ESD-TR-67-222

(Final Report)

AN INVESTIGATION OF SEQUENTIAL SEARCH ALGORITHMS

H. Beiman
H. Eisner

January 1967

DECISION SCIENCES LABORATORY
ELECTRONIC SYSTEMS DIVISION
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE
L. G. Hanscom Field, Bedford, Massachusetts

Distribution of this document is unlimited.

(Prepared under Contract No. AF19(628)-5989 by Operations Research Inc., 1400 Spring Street, Silver Spring, Maryland 20910.)

LEGAL NOTICE

OTHER NOTICES

ESD-TR-67-222

(Final Report)

AN INVESTIGATION OF SEQUENTIAL SEARCH ALGORITHMS

H. Belman
H. Eisner

January 1967

DECISION SCIENCES LABORATORY
ELECTRONIC SYSTEMS DIVISION
AIR FORCE SYSTEMS COMMAND
UNITED STATES AIR FORCE
L. G. Hanscom Field, Bedford, Massachusetts

Distribution of this document is unlimited.

(Prepared under Contract No. AF19(628)-5989 by Operations Research Inc., 1400 Spring Street, Silver Spring, Maryland 20910.)

FOREWORD

This report, An Investigation of Sequential Search Algorithms, presents the results of a study done under Contract Number AF 19(628)-5989 for Decision Sciences Laboratory, Electronic Systems Division of the Air Force Systems Command, L.G. Hanscom Field, Bedford, Massachusetts. Dr. Ugo O. Gagliardi, ESRHT, was the Air Force program monitor. The investigation covered the period from 15 April 1966 to 31 December 1966 and was presented as an interim report in January 1967. Appendix II was written by Drs. R.D. Johnson, Jr., and S. Kneale of Operations Research Incorporated.

This technical report has been reviewed and is approved.

James S. Duva

JAMES S. DUVA
Technical Director
Decision Sciences Laboratory

ABSTRACT

Characterizations of sequential search processes and algorithms are developed. Representative sequential search algorithms are reviewed and interpreted within the framework of these characterizations in several fields including equipment diagnosis, signal encoding, radar systems, coin weighing, and human decision processes. Conclusions and recommendations for application are presented. Appendixes include selected mathematical techniques in programming and information theoretic search procedures, analyses of scoring systems, and a bibliography.

TABLE OF CONTENTS

		Page
	FOREWORD	ii
	ABSTRACT	iii
	LIST OF FIGURES	vi
	LIST OF TABLES	vii
I.	INTRODUCTION	1
II.	CONCEPTUAL FRAMEWORK FOR SEQUENTIAL SEARCH PROCESSES AND ALGORITHMS	3
III.	REVIEW AND INTERPRETATION OF SEQUENTIAL SEARCH ALGORITHMS	26
IV.	SUMMARY	40
	APPENDIX I: SELECTED MATHEMATICAL TECHNIQUES	43
	APPENDIX II: SCORING SYSTEMS	62
	APPENDIX III: BIBLIOGRAPHY	70
	REFERENCES	77

LIST OF FIGURES

		Page
1.	Preliminary Search Algorithm Framework	4
2.	A Simple Decision Tree	10
3.	Illustrative Decision Trees	11
4.	Equivalence of Parallel Structured Tree	12
5.	An Additive Function on a Path	16
6.	Measures of Effectiveness for a Decision Tree . .	17
7.	Completely Replicated Tree	19
8.	A Parallel Tree	20
9.	Part of the Parallel Tree	20
10.	Illustrative Equivalent Trees	21
11.	Relationship Between Binary Coding and a Simple Search Algorithm	32
12.	Relationship Between Shannon-Fano Coding and a Simple Search Algorithm	33
13.	Simple Decision Tree	51
14.	Decision Tree Illustrating Noisy Detection Device .	54
15.	Forcing Quality of Scoring Functions	67

LIST OF TABLES

		Page
I.	Second Weighing Alternatives, Probabilities, and Entropies	57
II.	Second Weighing Alternatives if Balance Not Achieved on the First Weighing	59

SECTION I

INTRODUCTION

1.1 This report documents the results of investigations of sequential search processes and their applications. The study establishes the unifying characteristics of a broad class of sequential search processes and algorithms that have been examined and developed in diverse fields and different contexts so that this body of formalisms can be applied in the future to problems that confront the Air Force.

1.2 To accomplish this broad objective, the following subsidiary efforts were set forth:

a. Studies and experimental investigations directed toward the consolidation and clarification of recent developments in coding theory, fault diagnostics, and search theory

b. The generation of a conceptual framework sufficiently general to encompass existing search algorithms, and identification of directions in which future efforts might most profitably be made.

1.3 This document deals with the full scope of the defined investigation, which if completed, would increase the depth of analysis and the emphasis on applications of sequential search.

Limitations in Scope

1.4 To establish a viable scope within the prescribed effort, various initial limitations were explicitly identified. The processes under investigation were restricted mainly to sequential search, although at times non-sequential allocation of effort processes were considered to add background perspective. The special class of sequential detection problems (looking for a signal in noise), so extensively dealt with in the statistical theory of communications, is also not considered primarily because they are usually not related to the classical sequential search processes. The additional special set of game theoretic search algorithms is also not examined in detail although many such algorithms were encountered in passing.

Approach

1.5 The defined tasks were approached using the concurrent definition of a conceptual framework for generalized sequential search processes and investigation of the specific search algorithms as an aid in refining this framework and determining the mathematical tractibility of the possible solutions. This investigation was not intended to develop new algorithms for sequential search processes. However, as requested by ESD, an analysis of scoring functions was carried out. The results of this analysis are presented in Appendix II.

Organization of Report

1.6 Section II presents a conceptual framework for both sequential search processes and algorithms and discusses the distinction made between the characterizations of the search processes and algorithms that were developed to optimize (or otherwise handle) these processes. Section III reviews and interprets the sequential search algorithms that have already been developed in various fields and different contexts. Section IV summarizes and presents conclusions and recommendations for further activities. Three appendixes are provided. Appendix I presents selected mathematical techniques applicable to sequential search processes, and Appendix II describes the analytic work on scoring systems. Appendix III contains a bibliographic list of related papers, reports, and books. For the sake of more complete documentation the scope of the bibliography is somewhat broader than that of the study as defined. The list of cited references is given as the last section.

SECTION II

CONCEPTUAL FRAMEWORK FOR SEQUENTIAL SEARCH PROCESSES AND ALGORITHMS

2.1 This section sets forth a unifying perspective for viewing sequential search processes and algorithms that may have been developed independently and in different contexts. This development will provide a means of consolidating and clarifying recent developments in diverse fields to permit a more effective solution to problems confronting the Air Force.

2.2 No claim is made for the uniqueness of this approach; in fact, a framework was sought that would be sufficiently general to include most sequential search processes and algorithms as special cases. The intent of the approach was to allow outstanding problems to be recognized within the context of this general framework as particular cases for which solutions may already be available.

SEQUENTIAL SEARCH PROCESSES AND ALGORITHMS

2.3 In this investigation, a distinction is made between sequential search processes and algorithms and the ways each can be characterized. The search process itself is characterized by a sequential decision tree, which represents a conceptual framework for the available system states and alternates inherent in the process. The algorithms themselves may be viewed as systems of boundary conditions and constraints imposed on the basic process. They embody the assumptions and restrictions that must be considered to structure the process into mathematically manageable proportions.

CHARACTERIZATION OF SEQUENTIAL SEARCH ALGORITHMS

2.4 A conceptual framework for sequential search algorithms is shown in Figure 1. That is, the algorithms relevant to this investigation can be characterized by a particular path through this figure.

Multistage Processes

2.5 The first dichotomous choice, at the top of Figure 1, breaks down search processes into single-stage (static) and multistage processes.

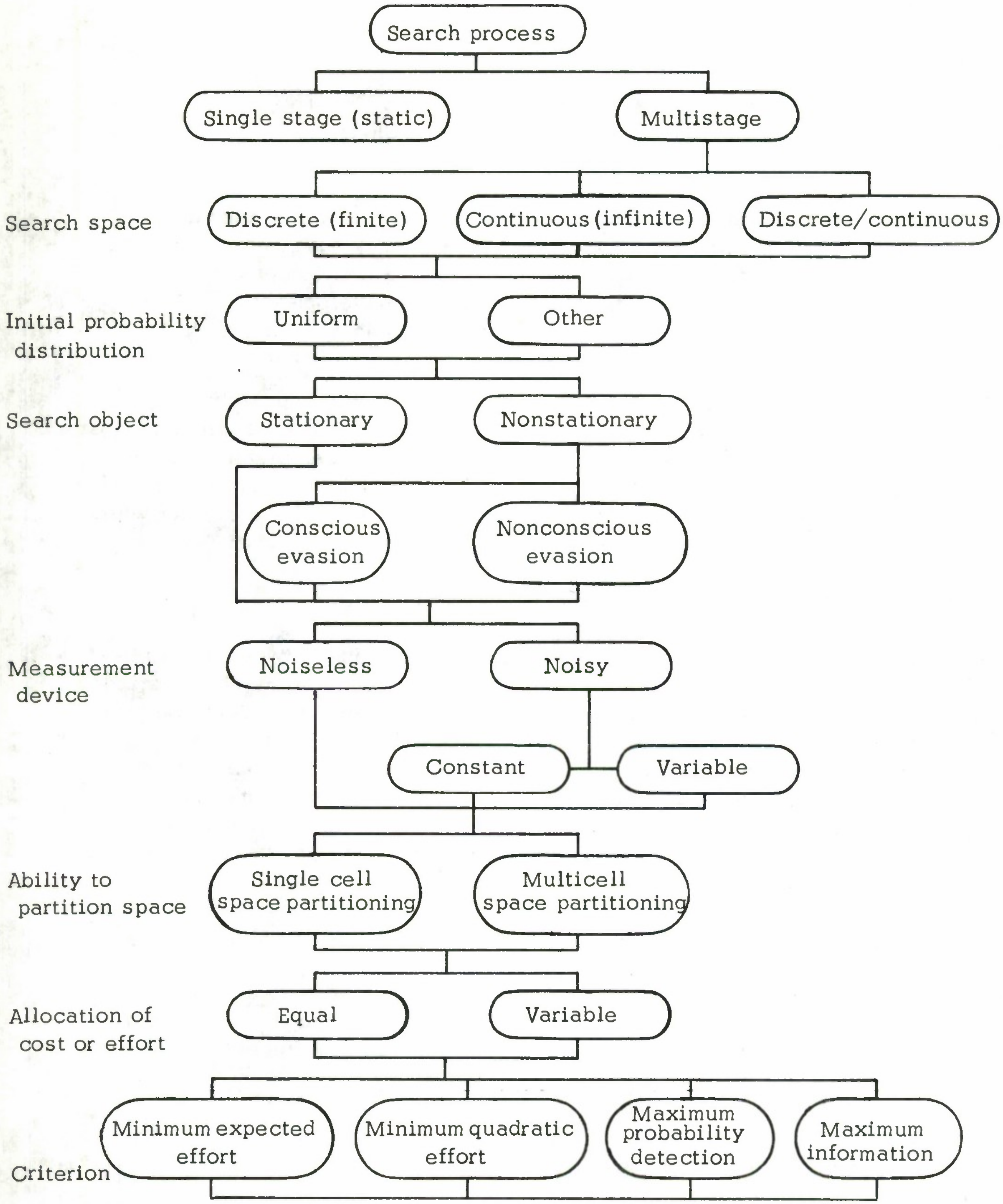


FIGURE 1. PRELIMINARY SEARCH ALGORITHM FRAMEWORK

Only the multistage process is considered in this investigation. This process is used synonomously with the sequential process.

Nature of Search Space

2.6 In any search process, one or more search objects are located in one or more cells. The complete set of such cells may be said to constitute a search space. As shown in Figure 1, the search space itself may be characterized as discrete (and finite), continuous, or some combination thereof. In virtually all cases of practical interest, the search space is discrete and finite. If not, the mechanism for carrying out the search process is limited in its resolution so that the search space is, in effect, reducible to the discrete finite case. For example, in a radar search for a missile or aircraft, the vector coordinates (e.g., position, velocity) of the search object may be described by a set of numbers that are infinitely dense (such as the position of a point on a line). But the radar's finite range, azimuth, and elevation resolution capability has the effect of limiting the search space to a finite number of discrete cells.

2.7 The search space may also be conceptual as opposed to a physical entity as in the radar search example. Consider the well-known coin weighing problem [1,2] in which there are 12 coins and a search is initiated, using an equal arm balance, for the one bad coin which is either heavier or lighter than all the other coins. Since there are 12 coins and the bad coin may be either heavier or lighter, the search space consists of 24 cells that do not exist as physical entities. The object of the search process, of course, is to isolate the cell describing the number of the coin and determine whether it is heavier or lighter than the others.

Probability Distribution Over Search Space

2.8 Inherent in the search problem is the definition of some probability distribution over search space. Figure 1 only distinguishes between a uniform and some other distribution at the initiation of the search. Several observations can be made in this regard.

2.9 Lacking other a priori information the assumption of a uniform distribution is usually made initially. There is some justification for this assumption, [3,4] which states that the search object is equally likely to be contained in any one of the search cells.

2.10 Clearly, as the search process itself continues, the probability distribution should depart from uniformity to the case in which the variance is forced to zero. This is reviewed in information theoretic terms as the sequential resolution of uncertainty, which is synonomous with the notion of decreasing the system entropy. In effect, this is the basis for using information theoretic approaches to search problems.

2.11 Radar search, diagnostics, information retrieval, and numerous other types of search problems clearly illustrate the idea of a successively more highly peaked distribution and resolution of uncertainty as to the location of the search object.

<u>Nature of the Search Object</u>

2.12 A first-order partitioning of characteristics of the search object or objects distinguishes between stationary and nonstationary cases. If the object remains in its initial search cell over time or alternately over all stages of the search process, it is considered to be stationary. In the nonstationary case, the further distinction can be made between conscious and nonconscious evasion. The former case implies a considered response, by the search object or an opponent, to the searcher's attempts to find or defeat it. Hence, game-theoretic formulations of search problems fall into this category as well as many conventional duels in which the search for an optimum strategy is sought in the face of action by a competitive adversary.

2.13 It is clear that the search problem becomes more difficult for the nonstationary object. Further, many different models for the behavior of the search object have been postulated, ranging from an attempted radar search for possible high velocity missiles to one where a search object shows up in a different cell on each trail, with complete independence from trial to trial.

<u>Measurement or Search Device</u>

2.14 The searcher has at his disposal some measurement device used to implement the search process. This is the mechanism used to decrease or resolve the uncertainty of locating the search object.

2.15 Such a measurement device can be treated as either noiseless or noisy. In the former case, by definition, there are no errors of either the first or second type. In the context of radar detection, for example, such a device would always detect a target in a cell under surveillance and would never indicate its presence if, in fact, there were no target. Noiseless measurement devices are idealizations useful in the context of some types of search problems. The coin weighing problem mentioned above assumes a noiseless measurement device in that the equal arm balance cannot provide a false indication of balance or unbalance.

2.16 The noisy measurement device, however, is often used and is the basis for much of the statistical theory of communications. In the radar example, it is more common to admit of nonzero false-alarm and failure-to-detect probabilities. Many search algorithms assume that there is some finite chance of not detecting the search object when observing the cell in which it is located and of detecting a nonexistent object in a cell.

2.17 Within the category of noisy measurment devices, the error probabilities can remain constant throughout the search or can change in some prescribed fashion, such as a function of the search space, time, stage of the search process, use of the measurement device, effort or resources devoted to the search of each cell. Although these considerations might appear to lead to insoluble problems, it is a fact that many search algorithms are based explicitly on such types of noisy devices. In the radar search, for example, the signal and noise power may vary with range so that each of the cells at different ranges has a different signal-plus-noise and noise-probability distribution leading to different failure-to-detect and false-alarm probabilities.

<u>Search Space Partitioning</u>

2.18 Another dimension to characterizing the sequential search algorithms involves the inherent ability of the searcher to partition the search space. This is important since it presents a basic constraint on the capability of obtaining information about the location of the search object. This point can be illustrated by the coin weighing problem.

2.19 Using an information theoretic approach (see Appendix I) shows that the initial uncertainty for 12 coins is of measure log (12)(2) = log 24 and the potential information gathered from three weighings is 3 log 3 = log 27. Since log 27 > log 24, it appears that three weighings are sufficient to find the bad coin. Although three weighings are, in fact, sufficient, the above condition is necessary but not sufficient. If the bad coin had to be found from among 13 coins, the potential information is still greater than the uncertainty, i.e., log 27 > log 26. However, the problem is not solvable in three weighings since it turns out that the maximum amount of information (log 3) cannot be obtained on each measurement. This results from the inherent inability to partition the search space, as reflected by the partitioning of the coins themselves, such that on each weighing the probabilities of the left side's being heavier, the right side's being heavier, and the two sides' balancing are equal.

2.20 The same types of search space partitioning constraints are imposed on radar search, diagnostics, information retrieval, and virtually all search problems. These constraints are often imposed jointly by the search object or space and the measurement device. For example, in the radar case, it is generally impossible to look at all combinations of search cells because of equipment limitations. It is also impossible to group test points arbitrarily in a diagnostic situation such that the measurement alternatives are equally likely.

Allocation of Effort

2.21 The next characteristic of sequential search algorithms, shown in Figure 1, involves the allocation of cost or effort over the search space. Many algorithms implicitly assume that this allocation is equal over all search cells; others allow this to be variable and, in fact, are seeking an allocation that maximizes or minimizes some figure of merit or measure of effectiveness.

2.22 As indicated previously, the probability of detecting a target in a particular cell may be an increasing function of the length of time or energy (effort) devoted to searching that cell. Based on the criterion that the overall probability of detection is to be maximized, an algorithm may be developed that will find the required time or energy allocations to each cell. Thus, the notion of allocation of effort (e.g., time and cost) is intimately tied to the criterion (measure of effectiveness and figure of merit) on which the algorithm is based. Further, the allocation at each stage may affect the search environment by changing any or all the above characterizations. [5/]

Criterion

2.23 The criterion for an algorithm is an explicit statement of the measure of effectiveness or figure of merit and the values that these should take. Figure 1 shows some of the more common criteria used: minimization of expected effort, minimization of quadratic effort, maximization of detection probability, and maximization of information obtained. These are often stated together with subsidiary constraints such as achieving a maximum probability of detection, given a false alarm probability, or a total available effort (which can be devoted to the search process).

SUMMARY OF SEARCH ALGORITHM CHARACTERIZATION

2.24 From this discussion of sequential search algorithms, it should now be possible to cite for each existing or possible algorithm a path through Figure 1 for which that algorithm is characteristic, whatever the field of application or motivation for its development. At this point note that specific methods of solution, such as dynamic or linear programming and areas of application, have not been considered. This is done in Section III and Appendix I.

2.25 The above characterization deals with the algorithms used for carrying out search processes and does not represent the search process itself. The remainder of this section discusses the latter case, i.e., the characterization of the process and the insights that can be gained by such a characterization.

DECISION TREES AS SEQUENTIAL SEARCH PROCESSES

2.26 A sequential search process is a process in which a sequence of decisions must be made in attempting to find something. After each decision, information is obtained that is relevant to the next decision. Consider first the case requiring a finite number of decisions and a finite number of alternatives for each decision. (The case requiring an infinite number of alternatives is not examined.) This process can be represented by a tree diagram as shown in Figure 2. Here the circles (or nodes) represent decision points, and the lines leading from them represent the alternatives. The square boxes represent the ends of decision sequences. Progress through the tree is made from top to bottom.

2.27 The fact that circles and squares appear at the same horizontal level in the diagram does not imply that the decisions they represent must be contemporaneous; they are drawn this way for convenience, and their appearance at the same level means that they are at the same relative positions in their sequences.

2.28 In Figure 2 for example, there are nine possible paths through the tree, each path ending at a square box. A complete path through the tree is referred to as a "strategy." A strategy is a plan for carrying out the search; it involves allocating effort, partitioning the search space, and using the measurement device (as discussed previously and shown in Figure 1) in accord with some stated criterion. The square boxes, representing the ends of the various possible search sequences, correspond either to finding the object sought or to the decision to end the

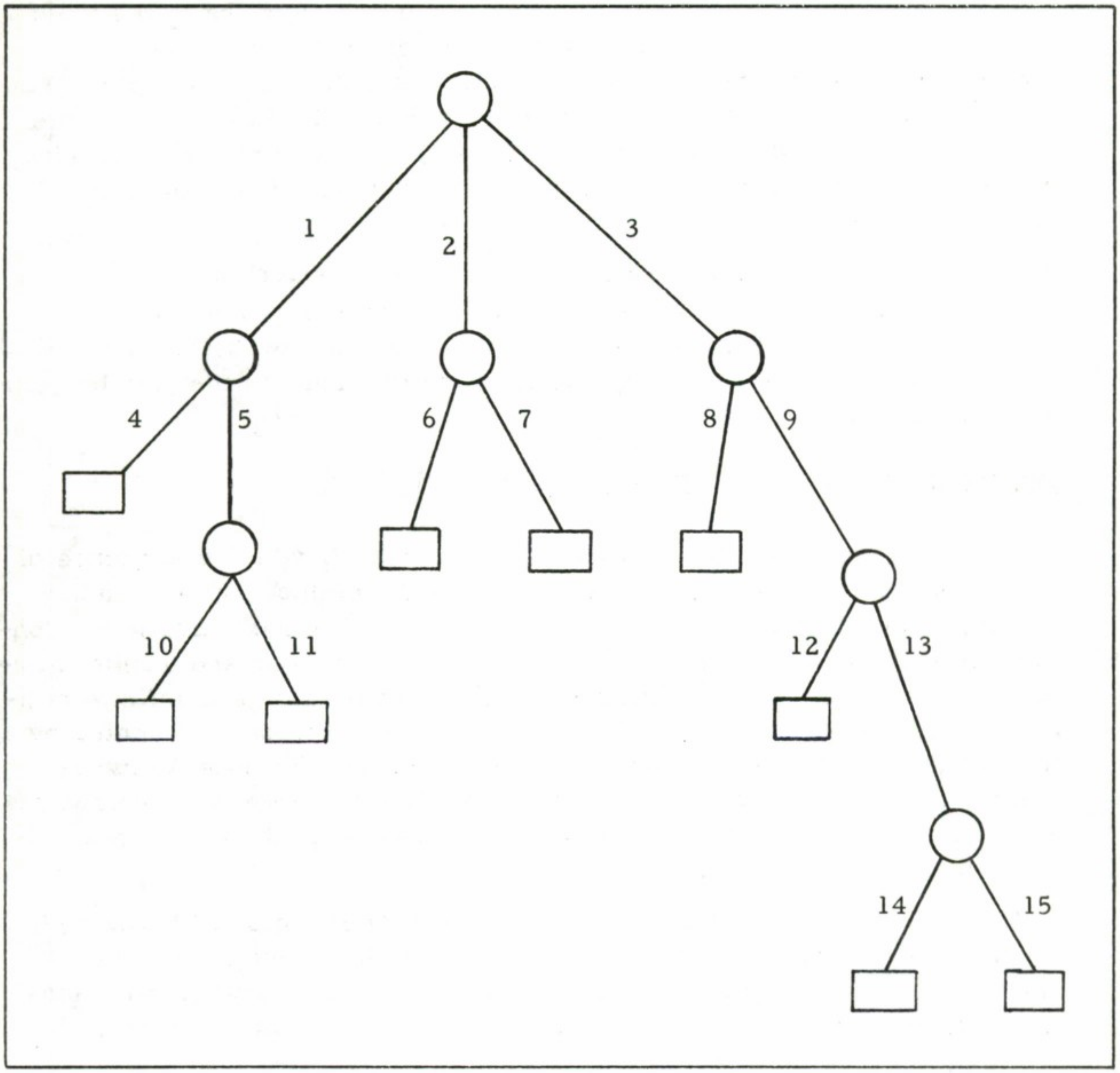
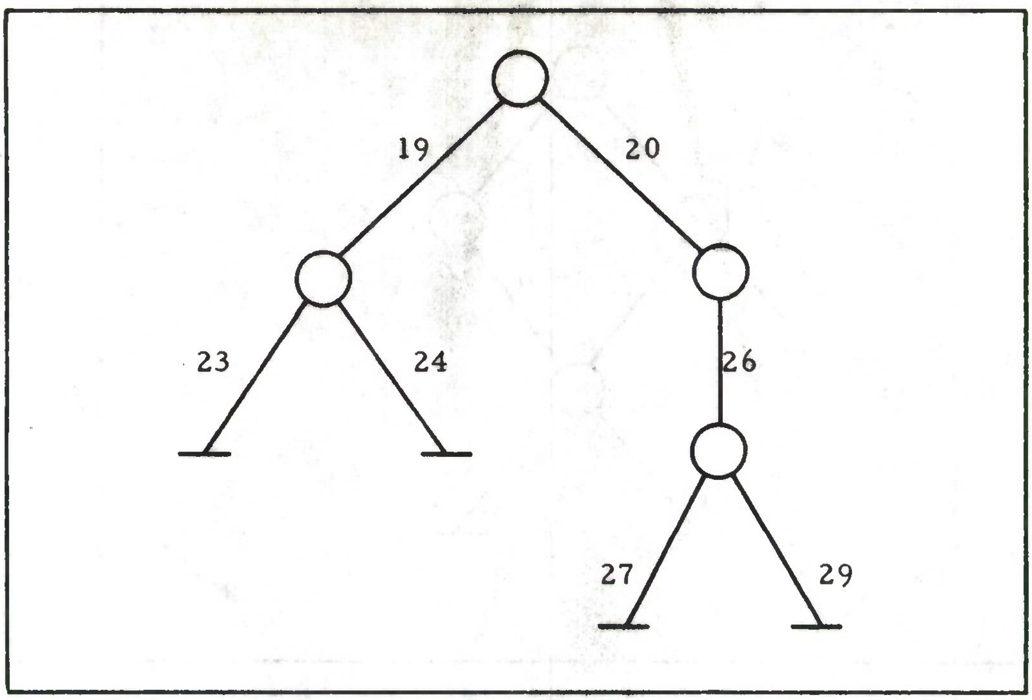


FIGURE 2. A SIMPLE DECISION TREE

a. Decision Node With Single Alternative



b. Equivalent Decision Tree

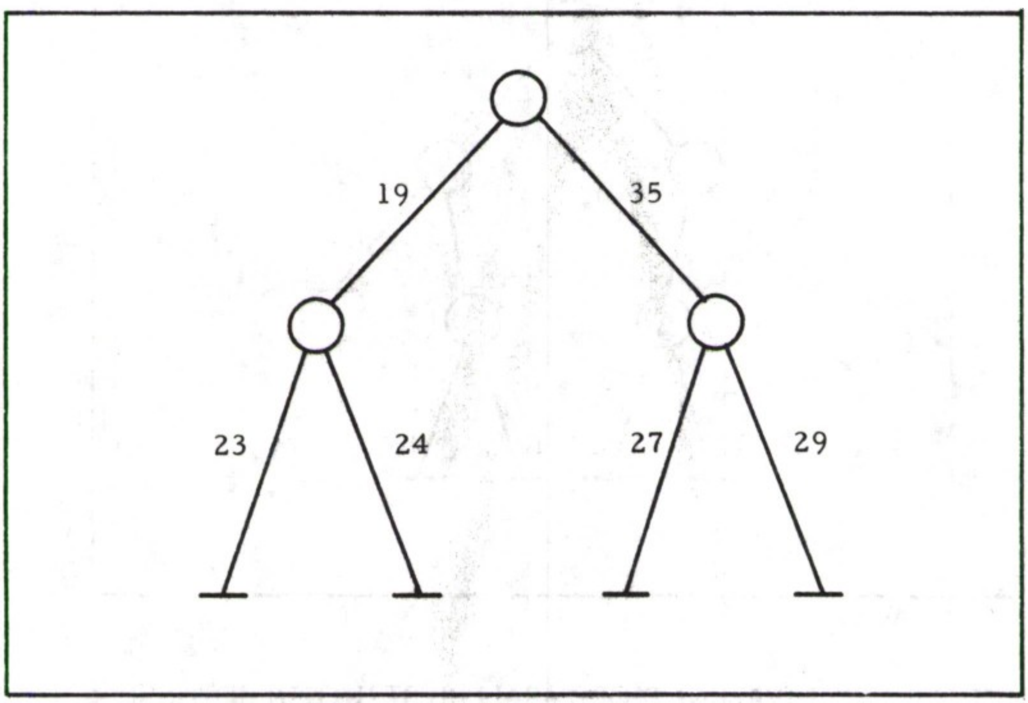
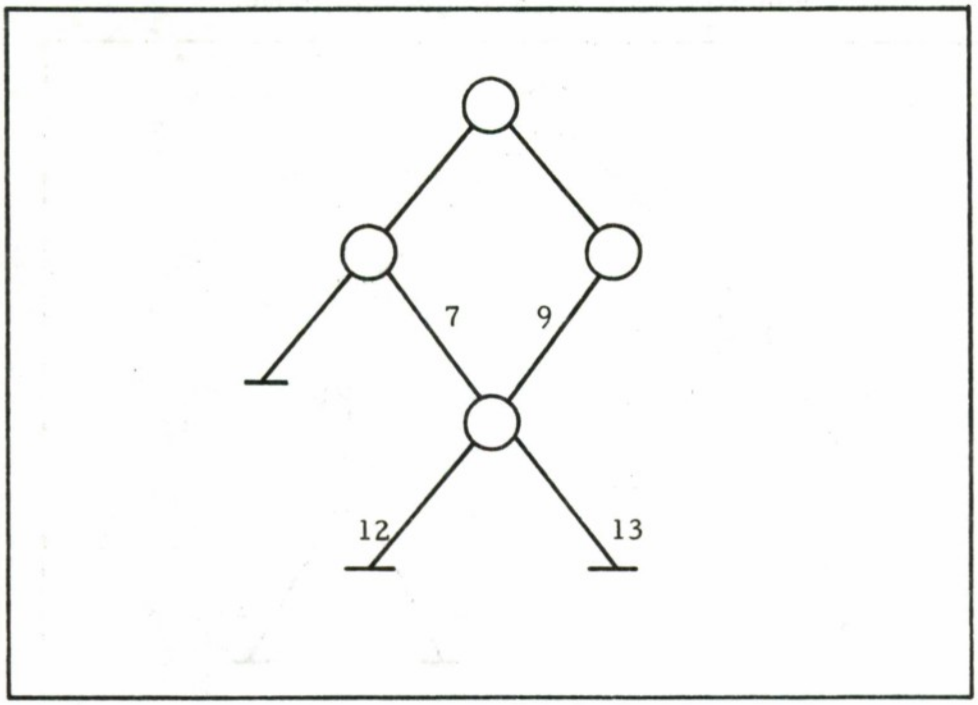


FIGURE 3. ILLUSTRATIVE DECISION TREES

a. Tree With Parallel Structure



b. Equivalent Tree

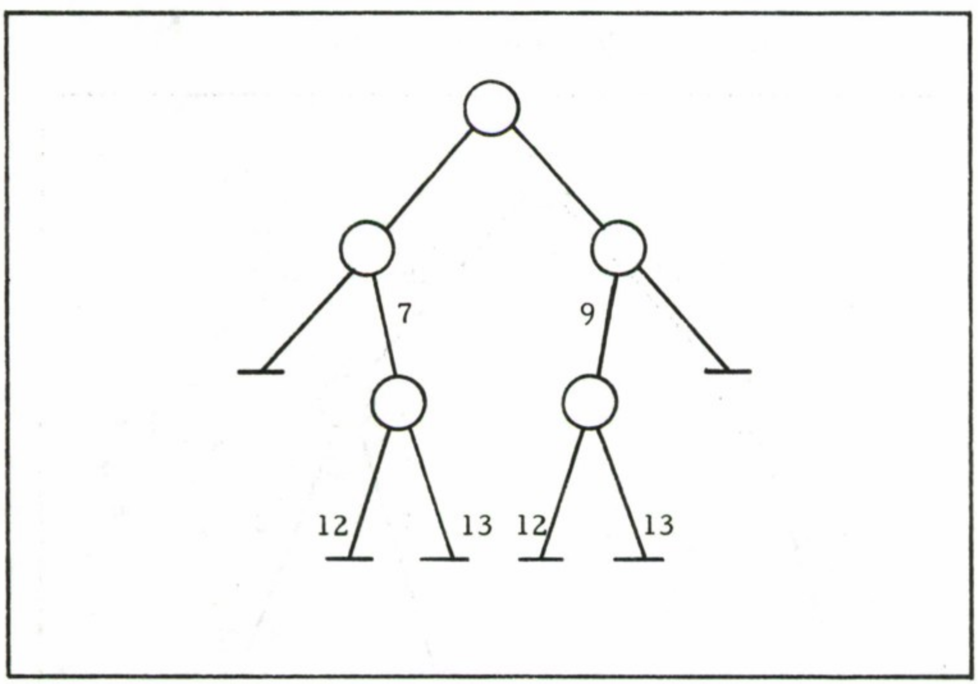


FIGURE 4. EQUIVALENCE OF PARALLEL STRUCTURED TREE

search process. Note that some strategies have more choices than others; thus the case is not excluded in which the number of decisions to be made depends on some of the decisions.

Properties of the Decision Tree

2.29 It is assumed in this discussion that each node has at least two lines leading down from it. This is no real restriction, since if only one line leads from a node it means that in choosing to go to that node the decision maker has also committed himself to take the only available path from that node, so that he has really made the double decision of taking both these steps. Thus, it is assumed that in all such cases the two paths have combined into a single one, representing the single alternative. Figures 3a and b illustrate this process. (The short horizontal line crossing certain paths indicates that the tree continues but is not reproduced here.) In Figure 3a, if path 20 is taken, path 26 also must be taken at the next node. Figure 3b shows the replacement, with paths 20 and 26 replaced by 35 and the node connecting them eliminated.

2.30 Assume also, at least temporarily, that every node (except the topmost) has exactly one line leading into it. The conditions under which this assumption can be removed will be discussed later. The reason for having at least one line entering each node is obvious; that for restricting the number to at most one is not so obvious. In any event, this assumption does not really restrict the type of problem that can be represented by a tree (although it could make the tree have more branches than necessary). For if two different lines are followed by the same set of subsequent nodes and lines, the tree structure following can be attached to each of them, rather than meeting them at some node. Figure 4a is an example of two lines leading to the same set of subsequent nodes and lines, and Figure 4b shows an equivalent tree.

2.31 Even in a relatively simple case requiring 10 decisions, each with 5 alternatives, there are 5^{10} or about 10,000,000 different strategies possible. Thus, the tree diagram is usually a conceptual rather than a realized entity. Nevertheless, it does provide a basic framework for examining sequential search processes.

Measures of Effectiveness

2.32 Deterministic Case. Each deterministic strategy leads to some final result, such as, "the pay-off is $5," or "8 radars will be inoperable," or "the object searched for is found after 7 trials." This result may be compared to a stated criterion. In each case a number is assigned to each possible strategy (actually to the outcome of that strategy). Even for the third case, where there are only two possible outcomes (each attained by many different strategies) "the object searched for is found" or "the object searched for is not found", the value "1" can be assigned to the first outcome and the value "0" to the second. Generally, the results possible for the various strategies must, at the very least, have an order relation established among them such that given any two strategies, either they are considered equivalent or one is better. Also, if this ordering relation is not to lead to logical contradictions, it is necessary for the ordering relation to be transitive; i.e., if strategy A is better than B, and B better than C, then A is better than C. This is equivalent to saying that to each strategy there is assigned at least one number, measuring the effectiveness of that strategy. A simple measure of effectiveness will be defined to be a scalar, rather than vector, quantity.

2.33 Probabilistic Case. It is also possible that the result of following a particular strategy can be expressed only in probabilistic terms. This will occur in cases for which the decision-maker has his choice of several probability distributions and the result of his choice is a sample from the distribution he has chosen. For instance, in the coin-weighing problem he may choose to weigh one coin against one, or two against two. The results of each of these choices are: (a) the coins balance, (2) the left side is heavier, (3) the right side is heavier. But, for each choice of how many coins to try, the probabilities of these results are different. Even if the measuring device is noisy, i.e., the balance sometimes gives incorrect results, this case can still be fitted into the scheme if the probabilities of all the incorrect results are known. Of course, the cost of doing this is that the probability distributions from which the decision-maker chooses become more complicated.

2.34 When the concept of a measure of effectiveness is added to the tree structure, the rationale for the restriction (in paragraph 2.30)—that only one path should lead into a node—is apparent. That is, even if the node and path structures following two different lines are identical, the measure for the total path, including the preceding nodes and lines, may be quite different.

Examples of Measures of Effectiveness

2.35 Coin-Weighing Problem. In the problem of finding the heaviest or lightest coin of 12 coins, with an equal arm balance, a measure of effectiveness could be the number of coins that have been determined to be true coins or the number of measurements made to find a true coin.

This is distinguished from the objective or criterion, which is to find the bad coin and determine whether it is heavier or lighter in no more than three weighings.

2.36 Frequency Assignment Problem. Suppose that 10 radars are to be operated in a region and that each has 15 frequencies to which it can be tuned. Suppose a frequency is needed for each radar so that the greatest number possible will operate satisfactorily. This can be thought of as a tree process in which the first step is to pick one of the 15 frequencies available for the first radar, then pick one from those available for the second. Thus, there are 15^{10} possible strategies. A measure of effectiveness for a strategy in this case could be the number of radars operating satisfactorily if that set of frequencies is used.

Types of Measures of Effectiveness

2.37 Measures of effectiveness can be subdivided into two classes: those for which a number can be assigned for partial paths through the tree and those for which this cannot be done. If the measure function is the first kind, it is possible to get an answer to the question "How am I doing so far?". If it is not, then in a strict sense the problem is not one of sequential search or decision, since if no information is available (at each decision point) about how well we have done or the worth of the alternatives available, then we might as well have decided on the whole strategy at one time. That is, in a true sequential search or decision process, each time a new decision point is reached, some new information should have reached the decision-maker. In this context, many allocation of effort search problems[6] and their extensions are not really sequential. The second type of measure function will not be abandoned completely. However, for the present the first type is the only one to be considered.

2.38 Measures of effectiveness of this kind—those that can be evaluated at every node—are called separable. Thus, in the problem of the assignment of frequencies to radars, any time we have assigned frequencies to a subset of the total population it is possible to decide how many radars considered so far would be operating satisfactorily, given these frequencies. The measure function would then be separable.

2.39 If the measure function is separable, it is always possible to replace it by an additive function, i.e., by a function whose value at node N is the sum of its values up to the last previous node plus the value of the path from the last previous one to this one. Leaving out some unnecessary mathematical notation, the idea is that the number attached to a line should be the value of the measure function at its ending node minus the value of the measure function at its starting

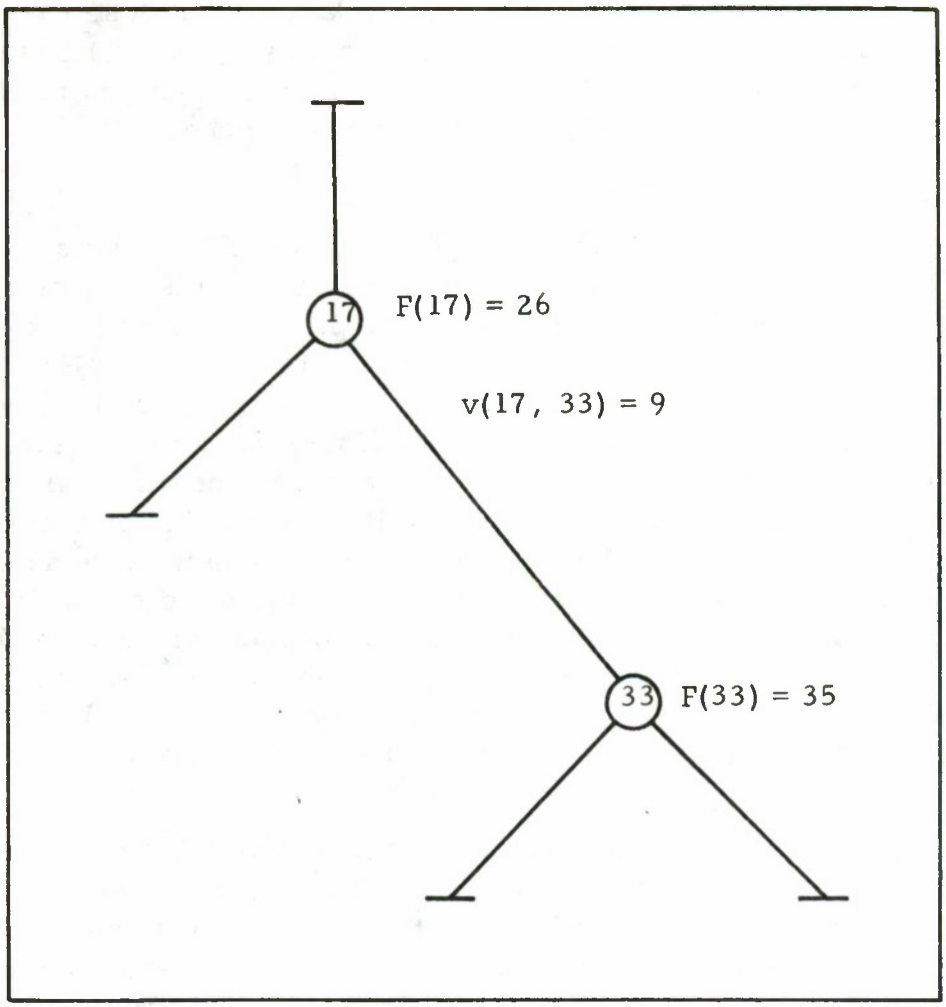


FIGURE 5. AN ADDITIVE FUNCTION ON A PATH

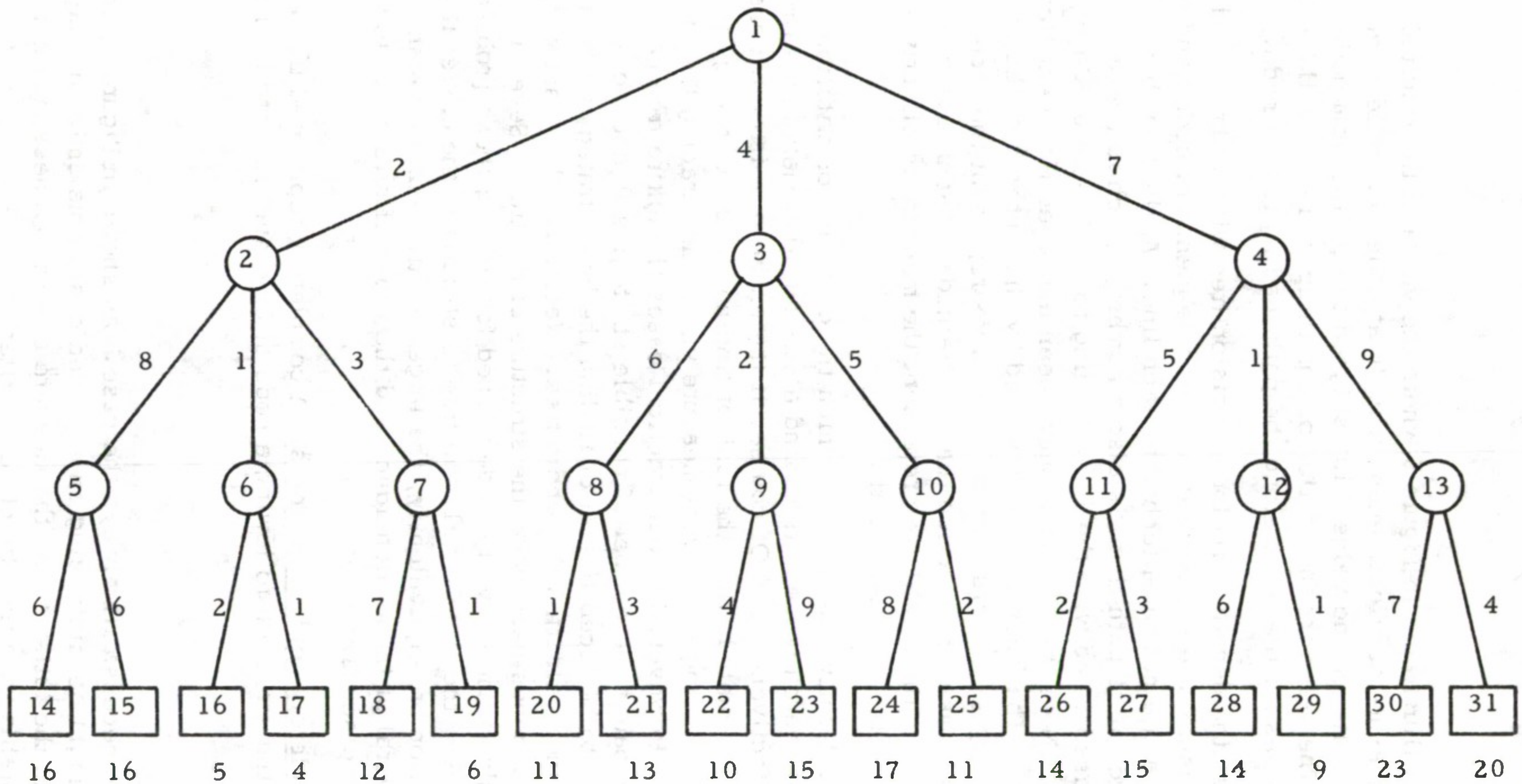


FIGURE 6. MEASURES OF EFFECTIVENESS FOR A DECISION TREE

node. This is illustrated in Figure 5 where the value of the original (separable) measure of effectiveness was 26 at node 17, while it was 35 at node 33. Hence the value 9 is assigned to the line from node 17 to node 33. The value for the portion of a path up to a node will be the sum of the values for the lines along the path leading to that node.

2.40 Thus, the general form for a decision tree with a simple and separable measure of effectiveness will be a sequence of nodes and connecting lines and a number attached to each line. At the final box at the end of each complete path there is also a number, which is the sum of the numbers associated with the lines leading to it. Figure 6 is a typical such tree. Note that the nodes have been numbered (for identification purposes). The lines can be identified by the numbers of the nodes they connect. The numbers alongside the lines represent the increment in the separable measure function that is attained if that line is used; the numbers under the final boxes represent the measure of effectiveness for the strategy that ends each path.

2.41 Consider now the case for which the objective or criterion in the sequential search process is to find a strategy that maximizes the measure of effectiveness.* Of course, if the entire tree is drawn and all strategies evaluated, then the task of finding the best strategy is done by simply finding the largest measure value and tracing the strategy that led to it. However, in most practical cases the entire tree cannot be constructed because the number of possible strategies is enormous. Thus, the task to be faced is that of finding the best strategy without constructing and evaluating the whole tree. Naturally, the possibility of doing this depends heavily on the structure of the tree. Several techniques (algorithms) have been developed for solving these problems. In a gross sense, the more tightly the tree is structured the easier it is to find a solution or, equivalently, the larger the tree size that can be handled. Several of these techniques and the types of structures to which they apply are discussed later.

2.42 Completely Replicated Trees. A completely replicated tree is one in which the structure following a node at any level is exactly the

---

* This is not necessarily always the case. As shown in Figure 1, the criterion might be minimizing an expected cost or maximizing a probability. In the latter case, the measure of effectiveness may not be commensurable with the objective or criterion.

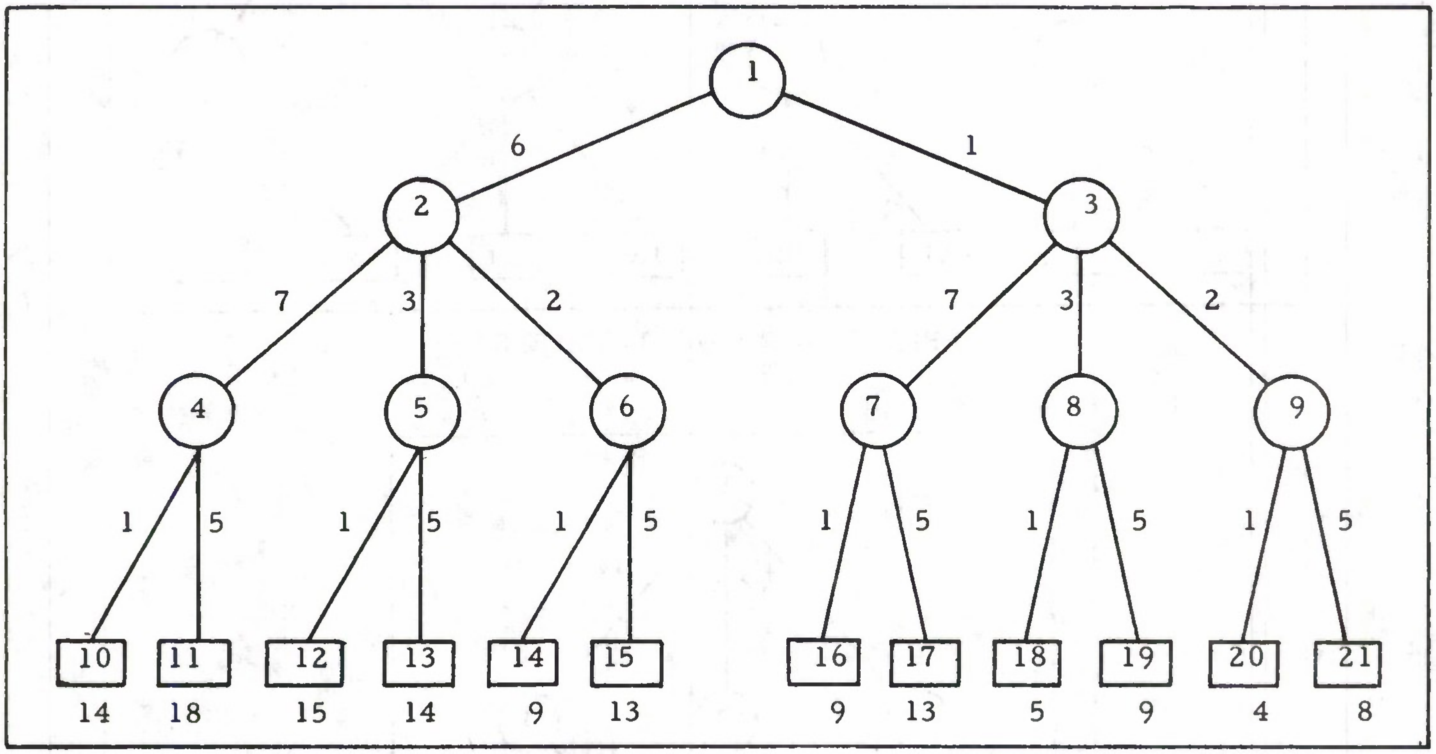


FIGURE 7. COMPLETELY REPLICATED TREE

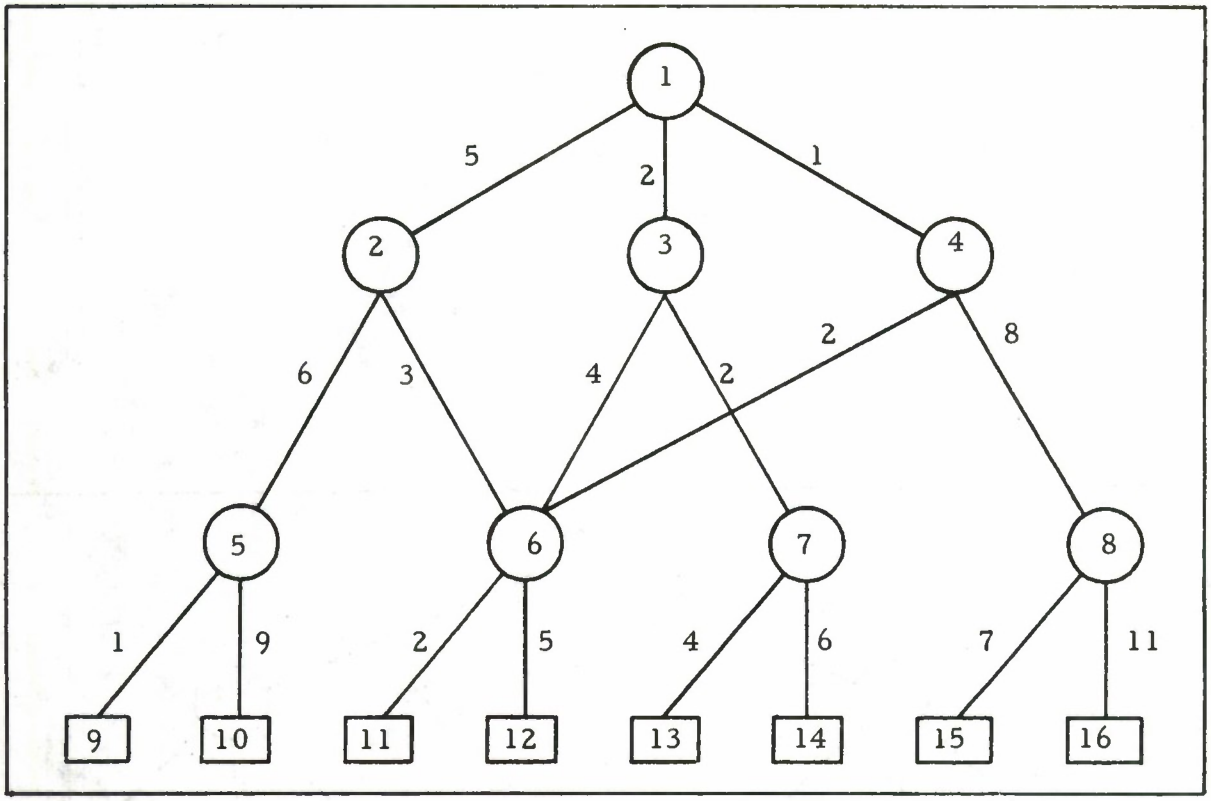


FIGURE 8. A PARALLEL TREE

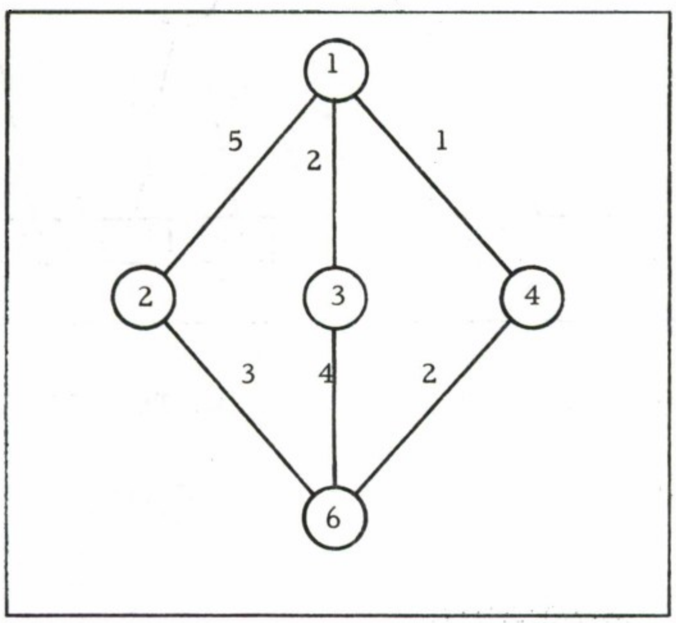
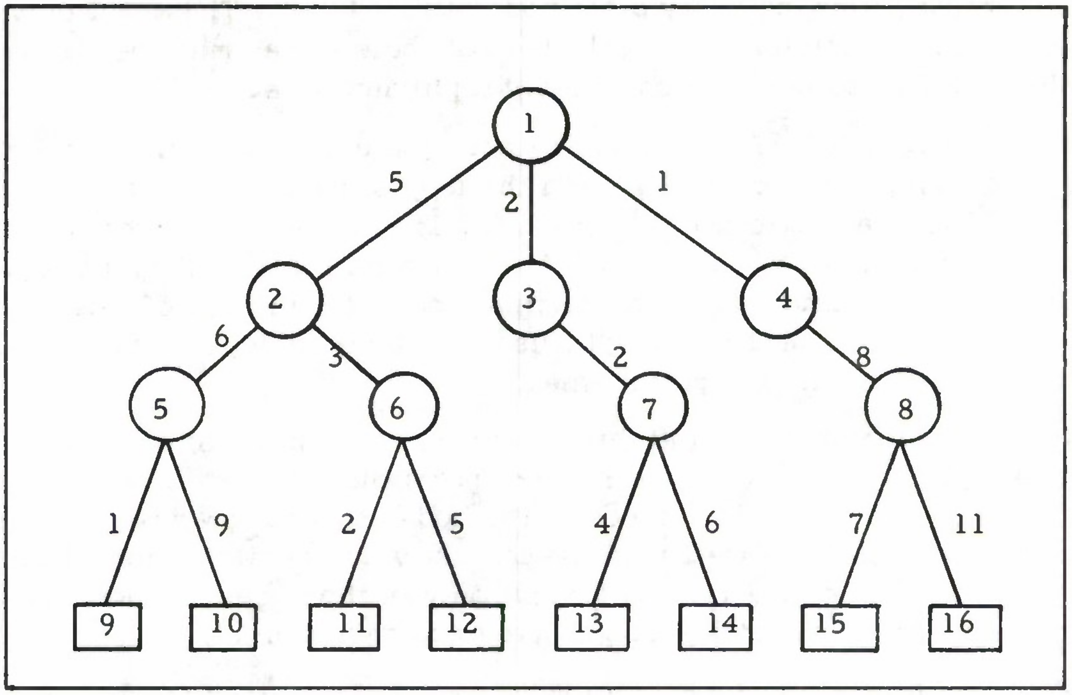


FIGURE 9. PART OF THE PARALLEL TREE

a. Tree Equivalent to Parallel Tree of Figure 8



b. Tree Equivalent to Tree in Figure 10a

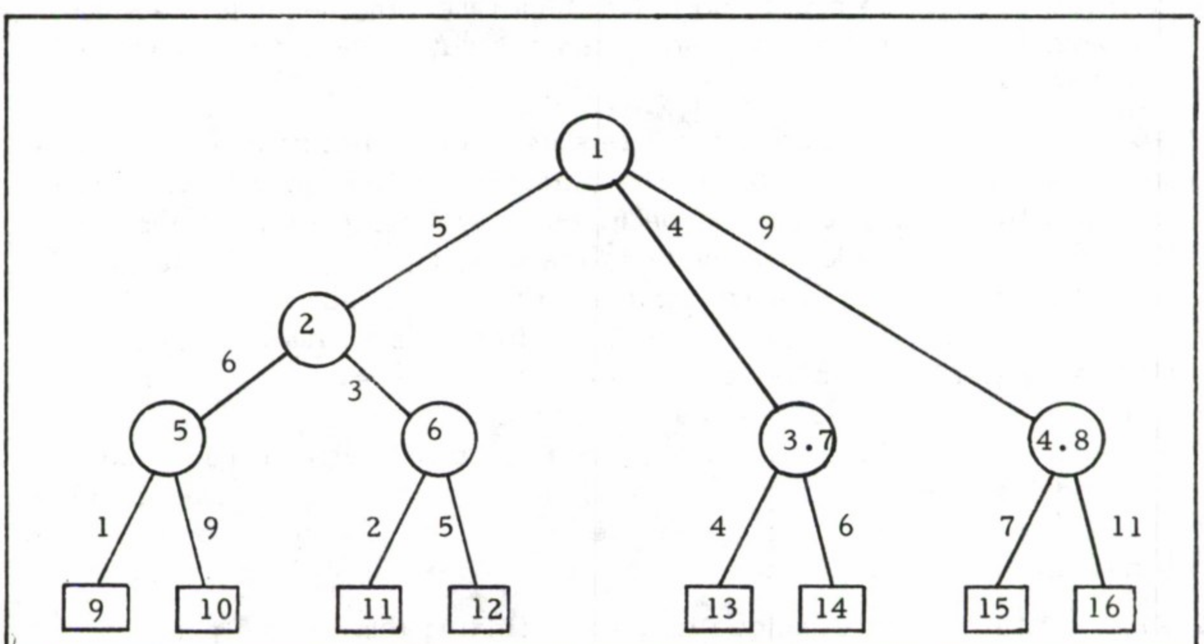


FIGURE 10. ILLUSTRATIVE EQUIVALENT TREES

same as that following every other node at that level. Thus, not only must the same decisions be available, but their values must be the same. Figure 7 is an example of a completely replicated tree.

2.43 In trees of this kind, at any node the decision would be to take the next step along the path offering the largest value. Thus at node 1, take the line leading to node 2, since $6 > 1$; at node 2, take the line leading to node 4, since the value of 7 is the greatest attained. Finally, at node 4, take the line leading to terminal node 11. In case of ties, any of the maximum-value lines may be used. This technique finds the best path by a single run through the tree.

2.44 The rationale behind this technique is simple. Since the structure is replicated, the decision-maker knows that no matter what his choice is at any node the choices at the nodes that follow will be identical. Therefore, he might as well take the largest value available for this choice. The replicated tree structure will occur only in those cases where the effect of a decision never depends on previous decisions.

2.45 Parallel-Type Trees. In some sequential search processes, certain decision nodes are followed by identical portions of the decision tree, with identical values for these decision lines. Under these conditions, it is allowable to simplify the tree structure by allowing more than one line to lead to a decision node. However, given this structure, the number under a final decision square no longer has a meaning because the same final square can be reached by more than one path. Figure 8 illustrates the situation.

2.46 Note that there are three strategies that end at node 11 and three more at 12. However, the key to the matter can be obtained by looking at the partial tree structure up to node 6, as shown in Figure 9. No matter what decisions are made after node 6, the final sum will be the partial sum up to node 6 plus the partial sum for the rest of that strategy. But to get to node 6, the lines from 1 to 2 and from 2 to 6 give a sum of 8, which is greater than the sums for the other two paths. So, no matter how the decision process proceeds, the optimal process will never use the subpath from node 1 to node 3 to node 6, nor will it use the subpath from node 1 to node 4 to node 6. Thus, the original decision tree could be replaced by the one shown in Figure 10a. If the one-decision nodes are combined, Figure 10b results, in which nodes 3 and 7 are combined.

2.47 This type of technique is used in finding the critical path through a PERT network. In the methods considered so far, the idea is to work down through the tree to each decision node that has more than one line leading into it, then find the subpath leading to that node that gives the maximum subtotal. All other partial paths leading to that node are then dropped from further consideration. Because of the large number of such interconnections in a PERT network, it is usually possible to evaluate all the strategies available, after eliminating those that contain dominated parallel paths or subpaths, to find the optimum.

2.48 Pseudo-Continuous Trees. The procedure outlined here applies in the case for which the same alternatives are available at all nodes at the same level. If the measure of effectiveness is separable and the incremental value for each alternative is the same, no matter what the preceding path, then this reduces to the completely replicated tree structure previously considered.

2.49 Suppose, as is often the case, that the choices at each node are really choices of some number, and that the value of this number has some meaning, not just as an identification procedure. For instance, in the problem of assigning frequencies to radars, the choices available at any node are the frequencies available to the radar currently being assigned. Also, it seems evident that the total interference in the environment depends, in a somewhat continuous fashion, on the frequency assigned to the radar; i.e., the total interference picture will not change much if the frequency of one radar is changed slightly. Under these conditions it is possible to talk about two strategies as being "close." Thus, if a strategy is thought of as an n-tuple of numbers, such as $(c_1, c_2, \ldots cn)$ where c_j is the choice at node j, then if two such n-tuples agree at all positions but one, and the values at that position are close numerically, the two strategies can be called "close."* The hope is that "good" strategies tend to cluster, so that if one finds a "fairly good" strategy, then a search around it is apt to find better ones.

2.50 To apply this procedure, one must have an initial "fairly good" strategy. Sometimes this is just the result of a guess or of an empirical procedure to be outlined later. Suppose the initial n-tuple is $(c_1, c_2, \ldots cn)$ where each c_i belongs to the class of numbers available for that decision.

---

* This is equivalent to imposing a metric on the class of n-tuples. The distance between strategy 1, given by $(c_1, c_2, \ldots cn)$ and strategy 2, given by $(d_1, d_2, \ldots d_n)$ is $d(S_1, S_2) = \sum_{i=1}^{n} | c_1 - d_1 |$.

The technique starts by evaluating the n-tuple obtained by replacing c_1 by its next higher value. If this improves, the value of the measure of effectiveness the next higher value is tried. This is continued until a value for the first number is tried which does not improve things.

2.51 Next, the search for better values is started in a downward direction from the initial choice. Of course, the standard of comparison now is the best value found so far. The search continues in this downward direction until a nonimproving change is tried. Then the same technique is tried with the second position of the n-tuple. This process is continued until the last position has been tried.

2.52 The search can now be restarted by varying the first element again. (Since some other variables have been changed, this will not be a repeat of the previous search but will be different combinations of inputs.) The technique can be repeated as many times as desired. However, if a complete run through the n-tuple is made with no change in variables, the process might as well cease, since the attempts made now will be merely repetitions of the previous ones.

2.53 If the measure function is separable, then an attempt may be made to obtain a good initial n-tuple by making a "one-step suboptimization" trip through the tree. This would mean that at each node, the best line leading from it is taken (this technique is used in many human decision processes). Of course, there is no guarantee that this procedure will find the optimal solution. Such a method is often called "heuristic."

Stochastic Sequential Search Processes

2.54 A more general type of sequential search process is one in which the results of choices are not deterministic but are samples from known probability distributions. Thus, in making a choice at any decision node, the decision-maker does not know exactly what the result of his decision will be. Rather, he can only associate probabilities with each of the alternatives. This is similar, but not equivalent, to the probabilistic case discussed in paragraph 2.33 in which the measurement device provides information that is not completely reliable.

SUMMARY: FRAMEWORK FOR SEARCH ALGORITHMS AND PROCESSES

2.55 This section has presented a tentative conceptual framework for both sequential search algorithms and processes. This framework is based upon the results of investigations to date. The processes are characterized by a decision tree describing the alternatives available at each stage. The applicability of search algorithms is shown to be dependent on the structure and constraints placed on the decision tree. In summary, the following quotation from Singh [7] is appropriate:

> "No computer, however rapid, could follow... through to success because the number of possible solutions...rises exponentially. It is the same with all other complex problems such as playing games, proving theorems, recognizing patterns, and so on, where we may be able to devise a recursive routine to generate possible solutions and a procedure to test them. The search fails because of the overwhelming bulk of eligible solutions that have to be tested."

> "The only way to solve such nonalgorismic problems by mechanical means is to find some way of reducing ruthlessly the bulk of possibilities under whose debris the solution is buried. Any device, strategem, trick, simplification, or rule of thumb that does so is called a heuristic.... In general, by limiting drastically the area of search, heuristics ensure that the search does terminate in a solution most of the time even though there is no guarantee that the solution will be optimal. Indeed, a heuristic search may fail altogether."

SECTION III

REVIEW AND INTERPRETATION OF SEQUENTIAL SEARCH ALGORITHMS

3.1 This section presents the methods used in the review and analysis of sequential search algorithms and some of the algorithms of particular significance that have been examined. This review and analysis is neither exhaustive nor complete, but represents a report of progress to date, highlighting those areas of special significance and utility. Conclusions and recommendations for further investigation are contained in Section IV.

SORTING OF ALGORITHMS

3.2 The basic approach to this investigation has involved the concurrent tasks of developing a general conceptual framework for characterizing the processes and algorithms (Section II). This section investigates specific realizations of these processes and algorithms. Since the literature supporting the latter is rather extensive, it was necessary to provide a mechanism for sorting and relating applicable algorithms so that common characteristics could be recognized and interpreted. This mechanism is a simple numerical coding in conjunction with the characteristics established in Section II (Figure 1). Although the actual coding and subsequent analysis of all algorithms reviewed (which are presented in the bibliography and in this section) are not completed, it has been possible to draw a number of significant inferences from the investigations.

Numerical Coding

3.3 The search algorithm characteristics are coded numerically in conjunction with the following scheme:

First Digit—Type of Search Process

1. Single-stage
2. Multistage

Second Digit—Search Space

1. Discrete
2. Continuous
3. Discrete/continuous

Third Digit—Initial Probability Distribution

1. Uniform
2. Other

Fourth Digit—Search Object

1. Stationary
2. Nonstationary
   - 2.1. Conscious evasion
   - 2.2. Nonconscious evasion

Fifth Digit—Measurement Device

1. Noiseless
2. Noisy
   - 2.1. Constant error probabilities
   - 2.2. Variable error probabilities

Sixth Digit—Ability to Partition Space

1. Single-cell partitioning
2. Multicell partitioning

Seventh Digit—Allocation of Effort

1. Equal
2. Variable

Eighth Digit—Criterion

1. Minimum expected effort (e.g., cost and time)
2. Minimum quadratic effort (e.g., cost and time)
3. Maximum probability of detection
4. Maximum information per stage
5. Maximum information per unit cost per stage
6. Other

This coding (itself a search problem) is not too cumbersome and allows for reasonable growth and reordering, if necessary. In addition, it was considered useful to characterize the methods used in the various search algorithms and their areas of application, as follows:

Ninth Digit—Methods

1. Dynamic programming
2. Information theoretic
3. Hypothesis testing
4. Conventional decision theory
5. Other

Tenth Digit—Applications

1. Human decision processes
2. Equipment diagnosis
3. Coding signals
4. Radar systems
5. Information retrieval
6. Coin weighing
7. Command and control
8. General detection process
9. Others

Dominant Algorithm Characteristics

3.4 For those algorithms that have been coded in this manner, there is a clustering according to certain characteristics such as the natural tendency toward refinement and extension of problems with particular characterizations. For example, the search space is generally discrete; when it is not, the succession of solutions deals more appropriately with allocation of resources rather than strictly sequential search processes. The initial probability distribution of the search object among the search cells is almost always uniform. Both nonstationary and stationary models are considered together with both noisy and noiseless measuring devices. More often only single-cell partitioning is possible, and the variable allocation of effort models tends toward those using a noisy measurement device. The basic reason for this is that greater effort allocations (e.g., more time, power) are assumed to increase the probability of detection by providing greater margin between "signal" and "noise."

3.5 The dominant criterion for sequential search processes is clearly minimization of expected effort. In some cases, several criteria are used either separately or jointly to provide different solutions or solutions that are invariant with the selected criteria. Three nonexclusive areas seem to be prevalent concerning methods of solution: information theoretic, mathematical programming (linear, dynamic), and statistical decision theoretic, including both Bayesian and non-Bayesian approaches. The following paragraphs examine some of the more significant algorithm characteristics and inferences that can be drawn from them.

INFORMATION THEORETIC METHODS

3.6 The information theoretic approaches use the conventional logarithmic measures of uncertainty and are generally coupled with the criterion of minimizing expected effort, measured by such factors as cost, time, number of questions, and digits. In some cases, a measure of effectiveness is established, e.g., obtaining the greatest amount of information or the greatest information per unit of effort at each stage of the process with the intent that expected effort will be minimized or close to a minimum.

3.7 Some of these methods are not on a firm mathematical foundation for several reasons. It is well known that for many types of unconstrained decision trees, which is the more general case, the "optimum" choice from among all alternatives at each stage of the search process does not necessarily optimize the entire search process. This is the underlying root of the difficulty of handling generalized search and decision processes (cf. Section II: decision trees).

Equipment Diagnostics

3.8 Several authors,[8-10/] apparently working together, have considered the application of information theoretic concepts to diagnostic problems.[8-10/] Their approach was to choose, at each step, from among all possible diagnostic tests, the one step that yielded the maximum amount of information per unit of cost. A figure of merit at each step was defined as the ratio of the entropy to the cost of performing the test. It was anticipated that such a sequential test procedure would be efficient in the sense of a low expected cost. Unfortunately, such a procedure provides no guarantees against a highly inefficient routine, as the following counter-example demonstrates.

3.9 Assume a piece of equipment, partitioned into four mutually exclusive cells, A, B, C, and D. Further, let the a priori probabilities of a defect (assuming only one defect) in these cells be

$$p(A) = \tfrac{1}{2},\ p(B) = \tfrac{1}{4},\ p(C) = p(D) = \tfrac{1}{8}$$

Also, let the costs associated with locating the defect be C_a, C_b, C_c, and C_d where C_a is the cost of making a measurement on cell A and determining whether the defect is in cell A. We are interested in locating the defect at least cost. This counterexample shows that the use of the figure of merit described above does not necessarily achieve this.

3.10 Let two methods of search be described by the following order of search:

Method 1 — A, B, C

Method 2 — B, A, C

Note that a measurement on cell D is not necessary. The expected cost, using Method 1, is

$$\overline{C_1} = \frac{C_a}{2} + \frac{C_a + C_b}{4} + \frac{C_a + C_b + C_c}{4} = C_a + \frac{C_b}{2} + \frac{C_c}{4} \qquad (3.1)$$

The expected cost, using Method 2, is

$$\overline{C_2} = \frac{C_b}{4} + \frac{C_a + C_b}{2} + \frac{C_a + C_b + C_c}{4} = C_b + \tfrac{3}{4}\, C_a + \frac{C_c}{4} \qquad (3.2)$$

3.11 In both cases, the total uncertainty is

$$H = - \Sigma\, p_i \log p_i = H\left(\tfrac{1}{2}, \tfrac{1}{4}, \tfrac{1}{8}, \tfrac{1}{8}\right) = 1.75 \text{ bits} \qquad (3.3)$$

However, using Method 1, we obtain 1 bit per measurement, and with Method 2, the information we obtain increases for each measurement. The expected number of measurements for Method 1 is 1.75, whereas for Method 2 it is 2. In both cases, of course, the expected information is equal to the total uncertainty of 1.75 bits.

3.12 The figure of merit approach dictates starting with the measurement for which the figure of merit is a maximum, e.g., choose

$$\max \left\{ \frac{H_a}{C_a}, \frac{H_b}{C_b}, \frac{H_c}{C_c} \right\} = \max \left\{ \frac{1}{C_a}, \frac{.81}{C_b}, \frac{.54}{C_c} \right\} \qquad (3.4)$$

Of the two methods, 1 and 2, the figure of merit approach says choose Method 2 if

$$\frac{.81}{C_b} > \frac{1}{C_a} \qquad (3.5)$$

or

$$1.23\, C_b < C_a \qquad (3.6)$$

In the region $[1.23\ C_b < C_a < 2C_b]$ the figure of merit approach claims that Method 2 should be used, whereas actually Method 1 should be used in this region to obtain a lower expected cost. Therefore, the figure of merit approach does not necessarily minimize expected cost, and there is no reason to believe that this approach is an efficient one in a least-cost sense.

The Coin Weighing Problem

3.13 The coin weighing problem is a classical example of a search process and is examined in some detail here and in Appendix I. With 12 coins, one of which is either lighter or heavier than the others and an equal area balance, it is possible to determine the bad coin in three weighings. The problem is to determine a search sequence that achieves this.

3.14 The greatest resolution of uncertainty is possible on the first measurement by weighing four against four.[2/] If they balance, then two courses of action provide the most information on the second weighing: two of the remaining possible bad coins against one of the possible bad coins and a true coin, or three of the possible bad coins against three true coins. Either of these choices is satisfactory. If balance is not achieved on the first measurement, then 13 possible weighings are available, any one of which provides the greatest information on the second weighing. By following any one of these sequences, the last weighing is trivial.

3.15 The point here is that an information measure (entropy) is used as the measure of effectiveness at each stage of the process in the hope that by obtaining the greatest resolution of uncertainty at each stage, the overall uncertainty can be resolved in three (or a minimum number of) weighings. This is basically a heuristic technique since, even the resolution of the greatest uncertainty at one stage, in the general case, may lead to a set of alternatives that cannot be partitioned so that the overall process is accomplished in a minimum number of stages.

Sequential Search and Coding

3.16 Various coding algorithms bear direct relations to sequential search processes such as those described above. For example, consider the binary coding of the message set [m] below:

m_1	00	where	$p(m_1) = \frac{1}{2}$
m_2	01		$p(m_2) = \frac{1}{4}$
m_3	10		$p(m_3) = \frac{1}{8}$
m_4	11		$p(m_4) = \frac{1}{8}$

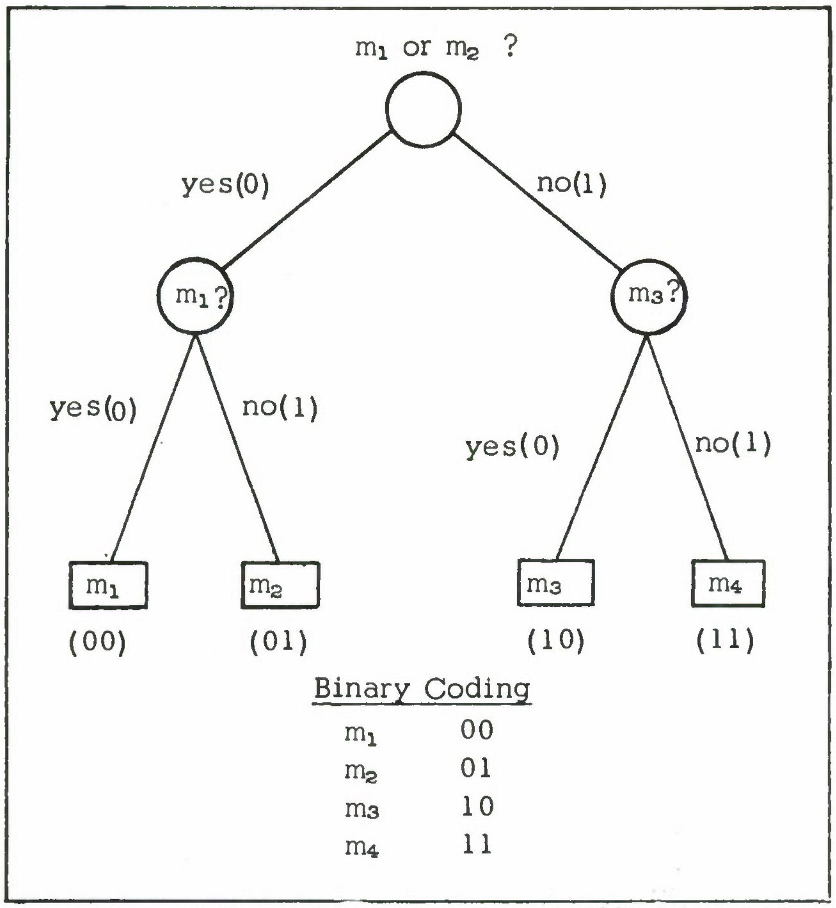
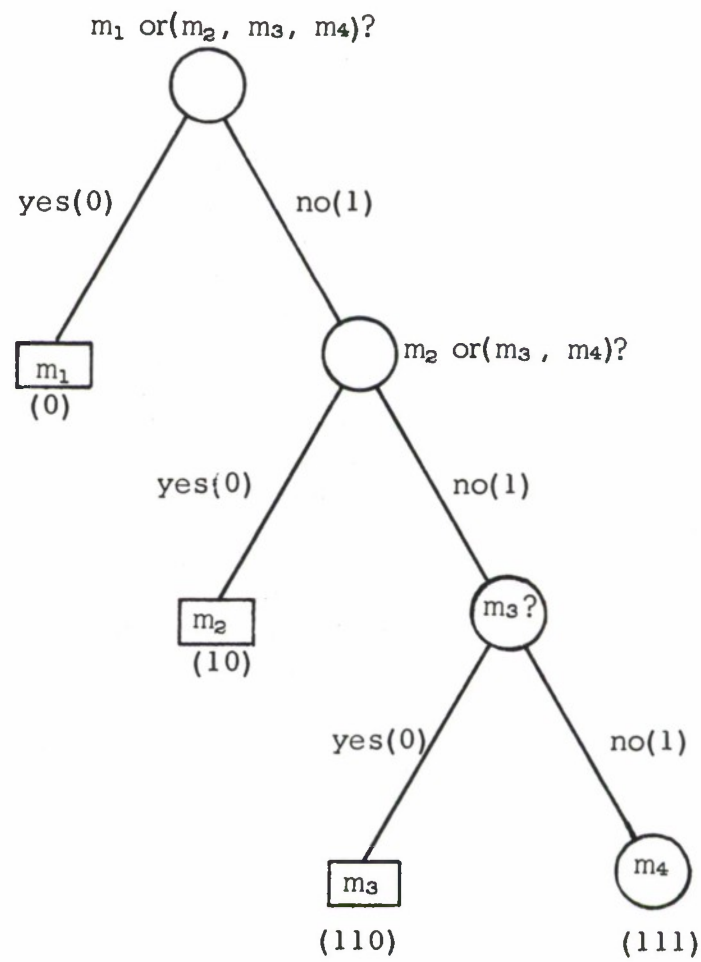


FIGURE 11. RELATIONSHIP BETWEEN BINARY CODING AND A SIMPLE SEARCH ALGORITHM



Shannon-Fano Coding

m_1	0
m_2	1 0
m_3	1 1 0
m_4	1 1 1

FIGURE 12. RELATION BETWEEN SHANNON-FANO CODING AND A SIMPLE SEARCH ALGORITHM

The average length $(\overline{L})$ of a message is two digits and the efficiency of the coding scheme is defined as

$$\text{Efficiency} \equiv \frac{H(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8})}{L \log D} = \frac{7/4}{2 \log 2} = 87.5 \text{ percent} \qquad (3.7)$$

where $H(\cdot)$ = message entropy

D = number of symbols in the coding alphabet .

3.17 This coding scheme implicitly represents a search algorithm. Consider a search for one of the four messages that is carried out by receiving answers to dichotomous questions. The zero for the first digit for both m_1 or m_2 can be interpreted as the question "Is the message either m_1 or m_2?". This process continues (see Figure 11) until the message is determined with certainty after two questions.

3.18 Although the search algorithm represented by the conventional binary coding resolves all uncertainty with two dichotomous questions, another code can be found which, on the average, resolves the uncertainty in fewer than two questions.

3.19 Consider the coding which partitions the set of messages into two equally likely groups each time an alphabet symbol is assigned. Such a code, sometimes referred to as the Shannon-Fano encoding procedure 11/, is

m_1	0	where, as before,	$p(m_1) = \frac{1}{2}$
m_2	10		$p(m_2) = \frac{1}{4}$
m_3	110		$p(m_3) = \frac{1}{8}$
m_4	111		$p(m_4) = \frac{1}{8}$

3.20 The average length of this code $(\overline{L})$ is $1\frac{3}{4}$ digits, which leads to a coding efficiency of 100 percent. This code reflects the search algorithm shown in Figure 12. Note that the code for each message is obtained by following each branch of the tree and adding coding digits as required by the decision at each junction.

3.21 If we accept the minimization of the expected number of questions or code length as a criterion for effectiveness of the search strategy, then the Shannon-Fano technique is more efficient than straight binary encoding. However, the variance is zero for the latter case and nonzero for the former. For example, the probability is one quarter that three questions will have to be asked to resolve the uncertainty. This simple example shows the classical differences between using first and second moments as criteria for effectiveness of a search algorithm.

3.22 Although it is true that the Shannon-Fano encoding procedure provides a smaller average length code than the binary encoding, it is not the most efficient procedure. The Huffman code [12] is, in general, a code that minimizes the average code length. Actually it is a minimum redundancy code with the irreducibility property.

3.23 These coding examples illustrate the rationale for distinguishing between a search process and a search algorithm and the formal correspondence between coding and search theory. Figures 11 and 12 present particular algorithms for isolating a message. Each particular algorithm is, in effect, a path through a decision tree that represents the set of all possible algorithms for isolating the search object.

Progression of Noiseless Coding Algorithms

3.24 The above example shows the relationship between minimum redundancy coding for the discrete noiseless channel and a search process. The progression of such coding algorithms started with Shannon [13] and Fano [11] for near-optimum codes for which all code symbols have equal cost. As indicated above, Huffman [12] developed a combinatorial algorithm that does yield an optimum code (or search process) for this equal cost case. The Shannon-Fano technique was extended to the nonequal cost case by Blachman [14], analogous to a search process in which the choice at each stage may, for example, take different times or cost different amounts as in the above diagnostic counterexample. The Blachman technique was improved by Marcus [15] using some of Huffman's results. A further extension was developed by Karp [16] for which the symbols are not necessarily equally probable and the costs not necessarily equal. In his development, Karp uses an integer programming algorithm. Also the codes of interest all have the prefix property, i.e., it is not possible to obtain another member of the code group by adding digits to any given member of the code group.

Radar Search Processes and Reconnaissance

3.25 A large body of literature is devoted to the applications of information theory to radar detection problems, including both single-stage and sequential detection. Such analyses are not examined here since they are outside the scope of the sequential search processes defined in this investigation. However, the distinction is sometimes difficult to discern and a selected number of such algorithms are examined for general background and perspective.

3.26 In 1964, Machol [17] examined information-theoretic and other limitations on the operation of a radar system. However, his paper stresses fundamental limitations rather than explicit search algorithms. Danskin [18] also uses an information-theoretic measure of performance in examining reconnaissance, but he does not set forth explicit sequential search procedures.

3.27 Novosad [19] examines search problems and information theory largely in the context of a radar search for a missile. He considers two criteria:

a. Maximizing the probability of detection and

b. Minimizing the uncertainty that can be expected at the conclusion of the search.

He shows that if the entropy is used as a measure of uncertainty (item b), the two criteria above can lead to different search strategies. In item a, one optimum strategy divides the number of examinations equally between two search areas. In item b, all the search effort is devoted to one area.

MINIMUM EFFORT ALGORITHMS

3.28 Many sequential search processes are posed in generalized terms without emphasis on particular applications. It is useful to group the algorithms developed in these cases in terms of the criteria that are used to establish the search strategy. Most of these investigations deal with the criterion of minimizing expected effort, measured by cost, time, or some function of effort. Note also that several of the information-theoretic approaches utilized the criterion of minimizing the expected number of steps in the process, or the expected length of a code, which are entirely equivalent. The most generalized approach was Karp's algorithm, which used integer programming techniques for solution.

3.29 Two related papers [20,21] deal with a search problem that is historically significant in that they were motivated by Koopman's work [6,22] on search, which is more appropriately an allocation of resources problem. However, in Blachman's work, an algorithm is developed which minimizes the expected delay between the appearance and detection of an object, where the time of appearance is distributed uniformly and the probability of appearance in the ith location is p_i. A conventional Lagrange multiplier approach is used. In the second paper by Blachman and Proschan [21] the object is to maximize a gain function that is nonincreasing in the delay between arrival and the beginning of the detecting look. In this case, objects arrive in accord with a Poisson process. Here again, conventional but complicated maximization-of-functions techniques are used to find a solution.

3.30 In a 1963 paper by Dobbie[23], a sequential approach to search theory is presented that is also based on the work of Koopman [6] and others [24-26]. Dobbie considers two criteria: minimizing the expected effort to attain a given probability of detection and maximizing the detection probability with a given effort. He concludes that

> "...the expected effort is minimized by always distributing the effort to maximize the probability of detection with the effort expended thus far. This is a distribution requirement that is very difficult to meet. By contrast, the distribution that maximizes the probability of detection with a given effort can be non-optimal for all values of the effort less than the total effort; the effort can be applied by any schedule that finally attains the required distribution when all the effort has been applied."

In developing his solutions, Dobbie uses the principle of optimality of dynamic programming.

3.31 Several studies conducted by MIT students have been directed toward minimum expected cost search processes.[26-29] In the 1962 study [28] it is shown that several criteria of optimality lead to the same search policy. Pollock's work [27], the most extensive and most recent, involves a stationary target and a noisy measurement device such that probability density functions can be established for cases in which a target either is or is not present. A modified Wald sequential probability ratio test is employed and a search strategy developed by solving a functional equation of the dynamic programming type. This is another example in which mathematical programming was the basic algorithm used to solve a search problem in which minimization of expected effort was the criterion.

Minimum Effort Diagnostic Algorithms

3.32 Several algorithms have been developed concerning diagnostic routines based on minimization of effort. Gluss [30] considers policies that minimize the expected amount of time consumed and penalties paid, and his analysis is based on dynamic programming concepts. Several years later[31], Chew considered a diagnostic routine (searching plan) for minimizing the expected time or cost. Such a plan "instructs the searcher to inspect at each stage the box in which the object is most likely to be found." His basic approach is Bayesian and his work is shown to be an extension of other allocation of effort and search algorithms.[6,32,33]

Minimum Time Radar Search

3.33 Posner [34] has considered a radar search for a satellite lost in a region of the sky and is concerned with minimizing the expected search time. A preliminary and final search are postulated, with the former resulting in a ranking of various portions of the sky and the latter examining the regions of greatest likelihood. A preliminary narrow beam search is proposed as best.

3.34 This is the case of a noisy measurement device whose characteristics change with the allocation of effort to each search cell. Results are obtained for the two-stage case (preliminary and final search) and extended to the multistage case. Reference is also made to a paper in preparation which proposes the following optimal strategy for minimizing expected search time: search the most likely cell until it is no longer the most likely; then search the cell that has become the most likely. Note that Chew's diagnostic routine [31] sets forth a similar search strategy.

HUMAN SEARCH AND DECISION PROCESSES

3.35 Part of the motivation for this investigation is to obtain a clearer understanding of the manner in which a person carries out search and decision processes. The search algorithms discussed previously are tools that would enable the rational human to implement a search in an optimal manner. However, as an imperfect collector, observer, and processor of information, these so-called optimal policies are rarely implemented, particularly when such operations take place under stress.

3.36 Since the literature on this subject is massive, the intent was simply to provide the serious investigator of human search and decision processes with a concise characterization of those processes and algorithms and a sounder foundation for further theoretical or experimental work. Further, it was intended that such a characterization would place in evidence some obvious missing links through an analysis of the morphological structure of these processes.

3.37 The studies that have been examined are cited in the bibliography. They are quite diverse and deal with all aspects of the categories of Figure 1, particularly the human as a measurement device and the problem of identifying criteria under which the human operates. Appendix II contains some analyses carried out in response to specific queries by the Air Force concerning scoring systems that tend to force a human subject to act in accord with his subjective probability estimates.

3.38 A good example of the relationship between the previous characterization of search processes and human decision processes is a study of a multistage decision task by Rapoport [5]. In this investigation, a brief classification of decision theory is provided together with a model for a particular multistage decision task.

3.39 The model is set up such that the solution, in terms of an optimal set of sequential decisions, is provided by existing dynamic programming methods. Experiments are then carried out with human subjects, the results of which are compared with the optimal solution.

3.40 What is significant about Rapoport's study in this investigation is that it is closely allied to sequential search processes and demonstrates the heuristic methods that are used by humans for moderately large or complicated decision trees. Interpreted somewhat differently, given enough time and technical expertise, the optimal solution to the posed multistage decision process could have been found by the subjects. In the absence of this, heuristic methods were adopted that provided results approximating this optimum in varying degrees. For the multidimensional search and decision process (Section II) for which no solutions exist at this time, heuristic methods are also used to obtain approximate solutions. It is also significant that dynamic programming provides a firm theoretical foundation for the solution to a broad class of sequential search problems.

SUMMARY

3.41 This section has provided a brief review and interpretation of sequential search algorithms that have been examined. The process of sorting and characterizing these algorithms was also discussed. Section IV provides tentative conclusions and recommendations, the results of this investigation. A brief description of some useful techniques for solving sequential search processes is provided in Appendix I. Appendix II presents an analysis of scoring systems as specifically requested by ESD.

SECTION IV

SUMMARY

4.1 This section summarizes the results of this investigation of sequential search processes in the form of conclusions, followed by recommendations for continued analyses and future applications that appear to be of significance to the Air Force.

CONCLUSIONS

4.2 Sequential search processes can be characterized as decision trees with various structures and properties. The utility of sequential search algorithms depends on the structure, properties, and constraints they impose on the tree as well as externally imposed measures of effectiveness and criteria (cf. Figure 1).

4.3 A particular search algorithm is the means by which one or more "best" paths in the decision tree are chosen, although the realization of that algorithm may itself be represented by a decision tree.

4.4 A broad class of sequential search algorithms may be effectively characterized by Figure 1 and the tentative categories in Section III. For purposes of further applications, with perhaps some modification and extension, this is an adequate and concise characterization of such algorithms.

4.5 An exhaustive search through a decision tree is impossible in most practical problems because of the enormous number of possible paths. Computationally tractible techniques do not exist for handling even moderately unconstrained decision tress. In attempting to simplify a problem so that is is computationally feasible, it is usually more effective to reduce the number of decisions to be made rather than the number of available alternatives at each decision point.

4.6 Any explicit coding procedure, developed from information-theoretic considerations or otherwise, can be put into a one-to-one correspondence with a decision tree. Algorithms for the formal development of codes, therefore, may be used to establish search strategies.

4.7 Aside from the area of coding and selected other applications that can be related to coding (e.g., the coin weighing problem), the utility of information-theoretic approaches to sequential search has been limited to characterizing and interpreting such processes rather than finding explicit algorithms for their solution.

4.8 The most significant advances in the development of sequential search algorithms have been based on conventional decision theory and mathematical programming. The technique of dynamic programming has been employed with particular success. (Additional details are presented in Appendix I.)

4.9 The search criterion most widely accepted is minimization of expected effort, where such effort is defined in terms such as cost, time, and number of stages. The reason for this is twofold: the mathematics of sequential search processes is generally most tractible for this case and, on independent grounds, there is reason to believe that the human carring out such a process most often considers this criterion to be most important. Another significant criterion is that of maximizing detection probability, which is often used in conjunction with a fixed limitation on total available search effort. Algorithms based on these criteria are available and have been applied in the areas of coding diagnostics, radar search, sequential detection, and warfare.

RECOMMENDATIONS

Completion of Present Investigation

4.10 Further activities recommended within the present scope of this investigation are:

a. Completion of the algorithm coding and subsequent sorting and interpretation by similar characteristics to provide a compendium of such algorithms and a mechanism for their retrieval

b. Establishment of an explicit correspondence between the formal structure of search and decision trees and applicable algorithms.

Future Activities

4.11 Further investigations of the structure and taxonomy should be carried out as an adjunct to specific applications of sequential search algorithms to problems confronting the Air Force at this time. Within the context of these applications, further developments should emphasize:

a. The nature of the constraints most often imposed on sequential search processes, e.g., maximizing expected effort, probability of detection, and stationarity of the search object

b. The use of the most applicable algorithms (e.g., dynamic programming) to obtain greater search efficiencies

c. Standardization of the presentation format of the sequential search processes and algorithms into a handbook to facilitate their application.

4.12 More explicitly, the areas in which sequential algorithms can most effectively be applied to Air Force activities are command and control, computation, surveillance, reconnaissance, communications, and human decision processes. Specific subareas to which the theory is applicable include the search for:

a. Aircraft or missiles by a radar system

b. Fixed or moving installations or platforms from an aircraft or satellite

c. Documents or files in a large-scale information retrieval system

d. Data bits in a generalized memory bank

e. Desired signals in an interference background

f. Malfunctions in an electronic system

g. Data, information, or signals by humans

h. Unoccupied lines in a communications switching system

i. Patterns on a surface.

APPENDIX I

SELECTED MATHEMATICAL TECHNIQUES

LINEAR PROGRAMMING

I.1 Linear programming is a technique for finding the values for the variables x_i, $i = 1, 2, \ldots, N$ to minimize

$$f(x_1, x_2, \ldots x_n) = \sum_{i=1}^{N} c_i x_i \qquad \text{(I.1)}$$

subject to the constraints $x_i \geq o$, all i, and

$$\sum_{i=1}^{N} a_{ij} x_i = b_j, \; j = 1, 2, \ldots M \qquad \text{(I.2)}$$

with $N > M$ (i.e., more variables than constraints) and the c_i, b_j, and a_{ij} are given constants. The function f is called the objective function. Note that if the problem were to maximize

$$g(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{N} d_i x_i \qquad \text{(I.3)}$$

this could be brought into the minimization format by seeking the minimum value for

$$h(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{N} (-d_i) x_i \qquad \text{(I.4)}$$

I.2 Since, unlike dynamic programming, linear programming requires a particular format, there exist computer programs to find these solutions. These programs can handle problems with several thousand variables, with several hundred constraint equations, not counting the positivity constraints. Since these programs exist, the details of the computational techniques would be of interest mainly to computer specialists.

However, a brief explanation of the general ideas will be given. In order not to burden this discussion with much cumbersome notation, let us take up the case where there are 10 variables and 6 constraints. The 10 numbers to be found will be considered as a point in a 10-dimensional space; this will be done merely to allow a language simplification.

I.3 It can be shown that if there is a unique, finite solution, then exactly 6 of the coordinates will be nonzero. Also, even if the solution is not unique, it will be attained at points whose coordinates satisfy these conditions, and every solution point can be expressed as a convex linear combination of points from this set. Thus, the solution technique searches through points with exactly 6 nonzero coordinates, satisfying the constraints.

I.4 The search starts with a first point, satisfying the constraints and having exactly 6 nonzero coordinates. There are many techniques for finding such a point; let us assume that one has been found. The program evaluates the objective function at this point, and then looks over the "neighboring" points (this term will be defined presently) to see if it can find one that satisfies the constraints, and still get a smaller value for the objective function. If no such point is found, it can be shown that the present point is the optimal one. If such a point is found, a search is made of its neighboring points, under the same conditions. Note that, once a point is left, it can never be returned to, since the value for the objective function must be less than the current best value.

I.5 The initial point contained 6 nonzero coordinates and satisfied the constraints. It can be shown that this is the only point satisfying the constraints and having nonzero coordinates in these positions. A neighboring point is a point also having exactly 6 nonzero coordinates, of which exactly 5 occur in the same positions as in the original point. Thus, if the original point is

$$(x_1, 0, 0, x_4, 0, x_6, x_7, x_8, x_9, 0) \qquad (I.5)$$

a neighboring point would be

$$(0, y_2, 0, y_4, 0, y_6, y_7, y_8, y_9, 0) \qquad (I.6)$$

Note that the values in the common positions are not assumed to be the same; instead, the property of having a nonzero entry remains the same in five of the positions. Again, there will be only one point satisfying the constraints and having nonzero coordinates in these particular positions.

I.6 The program thus proceeds from the initial point to a neighboring point and continues through other neighbors, until the optimal point is found. Computational experience has shown that the number of such steps, or "iterations" as they are called in the field of linear programming, is always less than 3 times the number of constraining equations.

I.7 Whereas this technique will solve any type of linear programming problem, certain special types of problems are solved by special techniques that work for larger numbers of variables or constraints, or obtain solutions in a shorter period of computer time. The so-called "Transportation Problem" falls into this category.

BACKTRACK PROGRAMMING

I.8 "Backtrack programming" is a method for finding the best path through a decision tree with a separable objective function which, under certain conditions, will involve a less-than-exhaustive search. Like dynamic programming, and unlike linear programming, it is more a philosophy than a definite procedure. Basically it consists of a bookkeeping procedure which omits examining certain paths through the tree when it can be seen, from prior knowledge, that they will not be optimum. There seems to be only one paper in the literature that deals with this, 35/ and in this paper is the statement "Backtrack is no more than an educated exhaustive search procedure."

I.9 To illustrate by means of an example, suppose we are trying to find nonnegative integer values for the variables $x_1, x_2, \ldots x_8$, subject to the constraint

$$\sum_{i=1}^{8} x_i = M \qquad (I.7)$$

for some known constant M, such that

$$\sum_{i=1}^{8} f_i(x_i) \leqq T \qquad (I.8)$$

where T is a given constant, $f_i(x_i)$ are known functions, and $f_i(x_i) \geqq 0$ for all f_i and for all allowable values for each x_i.

I.10 Now suppose a search is started by fixing a value for x_1 and trying all values for x_2 with this. To be more particular, suppose we are trying to find a solution with $x_1 = 3$. Then, in trying to find a value for x_2, the integers to try for this variable are those between zero and M-3. Suppose that, for all of these values for x_2,

$$f_1(3) + f_2(x_2) > T \tag{I.9}$$

I.11 Then it is obvious that no solution can be found using $x_1 = 3$ since the $f_i(x_i)$ are all nonnegative. Henc the number of possibilities has been reduced.

I.12 As the paper points out, some sort of balance must be achieved between the search procedure and the bookkeeping procedure, because if the bookkeeping gets too complicated it becomes more time-consuming than the straight exhaustive technique.

DYNAMIC PROGRAMMING

I.13 "Dynamic programming" is a way of looking at problems of maximizing or minimizing functions of several variables, satisfying constraints, that sometimes reduces an N-dimensional problem to a sequence of one-dimensional problems.

I.14 As an example of this philosophy (because, unlike linear programming, it is not a well-defined technique), let us look at the following problem: Maximize the product of N variables subject to the constraints that all N variables be positive and their sum be less than or equal to a fixed amount B. In mathematical notation, this can be written:

Maximize $$F(x_1, x_2 \ldots x_N) = \prod_{i=1}^{N} x_i \tag{I.10}$$

subject to $$x_i \geqq 0, \text{ all } i, \text{ and } \sum_{i=1}^{N} x_i \leqq B \tag{I.11}$$

I.15 This is a generalization of the elementary calculus problem of finding the rectangle of maximum area with given fixed perimeter. If $N = 2$, then $x_1 = B/2$, $x_2 = B/2$ gives the maximum value. Note that this solution did not depend on the particular value of B. No matter how much is available, the best way to solve the two-variable problem is to divide the available amount into two parts.

I.16 Suppose we look next at the three-variable problem. For any choice of a value for the third variable, we know how to proceed to maximize the product of the other two. Thus, if we allocate an amount y for the third variable, there will be B-y remaining for the other two, and the best way to divide this amount among them is to split it equally, with (B-y)/2 for each variable. For each choice for the third variable, then, the conditional maximum value (being dependent on the value for y) is given by

$$G_3(y) = y \cdot \frac{B-y}{2} \cdot \frac{B-y}{2} \tag{I.12}$$

I.17 The problem is: What would be the best value for y? But using the standard method of elementary calculus, we solve the equation

$$G_3'(y) = 0 \tag{I.13}$$

so that

$$\frac{B^2 - 4By + 3y^2}{2} = 0 \tag{I.14}$$

This gives y = B/3 and y = B. However, this second value gives a minimum. Thus $x_3 = B/3$, leaving $x_1 = x_2 = 1/2\ (B - B/3) = B/3$. Again, the answer to the three-variable problem can be expressed as a policy: Divide the available amount into three equal parts.

I.18 Continuing one more step, let us try the four-variable problem. Again here, for every choice of a value for the fourth variable, we know how to proceed with the other three to maximize their product. Thus, for every choice y for a value for x_4, if we continue by making the "best" allocation of the rest, we get

$$G_4\ (y) = y \cdot \left(\frac{B-y}{3}\right)^3 \tag{I.15}$$

I.19 Again, finding the "best best" by ordinary calculus, we find y = B/4, so the optimal allocation for the four-variable problem is

$$x_1 = x_2 = x_3 = x_4 = B/4 \tag{I.16}$$

I.20 If we wanted to continue this way, we could solve for N = 5,6.. etc., each time using the previous solution to solve the new problem.

Thus we have replaced the original problem of maximizing a function of N variables by a sequence of N-1 problems, each involving the maximizing of a function of one variable. (In this particularly simple problem, it should be evident that the optimal solution for the problem with K variables is to allocate B/K to each variable; this can be proved by using mathematical induction.)

I.21 Let us examine the qualities of this problem that enabled us to apply this technique for solution. The key feature was that the function to be maximized in the problem with K variables contained the function used for the problem with K-1 variables (with the constraints changed), and this was the only place the first K-1 variables appeared. In a more mathematical notation, if $F_K(x_1, x_2 \ldots x_K)$ is the function to be maximized in the problem with K variables, and $F_{K-1}(x_1, x_2, x_{K-1})$ was the corresponding function for K-1 variables, then

$$F_K(x_1, x_2, . x_K) = G\left[F_{K-1}(x_1, x_2, . x_{K-1}), x_K\right] \qquad (I.17)$$

I.22 F_K had to be a nondecreasing function of F_{K-1}, since when we wanted to find the conditional maximum for F_K, we made F_{K-1} as large as possible. We also needed to be able to express the maximum value of F_{K-1}, for any value of x_K that satisfied the constraints, as a function of x_K. Thus the solution of the problem of maximizing F_{K-1} must be a "policy," that is, a function of a parameter whose value can range from 0 to B. In the more general case where the constraint region is more complicated than the interval that was used in this model problem, for every value of x_K such that there are points in the constraint region having this number for their Kth coordinate, the optimal value for F_{K-1} over that set of points must be known.

I.23 For the technique to be practical, then, the optimal policy for F_{K-1} should be expressible explicitly as a function of x_K. For instance, if the optimal policy for F_{K-1} is known only to the extent that it is obtained by the simultaneous solution of a set of nonlinear equations involving the x_i, then the ensuing one-dimensional maximization problem may lead to grave computational difficulties. (In fact, the solutions may not even be continuous functions of the parameter.)

I.24 The same sort of ideas may be applied to problems in which only a discrete set of values is available for the variables. An example of this could arise in a problem in allocation of resources with a fixed budget, in which the resources come in unit size. To simplify our problem, assume that the costs per unit of the resources also are integers. (This problem is based on one given in "Nonlinear and Dynamic Programming" by Hadley.[36] Suppose that the return from using x_i units of resource i is given by a known function $f_i(x_i)$. Then the problem is

$$\text{Maximize} \qquad F(x_1, x_2, \ldots x_N) = \sum_{i=1}^{N} f_i(x_i) \tag{I.18}$$

subject to the constraints

$$x_i \geqq 0, \quad \sum_{i=1}^{N} a_i x_i \leqq B \tag{I.19}$$

where the a_i, x_i, and B are all nonnegative integers.

I.25 The expression $\Sigma_{i=1}^{N} a_i x_i$, for any possible choice of K and the x_i, will always take on only nonnegative integer values, since both the a_i and the x_i are integers.

I.26 The philosophy for solving this problem is the discrete analog to that used in the continuous case. Suppose we have found the optimum policy for the problem involving K-1 variables, that is, for every integer M less than or equal to B we know the optimum allocation for $x_1, x_2, \ldots x_{K-1}$, subject to

$$\sum_{i=1}^{K-1} a_i x_i = M \tag{I.20}$$

We can think of these results as being kept in a table, and now we desire to make up a similar table for the sum of the first K functions. To do this, let us see how we would make up the entry in the new table for the integer M_o. Thus we want to find the maximum value for

$$\sum_{i=1}^{K} f_i(x_i) \tag{I.21}$$

subject to

$$\sum_{i=1}^{K} a_i x_i \leq M_o \tag{I.22}$$

all the other constraints holding as before.

I.27 Thus we have the problem of allocating M_o among x_1, $x_2 \ldots x_K$ in an optimal fashion. To do this, we look at all the possible ways of allocating part of the M_o to x_K and the rest to the other variables. So, suppose we want to see what would happen if we allocated x_K^o for x_K. (Of course, we must have $a_K x_K^o \leq M_o$.)

I.28 Then there is available $M_o - a_K x_K^o$ for the variables $x_1, x_2, \ldots x_{K-1}$. But the table for F_{K-1} already lists the best allocation for these variables, given $(M_o - a_K x_K^o)$, since it lists the best for every integer up to B, and $M_o - a_K x_K^o \leq B$.

I.29 Now, with this allocation of x_K^o we can compute max $F_K(x_K)$. Whichever value for x_K that maximizes this is the best amount for x_K; and this value, together with the values for x_1, $x_2 \ldots x_{K-1}$ just obtained, is the value for this particular way of splitting up M_o.

I.30 Using this method, letting x_K^o successively equal 0, 1, 2, ... up to the biggest possible value that x_K can take subject to $a_K x_K^o \leq M_o$, and then letting $(M_o - a_K x_K^o)$ be available to x_1, $x_2, \ldots x_{K-1}$ we pick the allocation that gives the biggest value—that is, the "best best." This becomes the table entry opposite M_o in the table for the sum of the first K variables.

I.31 Since it is evident how the table should be made up for x_1 all by itself, and since the method just outlined shows how to get the next table given the table previous, the table for any number of variables can be obtained.

I.32 Although this method looks complicated, and indeed a computing machine should be used for most problems, it can be shown that it is much more efficient than a straightforward exhaustion method.

COMBINING PROBABILISTIC AND DETERMINISTIC EVENTS

I.33 To illustrate the problems involved and the various choices of criteria possible, let us set up a model problem. Suppose that there are two urns on a shelf. Urn number one we know to contain 99 red marbles and 1 blue, while urn number two contains 60 green marbles and 40 yellow. Knowing the composition of each of these urns, you choose the one you want and then choose, without looking, a marble from that urn. A sum of money is awarded you, depending on the color of the marble drawn, in accordance with the following scheme:

Marble Color	Payoff
Red	10
Blue	1000
Green	20
Yellow	25

I.34 The total process can be illustrated by a tree, in which a triangular box indicates a chance event. Figure 13 illustrates this. The numbers alongside the lines leading from the triangular boxes indicate the probabilities of taking those paths, and the numbers in the square boxes at the end give the values for these paths.

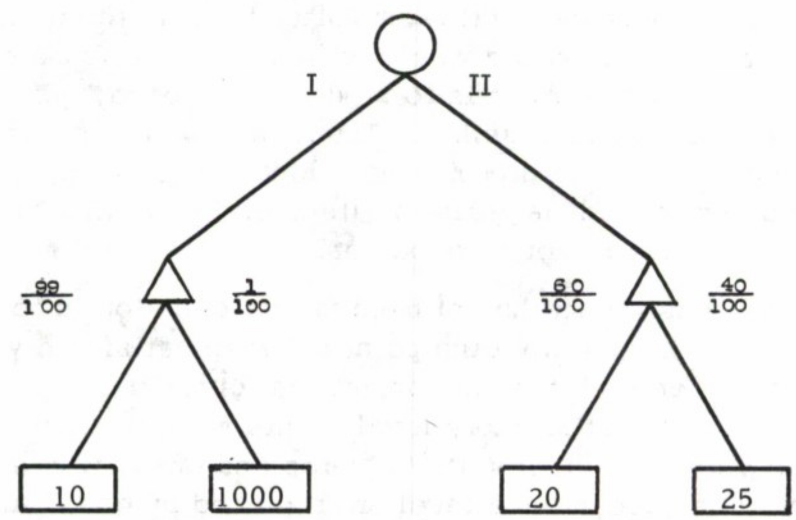


FIGURE 13. SIMPLE DECISION TREE

I.35 In this case, the decision-maker must decide whether he prefers urn one, which gives him a 99-percent chance of gaining $10 and a 1-percent chance of gaining $1000, to urn two, which gives him a 60-percent chance of gaining $20 and a 40-percent chance of gaining $25.

I.36 One commonly used method for making this choice is to take the path that offers the largest expected value. In this case, the expected value for urn one is

$$E[1] = \frac{99}{100} \cdot 10 + \frac{1}{100} \cdot 1000 = 9.90 + 10.00 = 19.90 \qquad (I.23)$$

while the expected value for urn two is

$$E[2] = \frac{60}{100} \cdot 20 + \frac{40}{100} \cdot 25 = 1.00 + 10.00 = 22.00 \qquad (I.24)$$

So, if this criterion is used, urn two will be chosen.

I.37 The reasoning behind the use of this criterion is the theorem that says if a probability experiment is performed N times, where N is a large number, the total return from these N trials will tend to be relatively close to NE, where E is the expected value of a single trial. Here we mean by "relatively close" that the difference between the actual sum and the quantity NE, when divided by NE, will tend to be close to zero. The larger N is, the more confident we are that this quotient will be small in magnitude. Thus, in this particular case, for urn one we should expect the total return after N trials to be close to 19.90 N, whereas for urn two we would expect the total to be close to 22.00 N. But, if the experiment is to be done only once, neither the result $19.90 nor $22.00 is possible. Thus, although this criterion may be quite useful when we are picking a long-range policy for an experiment to be performed many times, its usefulness when the choice is to be made only once is not so apparent.

I.38 Another possible method of making this decision is to set some level of return as being the dividing point between satisfactory and unsatisfactory, and then pick the experiment that gives the highest probability of attaining the satisfactory level. This is equivalent to replacing all the original values of the results by zeros and ones, where original values over or at the acceptance level are replaced by ones, and those under by zeros, then applying the expected value criterion to this zero-one problem. Thus, in this problem, suppose the acceptable level were set at $23. Then urn two would be chosen. However, if it were set at anything above $25, urn one would be chosen. This lack of continuity in the choice process might leave one a little uncomfortable about using it.

I.39 Another possibility would be to use some combination of these two methods. Thus, one could set a level below which returns would be worthless; returns above this level would have a value proportional to the difference between them and the acceptance level. Then, the expected value criterion could be used on the problem with this new set of payoffs. A closer look at this scheme indicates that the only real change from the original expected value scheme has been in the valuation of the possible results. In fact, most other methods for making this choice seem to involve nothing more than a change in the valuation of the payoffs, followed by the use of the expected value criterion on the new problem.

I.40 One important point is that in trees with probabilistic choices, certain criteria which were meaningful in the completely deterministic case are not meaningful. For instance, the criterion "find the path yielding the largest return" is certainly applicable in the completely deterministic case, but it is meaningless in the probabilistic case.

I.41 Finally, also note that the schemes for combining probabilities and outcomes that were outlined here are also applicable to problems in which more than one probabilistic stage is reached in the problem. The size of the tree involved in such a problem makes it difficult to draw one here, but a verbal explanation may help. Suppose that initially one of two paths must be chosen, and that each of these paths leads to two possible nodes, the node being chosen randomly, but with known probability. Now, from each of these four nodes, there are two choices, each of which gives rise to two possibilities with known probabilities, and the value of the final payoff for each of these 16 results is known.

I.42 In this problem there are two choices to be made: one initial choice and one after the first chance event. A strategy consists of a choice at the first level plus a scheme for making a second-level choice to go with each possible outcome from the first-level probabilistic event. With each of these strategies, however, the probability of attaining each of the payoff boxes is easily computed, so that any scheme for rating strategies by combining probabilistic and deterministic events is applicable.

Extension to Noisy Measurement or Detection Devices

I.43 As discussed earlier, in a stochastic decision process the decision-maker chooses from among a class of probability distributions.

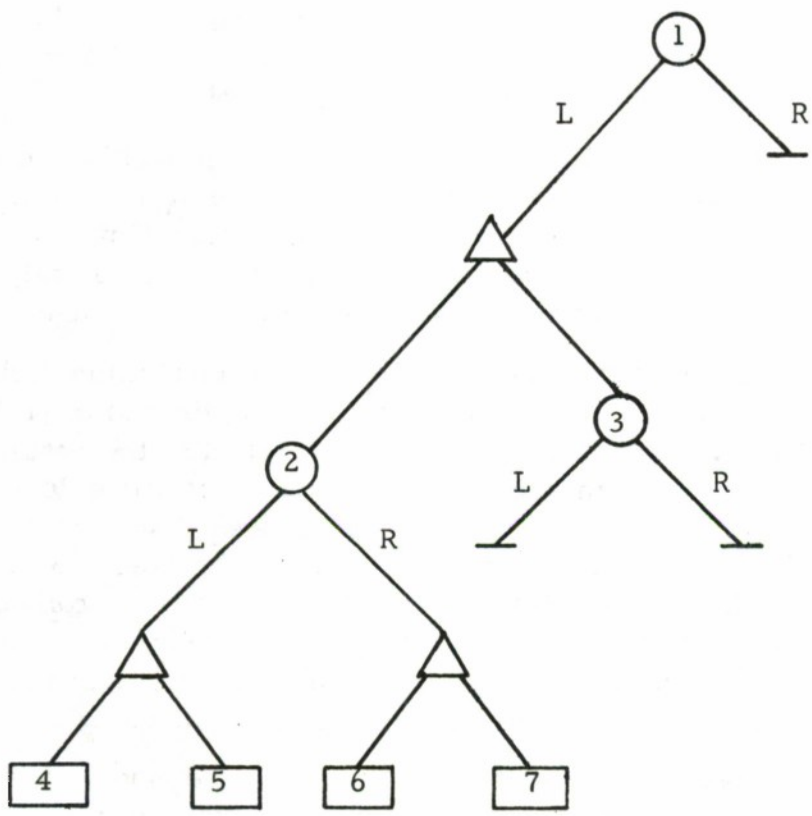


FIGURE 14. DECISION TREE ILLUSTRATING NOISY DETECTION DEVICE

A sample is then obtained from the distribution of his choice, and, having observed this sample, he then makes a choice from the class of distributions available at the next level, continuing this way until the final step in the process is reached. The objective, of course, is to find the best decision rule; i.e., he should have a first choice, and then, for each possible outcome of that first choice, a second choice, etc.

I.44 Suppose now that, instead of the decision-maker observing directly the samples from the probability distributions, the result is reported to him by a device that can be inaccurate, e.g., a radar; this will be referred to as a "noisy detection device." We assume, of course, that for each possible sample from each of the available distributions, the probability of each possible report is known. We will show that this extra complication can still be handled within the context of a stochastic decision tree.

I.45 To that end, let us examine the simplest possible problem to which this complication can apply. The extension of the idea to more complex problems should be evident. We will initially look at the problem as it would be with a perfect reporting device and then see what changes must be made if the device has the possibility of reporting incorrectly.

I.46 In this problem, there are two distributions to choose from at the first step. Each distribution has just two possible outcomes. Thus there are four possible starting points for the next level. Each of these, in turn, gives a choice from two distributions, with two possible samples from each distribution. Since the problem ends at this level, there are 16 possible ending points. Figure 14 is the portion of the decision tree ensuing from the choice of the left branch at the first decision level; the tree following the choice of the right branch is identical in structure, but of course the probabilities and payoffs may be different.

I.47 Now, suppose that the detection device is noisy. This means that when the device reports that we are at 2, there is a certain (known) probability that we actually are at 2, and also a (known) probability that we are at 3, with analogous statements applying in the other case. The courses of action available from node 2 and node 3 must be identical (although the probabilities of the chance events and the payoffs may be different); this follows from the observation that, if there were different courses of action available, then, for instance, if the decision-maker mistakenly thought he was at node 2, when he actually was at node 3, and if he tried to take a course of action not available at node 3, he would find this out and realize that he was at node 2.

I.48 So, if the decision-maker chooses "left" at the second level, he does not know whether he is choosing the distribution that follows the left choice at node 2, with outcomes 4 and 5, or the distribution that follows the left choice at node 3, with outcomes 6 and 7. But, he does know the probabilities that he is at node 2 and node 3 and, for each of these events, the probabilities of the various outcomes. Then, for instance, if the report on the outcome of the first choice is that he is at node 2, and if he chooses "left," there are four possible outcomes, and their probabilities are known. The possible outcomes are numbers 4,5,7, and 8, with the probability of obtaining 4 being the probability that he is at 2 multiplied by the probability of obtaining 4 given that he is at 2; similar calculations can be made for other possibilities. Then, the original tree can be replaced by one in which the possibilities at the second decision level represent reports of the noisy detection device and the probabilistic outcomes of the next choice are replaced by the more complicated ones obtained by combining the uncertainty of where you are in the original tree with that of what the outcome of your choice will be.

I.49 Finally, if the probabilities of the false reports are given in the form of the probability that the report will be node "i" given that you are at node "j," then Bayes' Rule must be used to obtain the probabilities that you actually are at node "i" and node "j."

INFORMATION THEORETIC SEARCH

I.50 An example of the utility of information theoretic methods with respect to sequential search can be found in a coin weighing problem.[37] Assume that there are 12 coins, of which 11 are true coins and one is false. The false coin may be either lighter or heavier than the others. It is desired to find the false coin and determine whether it is lighter or heavier. The available measurement device is a noiseless equal arm balance.

I.51 The uncertainty to be resolved is of measure log 24 since each of the 12 coins may be false and lighter or heavier. It is assumed that all the possibilities are equally likely (prior distribution is uniform). With three weighings, the maximum information that can be obtained is

$$3 \log 3 = \log 27 > \log 24$$

Therefore it may be possible to completely resolve the uncertainty.

I.52 If in the first weighing i coins are placed on each pan, then the

TABLE I. SECOND WEIGHING ALTERNATIVES, PROBABILITIES AND ENTROPIES

Index		Probability			Entropy,
i	j	P(B)[a]	P(R)[b]	P(L)[c]	H (dits)[d]
1	1	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	0.452
1	0	$\frac{3}{4}$	$\frac{1}{8}$	$\frac{1}{8}$	0.320
2	2	0	$\frac{1}{2}$	$\frac{1}{2}$	0.301
2	1	$\frac{1}{4}$	$\frac{3}{8}$	$\frac{3}{8}$	0.470
2	0	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	0.452
3	1	0	$\frac{1}{2}$	$\frac{1}{2}$	0.301
3	0	$\frac{1}{4}$	$\frac{3}{8}$	$\frac{3}{8}$	0.470
4	0	0	$\frac{1}{2}$	$\frac{1}{2}$	0.301

a/ P(B) = probability of balance.

b/ P(R) = probability right side heavier.

c/ P(L) = probability left side heavier.

d/ $H = \text{entropy} = -\Sigma p_i \log p_i$.

respective probabilities of balance, right side heavier or left side heavier are:

$$\text{Prob (balance)} = \frac{6 - i}{6} \qquad (I.25)$$

$$\text{Prob (right side heavier)} = \frac{i}{12} = \text{Prob (left side heavier)} \qquad (I.26)$$

I.53 To force these probabilities to be equal and thus obtain the maximum amount of information (log 3) on the first weighing, this value of i is chosen to be four. Now the pans may or may not balance. Consider first the case in which they do balance so that the false coin is known to be among those that were not used in the first weighing.

I.54 On the second weighing, place i of the four suspected coins in the right-hand pan and $j \leq i$ of these in the left-hand pan together with i-j of the true coins. Table I shows the various possibilities for values of i and j, the probabilities, and the overall entropy.

I.55 Since the entropy is greatest for (i,j) = (2,1) and (3,0), it is reasonable to consider the results of such choices. That is, the greatest resolution of uncertainty is obtained with either of these two measurements. Using the measurement (i,j) = (2,1), the symbolic representation of such a weighing will be (S_1, S_2) vs (S_3,T), where S_1, S_2, and S_3 are three of the suspected false coins and T is one of the known true coins. If balance is achieved then the false coin must be S_4. Its relative weight (heavier or lighter) can then be ascertained on the third weighing by a measurement against a known true coin, i.e., S_4 vs T. If the left side (S_1, S_2) is heavier, then S_4 is true, and either S_1 or S_2 is false and heavy or S_3 is false and light. The third measurement is then S_1 vs S_2. If balance is obtained, S_3 is false and light. If balance is not obtained, the heavier side has the false coin, which is heavy.

I.56 For the second measurement denoted by (i, j) = (3,0), a similar procedure can be followed. The measurement is represented by (T,T,T) vs (S_1, S_2, S_3). If balance is obtained, S_4 is measured against a known true coin. If, on the other hand, the right side is heavier, then S_1 or S_2 or S_3 is false and heavy. The third measurement could then be S_1 vs S_2. For balance it is concluded that S_3 is false and heavy. Otherwise, the heavier side is false and heavy.

I.57 Thus it is seen that for balance after the first weighing of four against four, two alternate schemes are available for the second and third weighings, each of which can completely resolve the uncertainty. These

TABLE II. SECOND WEIGHING ALTERNATIVES IF BALANCE NOT ACHIEVED ON THE FIRST WEIGHING [a/]

Index				Probability		
i_1	i_2	j_1	j_2	P(B)[b/]	P(R)[c/]	P(L)[d/]
2	1	2	1	$\frac{1}{4}$	$\frac{3}{8}$	$\frac{3}{8}$
2	1	2	0	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{3}{8}$
2	1	1	1	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{4}$
1	2	1	2	$\frac{1}{4}$	$\frac{3}{8}$	$\frac{3}{8}$
1	2	1	1	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{3}{8}$
1	2	0	2	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{4}$
3	1	1	0	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{4}$
2	2	1	1	$\frac{1}{4}$	$\frac{3}{8}$	$\frac{3}{8}$
2	2	1	0	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{3}{8}$
2	2	0	1	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{4}$
1	3	0	1	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{3}{9}$
3	2	1	0	$\frac{1}{4}$	$\frac{3}{8}$	$\frac{3}{8}$
2	3	0	1	$\frac{1}{4}$	$\frac{3}{8}$	$\frac{3}{8}$

a/ H = entropy = $-\Sigma p_i \log p_i$, and is equal to 0.47.

b/ P(B) = probability of balancing.

c/ P(R) = probability right side heavier.

d/ P(L) = probability left side heavier.

acceptable alternates for the second weighing were determined by examining the entropies associated with the alternatives and choosing those for which the entropy was a maximum. This afforded the maximum resolution of uncertainty on the second measurement. The third measurement was trivially determined.

I.58 Consider now the case in which balance is not achieved on the first weighing. Assume that such a weighing was (S_1, S_2, S_3, S_4) vs (S_5, S_6, S_7, S_8) and further that the right side was heavier. (A completely analogous argument holds if the left side was heavier.) The conclusion that can be drawn is that either one of the four coins (S_1, S_2, S_3, S_4) is false and light or one of the four coins (S_5, S_6, S_7, S_8) is false and heavy.

I.59 On the second weighing consider the placement of i_1 of S_5, S_6, S_7, S_8 and i_2 of S_1, S_2, S_3, S_4 in the right-hand pan with j_1 of S_5, S_6, S_7, S_8 and j_2 of S_1, S_2, S_3, S_4 in the left-hand pan together with $(i_1 + i_2) - (j_1 + j_2)$ of the coins that were determined to be true in the first weighing. Further, $i_1 + i_2 \geq j_1 + j_2$. The total number of admissable possibilities can be limited further by noting that the third weighing provides at most log 3 units of information. Hence no more than three possibilities may remain after the second weighing. Therefore the number of suspected coins not used in the second weighing cannot be greater than 3. It is concluded that

$$8 - (i_1 + i_2 + j_1 + j_2) \leq 3$$

or

$$i_1 + i_2 + j_1 + j_2 \geq 5 \qquad \text{(I.27)}$$

I.60 Since $i_1 + i_2 \leq j_1 + j_2$, then $i_1 + i_2 \geq 3$. If the right-hand pan in the second weighing is heavier, then either one of the i_1 coins on the right is false and heavy or one of the j_2 coins in the left is false and light. A similar argument can be made for the left hand pan being heavier. For these cases, the two further restrictions on the number of possibilities on the second weighing are $i_1 + j_2 \leq 3$ and $i_2 + j_1 \leq 3$. The remaining possibilities are now shown in Table II.

I.61 For each of the 13 alternatives the entropy is the same, i.e., 0.47 dits. This may be compared with the maximum of log 3 = 0.477 dits. Consider the case $(i_1, i_2, j_1, j_2) = (2,1,2,1)$ for which the following measurement is made: (S_7, S_8, S_2) vs (S_5, S_6, S_1). It is also recalled that the right side was assumed to be heavier on the first weighing, i.e., (S_1, S_2, S_3, S_4) vs (S_5, S_6, S_7, S_8).

I.62 If balance is obtained, then either S_3 or S_4 is false and light. On the third measurement, then, S_3 can be weighed against S_4 and the light side is false and a light coin. If, on the second measurement, the right side is heavier, then either S_5 or S_6 is false and heavy or S_2 is false and light. A third measurement of S_5 vs S_6 will resolve this uncertainty. If balance is obtained, then S_2 is false and light. Otherwise, the heavier side is false and heavy.

I.63 Thus it is shown that the uncertainty is resolvable for the case in which balance is not achieved on the first weighing. Furthermore, 13 alternate schemes for the second weighing are available, all of which yield identical resolutions of uncertainty. This indicates that many alternate strategies are available for solving this problem, which itself is only a special case of the set of sequential search algorithms discussed in the body of this report. In this special case, we have a multistage process in which the search space is discrete (24 cells), the initial distribution is assumed to be uniform, the search object is stationary, and the measurement device is noiseless. The partitioning is multicell by virtue of the freedom to group coins. The allocation of effort to each cell is, in general, variable, and the basic criterion is to carry out the search in no more than three steps. Satisfying this criterion is implemented by the essentially heuristic method of maximizing the information return per stage.

I.64 If some of these restrictions are removed, such as changing the number of false coins to some arbitrary value or introducing noise into the measurement device (the equal arm balance) so that it provides a true measurement only some percentage of the time, it may not be possible to completely resolve the uncertainty. However, as the number of measurements increases, it may be possible to successively decrease the variance of the distribution associated with the search space. This type of problem in its fullest generality has not as yet been solved.

APPENDIX II

SCORING SYSTEMS

II. 1 A request was made by ESD to examine the scoring systems defined in "Admissable Probability Measurement Procedures," by E. H. Shuford, Jr., A. Albert, and H. Edward Massengill,[38] with the objective of:

a. Determining which function of the class of derived scoring functions for the binary case is preferred and the reason for such preference.

b. Solving the m-nary case in its fullest generality.

This appendix presents the results of investigations to date in each of these areas.

Scoring Functions and the Confounding of the Motivations of Students

II.2 In a test, the student is confronted with a question to which he replies by wagering r on one possible answer and 1-r on the other (if there are only two choices). There is a scoring function, f(r), which then gives the student his grade on that question. The fundamental property required of scoring functions (by Shuford) is that they be "reproducing;" that is, that by choosing an r corresponding to p, his subjective probability of the truth of the first alternative, the student maximizes his expected score:

$$F(p,r) = p\, f(r) + (1-p)\, f(1-r) \leq p\, f(p) + (1-p)\, f(1-p). \qquad \text{(II. 1)}$$

A way to ensure that f(r) is reproducing is to take

$$f(r) = \int_0^r (1-t)\, g(t)\, dt$$
$$g(t) = g(1-t) \geq 0 \qquad \text{(II.2)}$$

where g(t) is an otherwise arbitrary function. It is convenient also to take f(1) = 1.

II.3 The use of the reproducing property depends on the student being motivated to maximize his expected return on the basis of a subjective probability. The student may be otherwise motivated. For example, suppose the student views the test as a game between the tester and himself, in which instead of subjective probabilities applying, the tester is selecting values of p to minimize the student's score. The student then selects a value for r to maximize his score in face of this kind of opposition. That is, he considers the expected value of

$$\max_{d(r)} \min_{p} F(p,r) \qquad \text{(II.3)}$$

where d(r) is a distribution of choices of r. On the other hand, the student might operate on a minimum regret basis. If he does, he considers the expected value of

$$\min_{d(r)} \max_{p} \{p\,[1 - f(r)] + (1 - p)\,[1 - f(1 - r)]\}$$
$$= \min_{d(r)} \max_{p} [1 - F(p,r)]\,. \qquad \text{(II.4)}$$

Since the roles (maximizing or minimizing) of p and r and the sign of F(p,r) are all reversed from those in Equation (II.3), the strategies for solution of this game will be the same as those for the solution of the previous game. Whatever f(r) is, these two motivations will be confounded. Therefore the minimum regret game can be put aside. But for some choices of f(r), the game and expectation motivations are also confounded.

II.4 Consider the quadratic scoring function

$$\left(\begin{array}{c} f(r) = 2r - r^2 \\ g(t) = 2 \end{array} \right) \qquad \text{(II.5)}$$

If the student treats the test as a game with his score as the payoff to himself, his optimum strategy is to take

$$r = \tfrac{1}{2} \tag{II.6}$$

and the payoff

$$M = p\,[f(\tfrac{1}{2}) + (1 - p)\,f(\tfrac{1}{2})] = f(\tfrac{1}{2}) = \tfrac{3}{4}\ . \tag{II.7}$$

Optimum tester strategy would be to take

$$p = \tfrac{1}{2}$$

since then

$$p\,f(r) + (1 - p)\,f(1 - r) = \tfrac{1}{2}\,(1 + 2r - 2r^2) \le \tfrac{3}{4} \tag{II.8}$$

although optimum tester strategy is not exactly germane. Thus with this scoring function, if a student gives an $r = \frac{1}{2}$ answer, it is not clear whether he is totally ignorant ($p = \frac{1}{2}$) and maximizing his expectation or is treating the test as a mathematical game—to mention only two of the possible motivations.

II.5 The fact that taking $r = \frac{1}{2}$ is an optimal strategy arises from the reproducing property itself. Since

$$\begin{aligned} F(p,r) &= p\,f(r) + (1 - p)\,f(1 - r), \\ F(p,\tfrac{1}{2}) &= f(\tfrac{1}{2}) \end{aligned} \tag{II.9}$$

independent of p, and this equality, with the inequality (II.1), assures that

$$r = \tfrac{1}{2},\ p = \tfrac{1}{2}$$

is a saddle point to the game with $F(p,r)$ as a payoff function. Therefore any scoring function that is reproducing will confound the expectation and game motivations.

Forceful Scoring Functions

II.6 The question has been raised of how to choose an $f(r)$ so that the student is forced, in some sense, to make a correct estimation of his own subjective probabilities. It is difficult to see how the idea of making a mistake in estimating subjective probabilities is at all tenable. However, it is easier to understand a student making a mistake in estimating his expected payoff, $F(p,r)$.

II.7 Suppose the student picks

$$r = p + \epsilon \,. \qquad \text{(II.10)}$$

That is, suppose that the student's error can be characterized by saying that he misses finding the maximum value of F by an error E(F). If F(p,r) is expanded about r = p, with

$$\frac{\partial F}{\partial r}(p,p) = 0 \,, \qquad \text{(II.11)}$$

then

$$F(p,r) \approx F(p,p) + \tfrac{1}{2}\,(r-p)^2 \frac{\partial^2 F}{\partial r^2}(p,p) \,. \qquad \text{(II.12)}$$

Thus

$$E(F) = F(p,r) - F(p,p) \approx \tfrac{1}{2}\,\Delta r^2 \frac{\partial^2 F}{\partial r^2} \qquad \text{(II.13)}$$

and Δr, the error in hitting the subjective probability, is made small for a given E(f) by choosing that function F(p,r) for which

$$\left.\frac{\partial^2 F}{\partial r^2}\right|_{r = p} = M(p) \qquad \text{(II.14)}$$

is as large as possible.

II.8 This second derivative is a simple function of the kernel g.

$$M(p) = pf''(p) + (1-p)\,f''(1-p) \qquad \text{(II.15)}$$

$$f''(p) = (1-p)\,g'(p) - g(p) \qquad \text{(II.16)}$$

$$f''(1-p) = -\,pg'(p) - g(p) \qquad \text{(II.17)}$$

since

$$g(1-p) = g(p), \qquad \text{(II.18)}$$

so that

$$-\,g'(1-p) = g'(p) \,.$$

Hence,

$$M(p) = p(1-p)g'(p) - pg(p) - p(1-p)g'(p) - (1-p)g(p) = -g(p) \; . \qquad \text{(II.19)}$$

Thus the "forcing" quality of a scoring function depends just on the kernel.

II.9 The kernels for the four scoring functions listed by Shuford are as follows:

a. $$g(u) = 2 \qquad \text{(II.20)}$$

b. $$g(u) = \frac{1}{(1-u)\ln 2} \qquad u < \tfrac{1}{2}$$
$$= \frac{1}{u \ln 2} \qquad \tfrac{1}{2} < u \qquad \text{(II.21)}$$

c. $$g(u) = [u^2 + (1-u)^2]^{-3/2} \qquad \text{(II.22)}$$

d. $$g(u) = \frac{2}{(1-u^2)\ln 3} \qquad u < \tfrac{1}{2}$$
$$= \frac{2}{[1-(1-u)^2]\ln 3} \qquad \tfrac{1}{2} < u \; . \qquad \text{(II.23)}$$

All these kernels give concave reproducing scoring systems. These functions are illustrated in Figure 15. Only half the range of the argument is shown since the functions are symmetric. The function a, of course, represents the highest value which can be maintained across the whole range of the argument. When in the testing procedure there is no interest in preferring precision at one level of the student's subjective probability p to that at another, this function seems ideal. When there is more interest in precision at some particular p, the other functions may look interesting. Functions b, c, and d, however, all put the extra precision in the region around $p = 0.5$.

II.10 A function which makes more precision available at extreme values of p is that labeled "e" in Figure 15. This function is a member of the family

$$g(u) = \{B(1-a, 2-a)[u(1-u)]^a\}^{-1} \qquad \text{(II.24)}$$

where $B(x,y)$ is the complete beta function. If $a = \tfrac{1}{2}$, this yields

e. $$g(u) = \frac{2}{\pi\sqrt{u(1-u)}} \qquad \text{(II.25)}$$

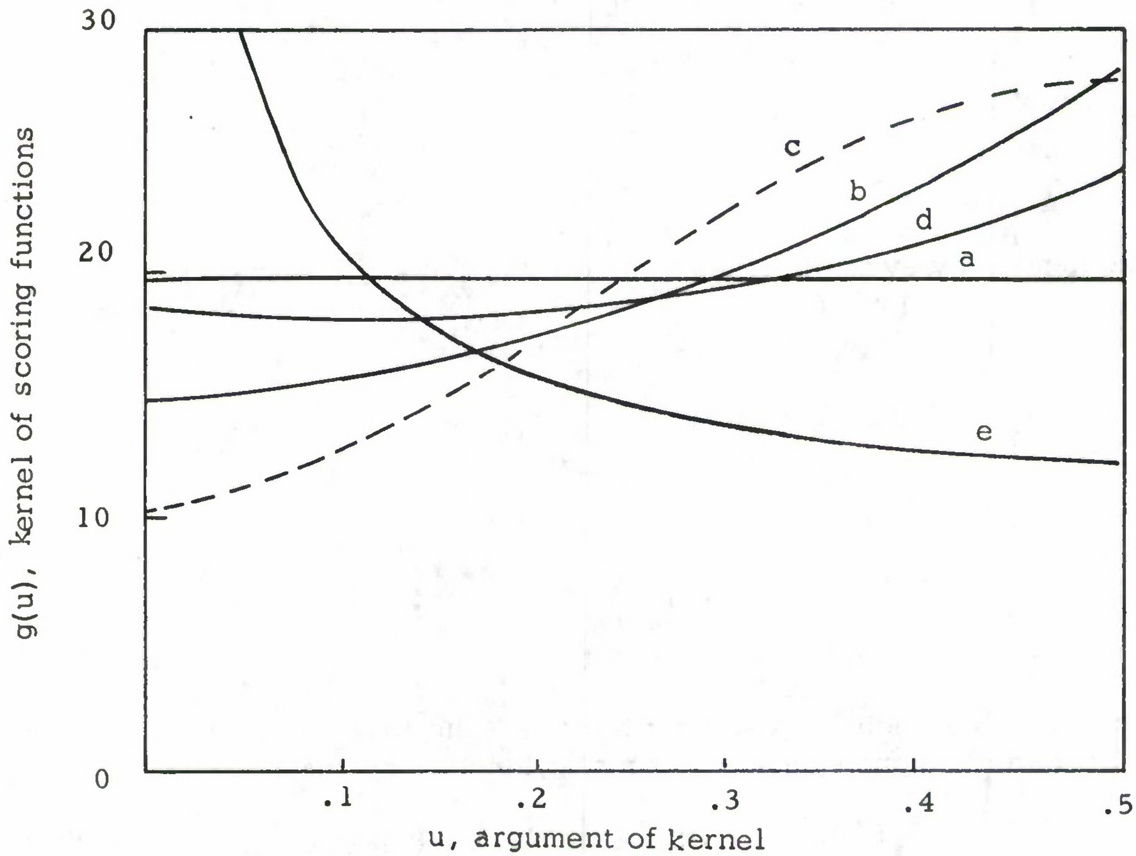


FIGURE 15. FORCING QUALITY OF SCORING FUNCTIONS

More General Scoring Functions

II.11 The problem of a general symmetric scoring function can be stated as finding a scoring function

$$f\ (x_1,\ x_2,\ .\ .\ .\ ,\ x_n) \qquad \text{(II.26)}$$

that is symmetric in the last n - 1 variables, where

$$\sum_{i=1}^{n} x_i = 1$$

and

$$x_i \geq 0\ .$$

The elementary symmetric functions are (x_2), (x_2x_3), $(x_2x_3x_4)$, . . . $(x_2x_3x_4\ .\ .\ .\ x_n)$ where

$$(x_2) \equiv \sum_{i=2}^{n} x_i$$

$$(x_2x_3) \equiv \sum_{i<j}^{n} x_i\ x_j \quad \text{etc.}$$

there are n - 1 such expressions; however, the first one $(x_2) = 1 - x_1$, so that it can be suppressed. The scoring function becomes

$$f\ [x_1,\ (x_2x_3),\ (x_2x_3x_4),\ .\ .\ .\ (x_2x_3 . . . x_n)]\ . \qquad \text{(II.27)}$$

II.12 In particular, for n = 3, the scoring function has the form f(x, yz). The expected payoff is

$$pf(x,\ yz) + qf(y,\ zx) + rf(z,\ xy)\ . \qquad \text{(II.28)}$$

The problem is to find f such that this expression is maximum for x = p, y = q, and z = r, subject to the constraint x + y + z = 1. Following Lagrange we add to the expected payoff λ (x + y + z) and set all partial derivatives to zero and x = p, etc. This yields the following equations.

$$pf_1(p, qr) + qrf_2(q, rp) + qrf_2(r, pq) + \lambda = 0$$

$$qf_1(q, rp) + rpf_2(r, pq) + rpf_2(p, qr) + \lambda = 0 \qquad \text{(II.29)}$$

$$rf_1(r, pq) + pqf_2(p, qr) + pqf_2(q, rp) + \lambda = 0$$

These imply that λ is a symmetric function of p, q, and r, so that each of the other parts of the equation must be symmetric. That is, the requirement is that

$$pf_1(p, qr) + qrf_2(q, rp) + qrf_2(r, pq) \qquad \text{(II.30)}$$

be a symmetric function of p, q, r.

II.13 A general "quadratic" scoring function has been found which satisfies the added constraint that $f(1) = 1$, $f(0) = 0$. It is

$$x_1 [2 - x_1 + 2(x_2x_3) + 6(x_2x_3x_4) + \ldots + (n-1)!(x_2x_3\ldots x_n)]. \qquad \text{(II.31)}$$

APPENDIX III

BIBLIOGRAPHY

Ackoff, R. L., "The Development of Operations Research as a Science," Operations Research, Vol. IV, June 1956.

Alderson, W., Search Theory in the Analysis of Consumer Shopping, Third National Meeting of ORSA, Boston, Massachusetts, 23-24 November 1953; invited paper, Abstract in Operations Research, Vol. II, February 1954, p. 78.

Arnold, R. D., Avoidance in One Dimension: A Continuous-Matrix Game, Operations Evaluation Group, OEG IRM-10 (AD 277 843), 11 January 1962.

Barlow, R., Probability Problems in Detecting Missile Salvos, Electronic Defense Laboratories, Sylvania Electric Products, Inc., Mountain View, California, EDL-M257, 12 February 1960.

Bellman, R., Adaptive Control Processes: A Guided Tour, Princeton, Princeton University Press, 1961.

———, Dynamic Programming, Princeton, Princeton University Press, 1957.

———, "Dynamic Programming, System Identification and Suboptimization," SIAM J. Cont., Vol. IV, February 1966, pp. 1-5.

———, On the Application of the Theory of Dynamic Programming to the Study of Control Processes, proceedings of the Symposium on Nonlinear Circuit Analysis, Polytechnic Institute of Brooklyn, Polytechnic Press, 1956, pp. 199-213.

———, and R. Kalaba, On Communication Processes Involving Learning and Random Duration, 1958 IRE National Convention Record, Part IV, pp. 16-20.

———, and R. Kalaba, Selected Papers on Mathematical Trends in Control Theory, New York, Dover Publications, Inc., 1964.

———, ———, and D. Middleton, Dynamic Programming, Sequential Estimation and Sequential Detection Processes, Nat. Acad. Sci. Proc., Vol. XLVII, March 1961, pp. 338-341.

Belzer, R. L., Solutions of a Special Reconnaissance Game, Rand Corporation, Santa Monica, California, RM-203, 10 August 1949.

Berlyne, D. E., "Attention, Curiosity and Decision," Predecisional Processes in Decision Making: Proceedings of a Symposium, Aerospace Medical Research Laboratories, TDR 64-77, December 1964.

Blackwell, D., and A. Girshick, Theory of Games and Statistical Decisions, New York, John Wiley & Sons, Inc., 1954.

Bram, J., A 2-Player N-Region Search Game, Office of Chief of Naval Operations, Operations Evaluation Group; OEG IRM-31 (AD 402 914), 17 January 1963.

———, and H. Weingarten, Estimation of Binomial Parameters from Search Data, Center for Naval Analysis, Research Contribution 3, 5 May 1964.

Bright, W. E., "A Particular Class of Generalized Search Problems" (Abstract), Operations Research, Vol. II, February 1954.

Coggins, P. B., Application of Information Theory to Search, memo for director of research, OEG(LO) 1975-52, 22 October 1952.

Danskin, J. M., "Reconnaissance," Operations Research, Vol. X, May-June 1962.

———, The Theory of Reconnaissance, Wright Air Development Center, Wright-Patterson AFB, WADC TN 59-409 (AD 233 998), November 1959.

———, The Theory of Reconnaissance, Aero Research Laboratories, AF Research Division, ARDC, Wright-Patterson AFB, ARL TR 60-337 (AD 233 998), November 1959.

———, "The Theory of Reconnaissance, Part II," Operations Research, Vol. X, May-June 1962, pp. 300-309.

———, On Koopman's Addition Theorem in Search Theory, Institute of Naval Studies, Franklin Institute, 1964, to be published in Operations Research.

———, The Theory of Max-Min, with Applications, Institute of Naval Studies, Franklin Institute, 1965.

———, A Max-Min Problem Associated with Strategic Deterrence Theory, Institute of Naval Studies, Franklin Institute, 1965.

DeGroot, M. H., "Uncertainty, Information and Sequential Experiments," Annals of Math. Stat., Vol. XXXIII, June 1962.

DeGuenin, J., "Les Fondements d'une Theore de la Recherche," Revue de Statistique Appliquée, Vol. VII, 1959.

Derman, C., "On Minimax Surveillance Schedules," Nav. Res. Q., Vol. VIII, December 1961.

Dieterly, D. L., Performance in Single-Stage vs. Multi-Stage Coding as a Function of Information Structure, M.S. thesis, The George Washington University, AD 617 058, June 1965.

Dobbie, J. M., "Surveillance of a Region by Detection and Tracking Operations," Operations Research, Vol. XII, May-June 1964, pp. 379-394.

Dresher, M., Games of Strategy - Theory and Application, Englewood Cliffs, New Jersey, Prentice-Hall, Inc., 1961, pp. 61-68, 175-178.

Dubines, L. E., "A Discrete Evasion Game, Annals of Mathematics Studies," in M. Dresher, A. W. Tucker, and P. Wolfe (eds), Contributions to the Theory of Games Vol. III, Princeton, Princeton University Press, 1957, pp. 231-255.

Edwards, W., "Dynamic Decision Theory and Probabilistic Information Processing," Human Factors, Vol. IV, 1962.

Eisner, H., "A Generalized Network Approach to the Planning and Scheduling of a Research Project," Operations Research, Vol. X, February 1962.

———, The Application of Information Theory to the Planning of Research, TIMS American International Meeting, 12-13 September 1963.

Engel, J. H., "Use of Clustering in Mineralogical and Other Surveys," Proceedings of the First International Conference on Operations Research, Oxford, 1957, The English Universities Press, pp. 176-192.

Enslow, P. H., Jr., A Bibliography of Search Theory and Reconnaissance Theory Literature, Stanford Electronics Laboratories, TR 1906-2, June 1965.

———, Observation Systems Employing Periodic Sampling, Stanford Electronics Laboratories, Report SEL-65-038 (TR 1906-1), June 1965.

Fano, R. M., TR 65, Research Laboratory for Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts, 1949.

Fine, N. J., Preliminary Study of the Interception Problem, Office of Chief of Naval Operations, Operations Evaluation Group, OEG Study 301 (ATI 28 837), 27 December 1946.

———, and F. W. Lamb, A Relation Between the Average Range of Direct Approach and Sweep Width, Office of the Chief of Naval Operations, Operations Evaluation Group, OEG Study 325, (ATI 28 829), 24 June 1947.

Firstman, S. I., and B. Gluss, Search Rules for Automatic Fault Detection, Rand Corporation, RM-2514, 15 January 1960.

Fishburn, P., Decision and Value Theory, New York, John Wiley & Sons, Inc., 1964.

Giammo, T. P., "On the Probability of Success in a Sudden Death Search With Intermittent Moves Confined to a Finite Area," SIAM Rev., Vol. V, January 1963, pp. 41-51.

Gluss, B., "Approximately Optimal One-Dimensional Search Policies in Which Search Costs Vary Through Time," Nav. Res. Log. Q., Vol. VIII, September 1961.

______, "The Minimax Path in a Search for a Circle in a Plane," Nav. Res. Log. Q., Vol. VIII, December 1961.

______, "An Alternative Solution to the 'Lost at Sea' Problem," Nav. Res. Log. Q., Vol. VIII, 1961, pp. 117-121.

Gross, O., A Search Problem Due to Bellman, Rand Corporation, RM-1603 (AD 87 962), 12 September 1955.

Gumacos, C., "Analysis of an Optimum Sync Search Procedure," IEEE Trans. Commun. Sys., Vol. XI, March 1963, pp. 89-99.

Hoehm A. J., and E. Saltz, "Mathematical Models for Determination of Efficient Troubleshooting Routes," IRE Trans., Reliability and Quality Control-13, July 1958.

Howard, R., Dynamic Programming and Markov Processes, Cambridge, The M.I.T. Press, 1964.

Hunt, J. A., The Optimization of Satellite Reconnaissance by the Application of Dynamic Programming Techniques, Mitre Corporation, ESD-TDR-63-168 (AD 402 810), April 1963.

Iaglom and Iaglom, Probabilité et Information, Dunod, Paris, 1959.

Isaacs, R., Differential Games, New York, John Wiley & Sons, Inc., 1965.

Isbell, J. R., "An Optimal Search Pattern," Nav. Res. Log. Q., Vol. IV, December 1957.

Johnson, S. M., "A Search Game," Advances in Game Theory, in M. Dresher, L. S. Shapley, and A. W. Tucker (eds), Princeton, Princeton University Press, 1964, pp. 39-48.

Karchere, A., and F. P. Hoeber, "Combat Problems, Weapons Systems, and the Theory of Allocation," Operations Research, Vol. I, November 1953, pp. 286-302.

Karush, W., and R. E. Dear, "An Optimal Procedure for an N-Stage Learning Process," SIAM J. Control, Vol. IV, No. 1, 1966.

Kimball, G. E., Simplified Theory of Sector Search, Office of Chief of Naval Operations, Operations Evaluation Group, OEG Study 282 (ATI 28 836), 28 June 1946.

———, The Ideal Search Theorem, Office of Chief of Naval Operations, Operations Evaluation Group, OEG Study 297 (ATI 28 831), 23 October 1946.

Kobzarev, U. B., and A. E. Basharinov, "On the Effectiveness of Search Algorithms Based on Samples of Controlled Duration (Sequential Detection)," Radiotekhnika, Elektronika VI:9, pp. 1411-1419; RM 2953-PR (AD 270 127), December 1961.

Kullback, S., Information Theory and Statistics, New York, John Wiley & Sons, Inc., 1961.

Langendorf, P. M., The Philosophy of the General Problem of Search and Detection, Rome Air Development Center, RADC-TN-59-130 (AD 213 583), May 1959.

Lawler, E. L., and D. W. Wood, "Branch and Bound Methods: A Survey," Operations Research, Vol. XIV, July-August 1966, pp. 699-719.

Luce, R. D., and H. Raiffa, Games and Decisions, New York, John Wiley & Sons, Inc., 1957.

MacDonald, D.K.C.,"Information Theory and its Application to Taxonomy," J. App. Phy., Vol. XXIII, May 1952.

Machol, R., and P. Gray, Recent Advances in Information and Decision Processes, New York, Macmillan Co., 1962.

MacQueen, J., and R. G. Miller, Jr., "Optimal Persistence Policies," Operations Research, Vol. VIII, May-June 1960.

McDonald, A.M.C., J. G. Fergusson,and R. W. Elliott, Theory of Search, The English Universities Press, 1962.

Mela, D. F., "Information Theory and Search Theory as Special Cases of Decision Theory," Operations Research, Vol. IX, November-December 1961.

Middleton, D., An Introduction to Statistical Communication Theory, New York, McGraw-Hill Book Co., 1960.

Miller, J. G., "Living Systems: Basic Concepts, Structure and Process, Cross-Level Hypotheses," Behav. Sci., July—October 1965.

Morse, P. M., "Mathematical Problems in Operations Research," Bul. Am. Math. Soc., Vol. LIV, July 1948.

Mryasawa, K., Information Amount as a Measure of Correcting One's Prior Probability, ONR, Astia (AD 617 326) July 1965.

Nemhauser, G. L., Introduction to Dynamic Programming, New York, John Wiley & Sons, Inc., 1960.

Neuts, M. F., "A Multistage Search Game," J. Soc. Indust. and App. Math., Vol. II, June 1963.

Nichols, R. E., and W. M. Whisler, A Technique for Analysis of Intermittent Search Operations Applicable to ASW, Boeing Airplane Co., D2-10868, 29 May 1961.

Norris, R. C., Studies in Search for a Conscious Evader, Massachusetts Institute of Technology, Lincoln Laboratory, TR 279, 1962.

Payne, B., "A Descriptive Theory of Information," Behav. Sci., Vol. XI, July 1966.

Pierce, J. R., Symbols, Signals and Noise, New York, Harper & Row, Publishers, Inc., 1963.

Potter, N. S. "The Optimization of Astronautical Vehicle Detection Systems Through the Application of Search Theory," Proc. IRE (Space Electronics Issue), Vol. XLVIII, April 1960.

———, "Programmed Search in Adaptive Systems," IRE Trans. on Mil. Elec., Vol. V, October 1961, pp. 363-369.

Raiffa, H., and R. Schlaifer, Applied Statistical Decision Theory, Cambridge, Harvard University Press, 1961.

Raisbeck, G., Information Theory, Cambridge, The M.I.T. Press, 1964.

Roberts, R. A., Theory of Signal Detectability: Composite Deferred Decision Theory, Cooley Electronics Laboratory, University of Michigan, Astia AD 614 480, February 1965.

Roby, T. B., "Belief States and the Uses of Evidence," Behav. Sci., Vol. X, July 1965.

Sandelius, M., "On an Optimal Search Procedure," Am. Math. M., Vol. LXVIII, February 1961.

Sherman, S., Total Reconnaissance with Total Countermeasures, Rand Corporation, RM-202, 5 August 1949.

Smallwood, R. D., "Optimum Policy Regions for Markov Processes with Discounting," Operations Research, Vol. XIV, July-August 1966, pp. 658-669.

Smith, W. E., "Various Optimizers for Single-Stage Production," Nav. Res. Log. Q., Vol. III, March - June 1956.

Takacs, Stochastic Process, Problems and Solutions, London, Methuen & Co., Ltd., 1960 .

Toda, M., and E. H. Shuford, Jr., "Utility, Induced Utility and Small Worlds," Behav. Sci., Vol. X, July 1965.

Wald, A., Sequential Analysis, New York, John Wiley & Sons, Inc., 1947.

———, Statistical Decision Functions, New York, John Wiley & Sons, Inc., 1950.

Bright, Jr., W. E., "A Particular Class of Generalized Search Problems," Operations Research, Vol. II, February 1954.

Wozencraft, J. M., and B. Reiffen, Sequential Decoding, Cambridge, The M.I.T. Press, 1960.

Zahl, S., "An Allocation Problem with Applications to Operations Research and Statistics," Operations Research, May—June 1963.

REFERENCES

1. M.J.E. Golay, "Notes on the Penny-Weighing Problem, Lossless Symbol Coding with Non-Primes," Professional Group on Information Theory, Vol. IV, September 1958.

2. S. Kullback, Notes on Information Theory, The George Washington University, Statistics 261-2, 1963.

3. E.T. Jaynes, Probability Theory in Science and Engineering, Socony Mobile Oil Co., Inc., October 1956.

4. M. Tribus, "The Use of the Maximum Entropy Estimate in the Estimation of Reliability," in R. E. Machol and P. Gray (eds), Recent Developments in Information and Decision Processes, New York, Macmillan Co., 1962.

5. A. Rapoport, "Human Control in a Decision Task," Behav. Sci., Vol. II, January 1966.

6. B. O. Koopman, "The Theory of Search: Part I, Kinematic Bases, and Part II, Target Detection," Operations Research, Vol. IV, 1956; "The Theory of Search: Part III, The Optimum Distribution of Searching Effort," Operations Research, Vol V, 1957.

7. J. Singh, Great Ideas in Information Theory, Language and Cybernetics, New York, Dover Publications, Inc., 1966.

8. J. D. Brule, R. S. Johnson and E. J. Kletsky, "Diagnosis of Equipment Failures," IRE Trans., RCC, April 1960.

9. R. A. Johnson, An Information Theory Approach to Diagnosis, paper presented at the Reliability and Quality Control Symposium, January 1960.

10. E. Kletsky, Diagnosis of Equipment Failures, Rome Air Development Center, TR-60-67B (AD-236189) April 1960.

11. R. Fano, Transmission of Information, New York, John Wiley & Sons, Inc., 1961.

12. D. A. Huffman, "A Method for the Construction of Minimum Redundancy Codes," Proc. IRE, Vol. XL, September 1952.

13. C. Shannon, "The Mathematical Theory of Communication," Bell Sys. Tech. J., July-August 1948.

14. N. M. Blachman, "Minimum Cost Encoding of Information," IRE Trans. Inform. Theory, Vol. III, March 1954.

15. R. S. Marcus, Discrete Noiseless Coding, M.S. thesis in E.E., Massachusetts Institute of Technology, Cambridge, Massachusetts, 1957.

16. R. M. Karp, "Minimum Redundancy Coding for the Discrete Noiseless Channel," IRE Trans. Inform. Theory, January 1961.

17. R. E. Machol, The Theory of Search, paper presented at Operations Research Society of America meeting, Honolulu, 14-18 September 1964.

18. J. M. Danskin, "The Theory of Reconnaissance, Part I," Operations Research, Vol. X, May-June 1962, pp. 285-299.

19. R. S. Novosad, Search Problems and Information Theory, Martin Denver, Working Paper No. 64, 5 June 1961.

20. N.M. Blachman, "Prolegomena to Optimum Discrete Search Procedures," Nav.Res.Log.Q., Vol. VI, December 1959.

21. —— and F. Proschan, "Optimum Search for Objects Having Unknown Arrival Times," Operations Research, Vol. VII, September-October 1959.

22. B. O. Koopman, Search and Screening, Office of Chief of Naval Operations, Operations Evaluation Group, OEG Report 56 (ATI 64 627), 1946.

23. J. M. Dobbie, "Search Theory: A Sequential Approach," Nav. Res. Log. Q., Vol. X, December 1963.

24. J. DeGuenin, "Optimum Distribution of Effort: An Extension of the Koopman Basic Theory," Operations Research, January-February 1961.

25. E. N. Gilbert, "Optimal Search Strategies," J. Soc. Ind. Appl. Math., Vol. VII, pp. 413-424, 1959.

26. S. M. Pollock, Optimal Sequential Strategies for Two Region Search When Effort is Quantized, Interim TR 14, MIT Contract DA-19-2-ORD-2684, 1960.

27. ——, Sequential Search and Detection, Operations Research Center, Massachusetts Institute of Technology, TR 5, May 1964.

28. W. L. Black, Sequential Search, M.S. thesis in E.E., Massachusetts Institute of Technology, Cambridge, Massachusetts, June 1962.

29. K. R. Scott, Optimal Multi-Region Discrete Search Under Linear Search Costs, M.S. thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts, 18 January 1963.

30. B. Gluss, "An Optimum Policy for Detecting a Fault in a Complex System," Operations Research, Vol. VII, pp. 468-477, July-August 1959.

31. M. C. Chew, Jr., An Optimal Searching Plan for a Problem in Maintenance Analysis, Carnegie Institute of Technology, Pittsburgh, Pennsylvania, ONR RM 130, October 1964.

32. D. Matula, "A Periodic Optimal Search," Am. Math. M., Vol. LXXI, January 1964.

33. A. Charnes and W. W. Cooper, "The Theory of Search: Optimum Distribution of Search Effort," Management Sci., Vol. V, October 1958.

34. E. C. Posner, "Optimal Search Procedures," IEEE Trans. Inform. Theory, Vol. IX, July 1963.

35. S.W. Golomb and L.D. Baumert, "Backtrack Programming," J.Assoc. Comput. Mach., Vol. XII, October 1965, pp. 516-524.

36. G. Hadley, Nonlinear and Dynamic Programming, Reading, Pennsylvania, Addison-Wesley Publishing Co., Inc. 1964.

37. G. Kimball, Detection and Tracking as a Markov Process, NATO ASW Meeting, Sienna, Italy, 1963.

38. E. H. Shuford, Jr., A. Albert, and H. E. Massengill, Admissable Probability Measurement Procedures, in manuscript (see Appendix II).

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Operations Research Incorporated 1400 Spring Street Silver Spring, Maryland 20910	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

AN INVESTIGATION OF SEQUENTIAL SEARCH ALGORITHMS

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Final Report (15 April 1966-31 December 1966)

5. AUTHOR(S) (Last name, first name, initial)

Beiman, H.
Eisner, H.

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
1967	82	38

8a. CONTRACT OR GRANT NO. AF 19(628)-5989 b. PROJECT NO. c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) ORI TM 103-67 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) ESD-TR-67-222

10. AVAILABILITY/LIMITATION NOTICES

Distribution of this document is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY Electronic Systems Division Air Force Systems Command, L.G. Hanscom Field, Bedford, Massachusetts

13. ABSTRACT

Characterizations of sequential search processes and algorithms are developed. Representative sequential search algorithms are reviewed and interpreted within the framework of these characterizations in several fields including equipment diagnosis, signal encoding, radar systems, coin weighing, and human decision processes. Conclusions and recommendations for application are presented. Appendixes include selected mathematical techniques in programming and information theoretic search procedures, analyses of scoring systems, and a bibliography.

DD FORM 1 JAN 64 1473

UNCLASSIFIED

Security Classification

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Sequential Search Algorithms						
Decision Processes						
Linear Programming						
Dynamic Programming						
Backtrack Programming						
Information Theory						
Probability Theory						

UNCLASSIFIED
Security Classification