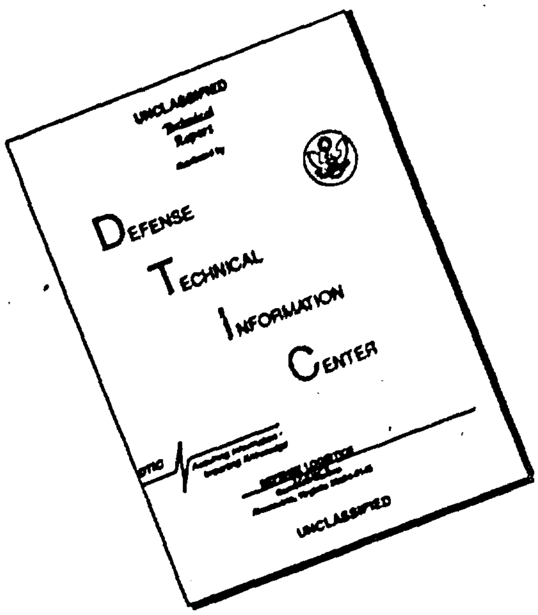
DISCLAIMER NOTICE



THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

OPERATIONS ANALYSIS PAPER • No. 5

THE APPLICATION OF STATISTICAL DECISION THEORY TO A MISSILE TESTING PROBLEM

WALTER L. DEEMER
Chief, Systems Analysis Team
Operations Analysis Office

and

JOHN P. MAYBERRY
Operations Analyst, Hq USAF

★ ★ ★ **DECEMBER 1961**

[redacted]

Operations Analysis
Office, Vice Chief of Staff
Headquarters, United States Air Force
Washington 25, D. C.

The Application of Statistical Decision Theory to a Missile Testing Problem*

ABSTRACT

Experiments are planned, performed, and analyzed for two main purposes: first, to discover facts; and second, to use those facts in making decisions.

Classical statistics typically considers these two purposes separately, with special emphasis on discovering (or estimating) the facts.

Statistical decision theory, on the other hand, considers each problem as a unified whole, so that the experiment and its analysis are greatly influenced (a) by the decision which is to be made, and (b) by the cost of the experiment itself.

As the price for treating more of the important factors explicitly, statistical decision theory requires more information than the classical methods. On the other hand, problems can be treated which are inaccessible to the classical methods, for example: "Shall we do any testing at all?"

*This paper is a revised and expanded version of a paper delivered at the World-Wide Operations Analysis Technical Conference - 12-13 October 1961

1. Introduction

In this paper, we wish to describe two approaches, each applicable to a wide variety of statistical problems. For expository purposes, we have chosen to apply both these approaches -- which we call, respectively, the classical approach and the statistical decision theory approach -- to a specific problem. In order to clarify the principles which we are presenting, we have sacrificed the realism of the example by making several simplifying assumptions.* Of course, both the methods discussed could be applied without making these simplifying assumptions, just as they could be applied to a wide variety of other real problems.

The central ideas of each of these approaches can be distinguished as follows:

a. Classical Statistics:
"What are the facts?"

b. Statistical Decision Theory:
"What shall we do?"

Because of the difference in these central ideas, there is a great difference between the two approaches in the parts of the total problem which are included in the mathematical model. Statistical decision theory, which focuses attention on the pragmatic question "What shall we do?", takes into consideration factors which are typically excluded from the classical approach.

*Two specific examples of this simplification: (a) We assume that all targets have the same value to the attacker and that all are equally hard; (b) we treat reliability alone, whereas real problems require that reliability and accuracy be treated together in the same model.

After reviewing both approaches, we will illustrate them by applying them to a specific example. The example concerns the question, "How many missiles should be expended in operational test firings?" It is important to realize that we are not concerned with R and D test firings, whose primary purpose is to improve missile design and reliability, but only with the later firings from operationally configured sites. The primary purposes of those later firings are (a) to evaluate missile reliability, and (b) to provide guidance for the allocation of missiles to targets.

2. Classical Statistics

Classical statistics is concerned with the theory of how to use the results of an experiment (a) to test hypotheses, or (b) to estimate parameters. (A hypothesis might be any statement whose truth is unknown, such as "Missile reliability is greater than 50%." A parameter might be any variable whose true value is unknown, such as the reliability of our missiles.)

A parameter estimate may be either (1) a point estimate, or (2) an interval estimate. (A point estimate is a statement such as "Based on these test results, reliability is 0.63." An interval estimate is a statement of the form "Based on these test results, we can say with 80% confidence that the reliability is between 0.53 and 0.71.") The latter type of estimate tells us how trustworthy the specific numbers may be.

a. Theory of Estimation.

Classical statistics emphasizes the development of theory to guide one in choosing functions of the sample observations for testing hypotheses, and for point and interval estimation. Not always ignored, but usually omitted from explicit treatment in the mathematical model, are considerations of testing costs and the loss associated with making a wrong estimate. There is no suggestion intended here that classical statistics could not take these things into account, implicitly or even explicitly; however, it is not usually done. For example, a brief search through several standard statistics texts reveals no mention at all of the cost of testing, except by implication in the discussion of sequential testing (since one of the stated advantages of sequential testing is that it decreases the average number of items tested).

b. Design of Experiments

We should also remark here that the large body of literature on the "design of experiments" does not (in general) refer explicitly to the cost of experimentation, but asks questions like "How much testing is necessary to achieve the precision and confidence required?" Of course, if the answer to this question implies an unreasonably costly experiment, the words "necessary" and "required" will be redefined and the question will be asked again. In this way the costs do have some effect on the classically-designed experiment, but in a subjective and imprecise way. In fact, most of the "design of experiments" is directed to the very different problem of avoiding undesired interactions between the things being measured or tested and random or irrelevant disturbances.

3. Statistical Decision Theory

Statistical decision theory, on the other hand, uses the same mathematical theory of probability which underlies classical statistics, but increases the scope of the mathematical model to include explicit consideration of (a) cost of testing, (b) decisions which may be influenced by test results, (c) cost of making the wrong decision, and (d) judgments on the likelihood of the various results (prior probabilities). One penalty for increasing the scope of the model is, of course, increased complexity. Another fact, often regarded as a penalty, is that statistical decision theory gives results which depend explicitly on the above four factors, so that the applicability of the results is usually limited to the one specific problem considered. However, we believe that those factors logically should have an effect on the amount of testing, and on the actions taken after observing the results, so that the generality obtained by other methods may be illusory.

The question, as we see it, is simply whether those four factors should be taken into account by explicitly including them in the model, or whether they should be taken into account by relying on the subjective judgment of the statistician. In the case of an experienced statistician, whose experience is relevant to the problem at hand, the classical approach will be more convenient; however, we can see no method of determing whether the statistician's experience is in fact relevant except by comparing his opinions with objective calculations which take the additional factors into account explicitly. For most operations analysis problems, and for our example in particular, the additional information needed for statistical decision theory can be made available by reasonable assumptions.

4. Our Specific Missile Testing Problem

The missile testing problem has two parts: first, "How many missiles should be fired in an operational test program?" and second, "How should the results of the test program be used as a guide in targeting?"

a. The Classical-Statistical Approach

The question of how many missiles to fire can be partially answered by the methods of classical statistics. Before explaining the method used, we will describe how confidence limits for reliability are determined.

(1) Confidence Limits

The confidence limits are based on the mathematical theory of probability and the binomial theorem, and are summarized in graphs of the type shown in Fig. 1.* The true reliability, R, is plotted on the ordinate and the observed fraction of successes in a sample, r_o, is plotted on the abscissa. Each pair of curves gives upper confidence limit R_U and lower confidence limit R_L, for a specific sample size (number of missiles tested) and a specific value of the confidence coefficient C, whose significance will appear in a moment. One enters the appropriate graph with the observed fraction of successes, r_o, draws a vertical line, and then reads R_U from the upper curve and R_L from the lower. Since R_U and R_L depend on r_o, the observed fraction of successes, which is of course a random variable, they too are random variables. The curves are so calculated that, if one determines R_U and R_L in this way from each of many experiments, using the same confidence coefficient C each time, then the random variables R_U and R_L will in fact bracket the true probability R a fraction C of the time.

(2) Choice of Confidence-Coefficient

The experimenter chooses C. Of course, he wants a large C; that is, he wants a high probability of covering the true value. He also wants a narrow band for R, that is he wants $R_U - R_L$ to be small. But, for a given sample size, as C increases, $R_U - R_L$ also increases, (as illustrated in Figure 1) so the experimenter must compromise. And it is here that the first weakness of classical statistical theory shows up. It offers no guidance as to how to resolve the conflict between the desire for large C and the desire for small $R_U - R_L$. Published aids for computing R_U and R_L are

*See Page 32.

typically for C = .8, .9, .95 and .99, and so a value from this set is usually chosen. But the practical implications of the choice of C and of $R_U - R_L$ are not ordinarily considered in the classical theory, and are very difficult even to state correctly.

(3) Choice of Number Tested

Further, if both large C and small $R_U - R_L$ are required, n must be large. Here again, the classical theory provides no guidance on how to resolve the conflict between the desire for large C and small $R_U - R_L$ on the one hand, and the cost of large n on the other. Past experience, judgment, and intuition must be used in making the choice. For example, a military headquarters could use its collective experience and judgment to select values of C and $R_U - R_L$ which were felt to be appropriate when $r_o = .5$. These chosen values could be used to determine the number of missiles to be tested, by means of confidence-limit curves of the kind shown in Fig. 1. If the values of C and $R_U - R_L$ that were chosen were appropriate for the headquarters' needs, then this would be a valid and satisfactory method of choosing the number to test.

(4) Difficulties in Making Those Choices

However, because the implications of a given choice are not obvious, there will surely be disagreement as to the most reasonable values. This problem is aggravated by the fact that we would like to employ the judgment of military officers who may not have statistical training, and who will find it difficult to direct their judgment towards the choice of C and $R_U - R_L$. It seems that a quantitative assessment of the various choices would have to employ some or all of the additional information used in statistical decision theory, and employ it in rather similar ways.

(5) Using the Classical Estimates

We should also notice here that the second question, "How to use the results of the test program as a guide in targeting" is not readily answered by classical theory, even after the experimenter has determined the number of missiles to test. The straightforward, uncritical use of an unbiased point estimate of the reliability as if it were the true reliability may lead to poor decisions, particularly if the cost function is biased. By using judiciously biased point estimates one may be able to compensate for biased cost functions, but classical theory offers little guidance on how to be judicious in this selection.

b. The Statistical Decision Theory Approach

We shall consider now the statistical decision theory approach. Since the decision theory model explicitly includes more than the classical model, we must gather more information before the model can be used.

(1) The Set of Possible Actions

For example, we must formulate the purpose of the testing in such a way that it refers to the actions that may be taken and to the choice between them. Thus we cannot simply state that we wish to "discover the reliability"; a suitable statement of the purpose of the testing would be "we test to determine how to allocate missiles to a target system." Thus the end result of our tests will be a choice of one allocation policy. In our example, we are attempting to choose the number of missiles to be allocated to each target actually attacked by these missiles, where we have previously decided to allocate the same number - at most six - to each target. Then our six possible policies are: allocate 1 missile to each target; allocate 2 missiles to each target;...; allocate 6 missiles to each target. (Note again that we are ignoring what the missile engineer may think of as the primary purpose of these tests, namely, to find hidden bugs in the system. In fact, of course, operational tests are never entirely separate from the development tests -- the purposes overlap a little -- but the present example restricts our attention to the operational-testing aspects.)

(2) Measure of Merit

The next step is to choose a measure of merit -- a "payoff-function" -- and to calculate its value for each allocation and each "state of nature".* In many practical problems this choice is the most difficult step -- one which requires and deserves a great deal of careful thought. The techniques of the analysis are independent of the validity of the choice of the payoff function; of course, the usefulness of the result depends critically on this choice.

For our example we have a chosen a measure of merit based on the idea that we wish to have a high assurance that those targets actually attacked are destroyed. We choose, therefore, the assurance level that we desire, and denote it by Q. Notice that this assurance level Q, unlike the confidence coefficient C, has an explicit operational significance. We give ourselves credit for a target destroyed only if the number of missiles assigned to the target is large enough to make the probability of destruction of the target at least Q.

For this example we make these simplifying assumptions: first, that the weapon yield and the missile accuracy are such that a missile that operates properly (i.e., a "reliable" missile) will reach the target and destroy it; second, that all targets are of equal value and we wish to achieve the same assurance of success Q for each one; third, that no other weapons are used on these same targets; fourth,

*A "state of nature" is simply a fact about the real world, which may be considered as a "choice" made by "nature". The concept is useful when we are dealing with an unknown quantity (such as the reliability in our example); a value, or range of values of that quantity may be called a "state of nature".

that a missile will not damage targets that it was not aimed at; and fifth, that all missiles have the same probability of operating properly (this probability is denoted by R and is called reliability).

The probability that a target will survive an attack by k missiles, each of reliability R, is then

$$(1 - R)^k,$$

and the probability that it is hit by at least one of those k missiles is

$$1 - (1 - R)^k.$$

Now, if we knew R we could simply calculate the smallest k for which

$$1 - (1 - R)^k \geqq Q,$$

and assign k missiles to each target.

It is clear that if testing cost nothing, we could test until (with any desired probability of any desired precision) we knew the actual value of R, and so could choose the correct value of k for the actual situation.

We will now show how statistical decision theory helps us take account of the fact that testing is not free. First we draw up a table showing the payoff -- the average number of targets killed per missile in the stockpile -- under various conditions. This payoff is of course a function of the number of missiles <u>required</u> per target (which depends on the "state of nature", or in this case on the value of the unknown R), and is also a function of the number of missiles actually <u>assigned</u> per target. For example, assume that we want 90% assurance of target destruction (or, in our notation, $Q = 0.90$); then, using the condition that $1 - (1 - R)^k$ must be at least Q, we can calculate the range of R-values for which 1 missile is required, 2 missiles are required, etc., and so generate the first two columns of Table 1.

The lowest value of R in Table 1 is 0.319 because we assume for this example that the results of the R and D test firings permit us to ignore the possibility of R being less than 0.319.

The zeroes in the lower left corner of the payoff matrix result from the assumption that no credit is given for attacking a target with fewer missiles than are required for a probability Q of destroying the target. The Commander wants to know with 90% assurance that certain specific targets were killed. To illustrate how the rest of the table is constructed we calculate the starred value in row 2, column 4. We are allocating four missiles to each target so that each four missiles suffice for one target. The average is 1/4 target per missile, and that number of targets will in fact be killed with probability greater than 90%, since two missiles would have been enough.

The reader may have misgivings about using this particular payoff function. We think that it is reasonable; but in any case it will serve as an example of a biased payoff function, since allocating too few missiles is much more costly than allocating too many.

(3) Prior Distribution

Statistical decision theory requires, in addition to the payoff matrix, a "prior probability distribution on the states of nature." This must be developed as the consensus of opinion of the statistician and the executive and should be an attempt to exploit whatever information and judgment are at hand. In our hypothetical missile-testing example, we might feel fairly certain that the reliability R lay in a certain range; if we could then agree on an estimate of betting odds for R lying in various intervals, and if we could agree on the general form of the distribution of R, we could make <u>an estimate</u> of the probability of each interval in the first column of Table 1, and could compute from those numbers the

TABLE 1

Payoff Matrix: Average Number of Targets
Correctly Identified as Killed with 90% Assurance (per missile allocated)

Range of R	Number of Missiles REQUIRED per Target	Number of Missiles ALLOCATED per Target 1	2	3	4	5	6
0.900-1.000	1	1	1/2	1/3	1/4	1/5	1/6
0.684-0.899	2	0	1/2	1/3	1/4*	1/5	1/6
0.536-0.683	3	0	0	1/3	1/4	1/5	1/6
0.438-0.535	4	0	0	0	1/4	1/5	1/6
0.369-0.437	5	0	0	0	0	1/5	1/6
0.319-0.368	6	0	0	0	0	0	1/6

Note: The zerces indicate cases where the number of missiles allocated per target was insufficient to achieve 90% assurance of success. Otherwise, the number of targets killed per missile is the reciprocal of the number of missiles allocated per target.

* Calculated in the text as an example.

probability of requiring various numbers of missiles. That estimated distribution, based on prior experience, is called the prior (or _a priori_) distribution.

The concept of prior probability distribution lies at the center of statistical decision theory; it is treated at length in sections 5 and 6 of this paper.

(4) Preposterior Analysis

Before we perform any experiment, or even decide what experiment to perform, we will make what Raiffa and Schlaifer (ref. 4), call the _preposterior analysis_. The purpose of that analysis is to evaluate possible experiments. In the course of the analysis we shall learn how to use the results of the experiments.

Consider for the moment some proposed experiment E, and some one possible result X of that experiment. (In our example, choosing the experiment consists simply of choosing the number $\underline{n}$ of missiles to be tested, and the result can be expressed by giving the number $\underline{m}$ of successful firings out of those $\underline{n}$ tests.) On the basis of that result X we will modify our prior distribution in the following way: each state of nature that would have made the observed results probable will be considered more likely than before (i.e., its posterior probability will be larger than its prior probability); and each state of nature that would have made the observed results improbable will be considered less likely than before. (In particular, any state of nature incompatible with the observed result X will be excluded from further consideration.)

(a) Bayes' Formula

The above ideas are made precise by Bayes' formula, which, in its general form is

$$P(A|X) = \frac{P(X|A) \cdot P(A)}{P(X)} \quad ,$$

where P(A) is the prior probability of the state of nature A, P(X) is the prior probability of observing the result X, and P(A|X) is the posterior probability of the state of nature A if the experimental result X has been observed.* The formula can be interpreted in the following way:

(i) the prior probability P(A) of the state of nature A is multiplied by the probability P(X|A) of the observed result occurring if A were true; the resulting number P(A).P(X|A) measures the <u>relative</u> weight to be given to A in the posterior distribution;

(ii) that set of relative weights (one for each state of nature A) must be divided by their sum, to get a set of weighting-factors (which must add up to unity);

(iii) the formula quoted above results as soon as we remark that the sum of those numbers, for all possible A and some fixed X, is just P(X); i.e.,

$$\sum_A (P(X|A).P(A)) = P(X). \; **$$

In our missile-testing example, the state of nature corresponds to some value or set of values of the unknown reliability R, and the test result will be <u>m</u>, the number of missiles (out of the total of <u>n</u> tested) which were successful. Bayes' formula now lets us calculate P(A|X), the posterior probability of the state of nature A after we have performed the experiment E and observed the test result X.

*The formula can be established by observing that the probability of the joint event A and X is expressible either as P(A).P(X|A) or as P(X).P(A|X); on equating these expressions and dividing by P(X), <u>Bayes' formula</u> results.

**The prior probability P(X) is the average (over all states of nature) of the probability P(X|A) of observing X if A is true, weighted according to the prior distribution P(A).

(b) Optimal Terminal Actions

The posterior distribution obtained above represents our judgment (as modified by the test results) of the plausibility of the various states of nature. For any one terminal action (i.e., in our example, for any one allocation policy), we have a set of payoffs, one for each state of nature; from that set of payoffs we can calculate the expected payoff (using the posterior probabilities on the states of nature as weights). Then we may select, for the proposed experiment E and possible result X, that terminal action which offers the greatest expected payoff, and record it as the optimal terminal action. We also record the expected value of the corresponding payoff.

(c) Expected payoff from an experiment

Consider now the other possible outcomes of the particular experiment E, and calculate for each of them the optimal terminal action and the expected value of the corresponding payoff.

When that process has been completed, we must consolidate those expected payoffs, which depend on the result X as well as on E. For this purpose we average them, using the prior probabilities P(X) of the various outcomes as weights, to obtain the overall expected payoff of the experiment E.

(d) Comparison of experiments

Now consider the totality of experiments under consideration, and perform the above expected-payoff calculation for each of them. Assess also the costs of each experiment. If the costs of the experiment can be expressed in the same terms as the payoffs, then the "design of the experiment" consists only in choosing that experiment which maximizes the difference (payoff minus cost). (The word "only" may be misleading, because an enormous amount of calculation may be required in this preposterior analysis.) If, on the other hand, the payoffs are incommensurable

TABLE 2

Calculation of Payoff Expected If We Test One Missile (n = 1)

State of Nature A	Value of R Representative of State of Nature A. R = P (success | A)	Number of Missiles Required Per Target	Prior Probability P (A)	Number of Missiles Allocated Per Target 1	2	3	4	5	6	Posterior Probs. If We Test One Msl: Fail (m = 0) P(A|0)	Succeed (m = 1) P(A|1)
1	.950	1	.210	1.000	.500	.333	.250	.200	.167	.050	.252
2	.850	2	.322	0	.500	.333	.250	.200	.167	.231	.346
3	.742	2	.276	0	.500	.333	.250	.200	.167	.340	.259
4	.610	3	.161	0	0	.333	.250	.200	.167	.300	.124
5	.487	4	.024	0	0	0	.250	.200	.167	.059	.015
6	.404	5	.005	0	0	0	0	.200	.167	.014	.003
7	.344	6	.002	0	0	0	0	0	.167	.006	.001
Expected Payoff, No Testing				.210	.404	.323	.248	.200	.167	choose to allocate 2 msls per target	
Expected Payoff, Test 1, Fail				.050	.310	.307	.245	.199	.167	choose to allocate 2 msls per target	
Expected Payoff, Test 1, Succeed				.252	.428	.327	.249	.200	.167	choose to allocate 2 msls per target	

Overall expected payoff if we do no testing is 0.404 targets per missile,

Overall expected payoff if we test 1 missile is

$0.791 \times 0.428 + 0.209 \times 0.310 = \underline{0.404}$ targets per missile.

with the costs of the experiment (e.g., the cost of an experiment is dollars, while the payoff is military effectiveness of one kind or another) it is necessary for the statistician to confer with the decision-maker, and employ the latter's judgment about the operational implication of the various payoffs, to decide which experiment to perform.*

Note that we have already made the calculations which tell us which terminal action to take when the experiment is complete and its result has been observed.

(5) Computations for the example.

In Table 2, a possible prior distribution (whose origin will be described in section 6 below) has been inserted as column 4. Table 2 represents the same situation as Table 1, with two changes: (a) the range of R, within which two missiles per target are required, has been divided into two smaller ranges in order to improve the precision with which later calculations can be made; (b) the midpoints of the ranges of R have been introduced in order to have a point value for use in later calculations.

We consider first the experiment n = 1, in which we test one missile. (Throughout this example we assume that the test missiles do not come from the stockpile, so that the number of missiles available to fire against targets is not reduced by the test program.) The possible outcomes are m = 0 (no successes, i.e., the missile fails)

*The concept of cost associated with testing may be a much richer concept than the term "cost" ordinarily suggests. Statistical decision theory permits it to include not only the dollar cost of production of test missiles, but also, for example, as negative costs, the advantages accruing from keeping production lines operating and from the crew training and experience which result from test firing. Insofar as this enrichment adds incommensurable elements to the cost, making cost a vector instead of a scalar, it complicates the later choice of an experiment, but the principles of the statistical theory are not affected.

TABLE 3

Calculation of P(X)

State of Nature A	Prior Probability P(A)	P(X|A) Conditional Probability of Success	P(X|A).P(A)
1	.210	.950	.199
2	.322	.850	.274
3	.276	.742	.205
4	.161	.610	.098
5	.024	.487	.012
6	.005	.404	.002
7	.002	.344	.001
		Total:	.791 = P(X).

and m = 1 (one success, i.e., the missile succeeds.) The posterior probabilities for m = 0 or 1 respectively are given in the two columns at the extreme right of Table 2, which are calculated directly from Bayes' formula (cf. paragraph 4b(4)(a)). For example, let us calculate the second entry in the last column. Bayes' formula is $P(A|X) = \frac{P(X|A).P(A)}{P(X)}$, where the A is state of nature No. 2, $0.800 \leq R \leq 0.899$; X is the result "the missile succeeds"; P(A) = 0.322 from column 4 of Table 2; P(X|A) is the probability of X (i.e., success) given that R is in the specified range, and we take 0.850 as a mid-range representative value of the reliability, so that P(X|A) = 0.850. Thus P(X|A).P(A) = 0.850 x 0.322 = 0.274. To calculate P(X) we must do the same calculations for all values of A, and form the sum as shown in Table 3. We find that $P(X) = \sum_A P(X|A).P(A) = 0.791$. Then $P(A|X) = \frac{P(X|A).P(A)}{P(X)} = \frac{0.274}{0.791} = 0.346$, as appears in the second place of the last column, Table 2.

The expected payoff (in targets per missile in stockpile), for each of the six methods of missile assignment considered, is shown at the foot of the column corresponding to that assignment in Table 2: each value in the row called "Expected Payoff, Test 1, Fail" is obtained by averaging the seven payoffs in the corresponding column, using the weighting-factors given by the posterior probabilities for m = 0; the values in the row called "Expected Payoff, Test 1, Succeed" are similarly calculated using the posterior probabilities for m = 1.

The maximum value in each row is underlined; it happens to be the value corresponding to "allocate two missiles per target," both for the "Fail" row and for the "Succeed" row. The decision-rule implied by these calculations, if we decide to test one missile, is to allocate two missiles per target, whether the missile succeeds or not. The overall expected payoff, in case we decide to test one

TABLE 4

Calculation of Payoff Expected if We Test Two Missiles (n = 2)

State of Nature A	Value of R Representative of State of Nature A. R = P (success | A)	Number of Missiles Required Per Target	Prior Probability P (A)	Number of Missiles Allocated Per Target 1	2	3	4	5	6	Posterior Probabilities If We Test Two Missiles 0 Succeed P (A|0)	1 Succeed P (A|1)	2 Succeed P (A|2)
1	.950	1	.210	1.000	.500	.333	.250	.200	.167	.009	.006	.296
2	.850	2	.322	0	.500	.333	.250	.200	.167	.122	.274	.364
3	.742	2	.276	0	.500	.333	.250	.200	.167	.308	.353	.237
4	.610	3	.161	0	0	.333	.250	.200	.167	.411	.256	.093
5	.487	4	.024	0	0	0	.250	.200	.167	.107	.040	.009
6	.404	5	.005	0	0	0	0	.200	.167	.029	.008	.001
7	.344	6	.002	0	0	0	0	0	.167	.014	.003	.000
Expected Payoff Test 2 and	0 Succeed (m = 0)		Probability of m successes .060	.009	.220	.283	.239	.197	.167			
	1 Succeeds (m = 1)		.299	.066	.346	.316	.247	.199	.167			
	2 Succeeds (m = 2)		.641	.296	.448	.330	.250	.200	.167			

Overall expected payoff if we test 2 is:

$.060 \times .\underline{283} + .299 \times .\underline{346} + .641 \times .\underline{448} = \underline{.408}$ targets per missile

missile, is to allocate two missiles per target, whether the missile succeeds or not. The overall expected payoff, in case we decide to test one missile, is equal to the probability of failure times the payoff expected in case of failure, plus the probability of success times the payoff expected in case of success. This is the calculation at the bottom of Table 2; since, as shown in Table 3, the overall probability of success is 0.791, the overall probability of failure must be 0.209, and the expected payoff is 0.404 targets per missile.

Naturally, we include in the set of "experiments" to be considered the case where we do no testing. This is the case $n = 0$ (test no missiles), and the posterior distribution is of course identical with the prior distribution. The expected payoff for each method of allocation, if we take $n = 0$, is given in the row of Table 2 labeled "Expected Payoff, No Testing." Each of these values was obtained by averaging the payoffs in the corresponding column, using in this case the prior probabilities as weights. Again, the maximum value would have been obtained when two missiles were allocated per target, and the expected payoff would have been 0.404. We see that testing one missile would not increase the expected number of targets killed per missile. This is not surprising when we realize that we would take the same action (allocate two missiles) for each of the two possible outcomes of the single test; since no possible outcome of the experiment would affect our action, the test is not worth doing. (Of course, such a test might affect other actions not considered in the model, and it might be worth doing for those other reasons.)

We may make similar calculations to decide whether we wish to test more than one missile. For example, Table 4 shows the calculations for the proposal to take $n = 2$, (i.e., to test two missiles.) The expected payoff is then 0.408 targets per missile instead of 0.404. If 20 missiles are tested (the calculations are not exhibited in this paper) the expected payoff is 0.480 targets per stockpile missile. If 40 are tested, the expected payoff is 0.512 targets per missile. (For example, with a stockpile of 300 missiles, we find that testing two missiles increases the number of targets killed from 121 to 122, testing 20 missiles increas s the number of targets killed to 144, and testing 40 increases it to 154.) This information is summarized in Figure 3, page 34.

The results of calculations of this type, as presented in Figure 3, furnish the data needed to make a decision on how many missiles to test. However, the use of data of this kind to decide how many missiles to test may require military, political, and economic judgment. The statistician, as such, is not qualified to make the decision; although his advice should be given when asked for, the final decision should not be his.

(6) Employment of the Computational Results

Our exposition so far has shown how information from missile testing can be used to increase the number of targets destroyed by a given missile force. The choice of the number of missiles to test would appear to require simply a comparison of the cost of testing and the value of the extra targets destroyed. But a direct comparison of these two variables can only be made when they are both measurable in a common unit (such as dollars). We first give in paragraph 4.b.(6)(a) an example showing how the choice is made when such a common unit does not exist; then the example of paragraph 4.b.(6)(b) illustrates how to use such a common unit, in case it does exist.

(a) In the present section we show how the choice may be made without an explicitly stated common unit of measurement for cost of missiles tested and value of targets destroyed.

The essence of the method is a liberal application of intuition and judgment to numbers of the kind presented in Table 5, which are prepared by the statistician. The intuition and judgment are furnished by the executive, who examines the rows that correspond to possible stockpiles -- just one row if he knows what the stockpile is or will be. Let us assume that the stockpile is firmly established at 300 missiles. The executive examines the number of targets killed in the 300-missile stockpile row, noticing, for example, that the first 10 missiles tested buy 12 extra targets destroyed (from 121 to 133) and that ten missiles tested after 40 have already been tested buy only three extra targets destroyed (from 154 to 157).

In some tactical situations an extra three targets destroyed may be very important. If the executive predicts that this is the situation he will be faced with, he may decide to test 50 missiles. On the other hand, he may feel that the 150 targets killed

TABLE 5

No. of Msls tested	0	10	20	30	40	50
Msl Stockpile						
100	40	45	48	50	51	52
200	81	89	95	100	102	104
300	121	133	144	150	154	157
400	162	178	191	200	205	209
500	202	223	239	250	256	261

with 30 missiles tested is the right place to operate, since the extra four targets killed by increasing from 30 to 40 missiles tested are not sufficiently important to justify the cost of testing 10 missiles. If the actual stockpile is not firmly established, possible stockpiles are examined just as the 300 stockpile was examined; the choice of number of missiles to test is made by subjective weighting of the choices for each stockpile.

Of course, analysis of this kind implicitly involves a common unit for measuring cost of missiles tested and value of targets destroyed. Nevertheless, a rational executive may find this method of decision more congenial than one which requires him to begin by making an explicit assignment of a dollar value to targets destroyed.

(b) For this second example, we assume that the cost and value data can be expressed in a common unit. Specifically, we assume that the cost of testing n missiles is simply nC dollars, where C is the cost of testing one missile, and that the value of attacking one additional target can be expressed in monetary terms as KC dollars, where K is some factor (presumably greater than 1) which we must estimate. For example, if the value of attacking one additional target was considered equal to the cost of a test missile, we would assign the value of $K = 1$; if the value of attacking an additional target was considered to be ten times the cost of a test missile, we would assign the value of $K = 10$. Believing that a target attacked is worth much more to us than the cost of the missile attacking it, we may feel that K should be between 1 and 10.

Denote by $T(n)$ the average number of targets struck per missile, if we test n and make the optimal allocation; a graph of $T(n)$ against n is given in Figure 3. Then the total number of targets attacked, if we have a stockpile of S missiles and test n

missiles, is $ST(n)$, and the value to us of attacking these targets, at KC dollars per target, is $KCST(n)$. The cost of testing n missiles is nC, so we wish to maximize the difference $KCST(n) - nC$, which equals $KCS(T(n) - \frac{n}{KS})$. This will be maximized when the curve $T(n)$, plotted as a function of n as in Figure 3, (page 34), has slope $(KS)^{-1}$. Tangent lines are drawn on Figure 3 with slopes $\frac{1}{100}$, $\frac{1}{234}$, and $\frac{1}{1000}$, showing that:

> for $KS = 100$ -- either $K = 1$ and a stockpile of 100 missiles, or $K = 10$ and a stockpile of 10 missiles -- we should do no testing;
>
> for $KS = 234$ it is a matter of indifference whether or not we do any testing, but <u>if we do test</u> we should test about 11 missiles; and
>
> for $KS = 1000$, we should test about 44 missiles.

If we take $K = 3$ for example, we show in Table 6 how the number tested depends on the stockpile S. In this case a stockpile of 300 missiles would imply that we should test 40 missiles.

TABLE 6

S, Number of Missiles in Operational Stockpile	n, Number of Missiles to be tested
< 78	0
78	0 or 11
100	15
120	18
150	22
200	29
300	40

5. Another Objection

The key to the procedure that we have expounded above is the assignment of a prior distribution to the states of nature.

a. The strongest argument against the procedure is that the concept of a prior distribution of a parameter (in our example, R) is logically meaningless; it can be argued that R is not a random variable, but a fixed (though unknown) value. This question, which really concerns the foundations and philosophy of probability theory, has not received, and probably never will receive, universally accepted answers. Nevertheless, some counter-arguments can be made:

(1) First, for classical statistics to be useful at all, one must in some way weight the possible states of nature. Furthermore, the implicit weighting which characterizes most applications of classical statistics is subject to the same logical objections as Bayesian statistics, but fails to yield any of the benefits of the Bayesian methods. For example, one decides about the alternatives among which one is going to choose, and one thereby <u>completely excludes</u> from formal consideration a host of imaginable alternatives. As another example, in setting up an unbiased estimator as a desirable type (as is normally done in classical applications) one is implicitly assuming that the expected cost of an error in one direction is equal to the expected cost of an error in the other direction -- which can only be true for one ratio of the relative probabilities. Calling the prior distribution on the states of nature a weighting-function (which might be subjective), rather than a probability distribution, is permissible (since both satisfy the postulates of probability theory) if that soothes the conscience of the logical objector.

(2) Second, any statement of probability can be regarded as an expression of ignorance. Before we flip a coin, most people will permit us to say that the probability of a head is 1/2 (or perhaps some other number near 1/2). After the coin has been flipped, but

before we have seen it, the probability of a head is surely either 1 (if it is a head) or 0 (if it is a tail). However, it may not be wholly unreasonable to maintain that the (subjective) probability, until we obtain some information about the result, should be considered 1/2. Setting a prior distribution on a fact (namely, the actual but unknown value of R) seems to be similar in spirit.

(3) Third, modern physics forces us to admit the possibility that length, position, velocity, etc., are primarily probability distributions, which only superficially appear to be definite values, rather than the converse. And the reliability of a missile, or the truth of a hypothesis, can hardly be considered more concrete and deterministic than the position of a particle.

b. Another difficulty, even if one admits that the concept of a prior distribution can be meaningful, is that the determination of such a distribution may be arbitrary, subjective, and unscientific. Several remarks must be made in answer:

(1) First, it may happen that the prior distribution actually has little or no effect on the decisions we make. In our missile-testing example, the influence of the prior distribution on our allocation of missiles will be less important if large numbers of missiles are tested; then prior distributions that appear very different may lead to decisions that are very similar.

(2) Second, the existence of a power known as "judgment" shows that subjective opinions need not be arbitrary and may be scientific. Such opinions may be based on a body of facts and experience which are valid even though we cannot quote them explicitly.

(3) Third, if it is true (as we maintain) that one cannot draw any conclusions from an experiment without some prior assumptions, then the difficulty of making good prior assumptions is regrettable but irrelevant. For example, Occam's Razor ("Prefer the simplest

hypothesis consistent with the facts") can be regarded as a prejudice in favor of simple physical laws; and modern science could hardly exist without that unscientific prejudice.

(4) Finally, almost any experiment must be incomplete in the sense that its subject is connected to other questions, on which experiments might also be performed. To make a problem finite, we must ignore many of those connections. However, the posterior distributions which resulted from experiments previously performed may influence (or even determine) the prior distribution for the present experiment -- and the posterior distribution of the present experiment may influence the prior distribution for subsequent experiments. (In other words, a parameter estimated by one experimenter should be used by later experimenters with due regard to the variability of the estimate). If the result of any particular experiment is very surprising, we may be forced to retract our prior distribution, saying, "This is a paradox". When the paradox has been explained, a quite different prior distribution may be postulated for the next experiment; if the results of this new experiment are then within the new expected range, we say, "A new theory has been confirmed".

6. A Practical Problem

If we accept the principle of a prior probability distribution, we are left with the practical problem of estimating its numerical values. This is never easy, but frequently one can choose the general form of the distribution from first principles, and employ what knowledge we have to estimate the actual distribution.

For example, the numbers given in column 3 of Table 2 were obtained in five steps. First, we assumed that a binomial distribution with some (unspecified) mean would give a distribution of the right general shape. Second, we assumed that R, the reliability, was not less than 0.32 -- i.e., we assumed that not more than 6 missiles would be required to give 90% assurance of a kill. Third, on the basis of hypothetical research and development tests, aided by opinions on the effects of the changes made during development, we guessed that R looked like 0.8 or so. Fourth, a binomial distribution with mean 0.8 was selected which gave a reasonable spread of values for R; we actually chose the binomial distribution of 10 trials. Fifth, we drew a smooth curve through the points representing that distribution, as shown in Figure 2,* and drew vertical lines separating the ranges of R in which 1, 2, ..., 6, missiles were required. Sixth, by integrating the various regions under the curve, we estimated the probability of R lying in each of the ranges considered. Those prior probabilities were given in column 3 of Table 2.

7. Summary

We see that the Bayesian approach has done the following for us:

a. It has permitted a direct comparison of tangible, understandable values to decide whether testing pays, and to permit us to choose the number to test. (As discussed in paragraph 4c and footnote, the direct comparison may nevertheless involve a high order of judgment).

b. For whatever number we decide to test, it tells us explicitly how to use the test results to allocate missiles to targets.

c. It permits an experiment, considered as a process for gathering information, to be regarded from a unified viewpoint. Instead of beginning in a vacuum (with no information about the objects under test) and ending with a confidence-interval (whose interpretation in practical terms is very difficult), we begin with opinions (in the form of the prior distribution) as to the relative likelihood

*See Page 33.

of the various possibilities, and end up with modified opinions (in the form of the posterior distribution) on those likelihoods. This makes it much more reasonable to split up a complex set of experiments into sub-experiments, using the posterior distribution from one experiment as the prior distribution for another, and makes it possible to relate distinct experiments on the same subject in a consistent way.

If the possible number of missiles one might test is large, hand calculations of the kind presented in our examples for $n = 0, 1,$ and 2, would be very arduous. Two alternatives are available: for some payoff functions and some prior probability distributions, closed-form solutions (formulae for quickly calculating the answers) can be obtained; when closed-form solutions are not possible the numerical results may be generated on a high-speed computer. Research on the former attack is going forward at Stanford University and the University of Chicago under AFCOA sponsorship. In addition, a program for the IBM 7090 has been written at AFCOA which permits a wide choice of prior distributions, and which calculates the payoff in accordance with the methods of this example. A subsequent OA paper will describe the program in detail.

8. Conclusion

It is our opinion that the information for applying statistical decision theory, including a prior distribution, can often be estimated and made available. In many practical cases, more operationally meaningful results can be obtained by using statistical decision theory than by using the classical statistical estimation theory.

REFERENCES

1. David Blackwell and M. A. Girshick,
   <u>Theory of Games and Statistical Decisions</u>
   <u>Wiley, 1954</u>

2. Herman Chernoff and Lincoln E. Moses,
   <u>Elementary Decision Theory</u>
   <u>Wiley, 1959</u>

3. R. D. Luce and Howard Raiffa,
   <u>Games and Decisions</u>
   <u>Wiley, 1957</u>

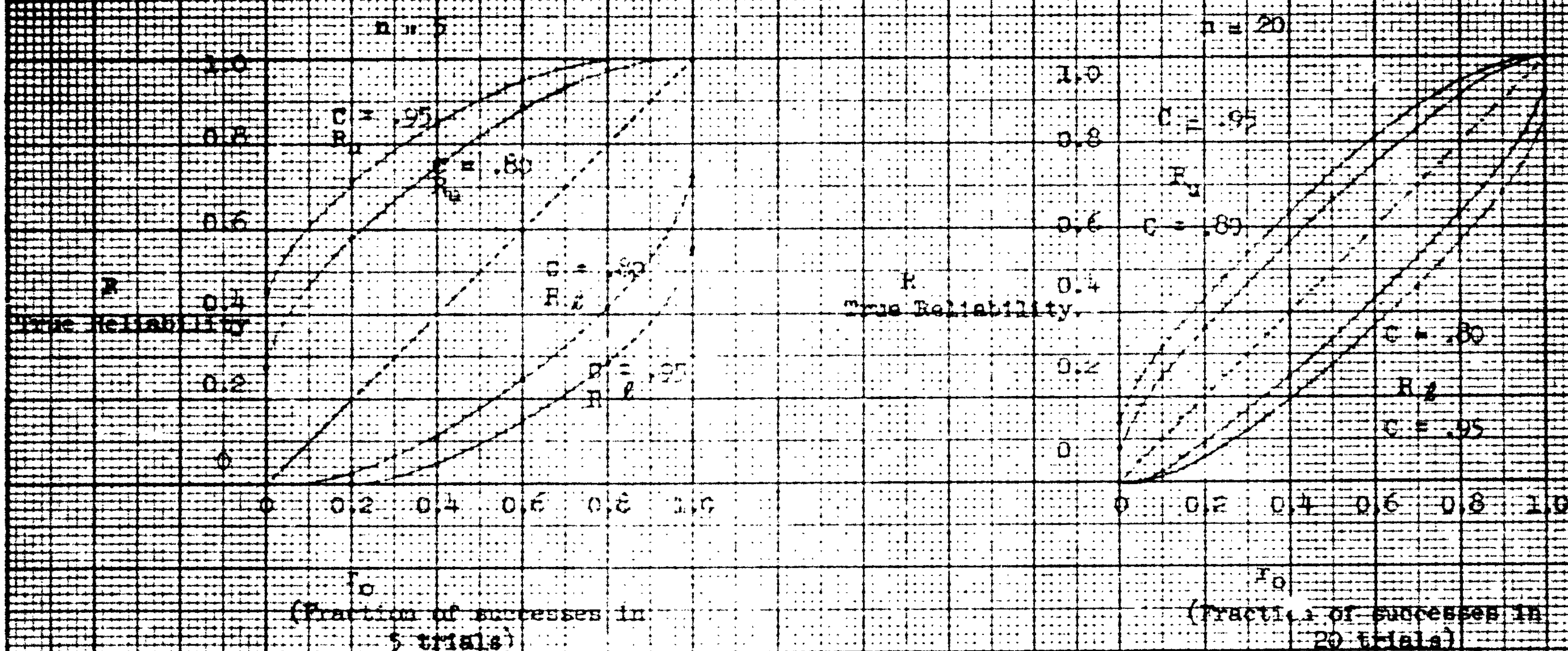
4. Howard Raiffa and Robert Schlaifer,
   <u>Applied Statistical Decision Theory</u>
   Graduate School of Business Administration
   Harvard University, 1961

5. Robert Schlaifer,
   <u>Probability and Statistics for Business Decisions</u>
   McGraw-Hill, 1959

6. Abraham Wald,
   <u>Statistical Decision Functions</u>
   Wiley, 1950

7. G. W. Tyler, G. E. Nicholson, Jr., D. F. Votaw, Jr., J.E. Condon,
   <u>Operations Analysis Working Paper No. 95</u> - <u>"The Testing of Missile Systems"</u>
   Hq, USAF

FIGURE 1



The straight lines show a point - estimate, R_{est}. The curves depict confidence limits, R_u above the line and R_ℓ below.

BINOMIAL CONFIDENCE INTERVALS

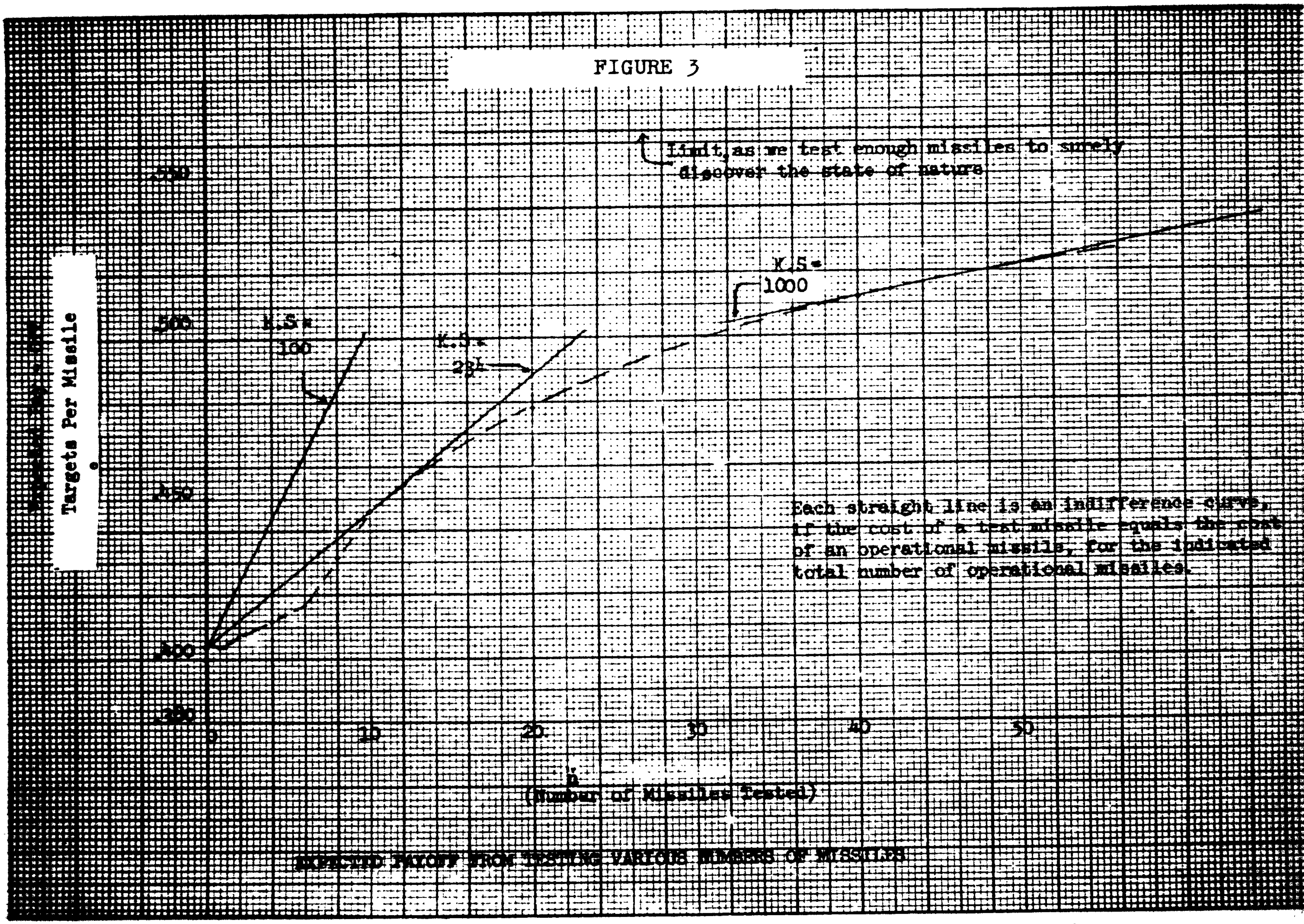
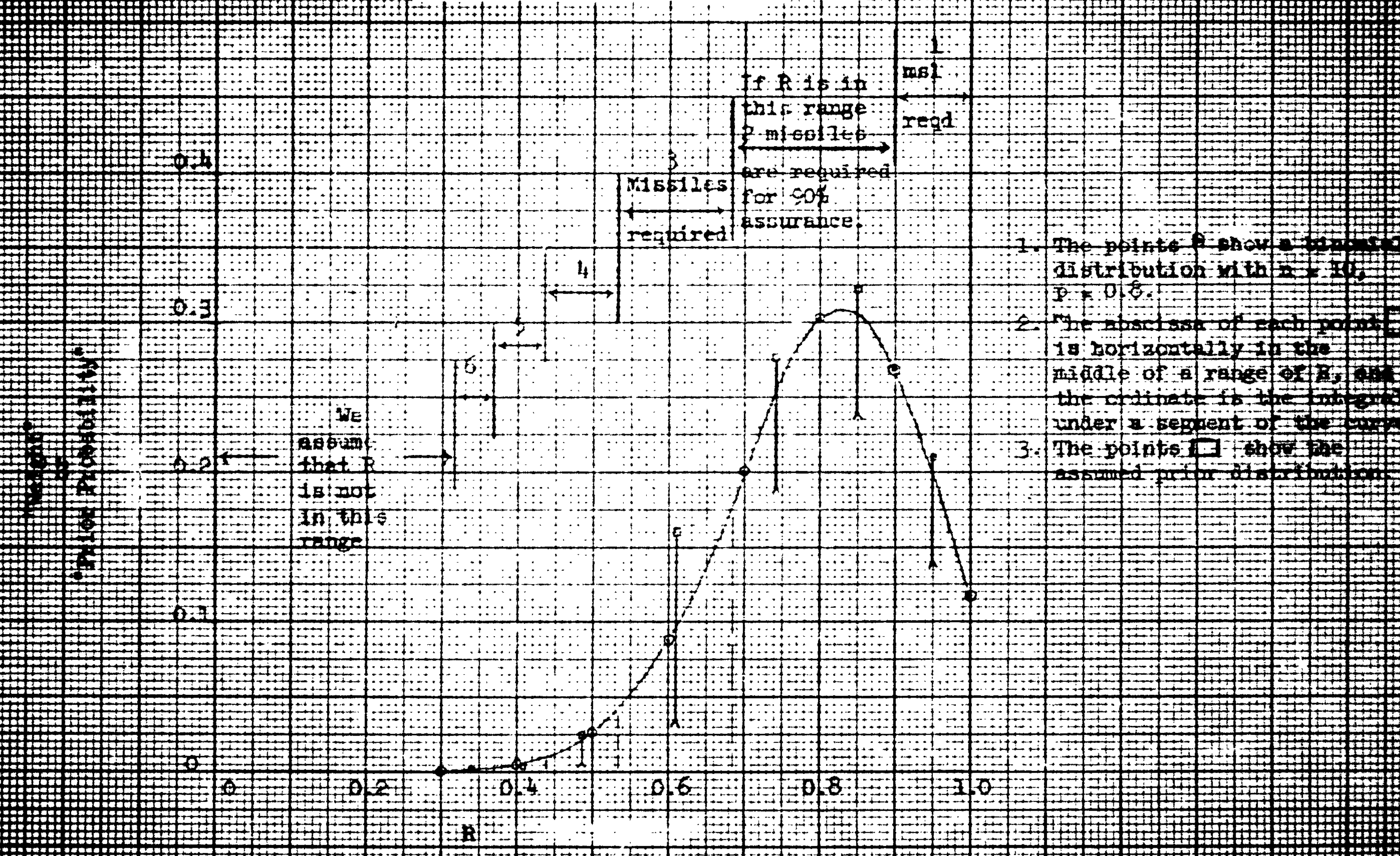


FIGURE 3
Limit, as we test enough missiles to surely discover the state of nature
K.S = 1000
K.S = 100
K.S =
Targets Per Missile
.500
.450
.400
0
10
20
30
40
50
n
(Number of Missiles Tested)
Each straight line is an indifference curve, if the cost of a test missile equals the cost of an operational missile, for the indicated total number of operational missiles.
EXPECTED PAYOFF FROM TESTING VARIOUS NUMBERS OF MISSILES

FIGURE 2



Constructing a reasonable prior distribution

DISTRIBUTION LIST

	No. of Copies
DEPARTMENT OF DEFENSE	
Director of Defense Research and Engineering	1
Deputy Director (Research)	1
Deputy Director (Weapon Systems)	1
Weapons Systems Evaluation Group	
Director	1
Director of Research	1
Senior Air Force Member (Maj Gen N. O. Ohman)	1
Assistant Secretary of Defense (Comptroller)	1
Directorate for Weapon System Analysis	1
HEADQUARTERS USAF	
Chief of Staff, USAF	
Chief, Operations Analysis (AFCOA)	25
Scientific Advisory Board	1
Comptroller of the Air Force	
Director of Status Analysis (AFAMA)	1
Director of Data Systems and Statistics (AFASC)	1
Deputy Chief of Staff, Research and Technology	
Director of Development Planning (AFRDP)	1
Director of Research (AFRDR)	1
Director of Advanced Technology (AFRDT)	1
Deputy Chief of Staff, Systems and Logistics	1
Assistant for Logistic Planning (AFSLP)	1
Deputy Chief of Staff, Plans and Programs	
Air Battle Analysis Division (AFXPD-ABA)	1
Deputy Chief of Staff, Operations	
Director of Operational Requirements (AFORQ)	1

AIR FORCE COMMANDS

	No. of Copies
Hq Command, USAF	
Air Force Intelligence Center, Arlington Hall Station, Arlington 12, Virginia	1
Air Defense Command	
ADC (ADOOA), Ent AFB, Colo	1
26 Air Div (2600A), Hancock Field, Syracuse 25, N.Y.	1
Air Force Logistics Command	
AFLC (MCFR), Wright-Patterson AFB, Ohio	1
Air Force Systems Command	
AFSC (SCAXC-Technical Library), Andrews AFB, Wash 25, D.C.	1
AFMTC (MTGO, Mr. Fennema), Patrick AFB, Fla	1
ASTIA, Arlington Hall Station, Arlington, Va	1
Air University	
Air War College, ATTN: Evaluation Staff, Maxwell AFB, Ala.	1
Strategic Air Command	
SAC (OA), Offutt AFB, Nebr	1
15AF (OA), March AFB, Calif	1
8th AF (OA), Westover AFB, Mass	1
2AF (OA), Barksdale AFB, La	1
1 STRATAD (OA), Vandenberg AFB, Calif	1
Tactical Air Command	
TAC (TPL-OAD), Langley AFB, Va	1
Alaskan Air Command	
AAC (OA), APO 942, Seattle, Wash	1
Alaskan Command	
Hq ALCOM (OA), APO 942, Seattle, Wash	1
Pacific Air Force	
PACAF (PFOOA), APO 953, San Francisco, Calif	1
5AF (5FOOA), APO 925, San Francisco, Calif	1
United States Air Forces in Europe	
Hq USAFE (OA), APO 633, New York, N.Y.	1

OTHER COMMANDS

	No. of Copies
Commander in Chief, Pacific Operations Analysis Office (J301), Attn: Mr. Roy F. Linsenmeyer, Box 17, FPO, San Francisco, Calif........	1
North American Air Defense Command CINC NORAD (NOOA), Ent AFB, Colo........................	1
Fourth Allied Tactical Air Force Operations Analysis Office, Hq 4th ATAF, APO 12, New York, New York....................................	1

OPERATIONS ANALYSIS STANDBY UNITS

Operations Analysis Standby Unit, University of North Carolina, P.O. Box 1146, Chapel Hill, North Carolina..............	1
Dept of Industrial Engineering, Operations Analysis Standby Unit, Iowa State University of Science and Technology, Box 62, Station A, Ames, Iowa..........................	1
Operations Analysis Standby Unit, Denver Research Institute, University of Denver, Denver, Colorado..................	1
Operations Analysis Standby Unit, Room 310 - Logan Hall, University of Pennsylvania, Philadelphia, Penna.........	1

DEPARTMENT OF THE NAVY

Operations Evaluation Group, Dept of Navy, Wash 25 D.C......	1
Commandant, Naval War College, Newport, Rhode Island........	1
Office of Naval Research, Dept of Navy, Wash 25, D.C. (Attn: Dr. Lundegard)................................	1

CONTRACTORS

Analytic Services, Inc., 1101 N. Royal St., Alexandria, Va..	1
Corporation for Economics and Industrial Research, Research Center, 1200 Jefferson Davis Highway, Arlington, Va....	1

	No. of Copies
CEIR, 11753 Wilshire Blvd., Los Angeles 25, California.........	1
General Electric Co., Santa Barbara, California................	1
International Business Machines Corp., 111 Connecticut Ave., N.W., Washington, D. C..................................	1
The MITRE Corporation, Bedford, Mass..........................	1
Research Analysis Corporation, 6925 Arlington Road, Bethesda, Maryland...	1
The RAND Corporation, 1700 Main Street, Santa Monica, Calif....	1
The RAND Corporation, 1000 Connecticut Ave., Washington, D.C...	1
Stanford Research Institute, Menlo Park, California............	1
Stanford University, Dept of Statistics, Palo Alto, California, (Attn: Prof. Solomon)................................	1
Systems Development Corporation, 2500 Colorado Avenue, Santa Monica, California.............................	1
University of Chicago, Laboratories of Applied Sciences, Museum of Science and Industry, Chicago 37, Illinois......	1
University of Chicago, Dept of Statistics, Chicago, Illinois, (Attn: Prof. Kruskal)................................	1
University of Georgia, Athens, Ga., (Attn: Prof. Cohen).......	1

	No. of Copies
CONSULTANTS	
Professor Holt Ashley, Massachusetts Institute of Technology Cambridge, Mass	1
Dr. Leroy A. Brothers, Dean, Dept. of Engineering, Drexel Institute of Technology, Philadelphia 4, Pennsylvania	1
Dr. Wendell Dwyer, Westinghouse, 1000 Connecticut Ave., N.W., Washington, D.C	1
Dr. Roland Larson, Applied Physics Laboratory, Johns Hopkins University, Silver Spring, Maryland	1
Dr. George E. Nicholson, Jr., Department of Statistics, University of North Carolina, Chapel Hill, North Carolina	1
Dr. Martin D. Schwartz, North American Aviation Corporation, Downey, California	1
Professor Jack Silber, Department of Mathematics, Roosevelt College, Room 418, Chicago 5, Illinois	1
Dr. W. Norris Tuttle, General Radio Corp., Concord, Mass	1
Mr. Roger Wilkinson, Bell Labs, New York, New York	1
OTHER	
U.S. Operations Research Groups, Liaison Representative U.K., Box 92, USN 100 FPO, New York, N.Y	1
Shape Air Defense Technical Center, (SADTC), The Hague, Netherlands	1
Assistant Secretary General for Scientific Affairs, NATO, Place du Marechal de Lattre de Tassigny (Porte Dauphine), Paris XVI, France	1
Auburn University, Mechanical Engineering Dept, Auburn, Alabama (Attn: Prof J. Grady Cox)	1