( r(x) + s(x) \geq 1 \). We note that relabeling the axes leaves \( R \) invariant. Applying lemma 2, any point obtained from \( x \) by rearranging the coordinates is an element of \( C(R) \). Since \( r(x) \leq n-1 \) and \( s(x) \leq n-1 \), it is possible to rearrange the coordinates to get a point \( y \) so that \( y \) differs from \( x \) in at least one coordinate which was previously \( m \) or \( M \). By convexity (postulate 7) \( z = \frac{n+1}{2} \in C(R) \). But \( r(z) + s(z) \leq r(x) + s(x) \).

By backwards induction we may obtain an element of \( C(R) \) for which \( r+s=0 \) and our lemma follows.

**Lemma 4.** If \( R \) is the representation of a problem of \( G_n \), \( C(R) \) contains at least those elements \( x \) of \( R \) which maximize \( \sum_{i=1}^{n} x_{i} \) (\( p_{2}, p_{3}, p_{5}, l_{1}, l_{3} \))

**Proof:** Let

\[
\begin{align*}
m &= \min_{x \in R} x_{i} \\
M &= \max_{x \in R} x_{i} \\
c &= \max_{x \in R} \sum_{i=1}^{p} x_{i} \\
R_{1} &= \{x: m \leq x_{i} \leq M \text{ for } 1 \leq i \leq n, \sum_{i=1}^{p} x_{i} \leq c\}, \\
R_{2} &= \{x: m \leq x_{i} \leq M \text{ for } 1 \leq i \leq n, \sum_{i=1}^{p} x_{i} = c\},
\end{align*}
\]
Note that $R_1$ and $R_2$ are representations of problems of $G_n$, that $R \subseteq R_1$ and $R_2 \cap R$ is non null. If $x$ is in $R_1$ and not in $R_2$, there is an element $y \in R_2$ such that $x_1 \leq y_1$ for $1 \leq i \leq n$ and $x_1 < y_i$ for some $i$. Applying postulate 5 and lemma 3, $C(R_1) = C(R_2) = R_2$. By postulate 4 applied to $R_1$ and $R$, $C(R) \Rightarrow R_2 \cap R$. Our result follows immediately from postulate 2a.

**Lemma 6.** If $R$ is the representation of a problem of $G_n$ and

$$\sum_{i=1}^{n} b_i < \max_{x \in R} \sum_{i=1}^{n} x_i, b \in R,$$ then $b \in C(R)$. (2', 4', 5', 7', 1', 14)

**Proof:** Let $a = (a_1, a_2, \ldots, a_n)$ be an element of $R$ which maximizes

$$\sum_{i=1}^{n} x_i.$$ Let $R$ be the line segment connecting $a$ and $b$. Clearly $R$ is the representation of a problem of $G_n$. By postulate 4, if $b \in C(R)$, $b \in C(R_1)$ and hence it suffices to show that $b \in C(R_1)$. In view of postulate 2b, we may assume that $b$ is not dominated by $a$ since the case where $b$ is dominated by $a$ is trivial (Postulate 2).

Let $c$ be a point (whose coordinates represent possible utilities) dominated by $a$ such that $\sum_{i=1}^{n} c_i > \sum_{i=1}^{n} b_i$. Let $d = \frac{a + b}{2}$ and $e = \frac{a + c}{2}$ (see Figure 1). Let $R_2$ be the set bounded by the triangle with vertices $a$, $b$, $c$. 16/

---

Figure 1.
Let \( R_3 \) be the set bounded by the quadrilateral with vertices \( b, c, d, e \).

Clearly \( R_2 \) and \( R_3 \) are representations of problems of \( G_n \). Note that \( R_2 \supseteq R_1 \), \( R_2 \supseteq R_3 \), and that every element of \( R_2 \) not on \( R_1 \) is dominated by a point of \( R_1 \).

Hence \( C(R_2) = C(R_1) \) [Postulate 5] and it suffices to show that \( b \in \ E^c \ C(R_2) \).

But if \( b \in \ E C(R_2) \), \( b \in \ E C(R_3) \) [Postulate 4]. However \( e \) maximizes \( \sum_{i=1}^{n} x_i \) on \( R_3 \) and by lemma 4, \( e \in \ E C(R_3) \). By convexity [Postulate 7], \( \frac{b+e}{2} \in \ E C(R_3) \). But this is impossible since \( \frac{b+e}{2} \) is dominated by \( d \).

**Lemma 6**: \( G_n \) is rationally solvable by the criterion \( \mathcal{C}(Q) = \{d: d \in D, d \text{ maximizes } \frac{1}{n} \sum_{s \in S} u(s, d)\} \).

**Proof**: All the postulates proposed including 10 are easily seen to be satisfied by the above criterion. (It should be noted that postulate 9 is vacuously satisfied in the sense that it can not be applied if \( G_n \) is the general problem.)

**Theorem 3**: \( G_n \) is rationally solvable by only the criterion

\[ \mathcal{C}(Q) = \{d: d \in D, d \text{ maximizes } \frac{1}{n} \sum_{s \in S} u(s, d)\} \text{ L}_4, L_5, L_6. \]

**Proof**: Theorem 3 is an immediate consequence of lemmas 4, 5, and 6.

We terminate this section by cryptically summarizing the definitions, lemmas, and theorems of this section. Next to the results we shall indicate the preceding results and postulates used in the proofs and also the complete set of postulates which were used to obtain the result.

**Definitions**.

10. \( G_n \) mixed problems based on \( n \) states of nature and a finite number of pure strategies.
11. Representation

\[ R = \{(x_1, x_2, \ldots, x_n) : x_i = u(s_i, d) \text{ d E D}\} \]

representatives of C(Q).

12. \( C_Q, \lambda (R) \)

Results.

Theorem 1. Converse of Convexity

| \( P_3 \) | \( P_4 \) | \( P_8 \) | \( P_3 \) | \( P_4 \) | \( P_8 \) |
| \( P_6 \) | \( P_8 \) |

Theorem 2. Relevance of Regret

| \( P_8 \) |

Lemma 1. \( C_Q, \lambda (R) = C(R) \)

| \( P_2 \) | \( P_3 \) | \( P_6 \) | \( P_2 \) | \( P_3 \) | \( P_6 \) |

Lemma 2. Relabeling axes.

| \( L_1 \) | \( P_2 \) | \( P_3 \) | \( P_6 \) |

Lemma 3. \( R = \{x : m \leq x_1 \leq M, \exists x \in C(R) = R \}

| \( P_1 \) | \( P_2 \) | \( P_3 \) | \( P_4 \) | \( P_5 \) | \( P_6 \) | \( P_7 \) |

Lemma 4. \( x \) maximizes \( \sum x_i \Rightarrow x \in C(R) \)

| \( P_2 \) | \( P_3 \) | \( P_4 \) | \( P_5 \) | \( P_6 \) | \( P_7 \) | \( P_8 \) |

Lemma 5. \( x \in C(R) \Rightarrow x \) maximizes \( \sum x_i \)

| \( P_2 \) | \( P_3 \) | \( P_4 \) | \( P_5 \) | \( P_6 \) | \( P_7 \) | \( P_8 \) |


Theorem 3. Only "equally likely" satisfies postulates.

| \( L_1 \) | \( L_2 \) | \( L_3 \) | \( P_1 \) | \( P_2 \) | \( P_3 \) | \( P_4 \) |

6. Interpretation of Main Results.

We shall now consider the implications of the theorems of the preceding section. Theorem 1 states that if a rational solution exists for our simplified formulation, randomized strategies are unnecessary. Certain aspects of randomization do not appeal to many statisticians. For example the notion, that after having observed the data one should toss a coin before deciding what action to take, is commonly considered objectionable but there are other aspects which lie at the foundation of statistics. Statisticians would generally be horrified if anyone should suggest that a poll be taken without the use of a random sample.
This contradiction between theorem 1 and common practice furnishes us with a first test of the existence of a rational approach to our problems. In other words we should investigate whether this contradiction implies (1) that no rational approach may be obtained via the decision function formulation, (2) that our formulation has been more or less oversimplified, (3) that some of our postulates used in proving theorem 1 are too strong, or (4) that randomization is not as vital in statistics as it has appeared to be. First let us examine some of the reasons for the acceptance of randomized strategies.

We shall begin our discussion with the role of randomized strategies in game theory where the von Neumann and Morgenstern formulation of the zero sum two person game is strongly related to the decision function formulation of statistical problems and where the min max criterion (which frequently involves randomized strategies) has been shown to have optimal properties. These optimal properties of min max in the zero sum two person game derive from the fact that the application of this criterion assures a player of an amount v while a clever opponent can prevent him from getting more than v no matter what strategy the original player selects. Since the player can not expect to do better than v, there is no reason why he should not be willing to settle for v. It must however be pointed out that this argument which establishes the optimal character of min max does not discredit the other available criteria. For example, the author feels that for the von Neumann and Morgenstern formulation, it is not irrational to apply the counterproposal of selecting any of the admissible (i.e., non dominated) pure strategies of which the min max strategy is a mixture. The author feels that acceptable objections to the counterproposal involve considerations omitted from the formulation of the game. The most pertinent objection is
based on the secrecy consideration. While secrecy may be important in some real games, it plays no role in the von Neumann and Morgenstern formulation nor in statistical problems. Another objection which may be raised to the counterproposal is based on the fact that with the min max strategy, the player knows his payoff will be \( v \) (unless his opponent is stupid in which case it may be more than \( v \)) while he may not have such information if he were to use a pure strategy. The author feels that this objection has some of the attributes of a delusion because a player applying a mixed strategy must toss a coin (or apply some other random device) before selecting a pure strategy. After the coin has been tossed, he knows no more about his ultimate payoff than he would have known had he applied the counter-proposal.

In statistical problems of testing hypotheses concerning discrete distributions one occasionally finds a randomized strategy applied to obtain a test with a specified significance level. This randomized strategy generally involves the basing of a decision on the irrelevant toss of a coin after the data has been observed. Since this practice is a source of uneasiness to many statisticians and no justification is ever given for the special advantage of one critical size over all others, we can not consider this practice as evidence that anything is wrong with theorem 1.

However, randomized strategies do play a major and widely accepted role in the selection of random samples and in some problems of design of experiments. The usual justification for the selection of a random sample is based on the notion that if a sample were selected in an arbitrary fashion there would be no protection against the existence of an unknown relationship between the true state of nature and the sample which would yield misleading results. For example suppose it were desired to estimate the average height of individuals
in a population and it were convenient to use for a sample the first five people who walked by a spot near a cliff top. The statistician would be misled if it happened that tall people tended to avoid cliffs and he did not know this. Even if he knew this fact, to properly handle his data he would have to introduce a complicated model indicating the possible ways in which height affects the probability that an individual would avoid a cliff. The introduction of this model might involve a great deal of computational labor. If he used a random sample, he might sacrifice some convenience in sampling, but he would gain in that forgetting or ignoring the relationship between cliffs and height would not force misleading results upon him. In the above example, any two states of nature $s_1$ and $s_2$, which differ only in the relationship between height and fear of cliffs, are equivalent. (i.e. $u(s_1d) = u(s_2d)$ for all $d$ involving random samples.)

The implication of postulate 9 would then cut down considerably on the class of states of nature. However $s_1$ and $s_2$ may not be equivalent if non random samples are permitted.

It would seem that the need for randomization depends on the statistician's limited computational ability to take advantage of the actual problem and his consequent need to oversimplify the statement of the problem. Generally the simplification has the effect of combining states of nature which are equivalent when random samples are insisted upon. Hence this widely accepted use of random samples does not seem to represent a contradiction of theorem 1, when one considers the simplifying assumptions made in the introduction.

Let us consider the possibility that the postulates used in proving theorem 1 were too strong. The proof made use of postulates 3, 4 and 8. As
was pointed out in footnote 17, postulate 3 could have been replaced by
a weaker one which is quite innocuous. Objections to Pi have been made
but the author has never understood them. Finally, Rubin's postulate P8
is often a source of objection. Most of these, however, take the point
of view that a postulate which is responsible for such results is too strong.
The author considers this type of objection as pertinent but insufficient.

In view of these remarks it seems reasonable to assume that the conflict
between theorem 1 and common statistical practice is based on our simplified
formulation and can not be held against the postulates used nor against the
possibility of a rational approach via the decision function formulation.

Theorem 2 which depends only on Rubin's postulate states that the
statistician may base his choice of strategy on the regret matrix. Not only
does this result fail to conflict with common statistical practice, but it
has been regarded by some authors as axiomatic (although no one has shown
that differences in utilities do in fact measure some feeling which may be
called regret).

Theorem 3 states that one should proceed as though each state of nature
were equally likely. This criterion, frequently attributed to Bayes and to
Laplace has been subject to considerable objections for many years. The chief
objection is based essentially on the natural conflict which derives from two
interpretations of a problem if these two interpretations differ on the
question of what constitutes a state of nature. Another objection that may
be raised to the result in theorem 3 is that this theorem shows little promise
of being extendable to the case of an infinite set of possible states of
nature.
In our development we have been deliberately vague about what constitutes a state of nature and we have purposely avoided applying postulate 9 by restricting ourselves to the general problem \( Q_n \). It would seem that the major difficulties facing us would be resolved if there were available a unique characterization of the class of all possible states of nature which involved only a finite number of states of nature.

It is of some interest that theorem 3 suggests that one may regard the postulates 1-8 as an axiomatization of the "principle of insufficient reason." Hence if we were faced with the problem of betting on the outcome of the toss of a coin which may be biased but about which we know nothing, it would be perfectly reasonable to consider heads and tails the two possible states of nature and to assume that each had an a priori probability of \( 1/2 \). While this interpretation of states of nature may not be commonly accepted it is one for which the postulates seem applicable and one may therefore invoke theorem 3. It should be emphasized that this method would fail if we were first allowed to observe the result of a previous toss of the coin. In that case the states of nature would have to describe the relationship between the previous and future toss of the coin (that is the probability of a head), and we would have a much more difficult problem involving an infinite set of possible states of nature.

Unless one may claim that our simplifications have eliminated the possibility of a solution to our problem, or that some of our postulates have been unnecessarily strong \(^{19}\) or that the above-mentioned attempt to characterize the states of nature may succeed, it seems natural to conclude that no rational criterion exists for the class of real problems via the use of the decision function formulation. In that event we must either seek a new formulation or admit that "scientific procedure" is and must be based in
part on the individual instincts of the scientist carrying out the procedure. Even if the above-mentioned attempt is successful, it will be necessary to modify our formulation to take some account of the effect of computational difficulties.

7. Miscellaneous Remarks and Results.

It is of some interest to indicate which postulates the min max risk criterion fails to satisfy. It may be recalled that in section 2, an objection to the min max risk criterion was based only on an uneasy feeling concerning some consequences of applying this criterion. It is readily seen that min max risk fails to satisfy only Rubin's Postulate. Of course min max regret will satisfy Rubin's Postulate but will not satisfy Pl.

A serious difference between the postulates in our paper and those used by other authors centers about the notions of partial ordering versus complete ordering and complete ignorance versus prior knowledge. The author feels that there is no sound a priori reason to insist upon a complete ordering of the available strategies. On the other hand in the discussion of postulate 3 reasons have been advanced to accept the notion of complete ignorance. Still it is of some interest to examine the consequences of weakening postulate 3 to eliminate the assumption of complete ignorance.

Consider

**Postulate 3**. If \( Q_1 = (S, D_1, u_1) \) and \( Q_2 = (S, D_2, u_2) \) are elements of \( G \) such that \( d_2 = g(d_1) \) is a one-to-one transformation of \( D_1 \) onto \( D_2 \) where

\[
u_2(s, g(d_1)) = u_1(s, d_1) \quad s \in S, \quad d_1 \in D_1
\]

then \( C(Q_1) = g[C(Q_2)] \).
Definition 13. Let \( G_s \) be the class of all mixed problems based on a finite number of pure strategies with \( s \) as the class of states of nature.

Theorem 4. If \( S \) has \( n \) elements \( s_1, s_2, \ldots, s_n \) and \( C(q) \) is a solution for \( G_s \) which satisfies postulates 1, 2, 3*, 4, 5*, 6-9 then there is a set of nonnegative numbers \( \gamma_1 \) adding up to one so that for any \( q \in G_s \)

\[
C(q) = \{d: d \text{ maximizes } \sum_{i=1}^{n} \gamma_i u(s_i, d)\}.
\]

A proof of this theorem will be sketched in the appendix. It is noteworthy that this theorem does not quite imply that the solution is equivalent to assuming an a priori probability on the states of nature. For example it is easily seen that for \( n=2 \) the postulates 1, 2, 3*, 4, 5*, 6-10 are all satisfied by the following criteria. Let \( S = (s_1, s_2) \)

\[
C_1(q) = \{d: d \text{ maximizes } \frac{1}{3} u(s_1, d) + \frac{2}{3} u(s_2, d)\}
\]

\[
C_2(q) = \{d: d \in C_1(q) \text{ and } d \text{ maximizes } u(s_1, d) \text{ for } d \in C_1(q)\}.
\]

\( C_2(q) \) represents a solution which may be interpreted as equivalent to assigning a priori probabilities \( \frac{1}{3} \) and \( \frac{2}{3} \) to \( s_1 \) and \( s_2 \) respectively.

Another result of some interest is the following which was derived in the original investigation (see [9]) at a point when the author had some qualms about the convexity postulate 7.

Theorem 5. A necessary and sufficient condition that \( C(q) \) satisfy postulates 1-6, 5-10 for \( G_s \) is that there is an interval \((p, 1-p)\) which may be open at both ends or closed at both ends, \( 0 \leq p \leq 1/2 \) [the interval must be open if \( n=0 \) and closed if \( n=1/2 \)] so that \( C(q) = \{d: \text{the representative } x \text{ of } d \text{ maximizes } \gamma x_1 + (1-\gamma) x_2 \text{ for some } \gamma \in (p, 1-p)\}\).
This criterion reduces to the one of theorem 3 if \( p = 1/2 \) and to the class of all admissible (i.e., non dominated) strategies if \( p = 0 \).

Our main result (theorem 3) can easily be extended to the larger class of problems where one may have available an infinite set \( D \) of pure strategies providing \( \frac{1}{n} \sum_{s \in S} u(s,d) \) attains a maximum. The case where \( \frac{1}{n} \sum_{s \in S} u(s,d) \) does not attain a maximum is of some minor interest. The difficulty that is raised in this case would exist even if there were no ignorance. In other words if the state of nature \( s \) were known and

\[
u(s, d_i) = \frac{1}{1 + i} \quad i = 1, 2, \ldots\]

there obviously is no optimal strategy. This situation is usually bypassed by assuming that in real life some additional effect (for example the effort in selecting \( d_i \) for a very large integer \( i \)) would modify the problem to the extent that a maximum is attained. It is of interest to note that this example presents another reason for limiting the class \( G \) of problems under consideration. Of course this example gives no reason for selecting for consideration the special class \( G_n \) which was so convenient for us to treat.

The author wishes to present two tentative directions for further research. One is the combination of statistics with games which results when two or more players engage in a game while in ignorance of the true state of nature. The other is a slight modification of the decision function formulation which arises when one replaces \( u(s,d) \) by a pair \( u_1(s,d), u_2(s,d) \). This pair is supposed to represent a range of possible values of the payoff in case \( u(s,d) \) is computationally difficult or impossible to obtain.
To prove theorem 4, it will be convenient to consider representations relative to one specific relabeling only. That is, we denote the representation $R$ of the problem $Q = (S, D, u)$ where $S = (s_1, s_2, \ldots, s_n)$ by

$$R = \{x = (x_1, x_2, \ldots, x_n): x_i = u(s_i, d) \quad i = 1, 2, \ldots, n\}.$$  

As in definition 12 we denote the solution of the representation $R$ of $Q$ by

$$C_Q(R) = \{x: x_i = u(s_i, d) \quad i = 1, 2, \ldots, n, \quad d \in C(Q)\}.$$  

As in section 4, we obtain

**Lemma 7.** $C_Q(R)$ depends only on $R$ and may be designated by $C(R)$, $(P_2^e, P_3^e, P_6^e)$.

We shall designate the convex set generated by the points $a_1, a_2, \ldots, a_n$ by $[a_1, a_2, \ldots, a_n]$. Let

$$U_a = \{x: C[a, x] = [a]\}$$

$$V_a = \{x: C[a, x] = [x]\}$$

$$W_a = \{x: C[a, x] = [a, x]\}.$$  

**Lemma 8.** With the exception of the point $a$ which is in $U_a, V_a,$ and $W_a$, each point $x$ which represents a possible payoff function is in one and only one of the sets $U_a, V_a, W_a$ $(P_1^e, P_7^e, T_1^e, T_7^e)$.

**Proof:** Suppose $x \in (U_a \cup V_a)$. $C[a, x]$ exists $[P_1^e]$. If $a$ and $x$ are both elements of $C[a, x]$, then by convexity $[P_7^e], C[a, x] = [a, x]$. Otherwise some interior point of the line segment $[a, x]$ must be an element of $C[a, x]$.

But then the converse of convexity $[T_1^e]$ implies that $a$ and $x$ are elements of $C[a, x]$.

The following lemmas will be introduced to show that $U_a$ and $V_a$ are convex sets which may be separated by a unique hyperplane because $W_a$ has no interior.
points. The translation theorem \([T_2]\) will tell us that the direction of
the normal to the hyperplane is independent of \(a\) and our main result will
easily follow.

**Lemma 9.** \(U_a\) is a convex set \((P_1, P_4, P_5, P_8, L_7)\)

**Proof:** Suppose \(C[a, x] = [a] \quad C[a, y] = [a]\) and \(z = p \cdot x + (1-p) y\quad 0 \leq p \leq 1\).

We must show \(C[a, z] = [a]\). The case where either \(x\) or \(y\) is \(a\) is trivial \([P_8]\).

If \(x \neq a\) and \(y \neq a\) then neither \(x\) nor \(y\) are elements of \(C[a, x, y]\) \([P_4]\).

Every point of \([a, x, y]\) except \(a\) is representable as a mixture involving \(x\) or \(y\).

Hence \(C[a, x, y] = [a] \quad [P_1, T_1]\). But then \(C[a, z] = [a] \quad [P_5]\).

**Lemma 10.** \(V_a\) is a convex set if \(-1 < a_i < 1\) for \(i = 1, 2, ..., n\) \((P_8, P_2, P_7, L_7)\)

**Proof:** The translation theorem \([T_2]\) and Rubin's postulate \([P_8]\) upon which
theorem 2 depends tells us that the solution of a representation is invariant
under translation and contraction or expansion. It follows that if \(x\) and \(y\)
represent possible payoff functions and if \(a\) is an interior point of \([x, y]\)
then \(x \in V_a\) if and only if \(y \in V_a\). The convexity of \(V_a\) follows immediately
from the convexity of \(U_a\) \((L_7)\).

**Lemma 11.** \(W_a\) contains no interior points (in the topology of the \(n\) dimensional
space) if \(-1 < a_i < 1\) for \(i = 1, 2, ..., n\) \((P_2, P_4, P_5, P_7, L_7)\).

**Proof:** If \(W_a\) had an interior point \(x\), there would be an element \(y\) of \(W_a\)
such that \(x \neq y\). We shall obtain a contradiction proving that \(a \in C[a, x]\) and
\(y \in C[a, y]\) implies that \(x\) can not dominate \(y\). By Rubin's postulate \([P_8]\) it
suffices to treat the case where \(x\) and \(y\) are close enough to \(a\) so that
\[ b = \frac{x+y}{2}, \quad c = \frac{a+y}{2}, \quad d = \frac{a-x}{2}, \quad e = \frac{a+x}{2}, \quad \text{and} \quad f = \frac{x+y}{2} \]

all represent possible utility functions. (See Figure 2.)

![Figure 2](image)

Let us assume \( a \neq y \). Then it is clear that \( y \neq b \neq y, \quad e \neq c, \text{ and } f \neq a \neq d, \)
and that \( [a,e,x] \) and \( [d,c,e] \) are parallel lines as are \( [b,e,f] \) and \( [y,c,a] \)
and also \( [d,a,f], [c,e] \) and \( [y,b,x] \). Since every point of \( [a,x,y] \) not on \( [a,x] \) is dominated by a point on \( [a,x] \) it follows that

\[ a \in C[a,x,y] \quad [P_5] \]

Similarly

\[ b \in C[b,d,f] \quad [P_5, T_2] \]

But then both \( a \) and \( b \) are elements of \( C[a,b,c,e] \) \( [P_7] \). By convexity \( [P_7] \)
it follows that \( \frac{a+b}{2} = \frac{a+e}{2} < e \) is an element of \( C[a,b,c,e] \) which contradicts

Postulate 2.

**Lemma 12.** If \( -1 < a < 1 \), \( U_a \) and \( V_a \) are convex sets separated by a unique
hyperplane \( \sum_{i=1}^{n} \gamma_i (x_i - a_i) = 0 \) which contains all points of \( W_a \). Furthermore,
one may normalize so that \( \gamma_i \) are non-negative, add up to one and are
independent of \( a \). \( (L_7, L_9, L_8, \ldots, L_{12}, T_2) \)
Proof: Lemmas 9, 10 and 11 imply the existence of a unique hyperplane containing all points of \( W_a \). Since \( V_a \) contains all points \( x \) which dominate \( a \), the coefficients of the hyperplane are all non-negative or all non-positive. Since they do not all vanish one may normalize so that they are all non-negative and add up to one. Finally, the translation theorem \([T_2]\) tells us that the coefficients do not depend on \( a \).

Lemma 13. If \( R \) is the representation of a problem of \( G_S \),

\[
C(R) \subseteq \left\{ x : x \text{ maximises } \sum \gamma_i x_i \right\} \quad (P_1', P_2, T_2, L_7, L_{12}).
\]

Proof: If \( x \) maximises \( \sum \gamma_i x_i \) and \( a \) does not and \(-1 < a_i < 1 \) for \( i = 1, 2, \ldots, n \), then \( x \in V_a \) \([L_{12}]\). Then \( C(a, x) = \{ x \} \), and \( a \in C(R), [P_4] \). The case where some components of \( a \) are equal to 1 may be treated by translation and contraction \([P_8', T_2]\). Theorem 4 follows immediately from lemma 13.
Bibliography


1. The author feels that he has profited greatly from his contact with Herman Rubin, L. J. Savage, K. J. Arrow, Jacob Marschak, Roy Radner and I. N. Herstein, although this feeling may not be shared by some of the above-mentioned. The author wishes to express his special gratitude to Herman Rubin for contributing postulate B and to L. J. Savage for arousing the author's interest in the problem. Marschak, Radner and Herstein have made many useful contributions to the clarity of the paper. The importance to this paper of the literature in Statistics and Inductive Inference, in particular the work of Wald, Von Neumann and Morgenstern, is so obvious as to not need mention.

This paper was prepared with the support of the Office of Naval Research under contract Nonr-358(01) and of the Cowles Commission for Research in Economics. This paper is to be included in Cowles Commission Papers, New Series No. It supersedes mimeographed papers by the same author issued in 1949-50 (Cowles Commission Discussion Papers Nos. 326, 326A, 346, 346A).

2. Von Neumann and Morgenstern [2] have shown that under relatively mild conditions, one may attach a utility u to a prospect, (the term prospect was used by Marschak [3]). The utility is a real valued bounded function of the class of prospects with the properties (i) \( u(P) > u(Q) \) when \( P \) is preferred to \( Q \), (ii) the prospect of getting \( P \) with probability \( p \) and \( Q \) with probability \( (1-p) \) has utility \( pu(P)+(1-p)u(Q) \). This latter property states essentially that the expectation of utility represents a utility. We shall use utility in place of negative risk.

3. Some statisticians, including Savage, do not like the term regret and use risk for regret and loss for risk.
4. If randomized strategies are permitted, the min max regret strategy would be a mixture of $d_1$ and $d_2$ where $d_1$ is selected with probability 98/99.

5. An analogous notion was used by Arrow [5] in a different problem.

6. The presence of a class of criteria is considerably different from the presence of a unique criterion which occasionally allows any one of a subset of the strategies to be applied.

7. We shall for simplicity assume that regardless of the true state of nature, the set of possible values of utility is the interval $(-1,1)$ which may or may not contain either endpoint. A utility function is a function whose range is this interval.

8. The $S$ and $D$ may vary for different problems of $Q$. We may denote the $S$ and $D$ of a problem $Q$ by $S(Q)$ and $D(Q)$.

9. The symbol $f(C(Q))$ represents the subset of $D$ which is obtained by taking $f(D_1)$ for each $d_1 \in C(Q)$.

10. As we have previously mentioned the dominance relationship is always in reference to a particular problem. In this case it is obvious that $Q_2$ is the problem referred to. We shall omit the reference to the problem when there is no ambiguity.

11. The symbol $\implies$ is read implies.

12. The finite set of strategies may vary for different elements of $Q_0$.

13. A vertex of a closed convex set is a point of the set not on the interior of any line segment in the set. The set of vertices generates a closed convex set and is a subset of any set generating a closed convex set.

14. By a translation $c$ we mean the rigid motion where $R$ goes into $R_2 = \{x + c : x \in R_1\}$. 
15. Extending the notion of domination to representatives, we say that \( s \) dominates \( y \) if \( x \not\succ y \) and \( x_i \succ y_i \) for \( 1 \leq i \leq n \).

16. Since \( c \) is dominated by \( a \) and \( b \) is not, it follows that \( a \), \( b \) and \( c \) are not collinear. Hence the triangle is not degenerate.

17. In such a discussion it is tacitly assumed that the postulates used in the proof of the theorem are not unreasonably strong. The proof of theorem 1 involves postulates 3, 4, and 8 although a considerably weaker form of postulate 3 (involving only the relabeling of the strategies) could have been used. The weakened form which we shall later call postulate 3* would be acceptable to most people who object to postulate 3.

18. The min max criterion has the property that even if the opponent discovers the player's strategy he can do nothing to improve his position.

19. We shall say a few more words about the possibility that our postulates are too strong in the next section.

20. As was pointed out in footnote 17, theorem 1 can be proved using only postulates 3*, 4 and 8.