ESD-TR-67-90

MTR-260

A SYSTEMATIC PROCEDURE FOR ASSESSING THE WORTH OF COMPLEX ALTERNATIVES

NOVEMBER 1967

J.R. Miller, III

Prepared for

EDP EQUIPMENT OFFICE

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

Project 8510
Prepared by
THE MITRE CORPORATION
Bedford, Massachusetts
Contract AF19(628)-5165

ESD-TR-67-90

MTR-260

A SYSTEMATIC PROCEDURE FOR ASSESSING THE WORTH OF COMPLEX ALTERNATIVES

NOVEMBER 1967

J.R. Miller, III

Prepared for

EDP EQUIPMENT OFFICE

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

Project 8510
Prepared by

THE MITRE CORPORATION
Bedford, Massachusetts

Contract AF19(628)-5165

FOREWORD

This report was prepared by The MITRE Corporation, Bedford, Massachusetts, for the EDP Equipment Office, Electronic Systems Division of the Air Force Systems Command, L. G. Hanscom Field, Bedford, Massachusetts, under Contract AF19(628)-5165.

REVIEW AND APPROVAL

This technical report has been reviewed and is approved.

for S. P. STEFFES
Colonel, USAF
Chief, EDP Equipment Office

ABSTRACT

This paper addresses itself to the problem of assessing worth. It is assumed that a decision context has been specified and that a fixed set of discrete alternatives has been produced. It then remains to assess the worth of each alternative, to estimate the resource drains required by each, and to combine these considerations, along with considerations of risk/uncertainty, so as to arrive at a final decision. The bulk of this paper is directed toward worth assessment.

To aid in the assessment process, a detailed procedure has been devised. The purpose of this procedure is set forth, and step-by-step instructions for its actual implementation are presented. A live instance of its complete application is also provided for illustrative purposes.

An experiment was performed whose purpose was to validate the procedure (i.e., to demonstrate that it could be carried out successfully by professional decision makers). Results drawn from the experiment are interpreted, and conclusions are drawn, along with additional implications for decision making. The procedure and the experiment are reviewed critically, and suggestions are made for further research.

Major conclusions are that the procedure can be carried out successfully--at least by professional decision makers in a laboratory setting--and that all phases of it exert an important impact upon the decision making process. Critical to its overall impact, and particularly to its success, are its quantitative aspects. Requirements to quantify worth notions induce decision makers to formulate and validate their preferences. These two consequences are received very favorably by the decision makers themselves. An additional consequence is to provide an explicit and logically consistent assessment structure which, if considered acceptable, may be used to guide a final decision.

The procedure is both general and flexible with respect to type of decision, type of alternative, and type of worth criterion. However, its application may be restricted by type of decision-making personnel and certain contextual factors.

PREFACE

This paper constitutes the final product of a research project which began in 1963. The purpose of this project was to develop a systematic procedure to aid in the assessment of worth. Most of the research reported herein was carried out by the writer and several of his colleagues at The MITRE Corporation, with generous advice and counsel provided by various members of the Electronic Systems Division, Air Force Systems Command. The experiment reported in Section 9 of this paper was carried out by the writer as part of his doctoral dissertation research at Massachusetts Institute of Technology. Toward this end the writer's doctoral thesis committee provided invaluable assistance.

Discussions presented herein do not attempt either to describe or to justify any currently practiced assessment techniques. Rather, the purpose of this paper is to address the many real problems--both theoretical and practical--involved in the task of worth assessment. The assessment procedure finally developed represents the end state of a lengthy evolutionary process wherein many issues were raised and debated, many alternative solutions were proposed, and many specific techniques were tried out. However, this is not the last word. On the basis of reported experimental findings, additional research along these same lines would likely prove quite fruitful.

TABLE OF CONTENTS

TABLE OF CONTENTS (CONTINUED)

TABLE OF CONTENTS (CONTINUED)

TABLE OF CONTENTS (CONTINUED)

TABLE OF CONTENTS (CONTINUED)

LIST OF EXHIBITS

SECTION 1

INTRODUCTION

Assessing the worth of "complex" alternatives in an "important" decision situation is generally regarded as difficult. However, this task constitutes but one phase in the still more difficult process of producing such alternatives and making a final decision among them. Some of the factors which make an alternative "complex", which make a decision "important", and which thereby render the overall decision process difficult are stated and illustrated in the paragraphs that follow.

Consider the case of a large government organization which has decided to automate a substantial portion of its day-to-day activities by acquiring and operating an electronic computer. Producing, evaluating, and finally selecting an alternative to implement this decision would very likely be difficult for several reasons.

First, it must be determined which activity or activities are to be automated and to what extent. This requires the decision maker to describe in some detail the job to be performed by whichever alternative is finally selected. In the case of a computer, such a job description would ordinarily be both lengthy and complex.

Second, having formulated an adequate job description, the next task is to explicate the overall purpose in automating whatever activities are described therein. But computers are multi-purpose rather than single-purpose instruments. They satisfy many different objectives simultaneously. For this reason, their overall worth to an organization cannot be reckoned on the basis of a single criterion. This suggests the need for responsible decision-making personnel to undertake the following tasks:

1. The several objectives which are to be satisfied by acquiring and utilizing a computer should be listed, and the list should be fairly complete.

2. From each listed objective should be derived a set of specific worth criteria in terms of which the physical performance of a computer may be assessed.
3. Some means should then be found to organize and integrate these multiple criteria into a consistent and meaningful assessment structure.

Third, both the acquisition and the operation of computing equipment have many important ramifications for an organization. There is no simple or unique consequence on the basis of which an entire decision can be made. There are many performance consequences which must first be ascertained with reasonable accuracy and then assessed meaningfully before a final decision can be reached. This suggests the need for a clear, systematic, and replicable procedure to insure that no important performance consequences are overlooked.

Fourth, since both multiple criteria and multiple performance consequences are present, some means should be found to establish worth connections between the two. But this is not as easy as it may seem. A single performance consequence may be related simultaneously to several worth criteria (e.g., excess core storage capacity over and above minimum requirements might be considered relevant simultaneously to computing capability, to expandability potential, and to providing a hedge against the disastrous consequences of underestimating job requirements). Conversely, many performance consequences may be related simultaneously to a single worth criterion (e.g., excess core storage capacity, estimated time required to complete a stated job, and various multi-processing features might all be considered relevant measures of total computing capability.)

Fifth, complex patterns of interaction may exist among various physical performance consequences due to the fact that computers are designed not as crude conglomerations of unrelated components, but rather as highly organized and integrated whole systems. This makes it difficult to understand and, therefore, to predict accurately an entire set of specific consequences.

Sixth, even if all performance consequences were known for certain, there may still exist complex patterns of interaction among the various worth criteria imposed by human beings upon this known performance. The structure of human worth notions is itself infested with intricate patterns of interdependence. Even worse, human beings often find it exceedingly difficult to distinguish in their own minds between interaction among performance consequences (a physical phenomenon characteristic of computers) and interdependence among imposed worth criteria (a psychological phenomenon characteristic of human beings). This renders still more difficult any attempt to understand, to assess, and to select computing equipment.

Finally, there is the question of resources expended to acquire and operate a computer. It is not always easy to predict with accuracy the amounts of manpower, materiel, and monetary resources which will of necessity be expended in order to implement any proposed automation alternative. Even if the resource implications of each proposed alternative were predictable, it must still be decided how much of each type of resource <u>should</u> be expended on the particular job under consideration. In other words, the relative importance of the job should be ascertained in advance, and this should be translated into specific amounts of each type of resource which might appropriately be expended to perform that job.

Historically, decision makers have attempted to cope with the above kinds of problems largely on the basis of subjective judgment and intuition. Subjective estimates have been used quite frequently to predict probable resource and performance consequences. Personal judgments have also been used both to assess the worth of different amounts of predicted performance and to effect trade-offs among various worth criteria. The twin problems of physical interaction among performance consequences and conceptual interdependence among worth criteria have been handled similarly--that is, on an intuitive basis. Now if the decision problem under consideration were simple, and if the consequences of making a poor decision were relatively inconsequential this might be the best way to proceed. The extra gains realizable from formalizing and systematizing the decision process would probably not justify the extra time, cost and effort required. However, when the problem becomes complex (e.g., in the senses just described), and/or when the consequences become important, strict reliance on subjective judgment becomes a dangerous gamble indeed. It seems unreasonable to permit such decisions to be made in the absence of factual evidence, logical discipline, and at least an opportunity to attain consensual validation. This suggests the need for a systematic procedure.

The central purpose of this paper is to develop an explicit, logically consistent and replicable procedure to aid in the assessment of worth.[1] In addition, the results of an experiment designed to measure the impact of this procedure upon professional decision makers will be reported. To place the procedure in perspective, it will also be compared with various alternative procedures discussed in the literature and practiced by practical decision makers.

---

[1] It should be made clear that explicitness, logical consistency, and replicability do not preclude the use of subjective judgment. Quite to the contrary, it is the writer's view that subjective judgment must be used both in assigning measures of worth to various performance consequences and in trading off worth among various criteria. Subsequent sections of this paper will be specifically devoted to supporting this point of view. Rather, what is being stipulated here is that, when used, subjective judgment should be made explicit, should be thoroughly scrutinized for logical consistency, and should be elicited by a uniformly applicable and replicable procedure. The writer can think of no better way to insure that personal judgments will be free of false assumptions than by stating these assumptions explicitly. Nor can he think of a better way to insure valid reasoning from assumptions to conclusions than by exposing the reasoning process to critical scrutiny. Nor can he think of a better way to elicit a cross-section of opinion and to establish a consensus of preferences than by means of a uniformly applicable and replicable procedure. Most important of all, the writer can think of no better ways than these to obtain feedback on the assessment process and, therefore, to provide a constant impetus to its improvement.

SECTION 2

STATEMENT OF THE PROBLEM

The overall problem, of which this paper will treat only a part, is seven-fold. Assuming that an important decision is to be made among complex alternatives, then the problem is:

1. to describe the job to be performed by whichever alternative is finally selected (i.e., to list the various activities which are to be carried out);

2. to formulate the overall purpose in making the decision (i.e., to abstract a specific set of job objectives from the job description);

3. to produce one or more feasible alternatives (i.e., to design and/or to solicit proposals for at least one alternative whose performance, if selected, would be viewed as satisfying minimum adequacy requirements);

4. to predict the worthwhile performance consequences associated with each alternative (i.e., to predict the types and amounts of worthwhile performance which would be realized from acquiring and utilizing each of the alternatives produced);

5. to assess the worth of these predicted performance consequences (i.e., to assess the extent to which the above-predicted performance would succeed in accomplishing stated job objectives);

6. to predict the resource consequences associated with each alternative (i.e., to predict the types and amounts of limited resources which would necessarily be expended to acquire and utilize each of the alternatives produced);

7. to <u>reach</u> a final decision (i.e., to match worth of performance received against limited resources expended on each of the alternatives produced so as to determine whether any of them should be selected and, if so, which one).

Actually, this paper will address itself almost exclusively to the problem of worth assessment (see 5. above). It will henceforward be assumed for discussion purposes that a job has been described, that an overall decision purpose has been formulated, that one or more feasible alternatives have been produced, and that the physical performance of each alternative has been adequately predicted. Nevertheless, despite these simplifications the residue of the problem is still very difficult. To illustrate the remaining difficulties, let us return to the case of selecting a computer and consider a highly simplified hypothetical example.

Suppose that the job to be automated is a real-time application of inventory control. Suppose, further, that the only performance consequences considered important are the time required to respond to an inquiry and the maximum useable size of the file of items which can be maintained. Suppose, also, that the only resource considered important is the initial dollar investment required to procure hardware. Finally, suppose that three alternative systems have been proposed by competing computer manufacturers and that validated estimates of their performance and cost consequences are as shown in Table 1.

TABLE 1

	Alternative I	Alternative II	Alternative III
Response Time	10 Minutes	20 Minutes	38 Minutes
File Capacity	100,000 Items	175,000 Items	150,000 Items
Investment Cost	$110,000	$125,000	$90,000

The decision maker would now be faced with the task of making trade-offs between different levels of response time and file capacity to arrive at some notion of the overall worth of each alternative, and then he would have to match overall worth against cost on all three. Comparing alternatives I and II, he would have to decide whether the degradation in response time from 10 to 20 minutes were at least offset by the increase in maximum file capacity from 100,000 to 175,000 items. If no, then it certainly would not be worthwhile spending the extra $15,000 to purchase alternative II. If yes, if the increased file capacity more than compensated for the inferior response time, then he would have to decide whether the net gain in worth derivable from selecting alternative II over I warranted spending the additional $15,000. But what about alternative III? It is much cheaper than the other two, its file capacity falls between the other two, and its response time is substantially higher (and, therefore, less desirable) than both of its competitors'. Comparisons similar to those described above would have to be made first between alternatives I and III, and then between alternatives II and III.

However, the above types of comparisons, even if carried out successfully, would not be sufficient to dispose of the problem completely. There still remain the twin dangers of "under-kill" and "over-kill". One or more of the three alternatives would provide "under-kill" if there existed either a maximum acceptable response time or a minimum required file capacity, and if estimated performance fell beyond either of these limits. This is an obvious kind of danger which can usually be detected with little difficulty - particularly if such mandatory performance requirements have been stipulated in advance on the basis of careful engineering and design considerations.

In contrast, the other kind of danger - the danger of "over-kill" - is far more subtle and much more difficult to detect. The reason is that "over-kill" is an economic rather than an engineering concept. Assessment of "over-kill" requires simultaneous consideration of both performance and resource consequences. The essence of "over-kill" lies not in the mere fact that more performance may be proposed than is necessary, but rather in the fact that whatever additional performance (over and above minimum requirements) is proposed may not warrant expending whatever additional resources are required to receive that additional performance. On economic grounds, it may be preferable to accept lesser performance - or even to accept zero performance (i.e., abandon the project) - and to expend the saved resources on some other project entirely. Returning to our example, it may be that alternative III, even with its relatively long (38 minute) response time, is more than adequate to meet the job requirements. Under such circumstances, it might be economically unwise to spend any more than $90,000 on this job. Alternatively, it may happen that even $90,000 is too much to spend. It may be that a far less efficient manual system costing only $25,000 would also be adequate, and that even the cheapest of the proposed automation alternatives (costing $90,000) would not justify spending the extra $65,000. That same money might better be spent on some other automation project or, perhaps, on some other project completely unrelated to automation. Before a final decision can be made, all of these issues should be considered, and the decision maker should be prepared to reject any (or even all) of the proposed alternatives if either "under-kill" or "over-kill" becomes apparent.

The difficulty of making the above types of trade-off decisions - first between different kinds of performance to arrive at an assessment of overall worth, and then between overall worths and their associated costs - is probably quite evident to the reader. And this was a highly simplified example. As the number of worth criteria and related performance consequences increases, the problem quickly reaches unmanageable proportions. If multiple resources are also considered (e.g., manpower and materiel as well as monetary resources), and if complex patterns of both physical interaction and worth interdependence emerge, then effective solution of the problem by sheer intuition becomes just about impossible. This suggests the need for a more formal approach.

The remainder of this paper will be oriented specifically toward the development of a more formal approach. Sections 3 through 8 will set forth a systematic procedure to aid in the assessment of worth. Section 9 will report the results of an experiment designed to validate the assessment procedure. Section 10 will integrate these results with the preceding discussion, and the entire procedure will be compared with alternative procedures suggested in the literature and/or practiced by practical decision makers.

SECTION 3

THE CONCEPT OF WORTH

For purposes of this paper, the worth of any object, activity, or situation is, roughly speaking, the extent to which such is perceived by a decision maker or group of decision makers as satisfying clearly articulated objectives. Thus, the worth of an alternative in a stated job context would be defined in terms of how well that alternative satisfied whatever objectives have been articulated regarding the job to be accomplished.

The above notion of worth is intentionally stated in very general terms. A detailed definition will be presented later. Specifically, step-by-step procedures for assessing worth will be developed in Section 7, and one purpose in setting forth these procedures is to provide an operational definition of the concept itself. For now, however, it will be useful to outline the intended meaning and scope of the worth concept--both to orient future discussion and to preclude the imputation of unintended meanings to the subject matter of this paper.

3.1 The Intended Meaning of Worth

From the above discussion, it is apparent that worth notions constitute an internal property of human decision makers--not an external property of the physical objects, activities, and situations whose worth is being assessed. Worth is here conceived as inherent within the perceptual apparatus of the decision maker himself. This represents an important departure from traditional approaches to the same problem which focus primarily upon the physical entities being assessed. The detailed procedures to be developed in Section 7 will clarify this distinction operationally.

Since worth is defined with respect to clearly articulated objectives, it is necessary that such objectives exist. Operationally, this usually means that a deliberate effort must be made to formulate and articulate clear objectives before worth may be assessed. It also means that worth notions will be multidimensional whenever multiple objectives and/or multiple performance measures are considered relevant (e.g., in the case of "complex" alternatives).

Finally, worth refers to the extent or degree to which some object, activity, or situation satisfies stated objectives. This suggests the need for establishing a definite scale in terms of which various degrees of goal satisfaction (and, therefore, imputed worth) may be expressed. Section 4 will address itself to establishing such a scale.

3.2 Implications for the Task of Assessing Worth

Having discussed briefly the meaning of worth, as it will be used in this paper, we shall now investigate some of its implications for the practical problem of assessing alternatives.

First, the act of formulating and articulating a clear set of objectives in terms of which worth may be assessed is not always easy to accomplish. Decision makers may be either unable or unwilling to formulate and display a complete list of objectives because of:

1. incomplete awareness of the problem at hand;
2. incomplete knowledge of the intricacies of the problem;
3. inability (due to time, money, and/or manpower constraints) to devote sufficient "thinking" effort to formulating a complete and explicit list of objectives.

Alternatively, they may be unwilling to formulate and particularly to display a complete list of objectives because of:

1. fear that some of the "real" objectives will be disapproved if layed bare to public scrutiny;
2. fear that some of the "real" objectives, even if tacitly approved, may not be easily defended in the political arena;

3. realization that some objectives, even if approved and defensible, may not receive complete consensual validation from all interested parties - particularly those who would suffer adverse consequences should the "real" objectives be satisfied.

These latter sources of unwillingness may attain particular motivational importance if decision makers are themselves imbedded in an organizational environment rife with threat, conflict, or a strong tradition of defensive conservatism.

Second, there is the issue of confirming worth judgments. Unlike allegations of fact or scientific predictions, worth judgments cannot be confirmed by empirical test. They are <u>in principle</u> untestable by ordinary scientific means. This is because worth judgments are stated in such a way as to be neither factually true nor factually false. They merely exist in the minds of human beings to be accepted or rejected either in whole or in part by other human beings (or, perhaps, by the same human being at a different point in time). In short, the acceptability of worth judgments is here conceived as a matter of informed opinion.

A third implication follows from the second. This involves the identity of decision makers. Different decision makers may very well have different objectives regarding the same situation, which renders the outcome of an assessment highly dependent upon who undertakes to perform that assessment. Stated a bit more simply, the outcome of an assessment depends critically upon <u>whose</u> values are adopted. One way out of this situation is to strive for consensus among potential decision makers, but this is not always possible (and perhaps even undesirable) to achieve. In any case, the worth concept is not here <u>defined</u> as requiring consensus.

A fourth implication involves the stability of worth judgments over time. Not only may there exist lack of consensus among separate decision makers at a given point in time, but there may also exist lack of agreement among separate worth judgements made by the same decision maker at different points in time. As additional experience is gained, one would expect (or at least hope) that a given decision maker would alter his worth judgments to account for whatever new insights this additional experience has brought about. Temporal instability is thereby created, but, possibly, in an entirely appropriate manner. In any case, the worth concept is not here _defined_ as requiring temporal stability either.

SECTION 4

CONSTRUCTING A MEASURE OF WORTH

Having committed ourselves to creating a formal assessment scheme, we must now tackle the problem of defining a uniform and convenient measure of worth. This suggest (although it does not require) reducing the problem to numbers. Why? Because numbers are familiar, widely used as tools of measurement, and easy to manipulate. However, lest there be any confusion on this issue, let it be understood at the outset that the measure of worth to be created, and particularly its numerical scale characteristics, constitute an ad hoc invention specifically designed for the assessment procedure to be developed in Section 7. No claim is being made that this worth measure or its scale characteristics derive deductively from any set of (normative) axioms. With this proviso firmly in mind, let us endeaver to construct a numerical measure of worth.

4.1 <u>The Basic Purpose in Constructing a Measure of Worth</u>

Perhaps the best way to initiate detailed discussion is with a formal statement of purpose. When a numerical measure of worth is used as a vehicle of assessment, the underlying rationale is that such numbers will be assigned to various objects and activities in such a manner as will reflect their perceived worth. That is, worth numbers will be assigned such that numerical relationships between assigned worth numbers will faithfully reflect perceived worth relationships between the objects and activities to which these numbers have been assigned.

In order to implement the above purpose, it is first necessary to specify the kinds of worth relationships which are to be reflected by means of numerical symbols. It is also necessary to specify the numerical conventions which establish a correspondence between numerical relationships and the perceived worth relationships which are being depicted thereby. The first of these tasks will be undertaken in

Section 4.2. The second will be undertaken in Section 4.3.

4.2 Worth Characteristics to be Reflected

The most fundamental characteristics of the worth concept to be reflected in our choice of a numerical measure are the three psychological states of preference, aversion, and genuine indifference.

A decision maker is said to possess a positive preference for some object or activity if and only if that object or activity elicits a positive affective response from him (e.g., joy, pleasure, interest, excitement, gratification, etc.). Thus, most people possess a positive preference for automobiles because they elicit all of the above positive responses. In addition, most people are willing to part with money (a scarce resource) in order to receive these benefits from an automobile.

A decision maker is said to possess a distinct aversion (negative preference) for some object or activity if and only if that object or activity elicits a negative affective response from him (e.g., distress, anxiety, shame, guilt, disgust, etc.). Most people possess a distinct aversion to death. The very thought of it arouses a great deal of distress and anxiety, and many people are willing to part with substantial amounts of money (i.e., purchase life insurance) in order to ameliorate its unwanted consequences.

A decision maker is said to feel genuinely indifferent toward some object or activity if and only if he possess neither a preference for nor an aversion toward that object or activity.

Returning to the concept of worth, this is usually thought of as related to positive preferences only. That is, when an object or activity is said to possess some worth, this usually means that somebody possesses a positive preference for it and/or its consequences. The concept of "negative" worth (referring to objects or activities toward which people feel aversive) is less well defined.

In light of these observations, certain numbers on the worth scale (to be created in Section 4.3) will be reserved to indicate positive preferences, and a single number not included in the above range will be reserved to indicate a state of genuine indifference. Negative preferences or aversions will be represented by negative analogs of the positive worth numbers.[2]

Another aspect of the worth concept to be reflected in our choice of a numerical measure involves its boundedness. Is it possible for something to be completely worthwhile? Can a decision maker be completely satisfied (or dissatisfied) with some object or activity? Although seemingly simple at first glance, this is a very subtle and important question. Let us investigate the issue more closely.

If asked to assess the worth of some object without specifying how or for what purpose that object is to be used, it seems difficult to conceive of any natural, logical outer bounds to the answer. Like the brightness of a color or the loudness of a sound, there exist no apparent natural limits. On the other hand, once a definite job has been specified and once a definite set of objectives has been defined, then the question appears in a somewhat different light. When asked to assess the worth of some object for performing some stated job in accordance with well defined objectives, it seems reasonable to talk in terms of the extent to which that object satisfies the stated objectives. Furthermore, since definite objectives have been stated, it seems reasonable to talk about the possibility, at least, of having those objectives completely satisfied. Thus, under these revised circumstances, there appears to be a natural outer bound to the assessment of worth. This will be reflected by placing numerical bounds on the

---

[2] More will be said in subsequent sections about negative preferences or aversions and their representation on the worth scale.

worth scale.

Still another aspect of the worth concept to be represented numerically is its continuity or divisibility. It would seem desirable to permit the expression of preferences and preference differences to range from infinitesimal magnitudes to large magnitudes. Although decision makers may not always wish to avail themselves of this flexibility, a continuous or everywhere-dense worth scale will be defined to accommodate such notions whenever they are felt.

Finally, there is the question of preference relationships between different objects or activities whose worth is being assessed. There are three kinds of basic relationships which will receive numerical representation by establishing appropriate scale conventions. These are:

1. same-difference relationship (i.e., whether two objects or activities are assessed as possessing the same or different worths);

2. greater than and less than relationships (i.e., whether one object or activity is assessed as possessing more or less worth than another);

3. comparative magnitude relationship (i.e., how many times as much worth one object or activity is assessed as possessing compared to another).

Let us now proceed to construct a numerical worth scale which will reflect all of the above characteristics.

4.3 Corresponding Scale Characteristics

A numerical worth scale will be established in accordance with the following ten scaling conventions.

1. Positive numbers will be assigned uniformly to situations assessed as possessing positive worth (i.e., toward which a positive preference is felt).

2. Negative numbers will be assigned uniformly to situations assessed as possessing "negative" worth (i.e., toward which a distinct aversion is felt).

3. The worth scale will be bounded from above by plus one and from below by minus one.

4. Plus one will be assigned only to those situations deemed completely successful in terms of accomplishing positive job objectives. Analogously, minus one will be assigned only to those situations deemed completely "successful" in accomplishing "negative" job objectives (i.e., to situations than which nothing worse is conceivable in the context of the stated job).

5. The number zero will be assigned uniformly to situations assessed as completely worthless (i.e., completely unsatisfactory - but not dissatisfactory; toward which genuine indifference - but not aversion - is felt.

6. All real numbers between plus one and minus one (inclusive) are permissible measures of worth.

7. Two situations will be assigned equal worth numbers if and only if they are assessed as possessing identical worth (i.e., a decision maker feels genuine indifference in choosing between them).

8. One situation will be assigned a higher worth number than another if and only if it is assessed as possessing more worth - that is, if and only if a decision maker prefers the first situation to the second.

9. Numbers between zero and plus one (exclusive) will be assigned to all situations assessed as partially successful in terms of accomplishing positive objectives. Worth numbers will be assigned to such situations according to their proportional or percentage accomplishment of the stated objectives. This defines magnitude comparisons in terms of their ratios.

10. Numbers between zero and minus one (exclusive) will be assigned to all situations assessed as partially "successful" in terms of accomplishing "negative" objectives (i.e., stated avoidance desires). Negative worth numbers will be assigned to such situations according to their proportional or percentage "accomplishment" of stated "negative" objectives.

4.4 Choice of a Unit of Worth: Arbitrary Points

The choice of a unit of worth has already been made implicitly by two previous decisions. First, it was decided to bound the worth scale from above and below (i.e., to restrict worth numbers to fall between plus and minus one). This precludes the use of dollars or any other unit whose range lacks intrinsic logical outer bounds.

Second, it was decided to assign a worth number to a situation according to the proportional accomplishment of stated objectives achieved by that situation. This means that worth numbers may be viewed as ratios between actually achieved satisfaction and maximum possible satisfaction of stated objectives. As such, no matter what units raw satisfaction might possess, any ratio formed would be dimensionless. That is, such raw units would cancel each other out in forming a ratio, and the result would not possess any explicit units at all. Worth numbers defined in this manner are like index numbers used by various rating schemes (e.g., The Consumer Price Index).

Despite the above considerations, there is an advantage to giving index numbers a definite label so that they may be easily remembered and conveniently discussed. Because such numbers have no physical dimensions and, therefore, no natural unit, any arbitrary label is permissible - so long as it does not suggest anything which possesses either dimensions or a natural unit. Henceforward, our worth numbers will be referred to as "points" or "worth points". This label is being adopted strictly for convenience. Worth points are completely arbitrary - by definition. Their significance is encapsulated within the ten scaling conventions outlined in Section 4.3. However, their entire significance may not be immediately obvious upon a single reading of these ten conventions. Consequently, the next section will be devoted to tracing out several important implications which may have escaped a cursory first reading.

4.5 The Significance of Worth Points

We are now in a position to answer two questions concerning the significance of worth point assignments. These relate to:

1. The legitimacy of performing basic arithmetic operations on worth points;
2. The legitimacy of assigning worth points to situations toward which no positive preference is felt.

These issues will be discussed in sequence.

First, before we can know which kinds of arithmetic operations may legitimately be performed on assigned worth points, we must investigate the process by which decision makers impute worth to situations. Specifically, we must ask whether decision makers can answer meaningfully a question of the following general form:

1. Given a stated objective;
2. Given a well-described situation which purports to accomplish the stated objective, at least partially;

   Then to what degree does the situation described succeed in accomplishing the stated objective, where degree is assessed in terms of a proportion or percentage-like number between zero and one?

If decision makers cannot answer the above type of question at all, then the issue of legitimate arithmetic operations becomes vacuous. There would be no worth points upon which to operate - either legitimately or illegitimately. (Note: The possibility of being unable to answer such questions is considered quite reasonable, and remedial measures will be discussed later.)

On the other hand, even if such a question can be answered, there is still the problem of meaningfulness. How meaningful are numerical point scores, when assigned? That is, how confident are decision makers in the interpretability of the numbers they give? If decision makers can muster a reasonable degree of confidence in their own ability to answer the above type of question (and only the decision makers themselves can make such a determination), then all four basic arithmetic operations may be performed on assigned worth points, except for a sign restriction. Worth points may be added, subtracted, and multiplied by a non-negative constant with complete freedom. However, attention must be paid to their sign when multiplying or dividing by one another. Only worth numbers of like sign may undergo these operations, and then only their absolute magnitude is relevant. In the language of scaling theory, each half of the worth scale constitutes a full-fledged ratio scale, with the negative half being treated as a "mirror-image" of the positive half.

The second issue concerns the legitimacy of assigning worth points to situations toward which no positive preference is felt. Ignoring the case of indifference, which receives a point score of zero, this includes both situations toward which a distinct aversion is felt and situations toward which neither a positive preference nor a distinct aversion are felt directly, but whose indirect consequences are such as to arouse a reduction in positive feelings.

An example of a situation generating direct aversive feelings would be the development of a new drug which proved highly efficacious in one way, but which also produced noxious side effects. Otherwise positive attitudes toward an effective contraceptive device would be substantially mitigated, if not completely overruled by the discovery that it induced permanent sterility. Sterility, here, would constitute a "negative" objective which most people would definitely prefer to avoid and toward which they would feel directly aversive. Negative worth points would be assigned to this situation depending upon the extent to which permanent sterility were induced.

However, there is another type of situation toward which decision makers may feel neither a direct preference nor a direct aversion. This is the situation wherein limited resources must be expended to complete a job. Unless a decision maker possesses miserly feelings, he has no direct aversion to spending money or committing workers or using up capital equipment *per se*. If the supply of such resources were truly unlimited relative to their demand, the resources themselves would have no worth at all - either positive or negative. Consequently, spending such unlimited resources could only be regarded with indifference. There would always be enough to go around - if the supply were truly unlimited. Resources only become valuable when their available supply falls below the total demand for their effective utilization. But even then, their worth is not intrinsic to the resources themselves. Rather, their worth derives from the fact that they may be diverted to some alternative application which, if carried out, would generate consequences perceived as worthwhile in their own right. Expending limited resources to complete one job precludes using the same resources to complete some alternative jobs, and the worth of completing the alternative jobs must, therefore, be forgone.

In view of these observations, we may now ask whether it is legitimate to assign worth points directly to the expenditure of resources. The answer is not usually. Worth points, as defined in this paper, can only be assigned to situations perceived as worthwhile in their own right because they succeed in accomplishing stated objectives. Although it would be possible to define "conserving resources" as a specific objective, it would be difficult to judge the worth of any given amount of conserved resources unless or until the alternative applications of the same resources had first been ascertained and assessed. Until this is accomplished, no meaningful point scores may be assigned to resource expenditures.

The above conclusions have two important procedural implications. Since worth points are generally not assigned to resource expenditures incurred in acquiring and utilizing a produced alternative, while worth points are assigned to other kinds of consequences related directly to stated job objectives, it is important to define at the outset just which consequences are to be regarded as resource-oriented and to distinguish these clearly from objective-oriented consequences. In addition, some alternative means must be found to reflect the worth implications of expending resources and to incorporate these explicitly into an overall decision making methodology. Section 5 will discuss briefly various ways of incorporating resource considerations into an overall decision making methodology.

4.6 Summary

To summarize the discussion in Sections 4.1 through 4.5, the following things have been concluded.

1. The basic purpose in assigning worth numbers to external situations (i.e., to physical objects, activities, and degrees of performance being assessed) is to reflect explicitly whatever worth characteristics and relationships are perceived by decision makers and imputed to such situations.

2. Numbers are used because they are familiar, well-understood, and easy to manipulate.

3. Worth numbers should be assigned in such a manner that numerical relationships existing among assigned numbers will correspond with and, therefore, serve to identify worth characteristics and relationships imputed to these external situations.

4. The most fundamental worth characteristics to be reflected by assigned worth numbers are the psychological states of preference, aversion, and indifference.

5. Although the concept of worth is generally conceived of as referring to situations for which a

positive preference is felt, "negative" worth has been defined in terms of "negative" objectives to take care of situations toward which a distinct aversion is felt. The concepts of "negative" objectives, "negative" worth, and their associated negative point scores are defined by "mirror-image" analogy to the corresponding positive concepts and associated point scores.

6. The assessment scheme to be developed later will apply only to situations for which definite objectives can be stated. Consequently, it makes sense to talk of the extent to which a given situation succeeds in accomplishing such objectives. This, in turn, suggests logical outer bounds to the scale of worth numbers corresponding to complete satisfaction of "positive" and "negative" objectives, respectively.

7. The scale of worth numbers is defined as continuous or everywhere-dense within its entire logical range.

8. Two situations are assigned identical worth numbers if and only if they are perceived as equally worthwhile (i.e., if and only if a decision maker feels genuine indifference in choosing between them).

9. One situation is assigned a higher worth number than another if and only if it is perceived as more worthwhile - that is, if and only if a decision maker prefers the first situation to the second. A reverse statement applies to assigning lower worth numbers.

10. The scale of worth numbers is restricted to contain all real numbers between plus one and minus one (inclusive). The number zero is assigned to any situation which is perceived as worthless (i.e., completely unsuccessful in accomplishing stated objectives) and, therefore, to which genuine indifference - but not aversion - is felt. Plus one is assigned to any situation which is perceived as completely successful in accomplishing positive objectives. Intermediate numbers are assigned to situations perceived as partially successful in accomplishing stated objectives according to the extent or degree (reflected as a proportion or percentage) to which they are successful. Negative worth numbers, when appropriate, are assigned analogously, with minus one being reserved for complete satisfaction of "negative" objectives.

11. Worth numbers are given the label "points". Worth points are arbitrary. They possess no physical dimension or physical unit. Their meaning and proper interpretation are completely specified by the scaling conventions outlined previously.

12. Addition, subtraction, and multiplication by a non-negative constant may be performed on worth points - provided the process by which they are originally assigned is such as to give them the full meaning specified by the ten scaling conventions. Whether or not they possess this full meaning must be decided by the decision makers

themselves. In addition, multiplication and division of the absolute value of worth points with the same sign is also permitted (to compute a geometric mean or a ratio, for example).

13. Worth points are not ordinarily assigned to the expenditure of resources, unless specific resource-conserving objectives have been stated for the job.

SECTION 5

RELATED CONCEPTS

Before moving to the development of a formal assessment procedure, the relationship of worth to the classical concept of utility deserves some brief attention. In addition, the roles of worth, utility, and resources in the overall decision making process deserve a few brief comments. Although neither of these topics will be treated extensively within this paper, a brief discussion of each will add perspective to our future discussions of worth assessment.

5.1 Risk, Uncertainty, and the Classical Concept of Utility

The worth concept is completely devoid of any risk and/or uncertainty considerations. In assessing the worth of a situation, activity, or performance consequence, it is assumed that such an outcome will occur for certain. Consequently, assigned worth numbers will not reflect the aversion which a decision maker may feel toward either risk or uncertainty regarding the actual occurrence of that outcome. Furthermore, the process of assigning worth numbers provides no mechanism for reflecting perceived trade-offs between the worth of an outcome, conditional upon its actual occurrence, and the variable risk or uncertainty surrounding its occurrence. The worth concept and the related worth-measuring and worth-assessing procedures are, therefore, incomplete in this sense.

In contrast, the classical concept of utility, as articulated by Von Neumann and Morgenstern and used by statistical decision theorists, does provide an explicit mechanism for reflecting perceived trade-offs between conditional worth, on the one hand, and risk or uncertainty on the other hand. However, the concept of utility ignores the problem of formulating and articulating a measure of conditional worth. It assumes that the decision maker has already formulated a worth measure and proceeds from there.

A complete assessment procedure should take account of both conditional worth and risk/uncertainty considerations. That is, it should provide a mechanism for assessing worth, conditional upon certainty, and then it should provide an additional mechanism to account for the decision maker's attitudes toward risk and uncertainty. This paper will address itself exclusively toward the former task (i.e., generating a conditional measure of worth). However, the numerical output of the worth assessment procedure to be developed herein is a perfectly legitimate input to the Von Neumann-Morgenstern utility assessment procedure. Worth point scores can be used as the numeraire to which utility numbers are assigned. More will be said later about this symbiotic bond between the worth and utility concepts.[3]

5.2 The Concept of a Decision Rule

A decision rule might be defined broadly as any uniformly applicable directive which indicates a clear choice among properly specified alternatives in a given situation. It is through the mechanism of a decision rule that decision makers specify the trade-offs they are willing to make among worth, risk/uncertainty, and resource considerations.[4] Examples of decision rules which are frequently used in selecting among alternatives include:

1. The Economy Rule, directing decision makers to select the least expensive feasible alternative (i.e., the least resource-consuming alternative which satisfies all stipulated mandatory performance

---

[3]The writer is grateful to Howard Raiffa for pointing out this explicit connection between the worth and utility concepts and for suggesting that worth point scores be used as the numeraire in a utility assignment procedure.

[4]The trade-off between conditional worth and risk/uncertainty is assumed throughout this discussion to be encapsulated in a utility index of the variety discussed in Section 5.1.

requirements and, possibly, physical resource limitations);

2. The Ratio Optimizing Rule, directing decision makers to select whichever feasible alternative maximizes a utility-to-cost (or, equivalently, minimizes a cost-to-utility) ratio;[5]

3. The Weighted Average Utility-Cost Rule, directing decision makers first to assign numerical weights to receiving valuable performance versus expending limited resources, then to assign explicit measures of utility both to received performance and to expended resources, and finally to select whichever feasible alternative maximizes the weighted sum of these separate utility indices.

Obviously, the above list does not exhaust all decision rules that have been or could be used to select alternatives, but it does provide a reasonable basis for discussion. In particular, it provides a reasonable basis for illustrating the primary role of a decision rule in integrating worth, risk/uncertainty, and resource considerations.

In choosing a decision rule, the decision maker must ask himself what he is really trying to accomplish when he finally selects an alternative. He may raise such questions as the following.

1. Assuming that at least one of the produced alternatives is feasible, must one of them always win the selection; or is it possible to reject all of the

[5]The word "cost" is used throughout this section to indicate the physical process of expending resources including, but not restricted to, monetary resources.

alternatives on the grounds that they _all_ provide "over-kill" and that the same resources might better be expended on some other project altogether?

2. Should each successive selection in which the decision maker is required to make a choice be considered separately, without regard to the consequences of that choice on subsequent selection decisions; or should the decision maker assume a broader viewpoint which embraces the _whole sequence_ of decisions he must make?

3. In what sense should valuable performance received be compared with resources expended? Is it worthwhile to expend additional resources in order to receive additional valuable performance over and above minimum requirements? If so, how much more and until what has been achieved?

Answers to these questions should help the decision maker choose a decision rule, or at least narrow substantially the field of candidates. To illustrate why this is so (i.e., how these questions are related to various decision rules), let us consider the implicit answers given to each by the economy rule, the ratio optimizing rule, and the weighted average utility-cost rule, respectively.

First, the economy rule requires that, if at least one feasible alternative is produced, then one of them must win. It is impossible, under the economy rule, to reject all feasible alternatives - even if the least costly alternative requires a staggering expenditure of resources. No protection against "over-kill" is provided.

Similarly, the ratio optimizing rule (in either of its two equivalent forms) provides no protection against "over-kill". It is quite possible to encounter a set of alternatives - all of which promise performance greatly in excess of what is required (or even desired) and which involve commensurately excessive resource expenditures. Nevertheless, that alternative with optimum ratio would still be defined and, unless the rule were enhanced with a budget constraint or some other protective device, "over-kill" would thereby be suffered.

The weighted average utility-cost rule also fails to provide any protection against "over-kill". As in the case of the ratio optimizing rule, all produced alternatives may promise excessive performance and require excessive resource expenditures. Without an explicit budget constraint or some other protective device, this rule is rendered equally helpless.

Regarding the second question, both the economy rule and the ratio optimizing rule focus attention exclusively on each successive selection considered by itself. No explicit consideration is given to the consequences of one selection decision on other such decisions. In particular, no recognition is given to the fact that, when the total supply of resources is limited (as it almost always is in real-world situations), what must be expended to choose an alternative in one selection cannot be expended on another selection. No limits or any other direct controls are placed on resource expenditures.

This is not the case, however, with the weighted average, utility-cost rule. In principle, a degree of control may be exercised over the amount of resources expended either by selectively altering the weight on resources or by choosing appropriate utility functions for resource expenditures. Although it may not be clear how to exercise these controls so as to achieve an appropriate allocation of resources over an entire sequence of selection decisions, at least a potential control mechanism exists.

Regarding the third question, the three decision rules give quite different answers. The economy rule rejects completely the notion that additional performance over and above minimum requirements might be worth spending additional resources to obtain. It chooses the cheapest alternative that does the job, even if performance is just barely satisfactory.

In contrast, both the ratio optimizing rule and the weighted average utility-cost rule recognize the potential worth of additional performance over and above minimum requirements, and both rules permit spending additional resources to obtain it. However, the extent to which each rule will spend additional resources and the apparent reasons for spending it are different. Under the ratio optimizing rule, the goal is to get the "best buy" (i.e., the most for whatever resources are expended as evidenced by either a maximum utility-to-cost or a minimum cost-to-utility ratio). Under the weighted average utility-cost rule, additional resources may be spent to obtain additional performance so long as the extra utility gained equals or exceeds the extra penalty suffered by spending more resources. The balancing of utilities is carried out by means of the utility functions on various performance measures and resources in conjunction with the weights placed on each.

The preceding discussion was intended to indicate that, even after a satisfactory measure of worth has been defined, and even after risk/uncertainty has been taken into account by means of a satisfactory utility index, there still remains the problem of integrating these considerations with a careful consideration of resource expenditures before a complete decision methodology can be achieved. Choice of a satisfactory decision rule constitutes a means of achieving integration, and, as the preceding discussion illustrated, this is not a simple task.

5.3 Some Concluding Remarks

In Sections 1 through 5 we have discussed the overall problem of designing and implementing a rational decision-making methodology.

An important part of this overall problem is the limited, but by no means simple task of constructing a systematic procedure to aid in the formulation, articulation, and measurement of worth notions. Once a worth measure has been achieved, it still remains to combine this with risk/uncertainty considerations (e.g., by defining a utility measure) and then to compare the result with whatever resources must be expended (e.g., by defining a satisfactory decision rule) before a final decision can be reached.

The great bulk of our discussion to date has been devoted to laying a conceptual foundation for the worth-formulating and measuring procedure to be developed in Section 7. In addition, an attempt was made to delimit the issue of worth determination from the related issues of risk/uncertainty, classical utility, resource expenditures, and an integrating decision rule. Let us now proceed to the development of a systematic procedure for formulating and measuring notions of worth.

SECTION 6

OUTLINE OF A PROCEDURE FOR FORMULATING AND MEASURING WORTH

In Section 1 of this paper it was pointed out that worth assessment is an especially difficult task in the case of "complex" alternatives due to:

1. Multiple objectives and assessment criteria to list and arrange in some organized fashion;
2. Multiple performance consequences to predict;
3. Multiple worth connections between listed assessment criteria and predicted performance consequences;
4. Physical interaction among performance consequences;
5. Worth interdependence among assessment criteria.

However, the difficulty of this task can and will be reduced somewhat by making two simplifying assumptions. First, it will be assumed that validated estimates are freely available for all relevant performance consequences associated with all produced alternatives. Naturally, both obtaining and validating such estimates constitute very real and highly important problems in their own right, but neither of these will be discussed here in any detail. Such omissions are purely for simplification.

Second, it will be assumed that our task is restricted to assessing a fixed set of discrete alternatives. The problem of producing alternatives (i.e., of designing, redesigning, or soliciting proposals for alternatives) will not be considered. This assumption reduces substantially any worries we might otherwise have had concerning physical interaction among performance consequences, since physical interaction is troublesome primarily because it renders prediction of performance difficult under differing design alternatives.[6]

Nevertheless, in spite of these simplifications, we must still worry about listing and arranging multiple objectives and assessment criteria, checking for worth interdependence among them, and establishing worth connections between these criteria and various performance consequences. The remainder of this section will address itself to these tasks.

6.1 Listing Overall Performance Objectives

The first step in making a formal assessment is to specify what is desired from whatever alternatives may be produced. This means listing objectives. At the outset, objectives may be (and should be) stated in very general terms. After all, the point is to be as all-encompassing as possible initially (to avoid omitting any important objectives which decision makers really possess and are willing to display), and then to work down through a process of successive elaboration to a very specific statement of desired performance. A very specific statement of intentions is required at the end of the process in order to carry out an actual assessment, but this need not concern us too heavily at the beginning. Rather, what should concern us is summarized below.

Any list of overall performance objectives should possess the following desirable properties.

1. The list should be complete and exhaustive. That is, all important performance objectives deemed relevant to the final decision should be represented by the

---

[6]The scope of our problem is greatly reduced by this assumption, but not to the point where it no longer possesses practical significance. After all, in any real decision situation, there comes a moment when the process of producing alternatives must be terminated, and an immediate choice must be made among whichever alternatives have already been produced. At that moment of decision it is reasonable to view the choice as among a fixed set of discrete alternatives.

items on the list. This is to guarantee that no important performance considerations are overlooked by the assessment procedure.

2. The list should contain mutually exclusive items. That is, no listed objective should be stated in such a way as to encompass (definitionally) or to be encompassed by (definitionally) any other objective either in whole or in part. This is to permit decision makers to view listed objectives as independent entities among which appropriate trade-offs may be established. This will also help prevent undesirable "double-counting" in the worth sense.

3. The list should be restricted to performance objectives of the highest degree of importance. That is, only overall objectives should be included. The purpose of this exclusion is to provide a sound basis or starting point from which lower-level criteria may subsequently be derived.

4. Finally, the list should contain objectives relatively independent in the worth sense. That is, for any pair of objectives on the list, decision makers should be willing to trade-off additional satisfaction on one objective for reduced satisfaction on the other at a rate relatively independent of the levels of satisfaction already attained on each. The meaning of independence

and the reasons for requiring objectives to be independent will be discussed in Section 6.6.

6.2 Generating a Hierarchical Structrue of Performance Criteria

Having established a list of overall performance objectives satisfying the above four logical requirements, the second step is to define more precisely what these highest-level objectives really mean. To accomplish this, each highest-level objective is subdivided into one or more lower-level criteria. The purpose of subdividing is to state explicitly (i.e., in terms of lower-level criteria) what is intended by or included within the meaning of each overall objective. But what, exactly, is the nature of this task?

Essentially, our task is to create a pictorial map of the structure of worth relationships existing within the mind of a decision maker. Just as a cartographer attempts to depict topographical relationships of distance, elevation, contiguity, etc., between masses of land and water in some specified geographical region, we are trying to depict worth relationships between overall performance objectives and successively lower levels of increasingly more specific performance criteria relevant to the selection of a specified alternative for some definite job. Just as the cartographer utilizes certain conventions such as contour lines and special coloring to convey information about the terrain he is describing, we shall adopt the convention of a tree-like array to convey information about the decision maker's worth structure.

Despite these similarities, however, there are a number of important differences between constructing maps of regional topography and maps of human worth structure. First, the cartographer attempts to describe various aspects of our physical environment. We, on the other hand, are attempting to describe various aspects of the inner minds of human decision makers. This suggests that the proper focus of our attention is not the "out-there" physical world of nature, but rather the "in-here" subjective world of human beings. It is to decision makers and their evaluative responses that we must look in constructing our map.

A second difference follows immediately from the first. Since the cartographer is attempting to map something physical and directly observable, he may utilize direct measuring devices such as compasses and other surveying tools. We, on the other hand, are attempting to map something non-physical and only indirectly measurable. We are therefore forced to utilize indirect measuring devices such as verbal questioning and behavioral observation. From these kinds of data we must infer the underlying structure of human preferences.

A third difference relates to the number and temporal stability of the entities being mapped. Whereas there is only one topographical region to be investigated by the cartographer (the particular region he is interested in mapping), there are frequently more than one decision makers to be investigated in mapping a worth structure (e.g., the group of decision makers responsible for making a selection decision). In addition, topographical features of our physical environment are apt to be highly stable over time, while attitudinal features of our assessment structure are apt to change over time with new learning and increased assessment experience.

A fourth difference, and by far the most important one, relates to the perturbing effect of the mapping process itself. The cartographer is concerned with depicting visually a territorial region which has already been formed by the forces of nature. His mapping process does not alter significantly the nature of the physical terrain being mapped. In sharp contrast, our mapping process has an enormous impact upon the worth structure being mapped. On the basis of an experiment to be reported later, it was concluded that the single most important consequence of the entire assessment procedure is to create a worth structure where one did not previously exist - at least not in conscious, well-defined, and easily articulatable form. Participation in this assessment procedure induces the decision maker to formulate a consistent worth structure. At the very least, this entails substantial clarification of what already existed in his mind. Typically, it induces him to alter substantial portions of his prior worth structure. At most, it induces him to create a structure which did not enjoy any prior existence at all in consciousness. Producing a pictorial map of the worth structure, once formulated, constitutes a separate and important consequence of the assessment procedure, but this is not the only consequence, nor is it the most important one.

More will be said later about the dynamic interrelationship between formulating and representing a worth structure. For now, however, we shall concentrate primarily upon the representational or mapping aspects of the process. By means of a step-by-step questioning procedure, a hierarchical, tree-like structure of increasingly more specific performance criteria is generated to represent what the decision maker desires from produced alternatives.

6.3 Selecting Physical Performance Measures

After generating a hierarchical tree of overall objectives and lower-level criteria, the third step is to select a single physical performance measure for each lowest-level criterion on the tree. The purpose of selecting physical performance measures is to give concrete, physical interpretations to their related lowest-level criteria. By this device, a bridge is constructed linking the subjective inner minds (i.e., the worth structures) of decision makers to the objective outer world of physical alternatives. Let us clarify this concept - particularly the distinction between performance criteria and physical performance measures - with further discussion.

A physical performance measure is any tangible reading or concrete observation that can be extracted from the real world. For purposes of assessment, it is any directly measurable attribute of a produced alternative. However, this is not the same thing as a performance criterion. Whereas stated criteria reflect what a decision maker desires from produced alternatives, performance measures reflect what an alternative can actually deliver. Performance criteria are attributes of decision makers, while performance measures are attributes of the physical alternatives being assessed.

Although this may sound like a mere academic distinction, it will be useful for very practical reasons to keep the two concepts clearly separated. There are three reasons for maintaining the distinction. First, the methods of approach and the people one talks to in formulating performance criteria are different from the methods and people involved in defining physical performance measures. Introspective reflection and discussions with fellow decision makers can help to formulate, to clarify, and to understand performance criteria. This seems like a reasonable way to define what is desired from an alternative. In contrast, inspection of physical alternatives and discussion with knowledgeable engineers would seem a more useful way to define physical performance measures. These reflect what an alternative will deliver (no matter what is desired).

A second reason for distinguishing between performance criteria and performance measures springs from the very different way in which they will be treated in the process of formal assessment. Once defined, physical performance measures will be used to describe each of the produced alternatives. The description of an alternative in terms of a set of physical performance numbers (and/or other descriptive symbols) will then be converted into equivalent worth point scores by means of a device called a scoring function (to be discussed in Section 6.4).

In direct contrast, worth scores attached by scoring functions to lowest-level performance criteria are not themselves run through scoring functions. Instead, they will be combined with other worth scores already attached to other performance criteria. Such combination will be carried out by means of a device called a weighting function (to be discussed in Section 6.5), and the result will be a single, overall index of worth associated with each produced alternative.

A third reason for maintaining the distinction involves the handling of interdependent performance criteria. It will be demonstrated at a later point in this paper that such criteria can sometimes be eliminated by defining a "high-order" performance measure. Although the meaning of this statement may not now be clear, subsequent discussion should clarify it.

Selecting physical performance measures must be done judgmentally. A decision maker must choose a well-defined and easily measurable physical attribute of an alternative which he feels serves to interpret, in phenomenological terms, the intended meaning of the lowest-level criterion under consideration. Thus, returning to the computer example, the performance criterion "printer speed" might be interpreted by means of the physical measure "maximum number of lines printed per minute, assuming no jamming or other form of breakdown." But this raises two questions. First, how does one come up with a candidate measure? Second, if more than one candidate arises, how does one choose from among them.

Coming up with candidate measures, just like generating subcriteria to fill out the hierarchical tree, requires ingenuity and informed judgment. Both tasks involve creative acts. However, both tasks will be much easier to accomplish if decision makers take the trouble to compile in advance a master list of reasonable candidates. This master list might contain all performance criteria and all performance measures that have ever been suggested and/or used on past decisions of a similar nature. Particular criteria and measures could then be extracted (or synthesized) from the master list as needed for each successive decision. In addition, the master list could be continually updated to include new criteria and measures as they are created.

As for choosing from among alternative candidate measures to associate with a given performance criterion, this also requires informed judgment. It may happen, for example, that certain printers are known to jam frequently under continued, high-speed operation. Under such circumstances, a better measure of "printer speed" might be "expected number of lines printed per minute" with jamming and other forms of breakdown taken into consideration by means of historical breakdown frequency data relating to each type of printer. However, this kind of choice must be made by decision makers on the basis of historical evidence and their own personal experience making assessments.

6.4 Establishing Specific Worth Relationships Between Lowest-Level Performance Criteria and Their Associated Physical Performance Measures: The Scoring Problem

The fourth step in constructing a formal assessment procedure is to establish specific worth relationships between each lowest-level performance criterion in the hierarchical structure and its associated physical performance measure. Selecting performance measures (the step just discussed in Section 6.3) serves to establish the existence of worth connections, but it does not serve to map out specific worth relationships. Specific relationships are established by means of scoring functions.

A scoring function is a mathematical rule which assigns a unique worth score in points to every possible value of some physical performance measure. It transforms raw performance (measured in terms of whatever physical unit is appropriate to the performance measure under consideration) into worth-of-performance (measured in terms of the worth points discussed in Section 4). Just as the selection of a physical performance measure serves to interpret concretely each lowest-level performance criterion and, therefore, to provide a bridge between the subjectively defined worth structure and the objectively defined physical characteristics of an alternative, the specification of a scoring function serves to define precisely the nature, shape, and particular parameters of this bridge.

In formulating a scoring function, it is temporarily assumed that the lowest-level performance criterion in question constitutes the only performance objective in the entire assessment. Then, the worth score assigned by the scoring function to any given amount of performance on the associated physical performance measure is supposed to indicate the extent to which that amount of physical performance actually satisfies the lowest-level criterion. To accomplish this, certain conventions or ground rules must be observed uniformly to insure that all worth scores thereby generated will be comparable with one another and subject to a uniform interpretation. Otherwise, the subsequent procedure by which individual worth scores assigned to separate criteria are to be combined cannot be meaningfully carried out. A set of scoring conventions designed to insure both consistency and comparability appears below.

1. The outputs of all scoring functions will be in terms of worth points.

2. Worth points will be as defined in the ten scaling conventions presented in Section 4.3.

3. All scoring functions will be formulated to cover the entire range of logically possible physical performance - not just the reasonable or expected range. This is to insure that a definite point score will be defined for every conceivable level of produced performance - no matter how unexpected it may be.

4. All scoring functions will be formulated independently of (i.e., without reference to) physical performance actually offered by produced alternatives. Knowledge of what is offered should not be necessary and may even bias the process of stipulating what is desired.

5. Hopefully, all scoring functions can be formulated prior to inspecting any performance date. This will help to guarantee that scoring is performed on an independent basis.

6. Most scoring functions will take the form of mathematical formulas and/or graphically depicted mathematical curves. However, some will not be expressed in these terms. Some will take the form of direct judgmental point assignments by decision makers without the aid of either formulas or graphs. In this latter case, scoring functions are thought of

as implicit within the minds of decison makers rather than explicitly stated in any precise mathematical form.

7. All scoring functions will be formulated by means of a single, uniform, and replicable procedure. A suggested two-phase procedure (embodying the above six scoring conventions) will be presented in Section 7.

6.5 Establishing a Procedure for Combining Worth Scores Assigned on the Basis of Separate Performance Criteria to Arrive at a Single, Overall Index of Worth: The Weighting Problem

In discussing scoring functions, it was suggested that one temporarily assume each lowest-level performance criterion under consideration to be the sole performance objective in the entire assessment. Obviously, this is an untenable assumption. There are many performance objectives to be satisfied as reflected in the hierarchical structure with its many lowest-level branches. This brings us to the fifth step in formal assessment - combining worth scores assigned on the basis of separate performance criteria to arrive at a single, overall index of worth. This step will be accomplished by defining a weighting function.

A weighting function is a conceptual device by means of which explicit recognition is given to the existence of multiple objectives and performance criteria. Whereas a scoring function is defined to indicate the extent to which any given level of measured performance succeeds in satisfying its related lowest-level performance criterion, a weighting function is defined to indicate the perceived relative importance of satisfying the criterion itself vis a vis other performance criteria. In this manner, the temporary assumption of a single criterion made in defining a scoring function is relaxed to reflect reality. Simultaneously, a means of combining worth scores assigned on the basis of separate criteria into a single, overall index is achieved. Let us illustrate these results by means of a very simple example.

Suppose that a computer is to be acquired and operated with two specific objectives in mind. These are:

1. to perform a current job clearly described in terms of workload and applications;

2. to expand at some future date beyond the currently stated workload and/or applications.

Suppose, also, that performance of the current job is to be measured by the time required (in hours) by each produced alternative to process a benchmark program simulating a day's workload and that an appropriate scoring function has been defined to convert all possible benchmark times into equivalent worth scores. Finally, assume that expansion potential is to be measured by immediate-access memory capacity (in number of words) unused by the benchmark exercise and that an appropriate scoring function has also been defined for this performance measure. Then, each produced alternative would be assigned two worth scores - one for performing the current job and another for demonstrated expansion potential. How can these two separate scores now be combined into an overall index of each alternative's total worth? This is the weighting problem.

One way to proceed would be as follows. Decision makers ask themselves which of the two performance criteria - doing the current job or providing expansion capability - should be considered more important. That is, if given the choice between satisfying either of the two criteria to the same extent, which one would they prefer to have satisfied? Alternatively, would decision makers feel genuine indifference in choosing between equal percentage satisfaction of the two criteria? If decision makers would prefer to have the current job criterion satisfied over having the expansion potential criterion satisfied to the same extent, then the former criterion must be considered more important than the latter. If genuine indifference is felt between having the two criteria equally well satisfied, then they must be regarded as equally important.

The next step is to be a bit more precise about the extent or degree of perceived relative importance. Just to say that doing the current job is more important than providing expansion potential is usually not sufficient to distinguish clearly between the overall worths of competing alternatives. The magnitude of this perceived relative importance must also be indicated. How much more important is it to satisfy the current job criterion than to satisfy the expansion potential criterion? Twice as important? Ten times as important? Representation of relative magnitudes once again suggests resorting to numbers.

Suppose that performing the current job were considered three times as important as providing expansion potential. Then, any pair of numbers standing in the ratio of 3:1 could be used to convey this information. In particular, the numbers 3/4 and 1/4 could be used. Then, whatever scores are attached by scoring functions to these two criteria could be combined by:

1. multiplying the score assigned to performing the current job by 3/4;
2. multiplying the score assigned to expansion potential by 1/4;
3. adding the two products to arrive at a weighted average score, using the importance ratios as constant weights.

The resulting sum of weighted scores might then be interpreted as an overall index of each alternative's total worth.

The above procedure has a definite appeal in its simplicity and directness. It seems to solve the problem of combining scores on separate criteria, and it seems to arrive at a single, overall index of worth. What's more, by requiring the set of constant weights to add internally to one (as was done in the example above), the resulting overall worth score (computed as the sum of weighted individual criterion scores) also lies between zero and one and may be subjected to the same interpretation as worth point scores assigned to individual performance criteria. This renders far more manageable the task of checking assigned weights for intuitive reasonableness and consensual validation. The same questions may be asked of weighted sums as are asked of individual criterion scores. Since worth scores cannot be validated by any other means (recall that they are in principle untestable by ordinary scientific techniques), uniform interpretability becomes an extremely important and valuable asset.

However, in spite of its simplicity and immediate appeal, the above procedure should be subjected to critical scrutiny before accepting it and incorporating it into a formal assessment scheme. It would be wise to inquire a bit more carefully into what this weighting procedure is really assuming about how decison makers view multiple assessment criteria, how they trade off worth among multiple criteria, and what procedural implications these assumptions have for the practical task of assessment. It will be shown that the key to understanding these issues lies in the concept of worth interdependence among separate performance criteria. This concept and its procedural implications will be discussed in the next section.

6.6 Identifying and Eliminating Worth Interdependence Among Separate Performance Criteria

The preceding example of combining worth scores by means of weighting and summing to arrive at a single index of overall worth assumes implicitly the following things.

1. The relative importance of satisfying separate performance criteria does not depend upon the various degrees to which each criterion has itself been satisfied. Rather, their relative importance is conceived as being constant in this respect.

2. The rate at which increased satisfaction of any given criterion contributes to overall worth is independent of the levels of satisfaction already achieved on that and other criteria. Rather, such rates are viewed as constant in this respect.

3. The rate at which decision makers would be willing to trade off decreased satisfaction on one criterion for increased satisfaction on other criteria so as to preserve the same overall worth is independent of the levels of satisfaction already achieved by any and all of the criteria. Such trade-off rates are viewed as constant in this respect.

These three logically interrelated statements, taken together, define the concept of worth independence.

To clarify further this concept of worth independence and, more particularly, to distinguish it from interdependence, let us consider two contrasting examples. First, we shall return to the example given in the preceding section and argue that performing the current job and providing expansion potential constitute worth-independent criteria. Then, we shall concoct a counter-example to illustrate worth-interdependence.

The two criteria previously considered were:

1. system performance potential with respect to the currently stated workload and applications;

2. expandability of the system beyond the currently stated workload and applications.

These two criteria are independent because decision makers would like to have both satisfied simultaneously and because the extent to which either is satisfied does not influence the extent to which the other is satisfied. Each of these outcomes is desirable in its own right. Under these circumstances, and only under these circumstances, it is reasonable to combine criterion scores by means of constant relative importance weights.

Now let us concoct an example of substantial worth interdependence. Suppose that, in terms of expanding the job beyond the currently stated workload and applications, decision makers perceive as valuable the presence of additional memory capacity. This might be provided either by additional core storage capacity, by additional disk capacity, by additional drum capacity, or even by additional tape capacity.

Suppose, further, that decision makers attempt to reflect this by subdividing the higher-level criterion, additional memory capacity, into four sub-criteria:

1. additional core storage capacity;

2. additional disk storage capacity;

3. additional drum storage capacity;

4. additional tape storage capacity.

Now these four sub-criteria would not be independent in the worth sense. They are highly interdependent. They constitute four alternative means of satisfying the same single objective - providing additional memory capacity. They are valuable only insofar as they accomplish this end. Because of this, satisfying any one of these four sub-criteria is a more or less acceptable substitute for satisfying any of the others, and the rate at which decision makers would be willing to exchange additional satisfaction on any one of the four for reduced satisfaction on any of the other three depends critically upon the extent to which any or all of them have already been satisfied. Consequently, combining worth scores on these four criteria by means of an additive scheme with constant weights would seem highly inappropriate.

Having illustrated the twin concepts of worth independence and worth interdependence and having shown that an additive scheme using constant trade-off weights only makes sense with worth-independent criteria, the time has come to ask a more fundamental question. What happens if constant trade-off weights are applied to worth-interdependent criteria? Numerical sums of weighted scores would be defined mathematically, but it would not be at all clear what these results signified (if anything) in terms of overall worth. This suggest that one of two remedial steps be taken under such circumstances. Either a different weighting function involving variable trade-off weights which depend upon the levels of criterion scores should be defined in place of the simpler function involving constant weights; or the simpler weighting function should be retained, and the hierarchy of performance criteria should be purged of worth-interdependent members. Alternatively, some combination of the above two remedies might be adopted.

In choosing a remedy, we are forced to ask a second fundamental question. How can decision makers really think about making worth trade-offs among separate criteria? How flexible are they in conceptualizing the issue? Judging from the frequent occurrence of additive combinatorial schemes using constant trade-off weights both in government and industry, it would appear that decision makers characteristically prefer to represent their trade-off notions in this very simple manner. Previous attempts on the writer's part to induce numerous decision makers to articulate their trade-off notions in a more complex manner all ended in confusion and failure. Now this does <u>not</u> mean that <u>all</u> notions of worth trade-offs are (or should be) viewed in terms of an independent and additive scheme. A counter-example of substantial worth interdependence has just been presented. It just means that this seems to be a <u>convenient</u> way to think and that more complex ways of thinking may be too difficult to articulate explicitly and, therefore, to incorporate systematically within a formal assessment scheme.

In light of these considerations, the writer has chosen to adopt a modified version of the second remedial strategy. That is, the formal assessment procedure to be developed herein will be based on an independent and additive weighting function. To make the procedure meaningful in terms of interpreting its numerical results, instances of substantial worth interdependence will be detected and purged from the hierarchical criterion structure.

The decision is an essentially pragmatic one. It is based on the assumption that decision makers would find it too difficult in practical situations to articulate trade-off relationships in a more complex manner. If subsequent research should indicate this to be a false assumption, the decision might appropriately be reversed.

On the other hand, complete elimination of all instances showing even the slightest trace of worth interdependence might reduce the hierarchical structure to almost nothing. This would serve only to eliminate from explicit consideration many important aspects of the assessment problem. Therefore, only instances of <u>substantial</u> worth interdependence will be purged from the hierarchy. A specific procedure for identifying and eliminating instances of substantial worth interdependence among performance criteria will be presented in Section 7.

6.7 The Meaning and Interpretation of Weights

Just as it was useful to establish by means of explicit scale conventions the meaning and proper interpretation of worth point scores, so also is it useful to establish a similar logical basis for numerical weights. This will be accomplished by stating and discussing briefly ten weighting conventions.

1. A set of numerical weights will be defined for every set of sub-criteria into which a higher-level criterion in the hierarchical criterion structure is subdivided. In the case of the highest-level of overall performance objectives, these are construed as "sub-criteria" of "overall worth" and, therefore, each of these will also receive a numerical weight. In all cases, a single weight will be defined for each such sub-criterion.

2. The numerical weight attached to each sub-criterion will be interpreted as an indication of the perceived relative importance of satisfying that sub-criterion in the context of the higher-level criterion within whose meaning it is alleged to be included. Relative importance means "relative to the other sub-criteria in the set."

3. Relative importance will be reflected in the <u>ratios</u> of any two weights assigned, respectively, to two separate sub-criteria in a given set. It is in such ratios that trade-off rates will be embodied.

4. Weights will be assigned <u>only</u> to sub-criteria perceived as <u>devoid</u> of substantial worth interdependence. A definite procedure will be presented to identify and eliminate sub-criteria displaying substantial worth interdependence.

5. Weights will be restricted to fall within the range of non-negative numbers. This is to indicate that the concept of relative importance possesses only "positive" connotations. Restricting weights to fall within the range of non-negative numbers guarantees that all trade-off rates (i.e., all ratios between pairs of weights) will be non-negative.

6. Theoretically, a weight of zero would be assigned to any sub-criterion in a given set of sub-criteria if and only if satisfying that sub-criterion were perceived as completely unimportant. In practice, however, a sub-criterion to which a zero weight might appropriately be attached will be ignored (i.e., such a sub-criterion will <u>not</u> be included in the hierarchical criterion structure), since, by the above definition, its satisfaction is viewed as totally unimportant. This definition is included only to provide a logical lower bound to the range of permissible weight numbers and to give the lower bound a definite interpretation.

7. All of the weights in any given weight set (corresponding to a given set of sub-criteria) will add to a finite positive constant, and the same positive

constant will apply to all weight sets. This serves to normalize assigned weights so that a given weight number will always have the same significance (i.e., indicate the same relative importance) in all weight sets. Consequently, the task of validating weight assignments by visual inspection becomes easier.

8. The finite positive constant to which all weights in any given weight set add will be one. Any such constant would be permissible, but setting this number equal to one has a certain conceptual appeal. Since all weights are non-negative and add to one, each weight must lie between zero and one. Hence, relative importance may be viewed as if it were a percentage or proportion, which decision makers may find to be a convenient and familiar conceptual aid.

9. Assigned weight numbers cannot exceed one, and a weight of exactly one will only be assigned in cases where a set of sub-criteria contains a single member. Then, that single sub-criterion must receive full-weight. As such, it must be interpreted as equivalent in the worth sense to its related higher-level criterion.

10. Any positive real number equal to or less than one will be a permissible weight. This will permit the formation of any desirable trade-off ratio by properly selecting pairs of weights.

6.8 Adjusting the Weights to Reflect the Relative Interpretability of Each Physical Performance Measure

Another issue, which has not yet been discussed, concerns the relative extent to which each physical performance measure previously selected to interpret (in physical terms) its associated lowest-level performance criterion does in fact succeed in providing an adequate interpretation thereof.[7] Decision makers might view "expected number of lines printed per minute" as an excellent measure of the lowest-level criterion "printer speed". This is because it reflects very well the intended meaning of "printer speed" within the context of the particular job under consideration. In contrast, "total number of discrete promises" found in a formal proposal submitted by a computer manufacturer to perform that job might be considered a poor measure of "manufacturer's good faith". This is because "manufacturer's good faith" refers to an attitude on the part of corporate executives, and this attitude may not be clearly reflected in the text of their formal proposal. Discussions with executives and review of their historical behavior in similar contractual situations should provide vastly superior measures of their good faith.

To the extent that wide differences emerge in the relative interpretive quality of various performance measures, this could have a seriously distorting impact upon the outcome of a decision. It is quite conceivable that a relatively important criterion (deserving a large numerical weight) cannot be interpreted with any measures of good quality because the decision maker is unable to articulate in explicit physical terms what he means by this criterion. The decision, therefore, should not be unduly influenced by such criteria, especially if other criteria - even though considered relatively less important - are much easier to interpret in terms of high-quality measures. In short, there should be some explicit mechanism for reflecting the relative quality of each criterion's interpretive measure as well as the relative importance of satisfying that criterion. A procedure will be presented in Section 7 to achieve this result.

---

[7] The writer is indebted to H. Martin Weingartner for originally raising and noting the importance of this issue. The writer is also indebted to Howard Raiffa for criticizing constructively the particular manner in which this issue is treated in the assessment procedure.

6.9 Summary

The first step in formal assessment is to define explicitly what is desired in the way of performance from produced alternatives to complete a stated job. This means listing overall objectives or major performance criteria and insuring that the list is:

1. complete (i.e., contains all criteria which decision makers are able and willing to formulate and display);
2. mutually exclusive (i.e., contains criteria which neither encompass nor are encompassed by other criteria on the list);
3. of major significance (i.e., contains only highest-level criteria);
4. free of worth interdependence (i.e., contains only worth-independent criteria).

Having established a list of overall performance objectives, the second step is to generate a hierarchical structure of successively more specific performance criteria. This involves breaking down or subdividing higher-level criteria into one or more lower-level criteria alleged to be included within the meaning thereof.

The third step is to select a single physical performance measure for each lowest-level performance criterion in the hierarchical structure. The purpose of selecting physical performance measures is to establish concrete connections between the hierarchical criterion structure (existing in the subjective minds of decision makers) and the outer world of physical alternatives.

However, merely establishing connections is not sufficient in itself to permit formal evaluation. Specific worth relationships must be mapped out between each lowest-level performance criterion and its related physical performance. This constitutes the fourth step. It is implemented by defining scoring functions which assign a unique worth score in points to every possible value of a physical performance measure. Scoring functions will be defined, either explicitly or implicitly, for every lowest-level criterion.

The fifth and final step is to combine worth scores assigned on the basis of separate performance criteria to arrive at a single, overall index of worth. This is accomplished by defining a weighting function. An additive weighting function with constant trade-off weights will be adopted for this purpose. This requires that sets of sub-criteria located at every branch of the hierarchical tree contain members relatively independent in the worth sense. In addition, weights must be adjusted to reflect the differential interpretive quality of various performance measures.

This completes the outline of the worth measuring procedure. Specific step-by-step means of implementation will be presented in Section 7. Section 8 will illustrate the overall assessment procedure by means of a complete example.

SECTION 7

A SPECIFIC STEP-BY-STEP PROCEDURE FOR ASSESSING WORTH

The procedure to be presented in this section is first of all intended to generate an assessment algorithm. This algorithm is supposed to encapsulate the worth notions of a particular decision maker (or group of decision makers) at a particular point in time with respect to a particular and clearly specified job. Once generated, the algorithm may then be applied to any feasible alternative produced to accomplish that job. Application of the algorithm to any one of the alternatives converts a description of that alternative, in terms of physical performance measures, into a single, overall index of that alternative's worth. It will be well to keep in mind the two-stage nature of the assessment procedure (i.e., first generate an assessment algorithm, and then apply the algorithm to generate a worth measure for each produced alternative). Otherwise, a confused interpretation may very likely result.

It is assumed that the following preliminary steps have been successfully completed prior to embarking upon the assessment procedure.

1. The job for which produced alternatives are being assessed has been adequately described.

2. From the job description a set of mandatory performance (and possibly resource) requirements has been extracted and recorded in physical terms.

3. At least one alternative has been produced.

4. The performance and resource estimates associated with each produced alternative have been validated (i.e., investigated for accuracy and truthfulness).

5. These validated estimates have been checked against stipulated mandatory requirements, and at least one alternative has been shown to be feasible.

6. Any alternatives which failed to satisfy one or more stipulated mandatory requirements have been deleted from further consideration.

7. The residue of feasible alternatives is to be assessed formally.

8. The first step in the formal assessment procedure has been completed, at least tentatively. A list of overall performance objectives has been formulated with reference to the job description and in accordance with the ground rules set forth in Section 6.1.

Let us now proceed to the task of formal assessment.

7.1 A Procedure for Generating Sub-Criteria: Filling Out the Rest of the Hierarchical Criterion Structure

As mentioned previously, it will be most helpful for implementing this and subsequent procedures if a master list of candidate performance criteria and performance measures has been compiled in advance. Although not necessary, experience has shown that reference to such a master list facilitates considerably the essentially creative process of filling out a criterion hierarchy and selecting performance measures. For purposes of discussion, it will be assumed that such a master list exists.

Beginning with one of the overall objectives, we ask what this means in the context of the stated job. To render the discussion concrete, let us return to the example of assessing a computer and select the major criterion, "system performance potential with respect to the currently stated workload and applications". With reference first to the job description, we might decide that the following sub-criteria are all intended by or subsumed under this overall objective:

1. capability to perform certain functions stated in the job description (e.g., process 1000 payroll records) within a stated time limit (e.g., weekly);

2. additional equipment capabilities related to performing the stated job.

Next, with reference to the master list, we might discover a third sub-criterion omitted from the formal job description, but which we also consider to be a part of system performance potential in the stated job context. This third item might be:

3. reasonable reliability in performing the stipulated functions.

If we feel that this more or less exhausts the intended meaning of system performance potential in the stated job context, we can put this aside temporarily and proceed to repeat the process on the three sub-criteria just defined.

Beginning with the first sub-criterion, we again ask ourselves the same question. What does capability to perform certain stated functions mean in the context of the job? We might decide to define the following sub-sub-criteria in response to this question:

1. capability to solve a prepared benchmark problem or set of problems within a certain time limit - preferably in less time;

2. capability to perform certain other standard data processing tasks not reflected within the benchmark exercise, but anticipated within the job environment.

At this point, we might decide that further conceptual subdivision is unnecessary. The time has come to select physical performance measures for each of these two sub-sub-criteria (e.g., observed number of hours required to perform the benchmark problem or problems and observed or estimated number of minutes required to perform a particular mix of standard data processing tasks).

Returning to the second sub-criterion (additional equipment capabilities), we might decide to subdivide this into:

1. additional speed capabilities;
2. additional capacity capabilities;

Then, additional speed capabilities might be further subdivided into:

1. instruction speed;
2. peripheral equipment speed.

For instruction speed, we might select a physical performance measure (e.g., average time in microseconds required to process a single instruction) and then subdivide peripheral equipment speed even further into:

1. card reader speed;
2. card punch speed;
3. tape speed;
4. printer speed.

For these four criteria we might select appropriate physical performance measures (e.g., maximum number of cards per minute, maximum number of characters per second, maximum number of lines per minute), which would close out these branches of the hierarchical tree.

The hypothetical example partially developed in the preceding paragraphs could be carried further, but the general idea should by now be clear. One starts at the highest level of the hierarchy with one of the overall objectives, asks himself what this means, defines one or more sub-criteria in response to this question, and then repeats the procedure with each of the defined sub-criteria. This process continues until it is decided that further subdivision is unwarranted. A physical performance measure is chosen, and that branch of the tree is considered filled out. A retreat is then made back up the tree to the first level containing incomplete branches. The process of successive subdivision is initiated at that point and carried out until another physical performance measure is defined. By so moving up and down the tree, an entire hierarchical structure may be generated. The final signal to stop occurs when no more incomplete branches exist (i.e., when physical performance measures have been attached to every branch of the tree).

Because the process just illustrated is recursive (i.e., because it involves successive reapplication of the same sequence of steps to move up and down the hierarchical tree), only the reiterated sequence of steps need be specified in any great detail to describe completely the entire process. A formal presentation of this reiterated sequence of steps follows immediately.

Step 1. Locate an incompleted branch on the hierarchial tree (i.e., any overall objective or sub-criterion without an attached physical performance measure). At the outset, incompleted branches will occur only at the top level of overall objectives.

Step 2. With reference to the job description and to the master list, decide whether the criterion under scrutiny is to be further subdivided or interpreted directly by means of a physical performance measure. If it is to be further subdivided, proceed to Step 3. If a physical performance measure is to be selected for it, proceed to Step 5.

Step 3. Again, with reference to the job description and to the master list, subdivide the criterion under scrutiny into one or more sub-criteria. That is, decide what sub-criteria are intended by or definitionally subsumed beneath the criterion under scrutiny. Each of these now constitute new incompleted branches of the hierarchy.

Step 4. Choose any one of the sub-criteria defined in Step 3 as a starting point and return to Step 1.

Step 5. With reference to the job description and to the master list, select a physical performance measure judged relevant to the criterion under scrutiny.

Step 6. Move backward up that particular branch of the hierarchy until the first level containing at least one incompleted branch is encountered. If this occurs at other than the top level of overall objectives, choose the incompleted branch (any one of the incompleted branches if more than one exists), and return to Step 1 with this as a new starting point. If no incompleted branches are encountered until reaching the top level, proceed to Step 7.

Step 7. Inspect the top level of the hierarchical tree. If all overall objectives have been satisfactorily specified (i.e., if all branches starting at the top level have been completed), the process is over. A complete hierarchical structure has been constructed. However, if one or more incomplete branches remain, choose any one of those remaining as a starting point, and return to Step 1.

This completes the procedure. A tentative criterion structure has been created and given concrete interpretation by means of the various physical performance measures attached to each of the lowest-level criteria. Subsequent procedures will test this tentative structure for worth interdependence and clarify the process of selecting physical performance measures (see Step 5).

7.2 A Procedure for Identifying Substantial Worth Interdependence Among Performance Criteria in the Hierarchical Structure

The preceding section outlined a procedure for generating lower-level performance criteria intended by or included within the meaning of a higher-level criterion. This procedure was presented in step-by-step form. Step 3 in the procedure is the exact point at which a higher-level criterion is to be so subdivided. The question now is, what guidelines can be provided to aid in this process of subdivision?

Perhaps the best way to answer the question is to look at the final use to which subdivided criteria will be put. After an entire hierarchical worth structure has been formulated and mapped, decision makers will investigate first the set of overall objectives and then each set of sub-criteria. For every such set, they will determine the relative importance of each sub-criterion as a component of its related higher-level criterion. The determined relative importance of each sub-criterion will then be reflected by a numerical weight assigned thereto. Finally, these numerical weights will be used to transform intermediate point scores assigned to the sub-criteria (one score to each sub-criterion) into a single point score to be assigned to their related higher-level criterion.

Now it was pointed out in Section 6.6 that use of an additive weighting function with constant weights is only legitimate when applied to performance criteria judged independent of one another in the worth sense. Therefore, whatever guidelines are developed to aid in the process of subdividing higher-level criteria should certainly include a means of identifying instances of substantial worth interdependence. Two specific questions are presented below to help distinguish worth-independent sub-criteria from those displaying substantial worth interdependence.

1. In comparing a candidate sub-criterion with its related higher-level criterion, which of the following statements better describes the apparent relationship between the two?

   A. The sub-criterion is intended by, included within the meaning of, or an integral part of the higher-level criterion.

   B. The sub-criterion is one alternative means of satisfying the higher-level criterion and important only insofar as it contributes thereto.

2. In comparing one candidate sub-criterion with another sub-criterion already judged as appropriately included within the same set, which of the following statements better describes the apparent relationship between the two?

   A. Willingness to accept reduced satisfaction on either sub-criterion in return for increased satisfaction on the other would not be influenced by the degree of satisfaction already obtained on each.

   B. Willingness to accept reduced satisfaction on either sub-criterion in return for increased satisfaction on the other would depend markedly on the degree of satisfaction already obtained on each.

In order to qualify for final inclusion in the hierarchical structure, every candidate sub-criterion must receive an "A" answer to both of the above questions.

A specific, step-by-step procedure incorporating the above pair of questions appears below. It is intended that a first pass be made at creating a tentative criterion structure by means of the procedure presented in Section 7.1. Then, this procedure may be applied to the candidate sub-criteria generated thereby. An alternative approach would be to perform this testing procedure every time a higher-level criterion is subdivided into a set of sub-criteria (i.e., after Step 3 in the procedure presented in Section 7.1). Either approach would work; however, the step-by-step procedures have been written under the assumption that they will be performed sequentially rather than concurrently.

Step 1. Begin with any set of candidate sub-criteria previously generated in filling out the hierarchical structure (see Section 7.1, Step 3, for the exact point at which a set of candidate sub-criteria is generated).

Step 2. Arrange them in a sequence. It makes no difference how they are arranged - any arbitrary sequence will suffice.

Step 3. Compare the first sub-criterion in the sequence with the higher-level criterion of which all the sub-criteria are alleged to be component parts. Ask which of the following two statements better describes the relationship between the candidate sub-criterion under scrutiny and its related higher-level criterion.

A. The sub-criterion is intended by, included within the meaning of, or an integral part of the higher-level criterion.

B. The sub-criterion is one alternative means of satisfying the higher-level criterion and important only insofar as it contributes thereto.

If statement A is selected as more descriptive than statement B, move to the next sub-criterion in the sequence, and repeat the same question regarding its relationship to the higher-level criterion. Continue in this manner until the entire sequence has been exhausted; then proceed to Step 4. On the other hand, if statement B is selected as more descriptive than statement A, the sub-criterion under scrutiny does not properly belong in the set. Delete this sub-criterion from the set, lay it aside temporarily, and reconsider it later. (Note: Suggested procedures for handling deleted sub-criteria are discussed later.) Move to the next candidate sub-criterion in the sequence and repeat the same question, continuing in this manner until the entire sequence has been exhausted.

Step 4. Select another set of candidate sub-criteria as yet unchecked for worth interdependence, and return to Step 2. If all sets of sub-criteria have been checked, proceed to Step 5.

Step 5. At this point, the entire hierarchical worth structure has been tested (at least partially) for worth interdependence. Quite possibly, some candidate sub-criteria have been deleted and set aside pending subsequent reconsideration. However, it will be useful to check the remaining structure to insure that all sub-criteria are really worth-independent. This can be accomplished by repeating Steps 1 through 4 on the entire hierarchy, but with a new question substituted for the old question in Step 3. A revised form of Step 3 is presented below to facilitate this "second pass" at testing the hierarchy.

Step 3 (revised). Compare every possible pair of sub-criteria in the sequence. (Note: If there are N sub-criteria in the sequence, there are $\frac{1}{2}N^2 - \frac{1}{2}N$ such pair-wise comparisons to be made.) Ask which of the following two statements better describes each pair-wise relationship.

A. Willingness to accept reduced satisfaction on either sub-criterion in return for increased satisfaction on the other would not be influenced by the degree of satisfaction already obtained on each.

B. Willingness to accept reduced satisfaction on either sub-criterion in return for increased satisfaction on the other would depend markedly on the degree of satisfaction already obtained on each.

If statement A is selected as more descriptive than statement B, move to the next pair of sub-criteria, and repeat the same question. Continue in this manner until all of the $\frac{1}{2}N^2 - \frac{1}{2}N$ pair-wise comparisons have been made; then proceed to Step 4. On the other hand, if statement B is selected as more descriptive than statement A, at least one of the sub-criteria in the pair-wise comparison does not properly belong in the set. Move to the next pair of sub-criteria, and repeat the same question. Continue in this manner until all pair-wise comparisons have been made. Then, by inspecting pairs which contain at least one improper member, delete and set aside those sub-criteria which do not belong in the set pending subsequent reconsideration.

This completes the identification procedure. Suggestions for ways of handling candidate sub-criteria which are identified as displaying substantial worth-interdependence will be presented subsequently.

7.3 A Procedure for Selecting Physical Performance Measures to Interpret Lowest-Level Performance Criteria

Let us now move to the task of selecting physical performance measures. After accomplishing sufficient conceptual refinement through successive subdivisions of higher-level criteria into sets of lower-level criteria, a single performance measure must be chosen to interpret concretely each of the lowest-level criteria in the generated hierarchical structure. In essence, our problem is to select for each lowest-level criterion some physically measurable attribute which is perceived as embodying or providing a concrete interpretation of that criterion. Thus, if one lowest-level criterion were job time, a suitable performance measure would be the time in hours required to complete a benchmark exercise. If more than one benchmark exercise were undertaken, the average benchmark time (averaged over the several benchmark exercises) would be even more appropriate. Alternatively, if it were known that a benchmark time understated the true on-site time to perform a stated job, and if the amount by which this true time were understated could itself be estimated, then an estimate of this downward bias could be added to each alternative's recorded benchmark time, and this would constitute the appropriate performance measure.

From the preceding illustrative discussion, the reader may be somewhat disturbed to see that more than one physical performance measure may be applicable to any given lowest-level performance criterion. Furthermore, where more than one performance measure appears applicable, it may not always be obvious which one to choose. In short, judgment on the part of the decision maker must again be exercised to select an appropriate measure just as it was in generating sub-criteria.

Another factor to consider in selecting performance measures is the question of their order (i.e., their degree of generality). An example of an extremely high-order measure would be the observed time (in hours) to complete a benchmark exercise. This would reflect numerous lower-order performance measures such as raw processing times (e.g., add time, multiply time, cycle time, etc.), and other more specific attributes of computer performance.

An example of a moderately high-order performance measure would be overall storage capacity provided either by core, by disk, by drum, or by tape. This high-order measure might be computed by adding the contributions (in bits or words) made by each of the above sources.

Contrast each of these two examples of high-order measures with the case of raw add time (in microseconds). Raw add time is a very basic, low-order measure. It is not easily decomposable into more elementary component measures. More importantly, it is not clear that raw add time would ever be a useful measure in and of itself.

Now the order of a physical performance measure is important for two reasons. First, as illustrated above, high-order measures are generally much more relevant to attempts at assessment than are low-order measures. Consequently, an effort should be made to select and/or concoct high-order measures whenever possible. Second, as illustrated by the overall storage capacity measure, creation of high-order measures out of lower-order measures can sometimes be used as a means of retrieving sub-criteria which have been temporarily deleted from the criterion structure and set aside due to worth interdependence. Such deleted sub-criteria can be replaced by a single higher-level criterion, and a single high-order performance measure can be selected to go with it.

In summary, what guidelines can now be provided for the selection of appropriate physical performance measures? Five guidelines are suggested.

1. Consult the master list to obtain a set of candidate measures.

2. Augment this set by inventing any additional measures not contained in the master list, but which seem appropriate in the context of the lowest-level criterion under consideration and the stated job.

3. Check candidate measures for practical feasibility (i.e., to insure that all data included in the measure can be conveniently and promptly gathered.

4. Attempt to combine candidate measures into higher-order measures, where possible and appropriate.

5. On the basis of judgment, select the seemingly most appropriate and highest-order of the practically feasible candidate measures.

A specific step-by-step procedure incorporating the above five guidelines appears below. It is intended that this procedure be implemented after completing the generating procedures outlined in Section 7.1 and Section 7.2.

Step 1. Begin with any one of the lowest-level performance criteria occurring at the base of the previously generated hierarchical structure.

Step 2. Consult the master list of performance criteria and physical performance measures. Looking only at the physical measures contained in the master list, identify those which are perceived as significantly related to the lowest-level criterion under consideration. This may be done by asking the following question about the relationship perceived to exist between the criterion and every physical performance measure on the master list.

> Would changes in the state or numerical value of the performance measure be capable of bringing about either significant increases or significant decreases in the extent to which the lowest-level criterion under consideration is satisfied?

If the answer to the above question is yes, then a significant relationship is said to exist between the lowest-level criterion and the physical performance measure. If the answer is no, then no such relationship is perceived.

Step 3. Add to the set of physical measures drawn from the master list and perceived as significantly related to the lowest-level criterion any additional measures which can be thought of and which also seem related. In this manner, decision makers can supplement the master list with their own imagination and experience.

Step 4. Looking now at all candidate measures generated by Steps 2 and 3, check to see whether each is practically feasible. That is, insure that all data necessary to form each measure can be conveniently and promptly gathered. Delete any candidate measures which are discovered to be practically infeasible.

Step 5. Inspect the residue of feasible candidate measures remaining after Step 4. Either choose one straightaway as the most appropriate single measure by which to interpret the lowest-level criterion or, if none of the feasible candidates seem really appropriate, attempt to construct a higher-order physical measure out of two or more individual measures.

Step 6. Proceed to another lowest-level criterion in the hierarchical structure, and repeat Steps 2 through 5. Continue in this manner until all lowest-level criteria have been assigned a corresponding physical performance measure.

This completes the procedure.

7.4 A Procedure for Attaching Numerical Weights to Hierarchically Arranged, Worth-Independent Performance Criteria

At this point, it is assumed that a complete hierarchy of worth-independent performance criteria has been constructed. It is also assumed that every lowest-level criterion in this hierarchy has been given a concrete interpretation by assigning to it a physical performance measure. It now remains to attach weights to all criteria in the hierarchy and to specify (by defining scoring functions) the precise worth relationships perceived as linking each lowest-level criterion to its assigned performance measure. The first of these two tasks will be performed in this section. The second will be performed in the next section.

The weight-setting procedure to be developed herein is divided into two sequential phases. In the first phase, an individual decision maker attempts to produce his own numerical weights corresponding to each of the sub-criteria contained in some specified set of sub-criteria appearing in the hierarchical structure. In the second phase, individual weight sets assigned by separate decision makers are compared, and lack of consensus among decision makers (if there are more than one) is resolved by an averaging technique.

The first phase of the procedure involves two major operations.

1. All sub-criteria subsumed under a given higher-level criterion are ranked in order of ascending perceived importance.

2. Then, starting with the most important pair of sub-criteria appearing at the head of the list, successive pair-wise comparisons are made between contiguous sub-criteria, and decision makers are asked to indicate in terms of a ratio the degree of perceived relative importance of the two. Stated alternatively, decision makers are asked to indicate the rate at which they would be willing to accept reduced satisfaction of one sub-criterion in return for increased satisfaction of the other.

A step-by-step procedure to implement this first phase follows immediately. The resulting individual weights generated by this procedure are all positive, they sum to one, and they are interpretable in accordance with the weighting conventions stipulated in Section 6.7. However, one word of warning seems appropriate. Although this procedure (if it can be carried out at all) guarantees that the resultant weights will possess certain desirable logical properties (i.e., consistency, transitivity, and preservation of the preselected importance ratios), the validity of the weights themselves still remains the responsibility of informed judgment on the part of decision makers. Neither this procedure nor any other procedure based solely on logical considerations can guarantee their validity. Only clearly articulated judgment can ever provide that.

Step 1. Begin with any set of sub-criteria subsumed under a higher-level performance criterion.

Step 2. List these sub-criteria in approximate order of relative importance, starting with the most important sub-criterion at the top of the list and the least important sub-criterion at the bottom. It is not necessary to have the sub-criteria perfectly ranked or ordered on this first pass, since subsequent operations will be performed to guarantee complete ordering.

Step 3. Compare the first two sub-criteria on the list.

a. If the first sub-criterion is deemed relatively more important than the second, proceed directly to Step 4.
b. If both sub-criteria are deemed roughly equal in importance, proceed directly to Step 4.
c. If the second sub-criterion is deemed relatively more important than the first, invert their positions on the list (i.e., place the first sub-criterion where the second used to be on the list, and vice-versa), and then proceed to Step 4.

Step 4. Compare the lower-ranked sub-criterion from Step 3 with the next sub-criterion on the list. Repeat the comparisons and stipulated operations in Step 3 on this new pair of sub-criteria. Continue in this manner all the way down to the end of the list until pair-wise comparisons have been made between all contiguous criteria.

Step 5. After the list has been completely exhausted, go back and determine whether any inversions (position changes) occurred.

a. If none occurred, proceed directly to Step 6.

b. If one or more occurred, return to the head of the list, and repeat the entire procedure described in Steps 3 and 4.

Step 6. Eventually, the list will become so arranged that successive pair-wise comparisons will generate no inversions. It may require several passes to achieve this result, but it will occur in the end (assuming that the decision maker's notions of relative importance among sub-criteria are both consistent and transitive). When the list has achieved an arrangement wherein no inversions occur, it will then reflect the decision maker's judgments of relative importance in terms of direction, but not yet in terms of magnitude. Relative magnitudes are determined by subsequent steps.

Step 7. Take the first sub-criterion on the rearranged list, and assign to it the number 1.0 or one hundred percent.

Step 8. Compare the second sub-criterion with the first, and assess their relative importance in terms of a ratio or fraction. That is, if satisfying the second sub-criterion seems only one-half as important as satisfying the first, assign the fraction 1/2 or its decimal equivalent .5 to the second sub-criterion. In like manner, fractions such as 3/4, 9/10, etc., or their decimal equivalents might equally well have been assigned. (Note: It may be difficult to set weights when the question is phrased in the above manner. An alternative form of the same question would be, "At what rate would reduced satisfaction of the first sub-criterion be acceptable in return for increased satisfaction of the second so as to maintain the same overall worth, considering satisfaction of both sub-criteria jointly?" The answer to this question, expressed in the form of a ratio, may then be assigned as before to the second sub-criterion.)

Step 9. Compare the second and third sub-criteria, assess their relative importance or trade-off rate in terms of either a fraction or a ratio, multiply the number assigned to the second sub-criterion by this fraction or ratio, and assign the resultant product to the third sub-criterion. For example, assuming that the second sub-criterion were assessed as being 1/2 as important as the first, while the third were assessed as being 9/10 as important as the second, the appropriate computation would be 1/2 x 9/10, and the number 9/20 would be assigned to the third sub-criterion.

Step 10. Repeat the above procedure for all successive pair-wise comparisons until the list of sub-criteria has been completely exhausted. Then, each sub-criterion will have been assigned a number equal to the product of its importance relative to the next higher sub-criterion times the number previously assigned to the next higher sub-criterion.

Step 11. Add the numbers assigned to all sub-criteria on the list, and then divide each one by the computed sum. This will serve to convert relative importance ratios into normalized weights. Each weight will be positive, and the whole set will add to one. In addition, the relative importance ratios will be preserved in the ratios of any pair of weights.

This completes the procedure.

Now there may not always be complete agreement among separate decision makers concerning the proper collection of weights to be attached to any set of sub-criteria. In fact, numerical differences, and perhaps even rank-order differences, are to be expected among separate decision makers - particularly if they set weights without first consulting one another. This lack of consensus would seem quite healthy, in the writer's opinion, and should be encouraged rather than discouraged. Unless any single decision maker is willing to claim that his weights are precisely correct and, therefore, that anybody who disagrees with him is necessarily wrong, then some method for combining group opinion would seem appropriate.

One way of combining group opinion would be to subject differences of opinion to open discussion in hopes of achieving greater consensus. This would be a particularly effective remedy for those situations where some decision makers possess greater knowledge and experience than others. By open discussion, the less knowledgeable and less experienced decision makers could benefit from their better endowed compatriots and thereby gain a sounder basis for assessment.

However, open discussion would _not_ be effective against genuine differences of opinion held by equally knowledgeable and equally experienced decision makers. _Nor_ would it be effective against whatever differences remain after open discussion has enlightened those decision makers who did not possess initially the same knowledge and experience as others, but who altered their opinions somewhat in the face of ensuing discussions. Some sort of compromise procedure would seem appropriate in these two instances.

One way of achieving a compromise would be by averaging individual weights across separate decision makers. That is, to each sub-criterion in a particular set, separate decision makers would assign their own individual weights. Then, an average weight would be computed for each sub-criterion by adding the weights assigned by separate decision makers and dividing the total by the number of decision makers. It can be shown that, if this averaging procedure is applied to each sub-criterion in a set, then the computed average weights assigned to each of the sub-criteria will sum to one. In addition, the resultant average weights would reflect group opinion instead of one single individual's opinion.

In actual practice, both of the above procedures would seem appropriate, if carried out in sequence. First, a group of decision makers would meet to discuss the relative importance of sub-criteria in some designated set. By open discussion, all decision makers would be accorded a similar basis for formulating their own individual opinions. Then, each decision maker would reflect his individual opinion in a set of numerical weights (generated via the step-by-step procedure just presented). Finally, remaining differences of opinion would be handled by averaging over individual weight assignments to arrive at a final set of group weights for each of the sub-criteria. In this manner, spurious differences of opinion arising from differences in knowledge and experience would be minimized, while genuine differences of opinion arising from genuinely different views of the situation would be adequately reflected in the final average weights.

A step-by-step procedure to implement the second phase of the weight-setting process, designed to average out remaining differences of opinion, is presented below.

Step 1. Collect whatever individual weights have been assigned by separate decision makers to a set of sub-criteria in the hierarchical structure.

Step 2. Suppose that there are N separate decision makers and M sub-criteria in the set to which individual weights have been attached. (Note: Both N and M are assumed to be greater than one. If N = 1, there would be no problem of lack of consensus. If M = 1, there would be only one sub-criterion in the set, and it would therefore have to receive a full weight of 1.0.)

Step 3. Lay out the individual weights assigned by separate decision makers in N parallel cclumns of M weights each. The resulting rectangular array may be though of as a matrix with M rows and N columns.

Step 4. Compute and record the sum of the weights appearing in each of the M rows of the above matrix. (Note: If it is considered desirable to weight some opinions more heavily than others, compute an appropriately weighted sum.)

Step 5. Divide each computed row sum by N. This gives an average weight, averaged across the N separate decision makers, for each of the M sub-criteria. (Note: The M average weights must add to one - except, perhaps, for small rounding errors. If they do not add to one, check the computations for algebraic errors.)

This completes the procedure.

7.5 A Procedure for Establishing Independent Scoring Functions to Link Lowest-Level Performance Criteria to their Assigned Physical Performance Measures

The next task is to formulate scoring functions by which each lowest-level performance criterion may be linked to its assigned measure of physical performance. Once formulated, these scoring functions may be used to convert measured physical performance into equivalent worth scores, and these worth scores may then be combined via the pre-established weights into a single numerical index indicating the overall worth of a proposed alternative.

The scoring procedure itself will be broken down into two major phases. The first phase will contain an ordered sequence of questions designed to determine the general nature and shape of whatever scoring function is to be formulated. The nature and shape of each scoring function will be inferred from answers to the following questions:

1. Is the physical performance measure to be scored discrete or continuous?

2. If discrete, how many measurement categories are contained in the physical performance scale; is there any inherrent order or sequence built into this scale; and are there any qualitative distinctions to be made concerning observations within each measurement category?

3. If continuous, is the physical performance scale bounded from above and/or from below?

4. If bounded, where do the boundaries of the physical performance scale fall?

5. With which points on the physical performance scale are zero and one hundred percent satisfaction of the related lowest-level performance criterion associated, respectively?

6. Does satisfaction increase or decrease with increases in measured performance?

7. Does the rate at which satisfaction increases or decreases with increases in measured performance ever change, or does it remain constant over the entire range of the physical performance scale?

8. If the above rate changes, does it always increase, or does it always decrease, or does it both increase and decrease over selected intervals within the range of the physical performance scale?

The second phase will contain a step-by-step procedure designed to select a specific scoring function of the general nature and shape indicated in the first phase. Actually, two alternative procedures will be presented to implement this second phase - one involving visual and graphic methods, the other involving numerical methods. The choice between these two alternative procedures will be left up to the discretion of decision makers.

7.5.1 Phase I of the Scoring Procedure: Determining the General Nature and Shape of a Scoring Function

The general nature and shape of a scoring function will be determined by answers to an ordered sequence of questions. In all, there are twelve possible questions which might be raised, but not every one of the twelve will be relevant to formulating any given scoring function. Consequently, all twelve questions will first be stated and clarified by example. Then, a step-by-step questioning procedure will be presented to indicate which of the twelve questions are relevant to any given scoring function.

Of the twelve questions, six refer to the scale of the physical performance measure itself for which a scoring function is to be formulated. These first six questions will appear, along with illustrative examples, in Section 7.5.1.1. The remaining six questions refer to relationships presumed to exist between the worth scale (calibrated in points ranging from zero to one) and the scale of the physical performance measure (calibrated in whatever physical units are appropriate). These six questions will appear, along with illustrative examples, in Section 7.5.1.2. Finally, Section 7.5.1.3 will present a step-by-step questioning procedure designed to carry decision makers in an orderly fashion to a final conclusion concerning the general nature and shape of the scoring function to be formulated.

7.5.1.1 Questions About the Scale of the Physical Performance Measure

Question 1. Consider the scale of the physical performance measure. Is it continuous or is it discrete? If all conceivable numbers within the range of the physical performance scale are possible measurements, then the scale is continuous. An example of a continuous performance scale would be "benchmark time." This scale is bounded from below by zero (i.e., it is logically impossible to record a negative benchmark time), but it does not possess any definite upper bound. Therefore, the logical range of this scale falls between zero and positive infinity. Any conceivable number of hours and portions thereof within this range constitutes a possible benchmark observation. For this reason, "benchmark time" possess a continuous or everywhere-dense scale.

On the other hand, if the physical performance measure can assume only certain specified values within its logical range, then the scale of that measure is discrete. An example of a performance measure with a discrete scale would be "number of on-call maintenance personnel provided by a computer manufacturer." This measure can only assume positive integral values or zero. Fractional values cannot occur, since human beings exist only in integral numbers. (Note: One might argue that a maintenance man can spend a portion of his time being on-call, but then the performance measure would be "number of on-call maintenance man-hours provided by a computer manufacturer." This measure possesses a continuous scale, but it is a different measure.)

Question 2. For a physical performance measure possessing a discrete scale, how many discrete categories or levels are included within its range? There must be at least two such categories; otherwise, the performance measure could never discriminate among alternatives (i.e., every alternative would fall in the same category, if only one category were defined, and in no category, if no categories were defined). However, it is possible for a discrete scale to possess two, three, or any higher integral number of categories or levels.

An example of a performance measure possessing a two-level discrete scale would be "provision of some specified analytical software routine" such as a canned regression program. Such a routine is either present or absent from any alternative.

An example of a performance measure possessing a five-level discrete scale would be "type of memory device" whose scale categories might include:

1. core
2. disk
3. drum

4. some combination of the above

5. other

Our previous example of "number of on-call maintenance personnel provided by a computer manufacturer" illustrates a many-level discrete scale. Here, any positive integer or zero constitutes a possible measurement value.

Question 3. Does the physical performance measure possess a scale which is neither purely discrete nor purely continuous, but is some hybrid or combination of the two? In other words, does the performance scale display discrete and continuous properties simultaneously? An example of such a hybrid measure would be "expected number of on-call maintenance personnel provided by a computer manufacturer." Here, a manufacturer might promise to provide no on-call maintenance men, or he might promise to provide at least one, but with the stipulation that such personnel provided would not always answer a call. If the manufacturer promises to provide at least one maintenance man on this conditional basis, and if the relative frequency with which each conditional man provided will in fact answer a call can be estimated, then the expected number of on-call maintenance personnel provided by that manufacturer may be computed by multiplying his conditional number provided by their estimated relative frequency of answering a call. The discrete aspects of this scale derive from the fact that a positive integral number or zero maintenance men may be promised, but perhaps only on a conditional basis. The continuous aspects of the scale derive from the fact that, where uncertainty exists concerning whether or not a call will be answered, this uncertainty is reflected by concocting a frequency-weighted average or expected value as a measure.

Question 4. For a physical performance measure possessing a discrete scale, do the various categories or levels contained within the range of that scale fall into some natural order or sequence, or are they strictly nominal in character? Our previous example of the "number of on-call maintenance personnel provided by a computer manufacturer" also illustrates this concept of natural ordering. Since zero men is less than one man is less than two men, and so forth, the discrete levels contained within this scale do fall into a natural sequence.

In contrast, our previous example of "type of memory device" does not possess a scale whose categories fall into any natural order. These categories are strictly nominal in character. The only thing we can say about core, disk, drum, some combination, and other is that they are different categories. There is no meaningful physical sense in which core is either less than or greater than disk. They are just different. The same conclusion may be drawn from all other pair-wise comparisons between scale categories associated with this performance measure.

Based on these contrasting examples, a more formal definition may be given for the conceptual distinction between strictly nominal discrete scales and naturally ordered discrete scales. If, by inspecting the discrete categories contained within the range of the scale, "greater than" and "less than" relationships are naturally defined in the physical sense, then these categories fall into a natural order, and the scale is said to be an ordered scale. If, on the other hand, inspection of the component scale categories permits only "same as" and "different from" distinctions, then these categories are strictly nominal in character, and the scale is said to be a nominal scale.

One word of caution seems appropriate before leaving the distinction between nominal and ordered scales. This distinction does not refer to the worth imputed by decision makers to various scale categories. Thus, core storage may be preferred to disk storage as a memory device due to the former's faster access time. However, the ordering here is a preference ordering imputed by decision makers, and it derives not from the physical identity of the two kinds of memory devices per se, but rather from an associated characteristic - access time. The only physical thing we can say about the sheer identity of core storage devices versus disk storage devices is that they are different. If we were concerned with memory devices because of their access time, then we should define "access time" directly as our performance measure, and we should ignore the nominal identity of alternative types of memory devices which provide this access time. Under these circumstances, "access time" would constitute a performance measure whose scale values fall into a natural order, but this is not the same measure as "type of memory device." On the other hand, where a performance measure with a discrete, but only nominal scale is appropriate (e.g., "presence or absence of some specified analytical software routine"), the existence of a preference ordering imputed by decision makers (presence is preferred to absence) must be clearly distinguished from the existence of a natural ordering inherrent in the physical measure itself (presence is just different from absence in the sheer physical sense).

Question 5. For a physical performance measure possessing a continuous scale, can either a logical lower bound or a logical upper bound or both be identified with that scale? Let us first consider the issue of a logical lower bound. Most continuous physical performance scales are defined in such a way as to exclude the possibility of negative measurements. It is impossible, for example, to conceive of performing a benchmark exercise in negative time or of acquiring a computer with one or more negative processing speeds. This means that most continuous performance scales do possess definite lower bounds and that such bounds fall either at zero or at some positive number. In light of this, and because negative performance measures introduce certain conceptual and computational complications, it will be desirable to insure that no performance scales can ever include negative measurement numbers. A procedure for accomplishing this objective will be presented later.

Having established by fiat that all continuous performance scales will be restricted in range to exclude negative measurement numbers, we have, in effect, established the existence of a lower bound. All such scales must have lower bounds falling either at zero or at some positive number. The next question is to determine whether zero or some positive number constitutes that lower bound.

In the vast majority of cases, logical lower bounds will fall naturally at zero. "Benchmark time," "slack time between benchmark time and some mandatory maximum daily, weekly, or monthly operating time," and various types of "processing speeds" all possess continuous performance scales with logical lower bounds falling exactly at zero. It is logically impossible to receive proposal data containing negative observations on any of these scales, but it definitely is possible to receive zero and various positive observations on all of them.

In some cases, however, logical lower bounds will occur at strictly positive numbers. Nevertheless, even in these (rare) instances, it will be convenient to re-scale the performance measure such that its logical lower bound falls exactly at zero. Re-scaling procedures will be presented later to accomplish this end.

Question 6. For a physical performance measure possessing a continuous scale, can a logical upper bound be identified with that scale, and, if so, at what measurement number does this upper bound fall?

An example of a physical performance measure possessing a continuous scale lacking any logical upper bound is "benchmark time." It is conceivable that any positive number of minutes or hours could be recorded from a benchmark exercise. This means, in effect, that the scale of "benchmark time" extends from zero to positive infinity.

Now what happens if some mandatory upper limit on "benchmark time" has been stipulated? Suppose, for example, that daily workloads must be processed within a single eight-hour shift in order that an alternative may be considered feasible. Does eight hours constitute a logical upper bound to "benchmark time"? No. It does not. Mandatory performance requirements are set by decision makers to reflect maximum or minimum levels of acceptable performance. They do not necessarily reflect the bounds of logically possible performance.

Alternately, it might be argued that benchmark times will very likely fall in a range from one or two hours to ten or twelve hours. However, twelve hours would not constitute a logical upper bound in this case either. It is conceivable, although unlikely, that some observed benchmark time would exceed twelve hours. In fact, no matter where we attempt to set an upper bound, so long as we set it at some finite positive number, it is still conceivable that at least one observed benchmark time might exceed it. The higher we set the upper bound, the more confident we can be that all observed benchmark times will fall below it; but we can never be absolutely certain. It is for this reason that the scale of "benchmark time" must be regarded as unbounded. Any positive number, no matter how unlikely, is still logically possible.

7.5.1.2 Questions About Relationships Presumed to Exist Between the Worth Scale and the Scale of the Physical Performance Measure

Question 1. With which levels on the scale of the physical performance measure are zero worth points and one worth point to be associated, respectively? Stated alternatively, what level of physical performance is to be regarded as completely unsatisfactory, and what level is to be regarded as completely satisfactory?

This question is easy to answer in the case of performance measures with discrete, two-level scales. One level must be completely unsatisfactory, and the other must be completely satisfactory.

Thus, absence of an analytical software routine would be completely unsatisfactory (receiving zero worth points), while its presence would be completely satisfactory (receiving one worth point).

In the case of performance measures with discrete, many-level scales, it depends upon whether the discrete scales are strictly nominal or ordered. If strictly nominal, then the question generally cannot be answered at all. A procedure to handle this type of situation will be presented later. However, if the discrete, many-level scale is ordered, then the setting of zero and one hundred percent satisfaction levels (i.e., zero and one worth point) depends upon the range (i.e., placement of logical lower and upper bounds) of the physical scale. The same is true of performance measures with continuous scales.

It has been established by fiat that all continuous performance scales will possess logical lower bounds falling exactly at zero. In most cases, this occurs naturally, but where exceptions arise, remedial procedures will be employed to make this occur. Similarly, discrete, ordered, many-level performance scales will be made to have logical lower bounds falling exactly at zero - if this does not occur naturally. The only remaining question, then, is whether or not such scales have logical upper bounds. Three examples will be drawn to illustrate the fitting of zero and one hundred percent satisfaction levels to these types of performance scales. Two examples will involve scales without logical upper bounds, and one will involve a scale with a definite logical upper bound.

Consider, first, the case of "benchmark time." The scale of this performance measure is bounded from below by zero, but is unbounded from above. Clearly, an observed benchmark time of zero (minutes, hours, days, etc.) would be one hundred percent satisfactory. Equally clearly, an infinitely large observed benchmark time would yield zero satisfaction. Intermediate levels of benchmark time would yield intermediate degrees of satisfaction. Consequently, we could associate a worth point score of one with zero benchmark time and a point score of zero with infinite benchmark time. But is this not a rather extreme - even peculiar - way to set zero and one hundred percent satisfaction levels? No real benchmark observations will fall exactly at zero time or anywhere near infinite time. In addition, we would be "almost" one hundred percent satisfied with benchmark times substantially above zero, and we would regard as "almost" completely unsatisfactory large benchmark times falling far below positive infinity. So why set such extreme limits? The answer to this question is that we should set as extreme limits as possible both to avoid the requirement of predicting the range of actual benchmark observations (which predictions are subject to error) and to recognize the fact that variations in benchmark time at both extremes of the time scale do yield variations in satisfaction - even though very slight. The scoring procedures to be presented in Appendices at the end of this paper will permit making variations in worth point scores as small as desired at the extremes of the performance scale; but we might as well construct scoring functions which are flexible enough to encompass all logically possible situations. This costs next to nothing in effort and may very well provide substantial savings in terms of future headaches.

Consider, next, the case of "number of on-call maintenance personnel provided by a computer manufacturer". This performance measure possesses a discrete, ordered, many-level scale bounded from below by zero, but unbounded from above. Zero satisfaction would be derived from zero maintenance personnel, and one hundred percent satisfaction would be derived from an infinite number of such personnel. Once again, we would set these end-points of the worth scale at the most extreme possible levels of the performance scale to insure logical completeness. Zero worth points would be assigned to zero maintenance personnel, and one worth point to infinite personnel.

As a third example, consider the case of "slack time between benchmark time and some mandatory maximum daily, weekly, or monthly operating time." The scale of this performance measure is continuous and doubly bounded between zero and the mandatory maximum time. Zero worth points would be assigned to zero slack time, and one worth point would be assigned to the maximum possible level of slack time.

In summary, fitting the end-points of the worth scale to a physical performance scale is usually a simple job of matching. Either zero worth points or one worth point is assigned to zero performance (depending upon the performance scale), and the other end of the worth scale is assigned either to the logical upper bound of the performance scale (if such an upper bound exists) or to infinite performance.

Question 2. For physical performance measures possessing either continuous scales or discrete, ordered, many-level scales, what is the direction of the preference relationship? Does a higher level of physical performance yield greater satisfaction, or does it yield less satisfaction? Additionally, is the direction of the preference relationship uniform over the entire logical range of the physical performance scale?

The first part of this question is very easy to answer. A higher observation on "benchmark time" is obviously less satisfactory and, therefore, deserving of a lower worth score than is a lower observation on "benchmark time." This same type of reverse preference relationship characterizes most physical performance measures expressed in terms of time required to do something. It is generally preferable to have things accomplished in a shorter rather than in a longer time.

In contrast, a direct preference relationship exists in the case of "slack time," "number of on-call maintenance personnel provided by a computer manufacturer," "efficiency of compilers," "number of machine reliability checks," and many other physical performance measures. In all of these cases, more physical performance is considered better than less performance. This type of direct preference relationship characterizes most capacity measures as well as many other operating measures.

The second part of the question is also easy to answer in most cases. Almost all physical performance measures have associated preference relationships which remain uniform in direction over the entire range of their physical scales. Thus, if a direct preference relationship exists in one region of a physical performance scale, it will generally hold in all other regions - although, perhaps, to a greater or lesser extent. If one reliability check is preferred to zero reliability checks, then one hundred such checks will be preferred to ninety-nine, and so on up the line. Similarly, reverse preference relationships usually hold uniformly over the entire range of performance scales. If a benchmark time of four hours is preferred to a time of five hours, then four minutes would be preferred to five minutes, and four days would be preferred to five days.

Occasionally, a physical performance measure may appear with a non-uniform preference relationship. That is, the preference relationship may change direction at some point or points within the range of the physical performance scale. For a while, more performance may be preferred to less; but after some point, less performance may be preferred to more. An example of this sort of situation occurs when one scratches an itching portion of his skin. For a while, continued scratching is preferable to discontinued scratching; but if the scratching process is continued too long or too intensively, it is preferable to scratch more lightly or to discontinue the process altogether. Special provisions will have to be taken in such instances, since the scoring procedures to be presented herein are not intended to cover them.

Question 3. For physical performance measures with uniformly directed preference relationships, how does the rate of increase (or decrease) of worth behave with increases in physical performance? Does the rate of change of worth remain constant over the entire range of the performance scale, or does this rate vary over different regions of the scale?

To clarify this question, let us define more explicitly what is intended by the notion of a rate of change of worth. First, we know that worth should change over various regions of a physical performance scale; otherwise, there would be no point in scoring the physical performance measure. Every possible level of performance would receive the same worth score under these circumstances, and such a measure would fail to discriminate among competing alternatives. Second, given that worth does change with performance, it is important to know the direction of change (i.e., the direction of the preference relationship discussed in Question 2). But knowing only the direction of change is not sufficient. We must also know the rate of change. Does worth change rapidly with changes in performance, or does it change only slowly. Finally, we must also know the acceleration of change. That is, does worth change with performance at a constant rate or at a variable rate (i.e., either at an accelerating rate or at a decelerating rate)?

The question posed here is about the acceleration of change in worth with increases in performance. If worth changes at a constant rate with increases in performance, then there is no acceleration or deceleration inherent in the preference relationship. A linear or straight-line relationship with a constant slope (indicating a constant rate of change) would be selected to depict this sort of situation. Equal changes in performance always receive the same increase or decrease in worth score throughout the entire range of the performance scale. However, if worth changes at a variable rate with increases in performance, there is some degree of acceleration or deceleration inherent in the preference relationship, and a non-linear scoring function would be selected to depict this type of relationship.

An example of a performance measure whose preference relationship displays a constant rate of change would be "slack time between benchmark time and some mandatory maximum daily, weekly, or monthly operating time". Assuming that slack time would be used to increase the workload by merely processing additional records (e.g., additional personnel records), and assuming that the worth of processing each additional record were the same for all records, then a simple, straight-line, direct relationship would exist between slack time and the worth of slack time. Under the above assumptions, the rate of increase of worth with increases in slack time would be constant, since each additional unit of slack time would be utilized to process additional records of equal worth.

An example of a performance measure whose preference relationship displays accelerating changes would be "frequency of machine breakdowns." Here, the direction of the preference relationship would be reverse (i.e., lower frequencies are preferred to higher frequencies), and the rate of this reverse preference increases as the frequency of breakdowns increases. Stated alternatively, the rate of decrease of worth with increasingly frequent breakdowns itself increases. Each breakdown generally requires backing up operations to some point in time prior to the actual moment of breakdown, which often ruins already accomplished work. This is particularly true of multi-processing systems.

An example of a performance measure whose preference relationship displays decelerating changes would be "number of on-call maintenance personnel provided by a computer manufacturer." The direction of this preference relationship would be direct (i.e., the more the better), but the rate of this direct preference decreases as more and more maintenance personnel are provided. The additional worth of each additional man decreases steadily, since there is just so much maintenance work to be done, and, after a point, providing additional personnel does not serve to get the work done either much faster or much better.

Question 4. In the case of either accelerating or decelerating rates of change of worth with increasing performance, does the preference relationship always accelerate or always decelerate, or does it do first one and then the other over different regions of the performance scale?

Our previous example of "number of on-call maintenance personnel provided by a computer manufacturer" illustrates an always decelerating preference relationship. Each additional man is worth somewhat less than his predecessor, and this is true throughout the entire logical range of the performance scale.

On the other hand, "frequency of machine breakdowns" illustrates a preference relationship which, although uniformly reverse, first accelerates and then decelerates. As discussed in Question 3, increases in the frequency of breakdowns serves to reduce the worth of the machine at an accelerating rate. But this does not continue forever. After a while, when the frequency of breakdowns gets high enough, the worth of the machine has already been reduced to such a low level that an even higher breakdown frequency will result in only a little additional deterioration. The situation is already so bad that it cannot get much worse - even if the machine were to shut down completely (an infinite breakdown frequency). This sort of situation would be depicted by a scoring function in the shape of a mirror-image "S".

Question 5. Does a preference relationship with uniformly changing rate always accelerate or always decelerate? The meaning of this question has already been illustrated in discussing Question 4.

Question 6. Does a preference relationship with non-uniformly changing rate first accelerate and then decelerate, or does it first decelerate and then accelerate? The meaning of this question has also been illustrated in discussing Question 4.

7.5.1.3 A Step-by-Step Questioning Procedure

In this section, the twelve questions presented in the two previous sections will be arranged in an ordered sequence. The purpose of providing such a step-by-step questioning procedure is to direct decision makers systematically toward a scoring function of particular nature and shape.

Step 1. Consider the scale of the physical performance measure. Is it continuous or is it discrete? If discrete, proceed to Step 2. If continuous, proceed to Step 7.

Step 2. Is the discrete scale purely discrete or is it a hybrid, containing continuous aspects as well as discrete aspects? If purely discrete, proceed to Step 3. If hybrid, treat it as if it were continuous, and proceed to Step 7.

Step 3. How many categories or levels are contained within the discrete scale identified in Step 2? If two, proceed to Step 4. If three, four, or five, proceed to Step 5. If more than five, proceed to Step 6.

Step 4. If the discrete, two-level scale identified in Step 3 is merely a case of presence or absence of some desirable attribute, proceed to scoring procedure 1 in Appendix I. If presence of the attribute is to be qualified by an additional measure of relative worth, proceed to scoring procedure 2 in Appendix II.

Step 5. Is the discrete scale identified in Step 3 strictly nominal or is it ordered? If strictly nominal, proceed to scoring procedure 3 in Appendix III. If ordered, proceed to scoring procedure 4 in Appendix IV.

Step 6. Is the discrete scale identified in Step 3 strictly nominal or is it ordered? If strictly nominal, proceed to scoring procedure 5 in Appendix V. If ordered, treat the scale as if it were continuous, and proceed to Step 7.

Step 7. Does the continuous scale identified in Step 1, Step 2, or Step 6 possess a logical lower bound? If yes, proceed to Step 10. If no, proceed to Step 8.

Step 8. It is very unlikely that a performance measure will have been selected whose scale is unbounded from below (i.e., where negative observations are possible and may range all the way to negative infinity). Therefore, ask once again whether the scale under scrutiny possesses a logical lower bound. If the answer is now yes, proceed to Step 10. If the answer is still no, look for a logical upper bound. If the scale possesses no logical upper bound either, the performance measure must be rejected. The scoring procedures presented herein are not equipped to handle doubly unbounded performance scales. Choose a new performance measure, and return to Step 1. However, if the scale does possess a logical upper bound, proceed to Step 9.

Step 9. Transform the scale identified in Step 8 by multiplying every number contained therein by minus one. This transformed scale will now possess a logical lower bound, but no logical upper bound. Proceed to Step 10, but keep in mind that the new transformed scale is just the reverse of the original scale. Consequently all subsequent questions about the transformed scale must be answered with this reversed aspect in mind.

Step 10. Does the logical lower bound fall exactly at zero? If yes, proceed to Step 12. If no, proceed to Step 11.

Step 11. Identify the numerical value of the logical lower bound. Transform the scale by subtracting this number from every number contained in the scale. Keep this transformation in mind, and remember that all subsequent questions will refer to the transformed scale. Proceed to Step 12.

Step 12. Does the scale possess a logical upper bound? If yes, proceed to Step 13. If no, proceed to Step 24.

Step 13. It has been determined that the performance scale is bounded from below by zero and from above by some finite positive number. What is the direction of the preference relationship? If direct, proceed to Step 14. If reverse, proceed to Step 19.

Step 14. Now fit the end-points of the worth scale to the logical lower and upper bounds of the performance scale. Assign zero worth points to zero performance and one worth point to the logical upper bound of the performance scale. Proceed to Step 15.

Step 15. Is the direct preference relationship identified in Step 13 uniform over the entire logical range of the performance scale? If yes, proceed to Step 16. If no, the scoring procedures presented herein are inadequate to handle such a performance measure. Define a new performance measure, and return to Step 1.

Step 16. Does the direct preference relationship identified in Step 15 maintain a constant rate of change of worth, or does it display a variable rate of change (i.e., either accelerating, decelerating, or both in sequence)? If constant, proceed to scoring procedure 6 in Appendix VI. If variable, proceed to Step 17.

Step 17. Does the variable rate of change of worth identified in Step 16 display uniform acceleration, uniform deceleration, or first one and then the other? If uniform acceleration, proceed to scoring procedure 7 in Appendix VII. If uniform deceleration, proceed to scoring procedure 8 in Appendix VIII. If first one and then the other, proceed to Step 18.

Step 18. Does the variable rate of change identified in Step 17 start by accelerating and then by decelerating, or does it start by decelerating and then end by accelerating? If the former, proceed to scoring procedure 9 in Appendix IX. If the latter, proceed to scoring procedure 10 in Appendix X.

Step 19. Now fit the end-points of the worth scale to the logical lower and upper bounds of the performance scale. Assign zero worth points to the logical upper bound of the performance scale, and assign one worth point to zero performance. Proceed to Step 20.

Step 20. Is the reverse preference relationship identified in Step 13 uniform over the entire logical range of the performance scale? If yes, proceed to Step 21. If no, the scoring procedures presented herein are inadequate to handle such a performance measure. Define a new performance measure, and return to Step 1.

Step 21. Does the reverse preference relationship identified in Step 20 maintain a constant rate of change of worth, or does it display a variable rate of change (i.e., either accelerating, decelerating, or both in sequence)? If constant, proceed to scoring procedure 11 in Appendix XI. If variable, proceed to Step 22.

Step 22. Does the variable rate of change of worth identified in Step 21 display uniform acceleration, uniform deceleration, or first one and then the other? If uniform acceleration, proceed to scoring procedure 12 in Appendix XII. If uniform deceleration, proceed to scoring procedure 13 in Appendix XIII. If first one and then the other, proceed to Step 23.

Step 23. Does the variable rate of change identified in Step 22 start by accelerating and then end by decelerating, or does it start by decelerating and then end by accelerating? If the former, proceed to scoring procedure 14 in Appendix XIV. If the latter, proceed to scoring procedure 15 in Appendix XV.

Step 24. It has been determined that the performance scale is bounded from below by zero, but that the scale possesses no logical upper bound. What is the direction of the preference relationship? If direct, proceed to Step 25. If reverse, proceed to Step 28.

Step 25. Now fit the end-points of the worth scale to the performance scale. Assign zero worth points to zero performance and one worth point to infinite performance. Proceed to Step 26.

Step 26. Is the direct preference relationship identified in Step 24 uniform over the entire logical range of the performance scale? If yes, proceed to Step 27. If no, the scoring procedures presented herein are inadequate to handle such a performance measure. Define a new performance measure, and return to Step 1.

Step 27. The following facts have been ascertained concerning the nature and shape of the scoring function for this performance measure.

1. The worth scale is bounded between zero and one (by convention).

2. The physical performance scale is bounded from below by zero, but it possesses no logical upper bound.

3. The preference relationship is uniformly direct over the entire range of the performance scale.

From these three facts, we must conclude that both a constant rate of change of worth and a uniformly accelerating rate of change of worth are logically impossible. The only remaining possibilities are uniform deceleration or initial acceleration followed by deceleration. If uniform deceleration, proceed to scoring procedure 16 in Appendix XVI. If initial acceleration followed by deceleration, proceed to scoring procedure 17 in Appendix XVII.

Step 28. Now fit the end-points of the worth scale to the performance scale. Assign one worth point to zero performance and zero worth points to infinite performance. Proceed to Step 29.

Step 29. Is the reverse preference relationship identified in Step 24 uniform over the entire logical range of the performance scale? If yes, proceed to Step 30. If no, the scoring procedures presented herein are inadequate to handle such a performance measure. Define a new performance measure, and return to Step 1.

Step 30. The following facts have been ascertained concerning the nature and shape of the scoring function for this performance measure.

1. The worth scale is bounded between zero and one (by convention).

2. The physical performance scale is bounded from below by zero, but it possesses no logical upper bound.

3. The preference relationship is uniformly reverse over the entire range of the performance scale.

From these facts, we must conclude that both a constant rate of change of worth and a uniformly accelerating rate of change of worth are logically impossible. The only remaining possibilities are uniform deceleration or initial acceleration followed by deceleration. If uniform deceleration, proceed to scoring procedure 18 in Appendix XVIII. If initial acceleration followed by deceleration, proceed to scoring procedure 19 in Appendix XIX.

This completes the ordered sequence of questions designed to determine the general nature and shape of the scoring function.

7.5.2 Phase II of the Scoring Procedure: Determining a Specific Scoring Function of the General Nature and Shape Determined in Phase I

Twenty individual scoring procedures have been developed to define a unique scoring function to reflect a decision maker's worth judgments. In Phase I of the overall procedure, the general nature and shape of this function was determined by proceeding through an ordered sequence of questions. On the basis of answers given to these questions, one of the first nineteen procedures in Phase II is to be selected. The steps contained in the twentieth procedure are common to many of the other nineteen. Therefore, it has been segregated to economize on space.

The reader is hereby referred to Appendices I through XX for these twenty individual scoring procedures.

7.6 A Procedure for Adjusting the Weights to Reflect Differential Interpretive Quality Among the Physical Performance Measures

The last step in formulating an assessment algorithm is to adjust the weights to reflect differential interpretive quality among the physical performance measures. A step-by-step procedure to accomplish this is presented below.

Step 1. Compute the "effective" weight associated with each lowest-level performance criterion. That is, identify the chain of weights linking each lowest-level criterion to the apex of the hierarchy, and compute the product of all weights in this chain. Then, each of the "effective" weights associated, respectively, with one of the lowest-level criteria will be positive, and they will sum to one.

Step 2. Now consider the relationship between each lowest-level criterion and its associated physical performance measure. Recalling the scoring function which has been defined for each of these linked pairs, assess the extent to which the performance measure serves to interpret, through its scoring function, the intended meaning of the lowest-level criterion. Assess its interpretive quality on a percentage scale, where zero means that the performance measure bears no relation at all to the performance criterion, and one hundred percent means that the performance measure interprets perfectly the intended meaning of that criterion.

Step 3. Assign percentage numbers to each linked pair at the base of the criterion structure.

Step 4. Multiply each "effective" weight by the corresponding percentage number assigned in Step 3.

Step 5. Add the products computed in Step 4.

Step 6. Divide each product computed in Step 4 by the sum computed in Step 5. The result is a set of "adjusted effective" weights.

This completes the procedure.

7.7 A Procedure for Computing Each Alternative's Total Worth Score

The process of formulating an assessment algorithm was completed in Section 7.6. This section describes the second stage in the complete assessment procedure wherein that algorithm is applied to an alternative to generate an index of its overall worth. A step-by-step procedure to accomplish application of the algorithm is presented below.

Step 1. Select one of the feasible alternatives.

Step 2. Select one of the performance measures in terms of which that alternative has been described.

Step 3. Referring to the scoring function associated with that performance measure, convert measured performance into an equivalent worth score.

Step 4. Multiply the equivalent worth score computed in Step 3 by the associated "adjusted effective" weight computed in Section 7.6.

Step 5. Repeat Steps 2 through 5 for all performance measures.

Step 6. Add the products computed in Step 5. The resulting sum constitutes an index of the selected alternative's overall worth.

This completes the procedure.

SECTION 8

A COMPLETE EXAMPLE

Previous sections of this paper have introduced and developed a procedure for assessing worth. The problem of assessing and eventually selecting computing equipment served as a convenient context within which various phases of the procedure could be illustrated. However, the procedure is by no means limited to computers. Quite to the contrary, one of the claimed advantages of the procedure is its generality with respect to type of assessment problem. To demonstrate this generality and, at the same time, to provide an integrated picture of the entire procedure at work, this section will describe its application in a completely different context--in assessing the worth of alternative employment opportunities.

One of the writer's acquaintances, a graduate student at Massachusetts Institute of Technology, became interested in the assessment procedure when he was faced with securing employment directly following graduation.[8] He had already solicited several job offers and, on the basis of preliminary analysis, he had reduced these to a set of four feasible and non-dominated alternatives. It was at this point that he undertook the task of formal assessment.

After reading completely an earlier draft of this paper and obtaining clarification on various methodological points from the writer, he set out to generate a criterion hierarchy, to establish weights, to define scoring functions, to adjust the weights, to assess the four alternative job offers, and, finally, to make a terminal decision. His progress through these sequential steps constitutes the subject matter of this section.

8.1 The Criterion Hierarchy

It would require too much space to present a complete historical record of this individual's progress through the various steps involved

[8] The individual involved has chosen to remain anonymous.

in generating a criterion hierarchy, purging it of worth-interdependent members, and selecting physical performance measures. He made at least four separate passes at creating and revising a hierarchy over a period of several weeks time. What will be presented instead is the end state of this process. The hierarchy of worth-independent criteria and associated performance measures which he finally selected as providing a satisfactory description of his job objectives is described below.

Four overall objectives or highest-level performance criteria were defined:

1. monetary compensation;
2. geographical location;
3. travel requirements;
4. nature of work

Monetary compensation was broken down to include:

1. immediate compensation;
2. future compensation.

Immediate compensation was further subdivided to include:

1. starting salary;
2. fringe benefits, which included -
    (a) insurance benefits;
    (b) retirement benefits.

Future compensation was subdivided to include:

1. anticipated salary in three years;
2. anticipated salary in five years.

His second major objective, geographical location, was broken down to include:

1. proximity to relatives;
2. degree of urbanity associated with the location;
3. climate.

His third major objective, travel requirements, was broken down to include:

1. daily commuting requirements to and from the place of work;
2. extended trips.

Extended trips was further subdivided to reflect:

1. proportion of time away from home;
2. duration of extended trips.

His fourth major objective, nature of work, was broken down to include:

1. immediate training requirements (undesirable if extensive);
2. continuing aspects.

Continuing aspects of the work were further subdivided to include:

1. personal interest in the technical content of the job;
2. degree of variety implicit in the job;
3. amount of training in management skills realizable from the job.

The above hierarchy contained fifteen lowest-level criteria, each one of which was interpreted by defining a single performance measure. These fifteen lowest-level criteria and their associated performance measures were as follows:

1. starting salary - locally adjusted after-tax annual dollars;[9]
2. insurance benefits - locally adjusted after-tax annual dollars;[9]
3. retirement benefits - locally adjusted after-tax annual dollars;[9]

---

[9]All dollar figures were adjusted to account for differences in average living costs associated with different geographical locations in the United States.

4. anticipated three-year salary - locally adjusted after-tax annual dollars;[9]
5. anticipated five-year salary - locally adjusted after-tax annual dollars;[9]
6. proximity to relatives - one way jet flight time in hours;
7. degree of urbanity - Standard Metropolitan Area population;
8. climate - direct worth estimate;[10]
9. daily commuting requirements - one-way travel time in hours;
10. proportion of time away from home - annual percentage;
11. duration of extended trips - maximum trip length in days;
12. immediate training requirements - required training time in months;
13. personal interest in the technical content of the job - direct worth estimate;[10]
14. degree of variety implicit in the job - direct worth estimate;[10]
15. amount of training in management skills realizable from the job - direct worth estimate;[10]

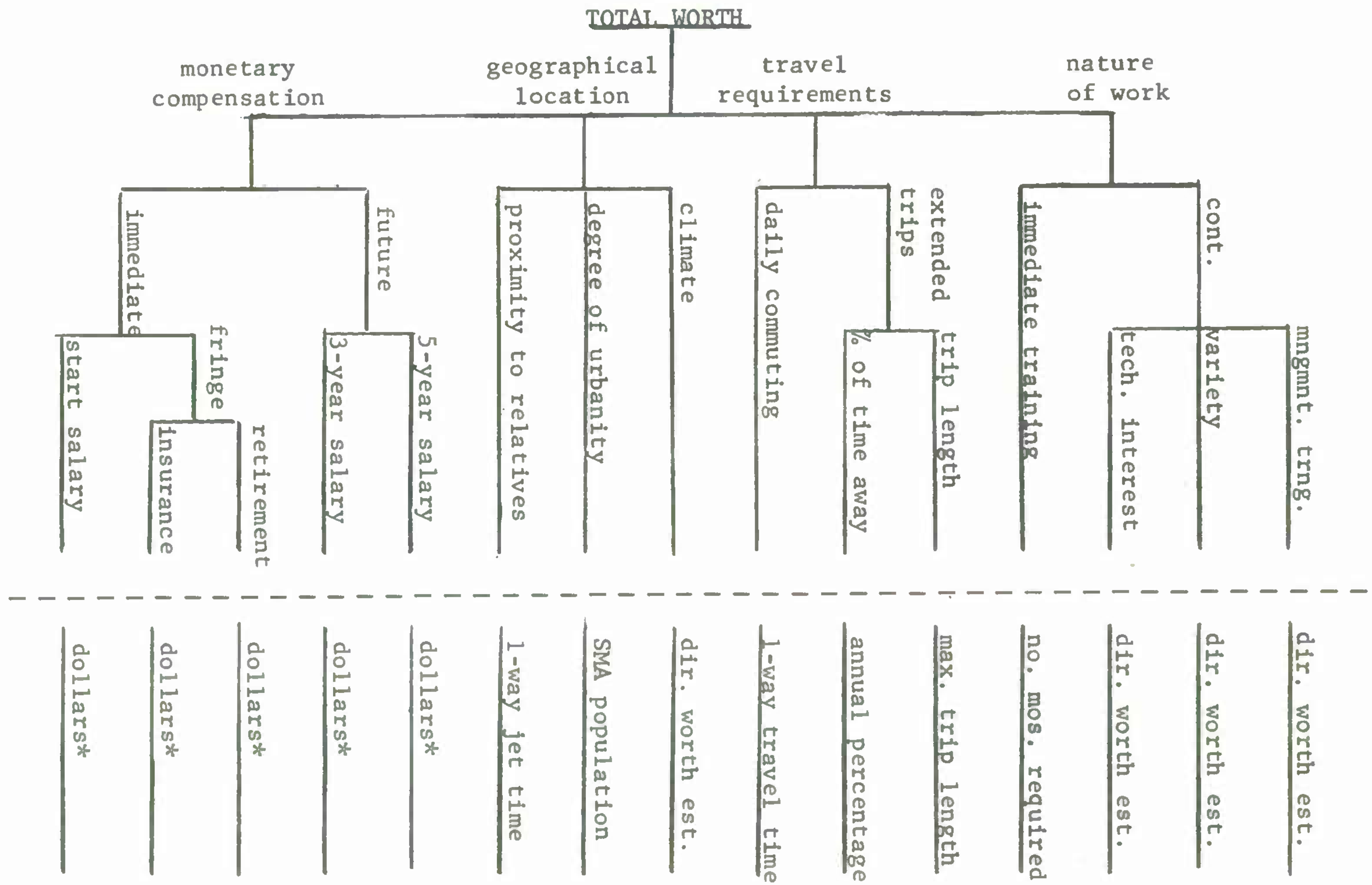
A pictorial display of this criterion hierarchy, complete with performance measures, is shown in Exhibit 1. The dotted horizontal line indicates the region of demarkation between performance criteria and performance measures. The reader will notice that abbreviations are sometimes used in Exhibit 1 to conserve space. However, review of the text should clear up any doubts about the meaning of these abbreviations.

---

[10]A direct worth point score was assigned subjectively to each alternative in this instance.

Exhibit 1

The Criterion Hierarchy



*All dollar figures locally adjusted after taxes.

8.2 The Weights

Numerical weights were then assigned to sub-criteria at every branching point in the hierarchy. For the overall criteria, this process yielded the following weights:

1.	monetary compensation	.33
2.	geographical location	.17
3.	travel requirements	.17
4.	nature of work	.33
	Total	1.00

Within monetary compensation, weights were assigned as follows:

1. immediate compensation			.70
(a) starting salary		.90	
(b) fringe benefits		.10	
(1) insurance benefits	.60		
(2) retirement benefits	.40		
Total	1.00		
Total		1.00	
2. future compensation			.30
(a) anticipated three-year salary		.65	
(b) anticipated five-year salary		.35	
Total		1.00	
Total			1.00

Within geographical location, weights were assigned as follows:

1.	proximity to relatives	.40
2.	degree of urbanity	.40
3.	climate	.20
	Total	1.00

Within travel requirements, weights were assigned as follows:

1.	daily commuting requirements		.20
2.	extended trips		.80
	(a) proportion of time away from home	.40	
	(b) duration of extended trips	.60	
	Total	1.00	
	Total		1.00

Finally, within nature of work, weights were assigned as follows:

1.	immediate training requirements		.40
2.	continuing aspects		.60
	(a) personal interest in the technical content of the job	.50	
	(b) degree of variety implicit in the job	.30	
	(c) amount of training in management skills realizable from the job	.20	
	Total	1.00	
	Total		1.00

The above assignment of weights lead to the following distribution of "effective" weights on each of the fifteen lowest-level performance criteria:

1.	starting salary	.208
2.	insurance benefits	.014
3.	retirement benefits	.009
4.	anticipated three-year salary	.064
5.	anticipated five-year salary	.035
6.	proximity to relatives	.068
7.	degree of urbanity	.068
8.	climate	.034
9.	daily commuting requirements	.034
10.	proportion of time away from home	.054
11.	duration of extended trips	.082
12.	immediate training requirements	.132
13.	personal interest in the technical content of the job	.099
14.	degree of variety implicit in the job	.059
15.	amount of training in management skills realizable from the job	.040
	Total	1.000

8.3 The Criterion Scores

Of the fifteen performance measures listed in Section 8.1 and displayed in Exhibit 1, only eleven were defined in such a manner as to require explicit scoring functions. In the remaining four instances, he decided to assign direct worth estimates to the relevant aspects of each alternative job offer. All eleven of the explicit scoring functions were sketched by a graphical technique similar to the one set forth in scoring procedure 20, Appendix XX of this paper.

Exhibit 2 below shows the estimated performance of each of the four alternatives on his fifteen performance measures.

Exhibit 2

Estimated Performance

Performance Criterion	Alt. I	Alt. II	Alt. III	Alt. IV
starting salary	$ 8,100/yr.	$ 8,250/yr.	$ 8,733/yr.	$ 8,550/yr.
insurance benefits	$ 475/yr.	$ 550/yr.	$ 475/yr.	$ 400/yr.
retirement benefits	$ 750/yr.	$ 1,000/yr.	$ 1,100/yr.	$ 875/yr.
three-year salary	$11,250/yr.	$ 9,500/yr.	$10,500/yr.	$10,500/yr.
five-year salary	$15,000/yr.	$10,500/yr.	$11,500/yr.	$11,500/yr.
proximity to relatives	0 hrs.	0 hrs.	5 hrs.	1 hr.
degree of urbanity	2.5 million	2.5 million	1.0 million	15.0 million
climate	*	*	*	*
daily commuting	.50 hrs.	1.00 hrs.	.25 hrs.	1.25 hrs.
% time away	0%	10%	0%	35%
extended trip duration	0 days	5 days	0 days	20 days
required job training	9.0 months	.5 months	1.0 months	.5 months
interest in job	*	*	*	*
variety	*	*	*	*
training in management	*	*	*	*

* means direct worth estimate was made

Exhibit 3 below shows the worth scores assigned either by graphical scoring functions or by direct worth estimation to the performance data associated with each alternative.

Exhibit 3

Assigned Worth Scores

Performance Criterion	Alt. I	Alt. II	Alt. III	Alt. IV
starting salary	.68	.70	.75	.73
insurance benefits	.60	.70	.60	.50
retirement benefits	.60	.80	.90	.70
three-year salary	.75	.63	.70	.70
five-year salary	.75	.45	.53	.53
proximity to relatives	1.00	1.00	.10	.50
degree of urbanity	1.00	1.00	.70	.80
climate	.70*	.70*	.85*	.60*
daily commuting	.60	.50	.90	.40
% time away	1.00	.70	1.00	.35
extended trip duration	1.00	.70	1.00	.50
required job training	.50	.90	.80	.90
interest in job	.40*	.60*	.75*	.85*
variety	.50*	.80*	.70*	.90*
training in management	.70*	.85*	.75*	.80*

* means direct worth estimate was made

8.4 The Adjusted Effective Weights

His next step was to adjust the "effective" weights according to the perceived interpretive quality of each performance measure. This lead to a set of "adjusted effective" weights which could then be applied to the worth scores shown in Exhibit 3. The original "effective" weights, the adjusting factors, and the final set of "adjusted effective" weights are shown below in Exhibit 4.

Exhibit 4

"Effective" Weights, Adjusting Factors, and "Adjusted Effective" Weights

Performance Criterion	"Effective" Weights	Adjusting Factors	"Adjusted Effective" Weights
starting salary	.208	1.00	.268
insurance benefits	.014	.95	.017
retirement benefits	.009	.95	.012
three-year salary	.064	.75	.062
five-year salary	.035	.75	.034
proximity to relatives	.068	.80	.069
degree of urbanity	.068	.75	.066
climate	.034	.90	.040
daily commuting	.034	.85	.037
% time away	.054	.50	.035
extended trip duration	.082	.85	.090
required job training	.132	.70	.118
interest in job	.099	.60	.076
variety	.059	.60	.045
training in management	.040	.60	.031

8.5 The Total Worth Scores

His last step was to multiply the assigned worth scores by their "adjusted effective" weights and add the products to determine each alternative's total worth score. The results of these computations are shown in Exhibit 5 below.

Exhibit 5

Total Worth Scores

Performance Criterion	Alt. I	Alt. II	Alt. III	Alt. IV
starting salary	.182	.187	.201	.195
insurance benefits	.010	.012	.010	.009
retirement benefits	.007	.010	.011	.008
three-year salary	.047	.039	.043	.043
five-year salary	.026	.015	.018	.018
proximity to relatives	.069	.069	.007	.035
degree of urbanity	.066	.066	.046	.053
climate	.028	.028	.034	.024
daily commuting	.022	.019	.033	.015
% time away	.035	.025	.035	.012
extended trip duration	.090	.063	.090	.045
required job training	.059	.106	.094	.106
interest in job	.030	.046	.057	.065
variety	.023	.036	.032	.041
training in management	.022	.026	.023	.025
Total Worth	.716	.747	.734	.694

Inspection of Exhibit 5 shows that Alternative II achieved the highest total worth score. As it turned out, Alternative II was selected.

SECTION 9

RESULTS OF AN EXPERIMENTAL VALIDATION OF THE PROCEDURE AS IMPLEMENTED BY PROFESSIONAL DECISION MAKERS

In the preceding sections of this paper, a systematic procedure to aid in the assessment of worth was first developed and then illustrated with a complete example. The purpose of this procedure, it will be recalled, is to help decision makers formulate and articulate a consistent assessment structure (really a complex algorithm) for assessing the worth of specified alternatives in a definite choice situation. Once formulated, this assessment algorithm may be applied to each specified alternative so as to generate a numerical index of its overall worth.

The experiment, whose results will be reported in this section, was designed to validate the assessment procedure - that is, to determine whether or not the procedure could be implemented by professional decision makers and, if so, with what degree of success.

9.1 A Brief Review of the Worth Concept

In order to recall the conceptual foundations of the assessment procedure and to motivate discussion of the experiment, five critical assumptions about the worth concept are restated below.[11]

1. Worth is an internal property of human beings. Worth notions exist within the perceptual and attitudinal apparatus of human decision makers - not as external properties of the physical objects and activities which human beings assess and to which they impute worth. To assess the worth of an object or activity, therefore, is to measure a decision maker's response (e.g., verbal assessments, behavioral choice, etc.) to that object or activity.

[11] The reader is referred to Section 3 of this paper for a more complete discussion of the worth concept.

2. In general, human notions of worth are multidimensional rather than unidimensional. This means two things:
   a. A given physical object or activity is perceived as relevant simultaneously to more than one human objective;
   b. A given human objective may be satisfied by more than one alternative object or activity.

3. An individual's notions of worth need not necessarily be shared by others (i.e., consensual validation is not a <u>definitional</u> requirement of legitimate worth notions), although some consensus can be expected, particularly within his reference group.

4. An individual's notions of worth need not necessarily be stable over time (i.e., temporal stability is not a <u>definitional</u> requirement of legitimate worth notions), although some stability can be expected, particularly where his more important values are concerned.

5. Worth notions do not usually exist in a conscious, clearly defined, and logically structured form within the minds of human decision makers. However, with some effort, a consistent assessment structure can be formulated to reflect an individual's notions of worth, so long as certain practical limitations on the ability to conceptualize are observed.

9.2 A Brief Review of the Assessment Procedure

The assessment procedure, it will be recalled, involves several sequential operations. These are outlined below.[12]

[12] The reader is referred to Section 7 of this paper for a more complete discussion of the assessment procedure.

1. Assuming that a job to be done and/or a set of activities to be performed has been described, formulate a list of overall job objectives by abstraction from the job description.

2. Refine each higher-level objective in terms of two or more lower-level, independent performance criteria which define more precisely what is intended by or subsumed under the meaning of the higher-level objective. Generate thereby a complete criterion hierarchy.

3. Interpret lowest-level criteria in terms of physical performance measures.

4. Specify individual worth relationships perceived as holding between each lowest-level criterion and its linked performance measure.

5. Establish an overall index of worth, considering all of the previously listed objectives and sub-criteria simultaneously.

If a decision maker can successfully complete the above five operations he will have created an assessment structure (really a complex algorithm) by means of which a single cardinal worth number may be assigned to any specified alternative in a given choice situation. Inputs to this assessment algorithm consist of various physical performance measures selected by the decision maker as describing the relevant measurable attributes of an alternative. The output of this assessment algorithm is a single cardinal number purporting to represent the worth imputed by the decision maker to that alternative.

9.3 The Purpose of the Experiment

As stated previously, the purpose of the experiment was to validate the assessment procedure. In particular, the following questions were raised concerning the impact of the procedure upon professional decision makers as they develop preferences for specified alternatives and eventually choose one of them.

1. Are professional decision makers both able and willing to undertake the complete assessment procedure in making a choice among specified alternatives?

2. If so, which aspects of the procedure are difficult to interpret and implement?

3. Does introduction of the procedure into the decision making process serve to clarify, to confuse, or to have no noticeable impact upon individual preferences for alternatives? If there is a noticeable impact, how great is it?

4. Does the procedure increase, decrease, or have no noticeable impact upon the number of preference discriminations made by decision makers among alternatives? If there is a noticeable impact, how great is it?

5. Does the procedure increase, decrease, or have no noticeable impact upon a decision maker's confidence in the accuracy of his indicated preferences?

6. How satisfied are decision makers with using the procedure? Specifically, do they consider it helpful in improving the quality of their final choices? If so, by how much?

7. Does implementation of the procedure serve to alter preferences for alternatives? If so, by how much and in what ways?

8. How aware are decision makers of the extent to which the procedure serves to alter their preferences?

9. To what extent do decision makers feel that any gains made in clarification, confidence, satisfaction, and/or appropriate alteration are worth the additional costs in time and effort expended (by implementing the procedure) to realize these gains?

10. To what extent and in what ways does implementation of the procedure serve to alter attitudes on the part of decision makers toward formal, quantitative techniques of assessment?

11. To what extent will decision makers spontaneously adapt various aspects of the procedure to other decision situations lying beyond the scope of the experiment itself (e.g., to situations more closely resembling the real world)?

These questions constitute the specific senses in which validation of the procedure were sought experimentally. The experiment itself, important results, and overall conclusions will be reported subsequently in summary form.[13]

---

[13] A more complete exposition of the experiment can be found in <u>The Assessment of Worth: A Systematic Procedure and its Experimental Validation</u>, an unpublished Doctoral Dissertation by James R. Miller III, Massachusetts Institute of Technology, September 1966.

9.4 The Context of the Experiment

Several years ago, the Department of Defense established a school at Wright-Patterson Air Force Base to train military and Civil Service personnel in the intricacies of modern weapons systems management. Military officers from all three branches of the Armed Services and Civil Service personnel from various defense-oriented government agencies (e.g., the National Aeronautics and Space Administration) are selected four times each year to participate in an eleven-week training course. A class consists of approximately sixty such individuals holding the rank of Colonel, Lieutenant Colonel, Captain (Navy), Commander, GS-14, GS-15, or the equivalent, and with at least some (in most cases, substantial) prior experience managing government projects. Since the purpose of the course is to train project managers, a large part of the curriculum is devoted to new techniques in scientific management--particularly those espoused by the Department of Defense. The eleventh and final week of the course consists of a computer-simulated game played by teams of five participants each. The computer is programmed to simulate contractor responses to various decisions made by each team as it progresses through the design, selection, installation, and eventual operation of a typical weapons system.

This eleven-week training course constituted the context of the experiment. The sixty military and Civil Service personnel being trained for duty as project managers comprised the sample of experimental subjects.

9.5 Specific Design Objectives

In designing the experiment, the following specific objectives were set forth.

1. First, it seemed essential to select a sample of experimental subjects who regularly make important

decisions among complex alternatives. After all, it is for precisely this kind of person that the assessment procedure was primarily (if not exclusively) designed. It is not clear that other kinds of people would be either willing or able to undertake such an arduous task.

2. Second, it seemed desirable to have each subject make a definite and clearly observable decision (i.e., choice among alternatives) concerning some issue which he would regard as meaningful and whose consequences would be directly and visibly related to his future well-being. By requiring each subject to make an observable choice, experimental measures of preliminary preferences (for the various alternatives) could be formulated and later tested for their ability to predict his final choice. By selecting an issue which he would perceive as both meaningful and bearing directly upon his future well-being, each subject could be expected to expend a reasonable amount of time and effort in formulating an assessment structure and applying this to the alternatives.

3. Third, to remain compatible with the assessment procedure, the choice had to be constrained to a fixed set of discrete and clearly specified alternatives.

4. Fourth, it seemed desirable to have all sixty subjects assess similar alternatives in an identical decision situation. This would permit making valid comparisons of results across subjects.

5. Fifth, it seemed desirable to have all sixty subjects make their assessments independently of one another. The focus of this experiment was upon individual (as opposed to group) decision making processes.

6. Sixth, it seemed desirable to make the decision situation relatively simple, relatively familiar, and restricted to a manageable number of alternatives. This would serve to economize time and effort both on the part of the experimenter and on the part of the experimental subjects (no prior training required).

7. Finally, to provide bases against which results of the complete assessment procedure could be compared, it seemed desirable to design experimental manipulations in such a way as to obtain similar measures of the impact of three alternative modes of assessment. These were:

   A. Spontaneous assessment with neither explicit information about the alternatives nor any explicit guidance on how to assess their worth or how to make a final choice;

   B. Assessment with the aid of raw information about the alternatives, but without any systematic guidance on how to utilize such information in assessing their worth or arriving at a final choice;

   C. Partially guided assessment (i.e., the first part of the complete procedure developed in Section 7 including only those operations designed to generate a criterion hierarchy and performance measures, but excluding the subsequent scoring and weighting operations).

9.6 The Decision Situation, the Alternatives, and the Final Choice

Recall that all sixty experimental subjects form teams of five participants each at the end of the training course. Through the medium of a computer game against simulated defense contractors they then proceed to test their newly-acquired knowledge. For purposes of the experiment, the decision which each individual subject had to make was to choose partners and thereby form a team to play the computer game.

Assuming five-man teams (of which there were twelve), an alternative consisted of a group of four other participants who, along with the individual making the choice, could constitute a complete five-man team.

If each subject were permitted to choose any four partners from among the entire remainder of the class, then he would have to consider over 455,000 alternative teams. This was obviously too many for any one person to handle. Consequently, a series of experimental devices had to be employed in order to reduce the alternatives to a manageable number.

The first device was to subdivide the sixty subjects into six sub-groups of ten each. Subdivision was performed prior to the beginning of the training course with the aid of a random number table. Then, each subject was asked to peruse a list of ten names (including his own) and to subdivide the remaining list of nine others into two sub-lists. The first sub-list contained six names of preferred candidates for inclusion in his final team, while the second sub-list contained the three remaining names. Subjects were asked to perform this latter subdivision after having had a few days to acquaint themselves with other participants in the training course. By means of these two devices, each subject then had only six other candidates from whom to choose four team partners. This served to reduce the number of alternative teams which each individual must consider to fifteen.

However, despite these experimental devices, there still remained the problem of giving each subject an independent choice to make. Except by unlikely accident, not every individual in a ten-man sub-group could have his complete choice of partners fulfilled. If two or more individuals included the same third individual in their most preferred team, but failed to include each other, then somebody would have to lose. Consequently, a third experimental device had to be employed to obviate this difficulty and to maintain the prospect of an independent decision for all sixty subjects. It was decided to announce at the outset of the experiment that one subject in each of the ten-man sub-groups would have his choice of team partners honored. The remaining five subjects not chosen by him would be grouped to form a second team. Exactly whose choices were to be honored remained undetermined until the end of the experiment, and a random number table was used to make this determination at that time. Therefore, each subject might proceed on the assumption that he would be making the final choice, for his chances would be just as good as anyone else's of having his choice honored.

9.7 The Experimental Procedure

Prior to the beginning of the experiment, all sixty subjects were assigned at random to three groups. Each group contained two of the ten-man sub-groups, making twenty subjects in all. One of these twenty-man groups performed the complete assessment procedure developed in Section 7. The second group performed part of the procedure (up to the point of generating a criterion hierarchy and selecting performance measures). The third group received only raw information about their alternatives, but no systematic guidance concerning its utilization. All three groups performed initial operations designed to measure the impact of neither information nor guidance.

A battery of written questionnaires, in conjunction with a schedule of personal interviews, was designed and administered during the first ten weeks of the training course. Through these instruments data were gathered concerning the impact of the various assessment procedures upon the decision making process. At the end of the tenth week, each subject made his final choice of team partners. Five-man teams were then formed on the basis of these choices, and all sixty subjects participated in the computer simulation exercise during the eleventh and final week of the training course.

9.8 Satisfaction of the Specific Design Objectives

The first objective--validating the assessment procedure on professional decision makers--was satisfied by the particular choice of experimental subjects and the experimental context. All sixty participants in the training course are sent to the school for the express purpose of receiving education in decision making. Most of them have had extensive practical experience in assessing and choosing among complex alternatives prior to coming. The curriculum focuses heavily upon decision making techniques, and the work-pace is intensive. Students live on the Air Base throughout the eleven-week period and are required to attend six hours of class each day. Consequently, on the basis of these personal background and contextual factors, it seemed reasonable to hope that both the subjects and the setting would provide an appropriate vehicle for validating the assessment procedure.

The second objective--having each subject make a definite and clearly observable decision--was satisfied by requiring everyone to choose four team partners at the end of the experiment, just prior to playing the computer game. The choice was definite. It was clearly observable by the experimenter (although not by the subject's fellow students). It could have an immediate impact upon his chosen team's performance in the game itself. Since the game was advertised in advance as competitive, and since previous participants in the game had demonstrated substantial personal committment and competitive zeal, it was reasonable to hope that subjects would take the experimental decision seriously.

The third objective--providing a fixed set of discrete alternatives--was satisfied by means of the first two experimental devices. Each subject had a fixed set of fifteen discrete teams from which to make a final choice (i.e., the fifteen logically possible combinations of six team partners taken four at a time).

The fourth objective--having all sixty subjects assess similar alternatives in an identical decision situation--was likewise satisfied by these two experimental devices.

The fifth objective--inducing each subject to make an independent decision--was satisfied by the third experimental device. By announcing in advance that an individual's choice of team partners would be honored if and only if his name were selected by a completely random mechanism and without regard to whom he chose or who chose him, it was hoped to discourage the formation of coalitions and the adoption of competitive bidding strategies. In addition, it was decided to give continual instructions to the subjects requesting that they refrain from discussing with one another their preferences, their assessment criteria, or their anticipated final choices.

The sixth objective--presenting a relatively simple and familiar decision situation--was satisfied by the nature of the required choice. Choosing partners for some group enterprise is a familiar decision made many times in almost everyone's lifetime. Choosing up sides for an athletic contest or parlor game, selecting new members for a social or business organization, and choosing a marriage partner are common examples.

Satisfaction of the seventh design objective--providing bases of comparison--was achieved by splitting the class into three groups of twenty each and having them undertake different modes of assessment.

9.9 A Summary of Results and Conclusions

The experiment yielded the following results and conclusions.

1. The complete assessment procedure developed in Section 7 was implemented in its entirety by all twenty of the subjects introduced to it. However, one subject stated in advance that he viewed the procedure as an empty ritual. His implementation was, therefore, only nominal and signified no real committment to its overall intent. A second subject chose to substitute alternative procedures of his own design for the scoring and weighting operations suggested in Section 7. From these results it was concluded that the procedure could be implemented by professional decision makers.

2. The one aspect of the procedure which consistently produced confusion and misunderstanding was the issue of independence among objectives and performance criteria. It required a fair amount of interpretive discussion to clarify the meaning of this concept. Hence, it was concluded that further efforts might profitably be expended upon this portion of the procedure.

3. The complete assessment procedure was judged superior to all three of the alternative modes of assessment included in the experiment. In addition, subjects introduced to the complete procedure tended to adapt it to another decision context (i.e., to making decisions during the course of the computer simulation exercise) to a significantly greater extent than did subjects introduced to the alternative modes of assessment.

4. Almost every conscious assessment activity which subjects perceived as relevant to making their choices served to clarify their preferences for alternatives. In particular, receiving factual information, being required to articulate and structure assessment criteria, and being required to quantify their preferences all had this effect. The mere realization that a choice had to be made, accompanied by preliminary efforts to structure the alternatives, had the same effect. However, when subjects did not perceive such activities as relevant, even though they were alleged to be, clarification did not occur. When clarification did occur, its magnitude varied with the particular type of activity engaged in. Of critical importance were those kinds of activity which challenged and thereby tested the validity of existing preferences (e.g., comparison of informal, subjective preferences with formally derived quantitative outputs of the assessment algorithm).

5. The number of preference discriminations spontaneously made by subjects among alternatives depended primarily upon individual factors. Changes thereto induced under alternative modes of assessment also depended upon individual factors.

6. Almost every conscious assessment activity perceived as relevant to the decision served to increase confidence in the accuracy of stated preferences. In particular, the four modes of assessment designed into the experiment had this effect. Irrelevant activities did not have this effect. Again, the

magnitude of this effect depended upon the particular type of activity.

7. The same results concerning clarification occurred in the case of satisfaction derived by subjects from undertaking various modes of assessment. Satisfaction, here, refers to the degree to which such activities were perceived as helpful to improving the quality of the final choice.

8. Although subjects did receive clarification, satisfaction, and additional confidence from undertaking various modes of assessment, this did not guarantee that they would overtly alter prior preference committments in light of newly-perceived implications. Once again, provision of a challenge or validity check (e.g., comparison of subjective preferences with numerical outputs of the assessment algorithm) was of critical importance. When such checks were performed, then overt committment generally did follow.

9. On the other hand, changes in preference occurred covertly following almost every conscious and relevant assessment activity, but did not occur (apart from random instability) unless such activity was perceived as relevant. The magnitude of such changes decreased steadily as confidence and clarification increased and as the moment of final decision drew near.

10. Without knowing precisely what their previous preferences were, subjects tended to underestimate their

temporal stability. They also tended to underestimate the magnitude of changes in stability over time. Both of these phenomena became less pronounced if they made a definite and overt committment to a particular preference structure.

11. The perceived value of engaging in various assessment activities compared to the time and effort expended depended critically upon the type of activity engaged in. When the activity was perceived as irrelevant, it was considered a great waste of time. However, even when the activity was perceived as relevant, it was not always considered sufficiently valuable to justify the time and effort expended. Once again, providing a challenge or validity check was particularly important in this respect.

12. The complete assessment procedure developed in Section 7 had a four-fold overall impact upon decision processes.

    a. Its primary impact was to induce subjects to formulate and validate a consistent assessment structure. Validation was provided by comparing formally derived with subjective preferences, and the quantitative aspects of the procedure were critical in this respect. Alternative modes of assessment, lacking quantitative aspects, did not, in general, produce this effect.

    b. In the process of formulating an assessment structure, preferences for alternatives were significantly altered. However, they were not altered randomly, but rather in a manner directed toward the final choice.

c. When the entire procedure was followed, particularly the final steps of quantitative assessment, a mechanism was provided to validate preferences. This, in turn, induced favorable reactions to formal assessment techniques. It also induced at least intermediate-term changes both in attitudes toward the procedure and in preferences for alternatives. On the other hand, when only part of the procedure or none of the procedure was followed, the reaction of subjects was nowhere near as favorable nor as permanent.

d. Another important impact was to measure and display assessment criteria, which can be useful both for purposes of normative decision making and for purposes of scientific description. The alternative modes of assessment did not produce this result--at least not to the same extent.

13. These results were obtained from adult, highly educated, highly competent, highly motivated, and well-trained professional decision makers. Furthermore, they refer to individual decision processes conducted under ideal laboratory conditions. Further research is required to extend their range of applicability either to different kinds of people or to different situations. The writer would be particularly skeptical about generalizing these results to situations characterized by substantial conflict, strife, or political combat. These results may also be inapplicable to individuals who do not characteristically possess the courage of their convictions or, perhaps, no convictions at all.

SECTION 10

THE PROCEDURE IN PERSPECTIVE

Having developed and illustrated the assessment procedure, it will now be placed in perspective. This will be accomplished in several ways. Section 10.1 will relate the procedure to similar research reported in the literature. Its specific relationship to earlier work by Churchman and his colleagues will be traced out, and several alternative approaches to the same problem will be discussed. This section will not contain a comprehensive review of the assessment literature. Only a few specific procedures will be described, and then only briefly. Those readers interested in obtaining greater familiarity with the relevant literature are directed to several basic reference works.

Section 10.2 will contrast the procedure with current assessment practices. Once again, a detailed discussion of current practices will be forgone in favor of a more general discussion of basic similarities and differences.

Section 10.3 will relate the procedure to normative decision theory--particularly to operations research and to statistical decision theory. The procedure's role in operationalizing such normative disciplines will be discussed.

A critical review of the entire procedure will be presented in Section 10.4, and the paper will close with overall conclusions in Section 10.5.

10.1 The Procedure and Previous Research

The problem of formulating objectives and selecting physical performance measures is certainly not new either to decision theorists or to practical decision makers. A classic theoretical discussion of this problem has been provided by Churchman et al. (1957) and by Churchman (1961). In fact, the assessment procedure developed in Section 7 of this paper might best be viewed as both an extension to and a refinement of Churchman's general approach. The specific senses in which our procedure differs from Churchman's are listed below.

1. Churchman does not provide any explicit procedure for formulating a criterion hierarchy, although provision for same is implicit within his discussion.

2. Churchman handles the problem of interdependence among worth criteria by means of an explicit test of the additivity of assigned weights. Our procedure includes explicit questions concerning independence and infers additivity therefrom.

3. Churchman does not treat the problem of differential interpretive quality among performance measures, although his weights could easily be adjusted by our procedure developed in Section 7.6.

4. Churchman provides no explicit procedures either for selecting performance measures or for parameterizing scoring functions. The single scoring technique which he presents for illustrative purposes is a special case of our procedure developed in Appendix II.[14]

---

[14] The reader is referred to Churchman, C. W., Prediction and Optimal Decision: Philosophical Issues in a Science of Values (1961) for a general discussion of worth assessment. A specific discussion of Churchman's methodological approach can be found in Churchman, C. W., et al, Introduction to Operations Research (1957), pp. 69-154. Additional methodological discussions of the same issues are cited on pages 153 and 154 of that book. An experimental application of Churchman's methodology is reported in Pound, W. H., "Research Project Selection: Testing a Model in the Field," IEEE Transaction in Engineering Management, 1964 (11), pp. 16-22. An experimental comparison of Churchman's weight-setting technique with five alternative techniques is reported in Eckenrode, R. T., "Weighting Multiple Criteria", Management Science, November 1965, pp. 180-192. Another empirically oriented investigation of weight-setting, particularly the effects of interdependence, is reported in Yntema, D. B., and Torgerson, W. S., "Man-Computer Cooperation in Decisions Requiring Common Sense," IRE Transactions on Human Factors in Electronics, March 1961, pp. 20-26.

A somewhat similar and much more detailed approach to worth assessment is suggested by Rosenthal (1964). Although this approach is oriented specifically toward assessing electronic data processing equipment, it is in principle generalizable to almost any context. The main differences between Rosenthal's approach and the procedure developed herein are as follows.

1. Rosenthal assesses alternatives relative to each other (i.e., worth scores are assigned either in direct or inverse proportion to measured performance differentials), whereas our procedure uniformly matches each alternative's measured performance against a pre-established standard of desired performance.

2. Rosenthal treats required resource expenditures (e.g., cost) in the same manner as received performance, whereas our procedure generally treats them differently.[15]

A substantially different approach is suggested by N. M. Smith, Jr. Here, decision makers are not required to articulate objectives in the explicit manner suggested by Churchman, by Rosenthal, and by this writer. Instead, worth is defined as a particular state of a complex system (outlined by Smith), whose behavior is influenced by choices among alternative courses of action.[16]

---

[15] The reader is referred to Rosenthal, S., "Analytical Technique for Automatic Data Processing Equipment Acquisition", 1964 Spring Joint Computer Conference, for a more complete discussion of this approach.

[16] An early exposition of this approach may be found in Smith, N. M., Jr., "A Calculus for Ethics: A Theory of the Structure of Value", Behavioral Science, 1956 (1), pp. 111-142 and pp. 186-211.

In addition to the foregoing references, which represent only a small sample of the assessment literature, a rather extensive review of assessment procedures in several fields is provided by Cronbach and Gleser (1965). Beside describing many such procedures and their applications in detail, this book also directs the reader to similar reviews made by other writers. No attempt will be made to review the same material in this paper. Instead, the interested reader is hereby referred to the above mentioned book.

10.2 The Procedure and Current Practice

The number of particular assessment procedures currently used in making practical decisions is obviously enormous. No attempt will be made to describe or even to enumerate them. As before, attention will be directed toward the important differences between our assessment procedure and various others.

Certainly the most striking feature of our procedure is its primary and explicit focus upon the decision maker himself. The more typical approach is to focus explicitly upon the decision context and the alternatives and only implicitly upon the decision maker. In this manner, an aura of objectivity is created, which many consider to be a safer, if not a more desirable way to assess alternatives. However, there is no real inconsistency between these two approaches. Quite to the contrary, they are mutually reinforcing. Let us develop this theme more fully.

A traditional objective in making business decisions is to maximize dollar profits. Although it is not always clear whether this is the only, the primary, or just another objective, it is a very common one. One substantial gain realized from defining such an objective is that many of the performance consequences of various alternatives may be expressed in terms of a single unit--dollars. This, in turn, permits comparing the alternatives themselves in terms of that same single index. Nevertheless, there frequently arise certain complications.

1. Not all important consequences of a decision can be measured meaningfully in dollar terms (e.g., consequences in terms of corporate image, good will, etc.).

2. Some consequences, although measurable in dollar terms, are important for entirely different reasons (e.g., the salary of an important executive would not reflect the true loss which would be suffered by a firm should he die--particularly if the firm could not continue without his unique skills).

3. Considering dollars both as a unit of worth and as a resource may lead to confusion. Resource dollars and worth dollars may not really be commensurable--particularly if decision makers in fact view them differently, as sometimes happens under complex budgeting and accounting systems.

Systems analysts obviate some of the above problems by defining a different unit of worth (usually called system effectiveness). It is recognized that profit maximization is not the objective at all (particularly for military and government organizations), and either a high-order performance measure (e.g., the kill probability of a missile) or a universal resource different from money (e.g., an equivalent number of man-hours of labor required) is used to render multiple performance consequences and, therefore, whole alternatives commensurable.

Now both of the above approaches are perfectly compatible with our assessment procedure. In the simplest case, there is logical equivalence. If either dollar profits or some single aspect of system effectiveness were considered the only objective in a decision, then a simple criterion hierarchy would be constructed containing only one criterion and only one performance measure. It is in more complicated cases, where both multiple objectives and multiple performance measures are considered relevant, that conceptual difficulties arise. It is also in such cases that the mutually reinforcing aspects of our procedure and alternative approaches become most apparent.

Review of the assessment procedure will show a point of tangency between it and both kinds of traditional approaches at the moment where physical performance measures are selected (see Section 7.3). Conclusions drawn by a cost accountant or by a systems analyst would be relevant in selecting such measures. However, achieving commensurability is accomplished in our procedure by an entirely different technique. Rather than seeking a very high-order performance or resource measure, and thereby achieving commensurability in phenomenological terms, an even higher-order measure is sought. Decision makers are asked to produce that measure themselves on the basis of their objectives, their experience, and their judgment. The measure in terms of which all performance consequences are rendered commensurable is worth points, and our procedure is the means of generating it. This difference in both focus (i.e., upon decision makers themselves) and emphasis (i.e., in achieving commensurability in the non-objective worth sense) is particularly important when broad policy issues are at stake.

There are other differences as well. These are not quite as striking as the one just discussed, but they deserve mention.

One aspect of our procedure which may not seem novel (but really is) is its simultaneous inclusion of both linear and non-linear elements.

Recall that linear weights are assigned to criteria throughout the hierarchy to indicate the relative importance of satisfying them. However, this does not imply linearity in the various performance measures. Since decision makers are free to define non-linear scoring functions linking performance measures to performance criteria, the overall index of worth can be non-linear in the measures themselves. Where linearity does seem appropriate, linear scoring functions may be chosen, but this is not essential.

Another difference relates to the distinction between desired performance and offered performance. Whereas many procedures confuse the two, our procedure takes great pains to keep them clearly separated. Conceptual separation is policed by manipulating performance criteria and physical performance measures in very different ways (i.e., weighting versus scoring).

Finally, our procedure is almost completely general--at least in principle. Only the judgment and experience of decision makers serves to restrict the range of jobs, of alternatives, or of situations to which it is potentially applicable. Most other procedures are developed for specified contexts and, therefore, are not so easily generalized.

10.3 The Procedure and Normative Decision Theory

As indicated by the experiment, the single most important impact of the assessment procedure is to elicit from decision makers an explicit statement of their objectives. Without clearly stated objectives, normative decision techniques are not applicable to real problems. In particular, operations research techniques have no practical application without a well-defined "objective function" to optimize. Statistical decision theory is of no practical use unless or until an explicit "loss structure" has been defined. To the extent that our procedure induces decision makers to formulate either an "objective function" or a "loss structure" it serves to operationalize these normative techniques. This, in fact, was the primary motivation for developing the procedure in the first place.[17]

10.4 Criticisms of the Procedure

The preceding sections discussed the procedure as a whole in relation to previous research, to current practices, and to normative decision theory. In this section, we shall investigate the procedure from within. Critical scrutiny will be directed toward the entire methodology in an attempt to pinpoint internal "soft-spots." Three types of criticisms will be made. First, important substantive issues omitted from discussion or accorded only cursury treatment will be recalled. Second, attention will be focused upon underlying assumptions which seem particularly questionable. Finally, some of the more difficult component procedures will be reviewed. On the basis of this discussion, overall conclusions will be drawn in Section 10.5.

The reader may have noticed that several important aspects of the complete task of assessment were either ignored completely or else given only a cursory treatment. Methodological issues falling into this category include:

1. the problem of describing adequately and accurately the job to be performed by whichever alternative is finally selected--this was ignored completely;

---

[17] The procedure's role in defining an "objective function" is discussed at length by Churchman in the previously cited works. A discussion of "loss structures" and the practical application of statistical decision theory can be found in Schlaifer, R., Probability and Statistics for Business Decisions (1959), and in Raiffa, H., and Schlaifer, R., Applied Statistical Decision Theory (1961). A specific application to oil and gas drilling decisions is documented in Grayson, C. J., Jr., Decisions Under Uncertainty: Drilling Decisions by Oil and Gas Operators (1960).

2. the problem of producing alternatives to accomplish the stated job--this was also completely ignored;

3. the problem of predicting both performance and resource consequences associated with each produced alternative--very little was said about this issue;

4. the problem of validating the descriptive accuracy of performance and resource estimates--this was ignored completely;

5. the problem of establishing feasibility constraints (i.e., mandatory performance and/or resource requirements) on alternatives--this issues was also ignored;

6. the problem of incorporating risk/uncertainty considerations--this was discussed only briefly;

7. the problem of defining a decision rule--this also was discussed only briefly;

8. the problem of selecting appropriate personnel to assess alternatives and to make a final choice--except to point out that final results could depend critically upon both the identity of decision makers and the point in time when an assessment is made, this issue was largely ignored.

Now it is <u>not</u> claimed that the above issues are unimportant. Quite to the contrary, they are all very important, and they deserve the same amount of attention accorded to worth assessment. However, the scope of this paper was not intended to cover these issues, except insofar as they provided a context in which to discuss worth assessment.

Three critical assumptions about the manner in which decision makers can be induced to conceptualize worth notions deserve special attention. First, is it easy for decision makers to formulate objectives and to map out a complete criterion hierarchy? From the experiment we know that high-calibre, professional decision makers can, without much difficulty, but this does not mean that everybody can.

Second, can decision makers conceive of and articulate meaningfully the notion of proportional satisfaction of a stated objective or criterion? Specifically, can they distinguish easily between this problem and the problem of comparing alternatives with one another and assessing them on a purely relative scale? The experiment demonstrated that this is a somewhat difficult distinction to make at the outset. It also demonstrated that the first problem is a more difficult one to solve than the second, even after they have been distinguished. However, at least for the professional decision makers studied, this is possible.

Third, can decision makers be comfortable with a linear weighting scheme? A typical first reaction to this question is negative on the grounds that strict linearity is too simple and too restrictive. However, after realizing that linearly weighted performance criteria do not necessarily imply an assessment algorithm linear in performance measures (recall that scoring functions can assume any desirable nonlinear shape), decision makers will generally retract their objection. At least this is what occurred during the experiement.

This brings us to the question of difficult assessment procedures. The most difficult one by far was shown (by the experiment) to be identification of worth-interdependence among performance criteria. Without independence, the linear weighting assumption is highly suspect. The writer will readily admit that this is a difficult question to understand and an even more difficult one to answer. The current procedure is not completely satisfactory in this regard. Additional research might profitably be directed toward reformulating and illustrating this issue more clearly.

Another difficult procedure involves definition and selection of physical performance measures. This requires some ingenuity and, therefore, may restrict the procedure's applicability to sophisticated decision makers. However, provision of a master list does help a great deal. This was demonstrated in the experiment (biographical information constituted the master list).

Adjusting the "effective" weights to account for differential quality among performance measures is not a difficult procedure to implement, but there is some question in the writer's mind whether this is the proper way to handle the problem. Critics of the procedure have been similarly skeptical. Additional research might profitably be directed toward this problem also.

10.5 Overall Conclusions

Subject to the preceding criticisms, the writer would draw the following conclusions about the assessment procedure.

1. It can be carried out successfully by professional decision makers in an ideal laboratory setting. Whether it can or should be carried out by other types of people or in other types of settings is not yet clear. This issue could (and in the writer's opinion should) be resolved by further research.

2. Although reactions to the procedure on the part of decision makers were shown to be quite favorable, no claim is made that alternative procedures could not produce the same effects. In other words, the particular procedure developed herein is not claimed to be unique. However, it is claimed to be sufficient in these respects.

3. The reader should bear in mind that the complete procedure has two distinct phases. The first phase induces decision makers to formulate, to articulate, to validate, and to map out their assessment structure. The primary advantages of the procedure lie in this first phase--particularly in its quantitative aspects--which permit internal validation of prior preferences. The second phase involves application of the assessment algorithm so generated to specified alternatives.

4. The procedure is not completely free of conceptual problems or difficult questions. Further research would probably alleviate some of these difficulties.

5. The procedure is definitely not "objective" in the sense of eliminating the need for human judgment. Quite to the contrary, its basic purpose is to systematize subjective judgment in ways which decision makers will find helpful.

6. The procedure is completely general with respect to type of decision problem and type of objective, criterion, and performance measure. As such, it is potentially applicable in many practical situations.

J. R. Miller III

J. R. Miller III

APPENDIX I

SCORING PROCEDURE 1

References in text: Section 7.5.1.3

The performance measure under scrutiny has been determined to have the following characteristics:

1. discrete scale;
2. two-level scale;
3. level 1 = absence of some desirable performance attribute;
4. level 2 = presence of that desirable attribute.

Step 1. Assign zero worth points to absence of the desirable attribute.

Step 2. Assign one worth point to presence of that desirable attribute.

This completes the procedure.

APPENDIX II

SCORING PROCEDURE 2

References in text: Section 7.5.1.3

The performance measure under scrutiny has been determined to have the following characteristics:

1. discrete scale;
2. two-level scale;
3. level 1 = absence of some desirable performance attribute;
4. level 2 = presence of that desirable attribute in conjunction with some qualitative measure of relative worth, when present.

Step 1. Assign zero worth points to absence of the desirable attribute.

Step 2. Identify all feasible alternatives which promise the desirable attribute.

Step 3. Assemble one or more decision makers.

Step 4. After discussing collectively the various merits of the desirable attribute - why it is important and what benefits its presence provides - have each decision maker make a separate and independent judgment of the extent to which each feasible alternative's promised attribute satisfies the related lowest-level performance criterion. All judgments will be recorded by assigning a number between zero and one indicating the proportional satisfaction provided by each feasible alternative.

Step 5. To determine each feasible alternative's score on this performance measure, assign either zero points (if the attribute is absent) or the arithmetic mean (possibly weighted) of the individual scores assigned judgmentally by separate decision makers.

This completes the procedure.

APPENDIX III

SCORING PROCEDURE 3

References in text: Section 7.5.1.3

The performance measure under scrutiny has been determined to have the following characteristics:

1. discrete scale;

2. three, four, or five levels on the scale;

3. all levels constitute strictly nominal categories.

Step 1. Since the scale of the physical performance measure is strictly nominal, a preference or worth ordering must be placed directly on the nominal categories. This will be accomplished by means of the same ranking procedure used in defining weights and presented in Section 7.4.

Step 2. Assemble one or more decision makers.

Step 3. After discussing collectively the various merits of nominal categories, have each decision maker perform a separate and independent rank-ordering of the various categories. This may be accomplished by performing Steps 4 through 7 below.

Step 4. List the nominal categories in approximate order of decreasing worth, starting with the category perceived as most valuable at the top of the list. The category perceived as least valuable should appear at the bottom of the list. It is not necessary to have the categories perfectly ranked or ordered on this first pass, since subsequent operations will be performed to guarantee complete ordering.

Step 5. Compare the first two categories on the list.

a. If the first category is perceived as more valuable than the second, proceed directly to Step 6.

b. If both categories are perceived as roughly equal in worth, proceed directly to Step 6.

c. If the second category is perceived as more valuable than the first, invert their positions on the list (i.e., place the first category where the second used to be on the list, and vice versa), and then proceed to Step 6.

Step 6. Compare the lower-ranked category from Step 5 with the next-lower category on the list. Repeat the comparisons and stipulated operations in Step 5 on this new pair of categories. Continue in this manner all the way down to the end of the list until pairwise comparisons have been made between all contiguous criteria.

Step 7. After the list has been completely exhausted, go back and determine whether any inversions (position changes) occurred. If none occurred, proceed directly to Step 8. If one or more occurred, return to the head of the list, and repeat the entire procedure described in Steps 5 and 6.

Step 8. Next, have each decision maker make a separate and independent judgment of the extent to which each ranked category satisfies the related lowest-level criterion. All judgments will be recorded by assigning a number between zero and one indicating the proportional satisfaction provided by each scale category.

Step 9. Finally, to determine each nominal category's point score, compute and record the (possibly weighted) arithmetic mean of the individual category scores assigned by separate decision makers in Step 8.

This completes the procedure.

APPENDIX IV

SCORING PROCEDURE 4

References in text: Section 7.5.1.3

The performance measure under scrutiny has been determined to have the following characteristics:

1. discrete scale;

2. three, four, or five levels on the scale;

3. the scale is ordered.

Step 1. List the ordered levels in a single column.

Step 2. Inspect the level appearing at the head of the column. Is that the most preferred or the least preferred level? If most preferred, proceed to Step 4. If least preferred, proceed to Step 3.

Step 3. Invert the column, and list the levels again - this time in reverse order. Proceed to Step 4.

Step 4. The discrete levels should now be listed in perfect order of descending relative worth. Inspect to verify that this is true. If so, proceed to Step 5. If not, check earlier steps to insure that no errors occurred. If no errors occurred, this particular performance measure cannot be handled by the scoring procedures presented herein. Define a new measure, and return to Step 1 of the questioning procedure in Section 7.5.1.3.

Step 5. Assemble one or more decision makers.

Step 6. After discussing collectively the various merits of nominal categories and (hopefully) agreeing on their rank-order, have each decision maker record a separate and independent judgment of the extent to which each nominal category satisfies the related lowest-level performance criterion. All judgments will be recorded by assigning a number between zero and one indicating the proportional satisfaction provided by each scale category.

Step 7. To determine each nominal category's score, compute and record the (possibly weighted) arithmetic mean of the individual category scores assigned by separate decision makers.

This completes the procedure.

APPENDIX V

SCORING PROCEDURE 5

References in text: Section 7.5.1.3

The performance measure under scrutiny has been determined to have the following characteristics:

1. discrete scale;

2. more than five levels on the scale;

3. all levels constitute strictly nominal categories.

Step 1. Since the scale of the physical performance measure is strictly nominal, a preference or worth ordering must be placed directly on the nominal categories. This will be accomplished by means of the same ranking procedure used in defining weights and presented in Section 7.4.

Step 2. Assemble one or more decision makers.

Step 3. After discussing collectively the various merits of nominal categories, have each decision maker perform a separate and independent rank-ordering of the various categories. This may be accomplished by performing Steps 4 through 7 below.

Step 4. List the nominal categories in approximate order of decreasing worth, starting with the category perceived as most valuable at the top of the list. The category perceived as least valuable should appear at the bottom of the list. It is not necessary to have the categories perfectly ranked or ordered on this first pass, since subsequent operations will be performed to guarantee complete ordering.

Step 5. Compare the first two categories on the list.

a. If the first category is perceived as more valuable than the second, proceed directly to Step 6.

b. If both categories are perceived as roughly equal in worth, proceed directly to Step 6.

c. If the second category is perceived as more valuable than the first, invert their positions on the list (i.e., place the first category where the second used to be on the list, and vice versa), and then proceed to Step 6.

Step 6. Compare the lower-ranked category from Step 5 with the next-lower category on the list. Repeat the comparisons and stipulated operations in Step 5 on this new pair of categories. Continue in this manner all the way down to the end of the list until pair-wise comparisons have been made between all contiguous criteria.

Step 7. After the list has been completely exhausted, go back and determine whether any inversions (position changes) occurred. If none occurred, proceed directly to Step 8. If one or more occurred, return to the head of the list, and repeat the entire procedure described in Steps 5 and 6.

Step 8. Inspect adjacent pairs of ranked scale categories. Locate that adjacent pair of scale categories which seem closest to one another in terms of their perceived worth (i.e., locate the most equally valuable adjacent pair of scale categories). Collapse these two categories into a single category.

Step 9. Repeat Step 8 as many times as is required to reduce the number of resulting categories to five. Then proceed to Step 10.

Step 10. Next, have each decision maker record a separate and independent judgment of the extent to which each ranked category satisfies the related lowest-level criterion. All judgments will be recorded by assigning a number between zero and one indicating the proportional satisfactions provided by each scale category.

Step 11. Finally, to determine each nominal category's point score, compute and record the (possibly weighted) arithmetic mean of the individual category scores assigned by separate decision makers in Step 10.

This completes the procedure.

APPENDIX VI

SCORING PROCEDURE 6

References in text: Section 7.5.1.3

The performance measure under scrutiny has been determined to have the following characteristics:

1. continuous scale;
2. bounded from below by zero;
3. bounded from above by some finite positive number;
4. direct preference relationship;
5. worth score zero assigned to zero performance;
6. worth score one assigned to performance at the logical upper bound;
7. constant rate of change of worth with increases in performance.

The above seven characteristics describe completely a linear scoring function passing through the origin and with positive slope equal to the reciprocal of the logical upper bound. The equation of this scoring function is

$$\text{worth score} = \frac{\text{Measured Performance}}{\text{Logical Upper Bound}}$$

A graphical picture of this scoring function appears below in Figure 1.

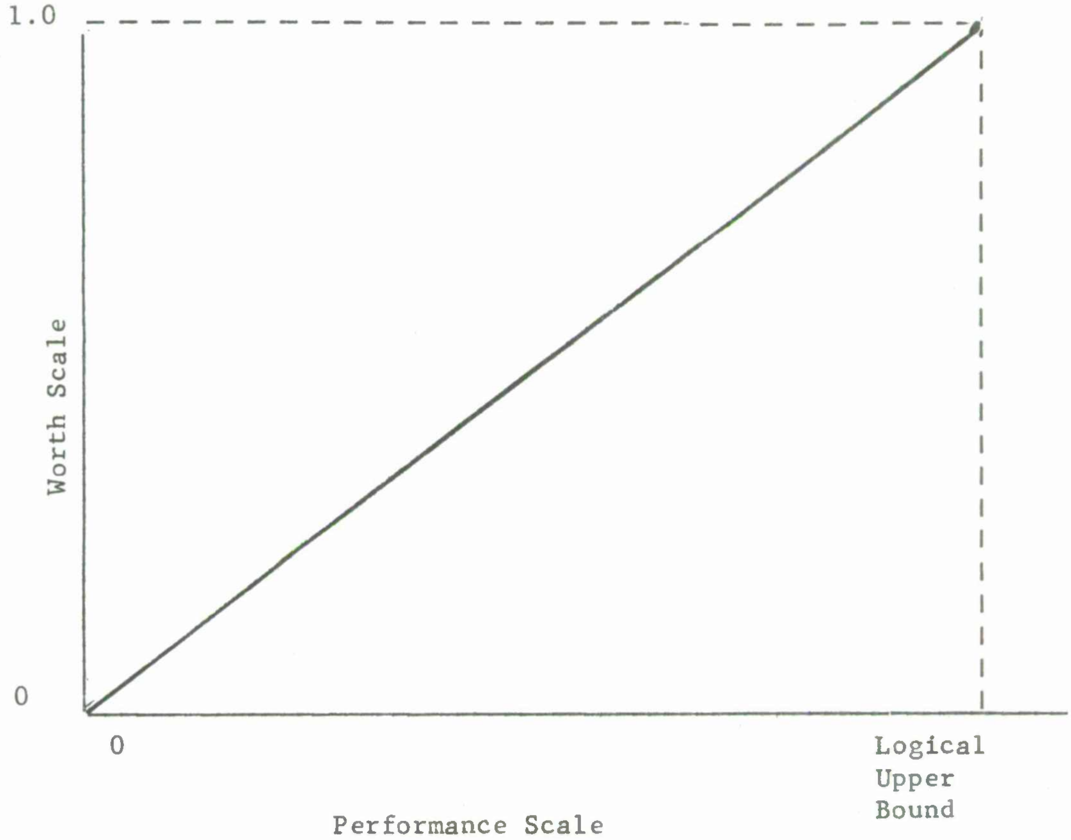


FIGURE 1

This completes the procedure.

APPENDIX VII

SCORING PROCEDURE 7

References in text: Section 7.5.1.3

The performance measure under scrutiny has been determined to have the following characteristics:

1. continuous scale;
2. bounded from below by zero;
3. bounded from above by some finite positive number;
4. direct preference relationship;
5. worth score zero assigned to zero performance;
6. worth score one assigned to performance at the logical upper bound;
7. uniformly accelerating rate of change of worth with increases in performance.

A graphical picture of this general shape of scoring function appears below in Figure 2.

Step 1. At this point, decision makers have two choices. The simplest procedure would be to fit a standardized quadratic scoring function to the performance measure under the following stipulated assumptions.

1. The scoring function is quadratic with positive second derivative (indicating uniform acceleration).
2. The minimum of the quadratic function falls exactly at the origin.
3. The upper tail of the scoring function passes through the point whose coordinates are (performance = logical upper bound, worth score = one).

These three assumptions completely determine a scoring function (see Figure 2) whose equation is

$$\text{worth score} = \left(\frac{\text{Measured Performance}}{\text{Logical Upper Bound}}\right)^2.$$

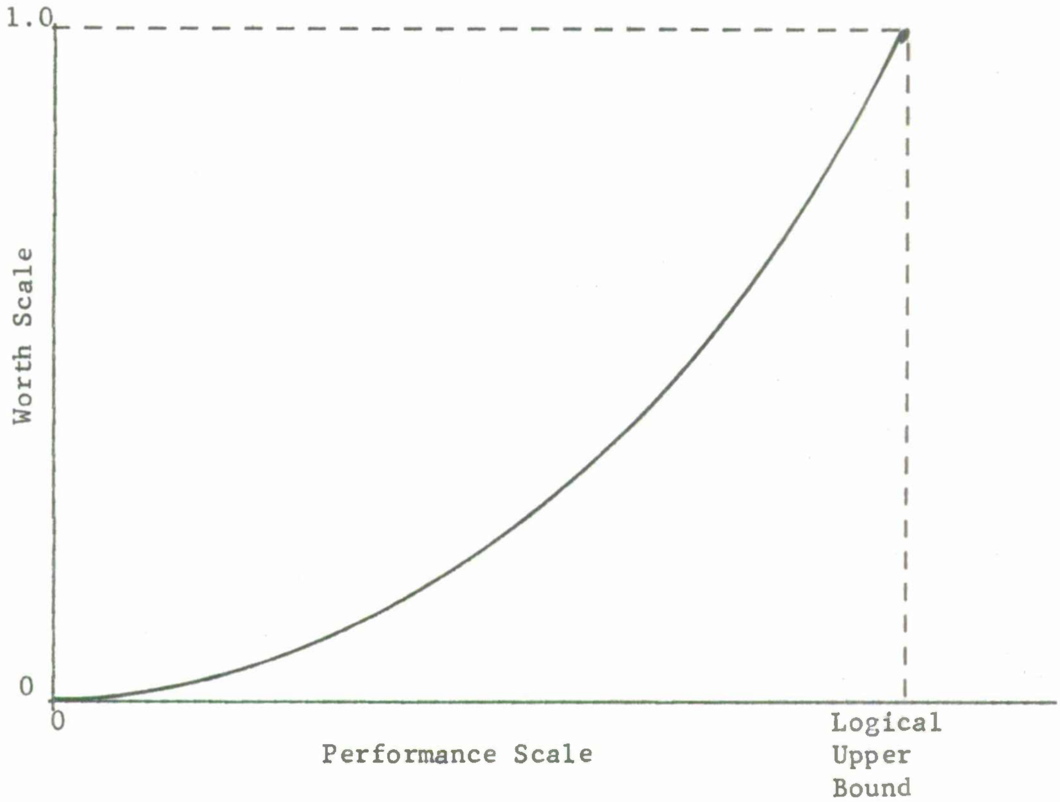


FIGURE 2

To determine whether or not this looks like an appropriate scoring function, it is suggested that a sheet of standard graph paper be procured and that the above equation be plotted thereupon. Five or six representative points should be sufficient to grasp the exact shape of the function and to decide whether or not it seems appropriate. If yes, this completes the procedure. If no, proceed to scoring procedure 20.

APPENDIX VIII

SCORING PROCEDURE 8

References in text: Section 7.5.1.3

The performance measure under scrutiny has been determined to have the following characteristics:

1. continuous scale;
2. bounded from below by zero;
3. bounded from above by some finite positive number;
4. direct preference relationship;
5. worth score zero assigned to zero performance;
6. worth score one assigned to performance at the logical upper bound;
7. uniformly decelerating rate of change of worth with increases in performance.

A graphical picture of this general shape of scoring function appears below in Figure 3.

Step 1. At this point, decision makers have two choices. The simplest procedure would be to fit a standardized quadratic scoring function to the performance measure under the following stipulated assumptions.

1. The scoring function is quadratic with negative second derivative (indicating uniform deceleration).
2. The maximum of the quadratic function falls exactly at the point whose coordinates are (performance = logical upper bound, worth score = one).

3. The quadratic function passes through the origin.

These three assumptions completely determine a scoring function (see Figure 3) whose equation is

$$\text{worth score} = 2\left(\frac{\text{Measured Performance}}{\text{Logical Upper Bound}}\right) - \left(\frac{\text{Measured Performance}}{\text{Logical Upper Bound}}\right)^2.$$

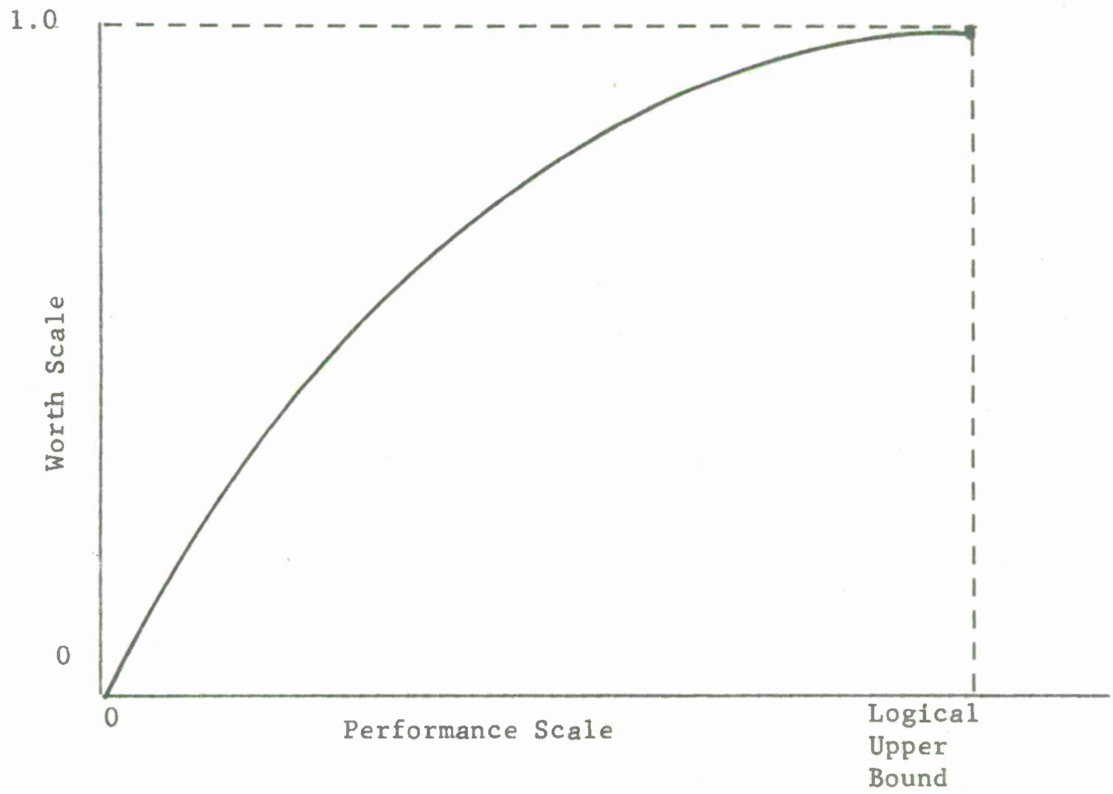


FIGURE 3

To determine whether or not this looks like an appropriate scoring function, it is suggested that a sheet of standard graph paper be procured and that the above equation be plotted thereupon. Five or six representative points should be sufficient to grasp the exact shape of the function and to decide whether or not it seems appropriate. If yes, this completes the procedure. If no, proceed to scoring procedure 20.

APPENDIX IX

SCORING PROCEDURE 9

References in text: Section 7.5.1.3

The performance measure under scrutiny has been determined to have the following characteristics:

1. continuous scale;
2. bounded from below by zero;
3. bounded from above by some finite positive number;
4. direct preference relationship;
5. worth score zero assigned to zero performance;
6. worth score one assigned to performance at the logical upper bound;
7. first accelerating, then decelerating rate of change of worth with increases in performance.

A graphical picture of this general shape of scoring function appears below in Figure 4.

<u>Step 1</u>. At this point, decision makers have two choices. The simplest procedure would be to fit a standardized cosine function to the performance measure whose equation is

$$\text{worth score} = \frac{1}{2} - \frac{1}{2}\,\text{cosine}\left[\pi\left(\frac{\text{Measured Performance}}{\text{Logical Upper Bound}}\right)\right],$$

where $\pi = 3.1416$,

and cosine values may be looked up in a trigonometric table (function expressed in terms of radians) or computed on an engineering slide rule.

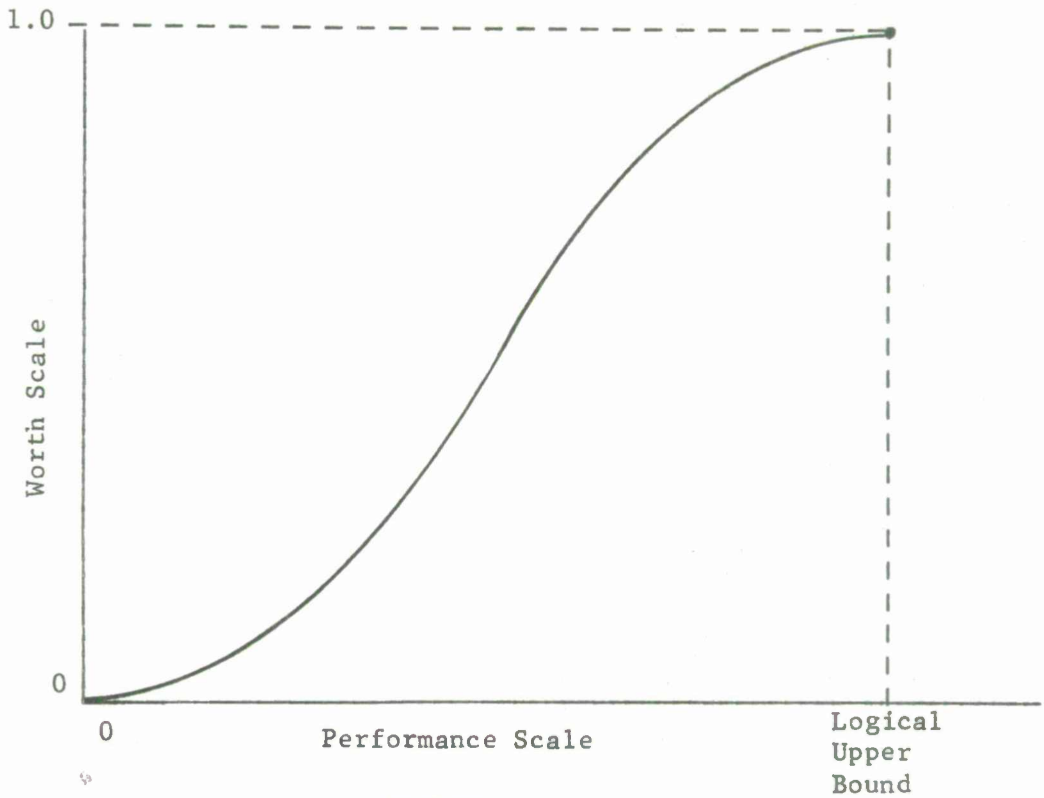


FIGURE 4

To determine whether or not this looks like an appropriate scoring function, it is suggested that a sheet of standard graph paper be procured and that the above equation be plotted thereupon. Five or six representative points should be sufficient to grasp the exact shape of the function and to decide whether or not it seems appropriate. If yes, this completes the procedure. If no, proceed to scoring procedure 20.

APPENDIX X

SCORING PROCEDURE 10

References in text: Section 7.5.1.3

The performance measure under scrutiny has been determined to have the following characteristics:

1. continuous scale;
2. bounded from below by zero;
3. bounded from above by some finite positive number;
4. direct preference relationship;
5. worth score zero assigned to zero performance;
6. worth score one assigned to performance at the logical upper bound;
7. first decelerating, then accelerating rate of change of worth with increases in performance.

A graphical picture of this general shape of scoring function appears below in Figure 5.

Step 1. At this point, decision makers have two choices. The simplest procedure would be to fit a standardized cosine function to the performance measure whose equation is

$$\text{worth score} = 2\left(\frac{\text{Measured Performance}}{\text{Logical Upper Bound}}\right) + \frac{1}{2}\,\text{cosine}\left[\pi\left(\frac{\text{Measured Performance}}{\text{Logical Upper Bound}}\right)\right] - \frac{1}{2},$$

where $\pi = 3.1416$,

and cosine values may be looked up in a trigonometric table (function expressed in terms of radians) or computed on an engineering slide rule.

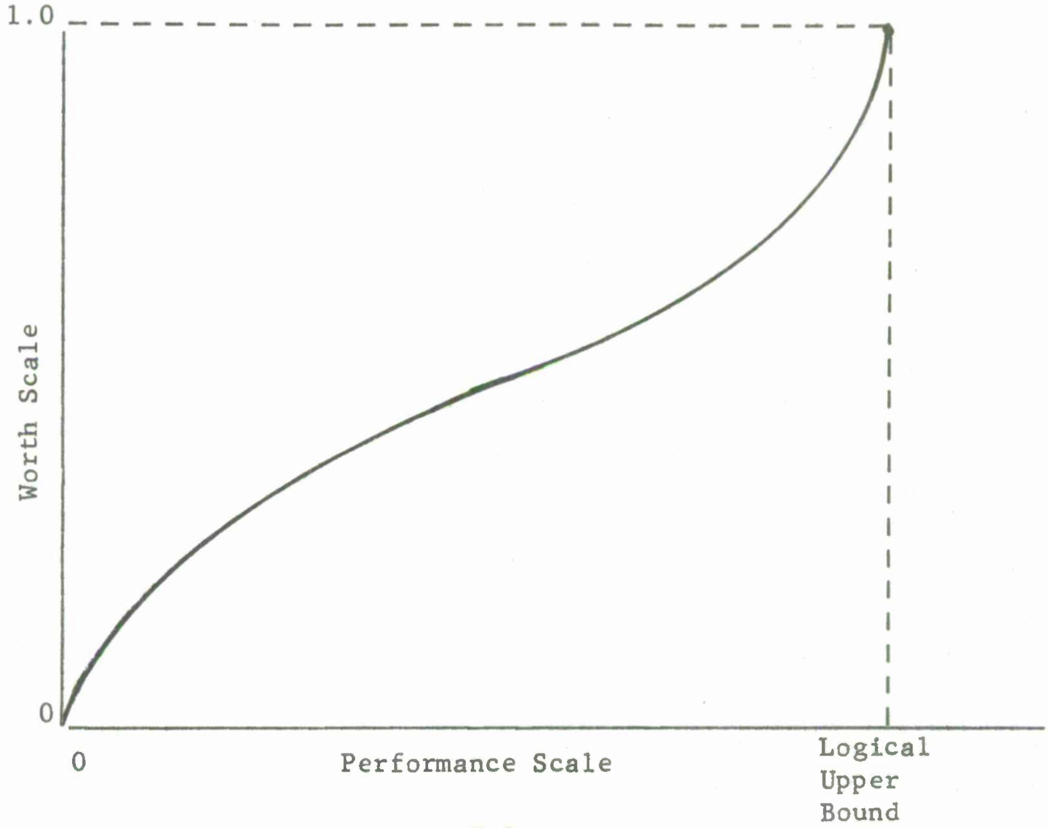


FIGURE 5

To determine whether or not this looks like an appropriate scoring function, it is suggested that a sheet of standard graph paper be procured and that the above equation be plotted thereupon. Five or six representative points should be sufficient to grasp the exact shape of the function and to decide whether or not it seems appropriate. If yes, this completes the procedure. If no, proceed to scoring procedure 20.

APPENDIX XI

SCORING PROCEDURE 11

References in text: Section 7.5.1.3

The performance measure under scrutiny has been determined to have the following characteristics:

1. continuous scale;
2. bounded from below by zero;
3. bounded from above by some finite positive number;
4. reverse preference relationship;
5. worth score zero assigned to performance at the logical upper bound;
6. worth score one assigned to zero performance;
7. constant rate of change of worth with increases in performance.

The above seven characteristics describe completely a linear scoring function passing through the point whose coordinates are (performance = zero, worth score = one) and with negative slope equal to minus the reciprocal of the logical upper bound. The equation of this scoring function is

$$\text{worth score} = 1 - \frac{\text{Measured Performance}}{\text{Logical Upper Bound}}$$

A graphical picture of this scoring function appears below in Figure 6.

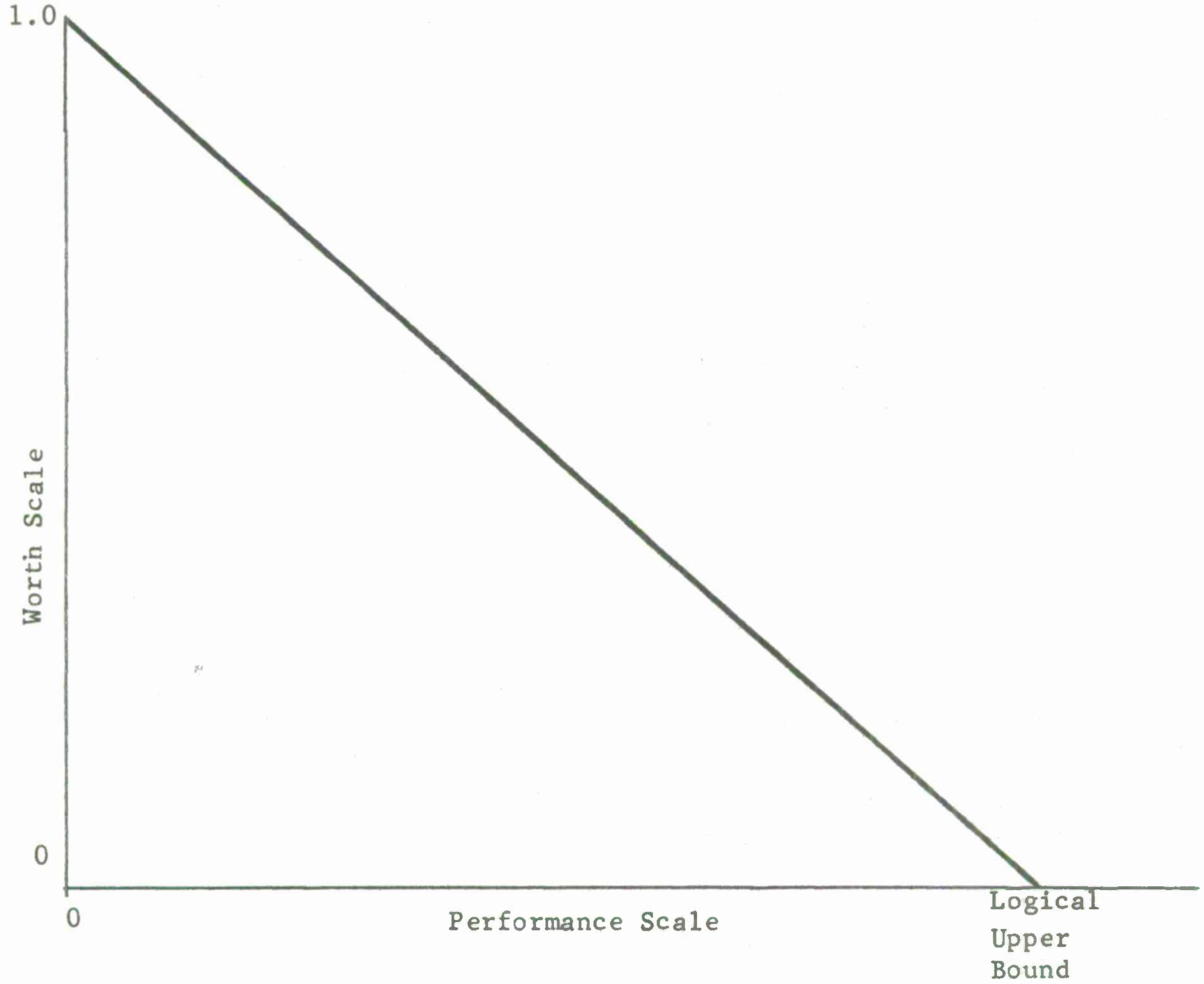


FIGURE 6

This completes the procedure.

APPENDIX XII

SCORING PROCEDURE 12

References in text: Section 7.5.1.3

The performance measure under scrutiny has been determined to have the following characteristics:

1. continuous scale;

2. bounded from below by zero;

3. bounded from above by some finite positive number;

4. reverse preference relationship;

5. worth score zero assigned to performance at the logical upper bound;

6. worth score one assigned to zero performance;

7. uniformly accelerating rate of change of worth with increases in performance.

A graphical picture of this general shape of scoring function appears below in Figure 7.

Step 1. At this point, decision makers have two choices. The simplest procedure would be to fit a standardized quadratic scoring function to the performance measure under the following stipulated assumptions.

1. The scoring function is quadratic with negative second derivative (indicating uniform acceleration).

2. The maximum of the quadratic function falls exactly at the point whose coordinates are (performance = zero, worth score = one).

3. The quadratic function falls to the point whose coordinates are (performance = logical upper bound, worth score = zero).

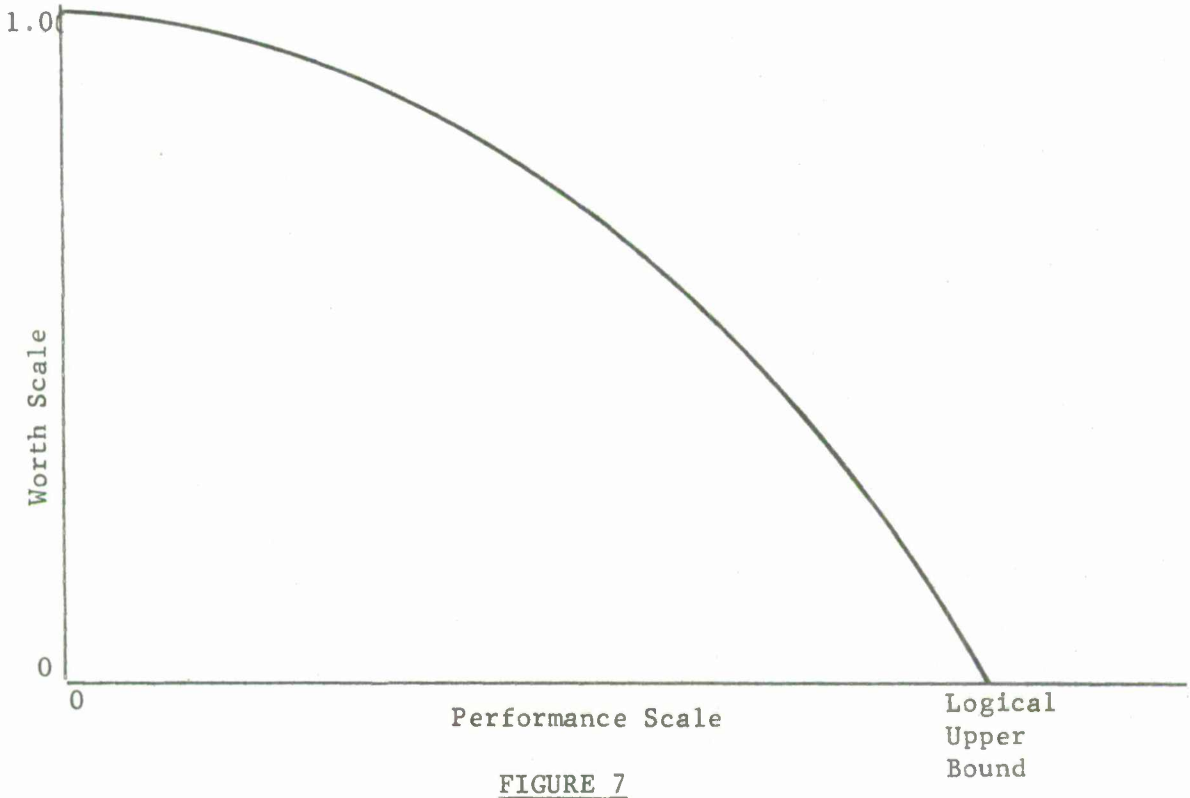


FIGURE 7

These three assumptions completely determine a scoring function (see Figure 7) whose equation is

$$\text{worth score} = 1 - \left(\frac{\text{Measured Performance}}{\text{Logical Upper Bound}}\right)^2 .$$

To determine whether or not this looks like an appropriate scoring function, it is suggested that a sheet of standard graph paper be procured and that the above equation be plotted thereupon. Five or six representative points should be sufficient to grasp the exact shape of the function and to decide whether or not it seems appropriate. If yes, this completes the procedure. If no, proceed to scoring procedure 20.

APPENDIX XIII

SCORING PROCEDURE 13

References in text: Section 7.5.1.3

The performance measure under scrutiny has been determined to have the following characteristics:

1. continuous scale;
2. bounded from below by zero;
3. bounded from above by some finite positive number;
4. reverse preference relationship;
5. worth score zero assigned to performance at the logical upper bound;
6. worth score one assigned to zero performance;
7. uniformly accelerating rate of change of worth with increases in performance.

A graphical picture of this general shape of scoring function appears below in Figure 8.

Step 1. At this point, decision makers have two choices. The simplest procedure would be to fit a standardized quadratic scoring function to the performance measure under the following stipulated assumptions.

1. The scoring function is quadratic with positive second derivative (indicating uniform acceleration).
2. The minimum of the quadratic function falls exactly at the point whose coordinates are (performance = logical upper bound, worth score = zero).

3. The upper left-tail of the function passes through the point whose coordinates are (performance = zero, worth score = one).

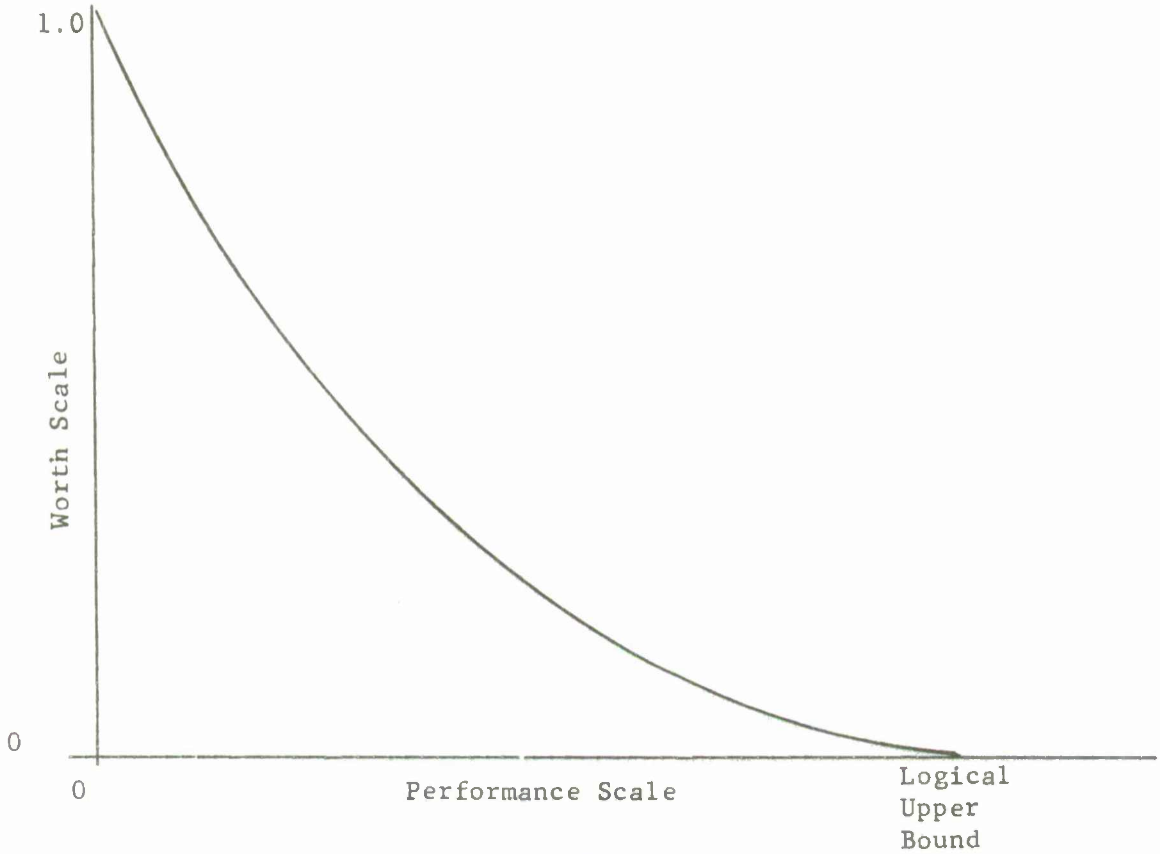


FIGURE 8

These three assumptions completely determine a scoring function (see Figure 8) whose equation is

$$\text{worth score} = 1 - 2\left(\frac{\text{Measured Performance}}{\text{Logical Upper Bound}}\right) + \left(\frac{\text{Measured Performance}}{\text{Logical Upper Bound}}\right)^2.$$

To determine whether or not this looks like an appropriate scoring function, it is suggested that a sheet of standard graph paper be procured and that the above equation be plotted thereupon. Five or six representative points should be sufficient to grasp the exact shape of the function and to decide whether or not it seems appropriate. If yes, this completes the procedure. If no, proceed to scoring procedure 20.

APPENDIX XIV

SCORING PROCEDURE 14

References in text: Section 7.5.1.3

The performance measure under scrutiny has been determined to have the following characteristics:

1. continuous scale;
2. bounded from below by zero;
3. bounded from above by some finite positive number;
4. reverse preference relationship;
5. worth score zero assigned to performance at the logical upper bound;
6. worth score one assigned to zero performance;
7. first accelerating, then decelerating rate of change of worth with increases in performance.

A graphical picture of this general shape of scoring function appears below in Figure 9.

Step 1. At this point, decision makers have two choices. The simplest procedure would be to fit a standardized cosine function to the performance measure whose equation is

$$\text{worth score} = \frac{1}{2} + \frac{1}{2}\ \text{cosine}\left[\pi\left(\frac{\text{Measured Performance}}{\text{Logical Upper Bound}}\right)\right],$$

where $\pi = 3.1416$,

and cosine values may be looked up in a trigonometric table (function expressed in terms of radians) or computed on an engineering slide rule.

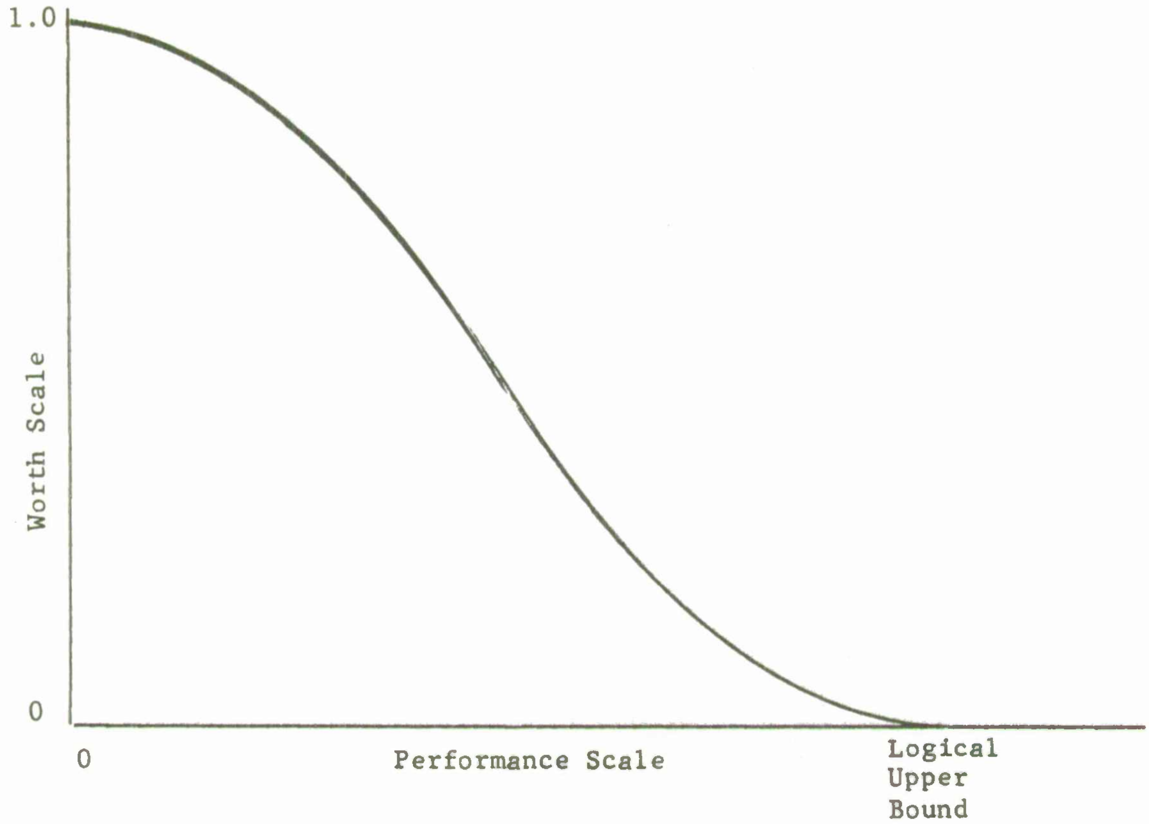


FIGURE 9

To determine whether or not this looks like an appropriate scoring function, it is suggested that a sheet of standard graph paper be procured and that the above equation be plotted thereupon. Five or six representative points should be sufficient to grasp the exact shape of the function and to decide whether or not it seems appropriate. If yes, this completes the procedure. If no, proceed to scoring procedure 20.

APPENDIX XV

SCORING PROCEDURE 15

References in text: Section 7.5.1.3

The performance measure under scrutiny has been determined to have the following characteristics:

1. continuous scale;
2. bounded from below by zero;
3. bounded from above by some finite positive number;
4. reverse preference relationship;
5. worth score zero assigned to performance at the logical upper bound;
6. worth score one assigned to zero performance;
7. first decelerating, then accelerating rate of change of worth with increases in performance.

A graphical picture of this general shape of scoring function appears below in Figure 10.

Step 1. At this point, decision makers have two choices. The simplest procedure would be to fit a standardized cosine function to the performance measure whose equation is

$$\text{worth score} = \frac{3}{2} - 2\left(\frac{\text{Measured Performance}}{\text{Logical Upper Bound}}\right) - \frac{1}{2}\,\text{cosine}\left[\pi\left(\frac{\text{Measured Performance}}{\text{Logical Upper Bound}}\right)\right],$$

where $\pi = 3.1416$,

and cosine values may be looked up in a trigonometric table (function expressed in terms of radians) or computed on an engineering slide rule.

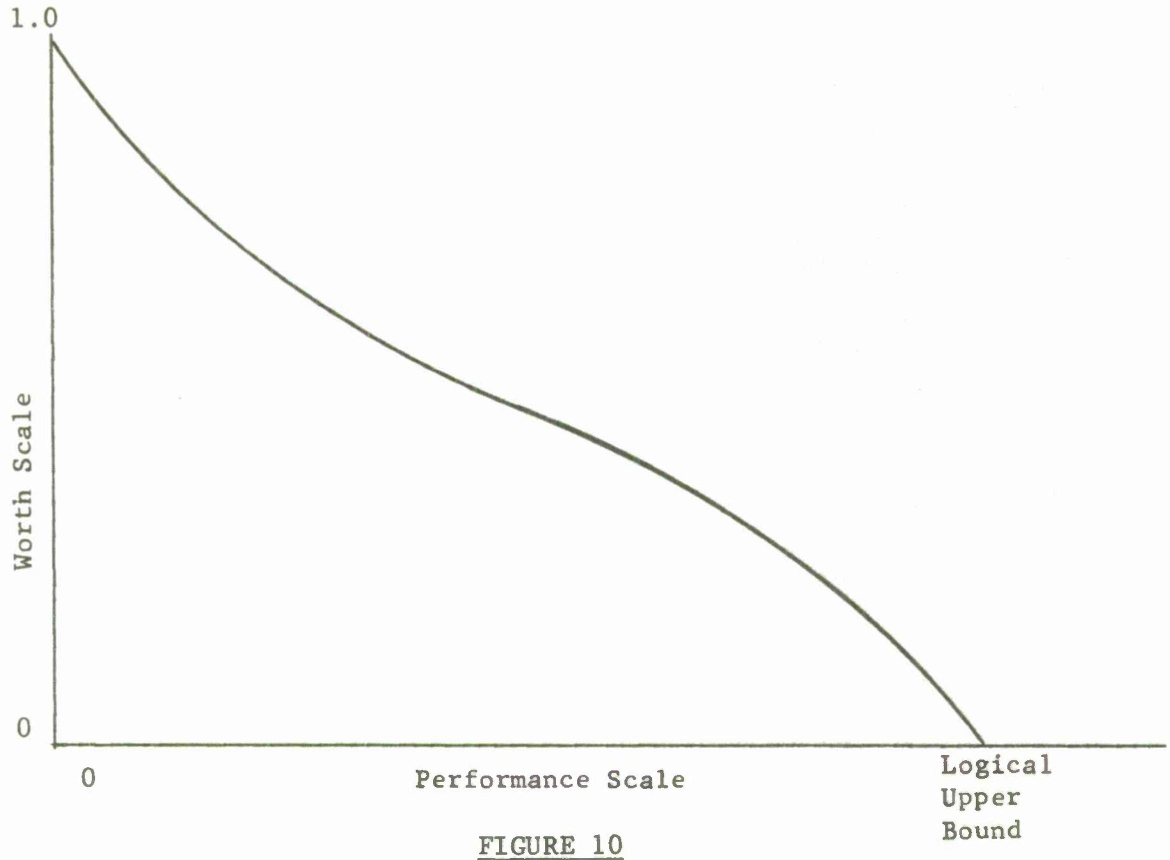


FIGURE 10

To determine whether or not this looks like an appropriate scoring function, it is suggested that a sheet of standard graph paper be procured and that the above equation be plotted thereupon. Five or six representative points should be sufficient to grasp the exact shape of the function and to decide whether or not it seems appropriate. If yes, this completes the procedure. If no, proceed to scoring procedure 20.

APPENDIX XVI

SCORING PROCEDURE 16

References in text: Section 7.5.1.3

The performance measure under scrutiny has been determined to have the following characteristics:

1. continuous scale;
2. bounded from below by zero;
3. no logical upper bound;
4. direct preference relationship;
5. worth score zero assigned to zero performance;
6. worth score one assigned to infinite performance;
7. uniformly decelerating rate of change of worth with increases in performance.

A graphical picture of this general shape of scoring function appears below in Figure 11.

Step 1. There is no simple, standardized equation to fit all situations of this type. Although the general shape of this scoring function is given by the equation

$$\text{worth score} = 1 - \exp\left[(-k)\ (\text{measured performance})\right],$$

where exp is the exponential function with basis e = 2.718, and k is a positive fitting constant,

still, the exact value of the fitting constant cannot be determined in a standard way for all performance measures. Consequently, proceed to scoring procedure 20.

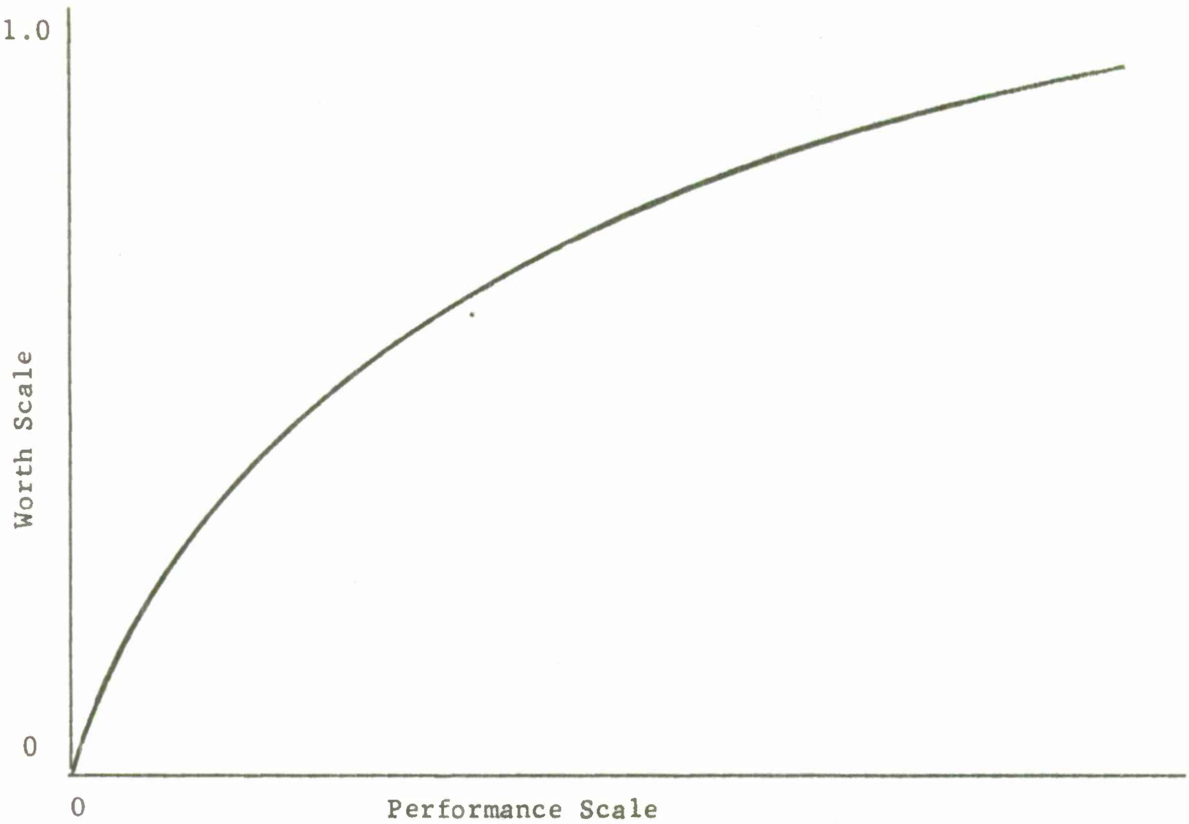


FIGURE 11

APPENDIX XVII

SCORING PROCEDURE 17

References in text: Section 7.5.1.3

The performance measure under scrutiny has been determined to have the following characteristics:

1. continuous scale;
2. bounded from below by zero;
3. no logical upper bound;
4. direct preference relationship;
5. worth score zero assigned to zero performance;
6. worth score one assigned to infinite performance;
7. first accelerating, then decelerating rate of change of worth with increases in performance.

A graphical picture of this general shape of scoring function appears below in Figure 12.

Step 1. There is no simple, standardized equation to fit all situations of this type. Although the general shape of this scoring function is given by the equation

$$\text{worth score} = \exp\left[(-a)\,(\text{measured performance})^{-b}\right],$$

where exp is the exponential function with basis e = 2.718, and both a and b are positive fitting constants ($b \geq 1$), still, the exact values of the fitting constants cannot be determined in a standard way for all performance measures. Consequently, proceed to scoring procedure 20.

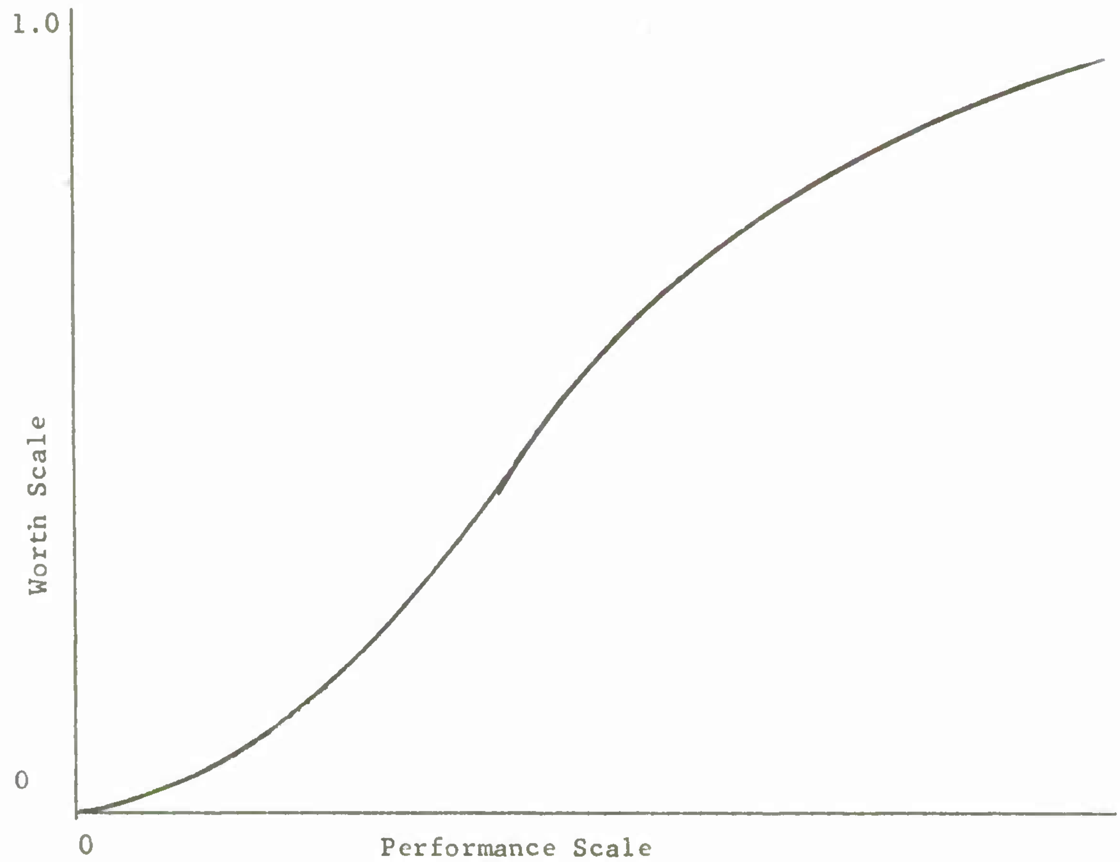


FIGURE 12

APPENDIX XVIII

SCORING PROCEDURE 18

References in text: Section 7.5.1.3

The performance measure under scrutiny has been determined to have the following characteristics:

1. continuous scale;
2. bounded from below by zero;
3. no logical upper bound;
4. reverse preference relationship;
5. worth score zero assigned to infinite performance;
6. worth score one assigned to zero performance;
7. uniformly decelerating rate of change of worth with increases in performance.

A graphical picture of this general shape of scoring function appears below in Figure 13.

Step 1. There is no simple, standardized equation to fit all situations of this type. Although the general shape of this scoring function is given by the equation

$$\text{worth score} = \exp\left[(-k)\ (\text{measured performance})\right],$$

where exp is the exponential function with basis e = 2.718, and k is a positive fitting constant,

still, the exact value of the fitting constant cannot be determined in a standard way for all performance measures. Consequently, proceed to scoring procedure 20.

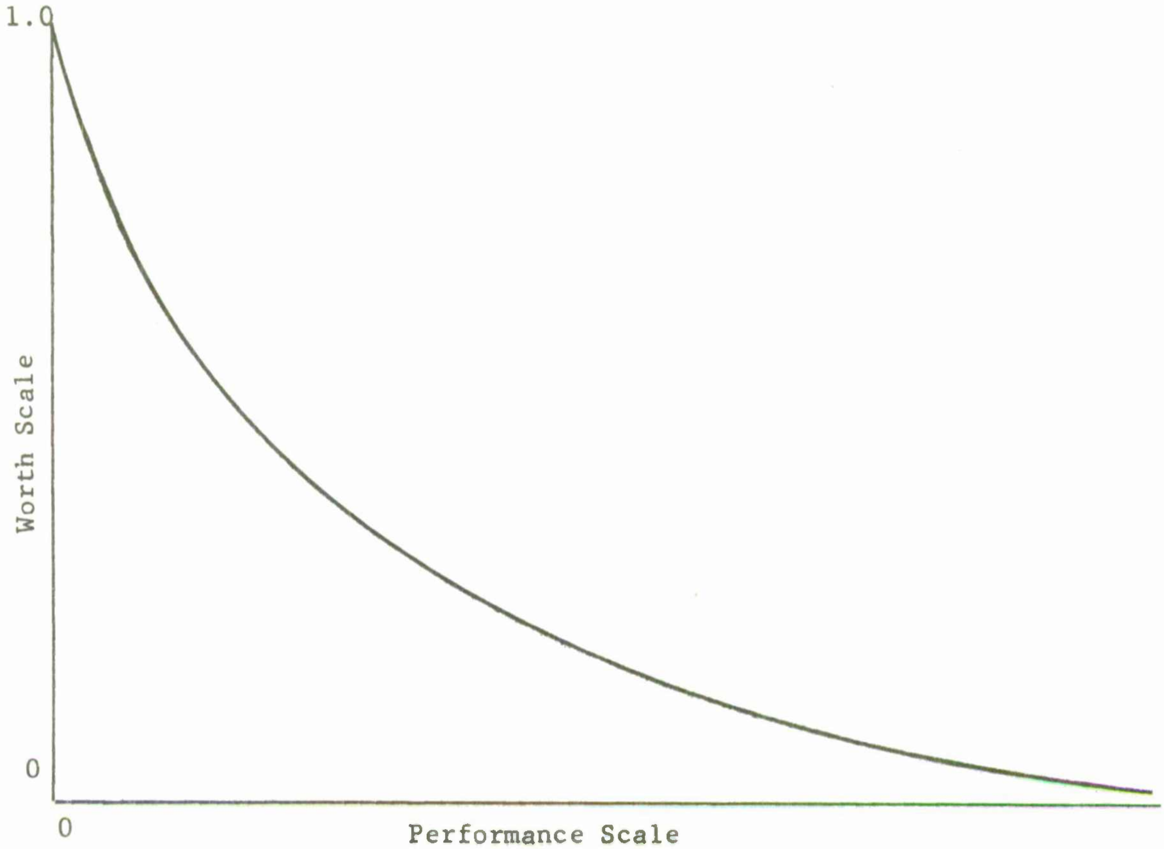


FIGURE 13

APPENDIX XIX

SCORING PROCEDURE 19

References in text: Section 7.5.1.3

The performance measure under scrutiny has been determined to have the following characteristics:

1. continuous scale;
2. bounded from below by zero;
3. no logical upper bound;
4. reverse preference relationship;
5. worth score zero assigned to infinite performance;
6. worth score one assigned to zero performance;
7. first accelerating, then decelerating rate of change of worth with increases in performance.

A graphical picture of this general shape of scoring function appears below in Figure 14.

<u>Step 1</u>. There is no simple, standardized equation to fit all situations of this type. Although the general shape of this scoring function is given by the equation

$$\text{worth score} = 1 - \exp\left[(-a)\,(\text{measured performance})^{-b}\right],$$

where exp is the exponential function with basis e = 2.718, and both a and b are positive fitting constants ($b \geq 1$),

still, the exact values of the fitting constants cannot be determined in a standard way for all performance measures. Consequently, proceed to scoring procedure 20.

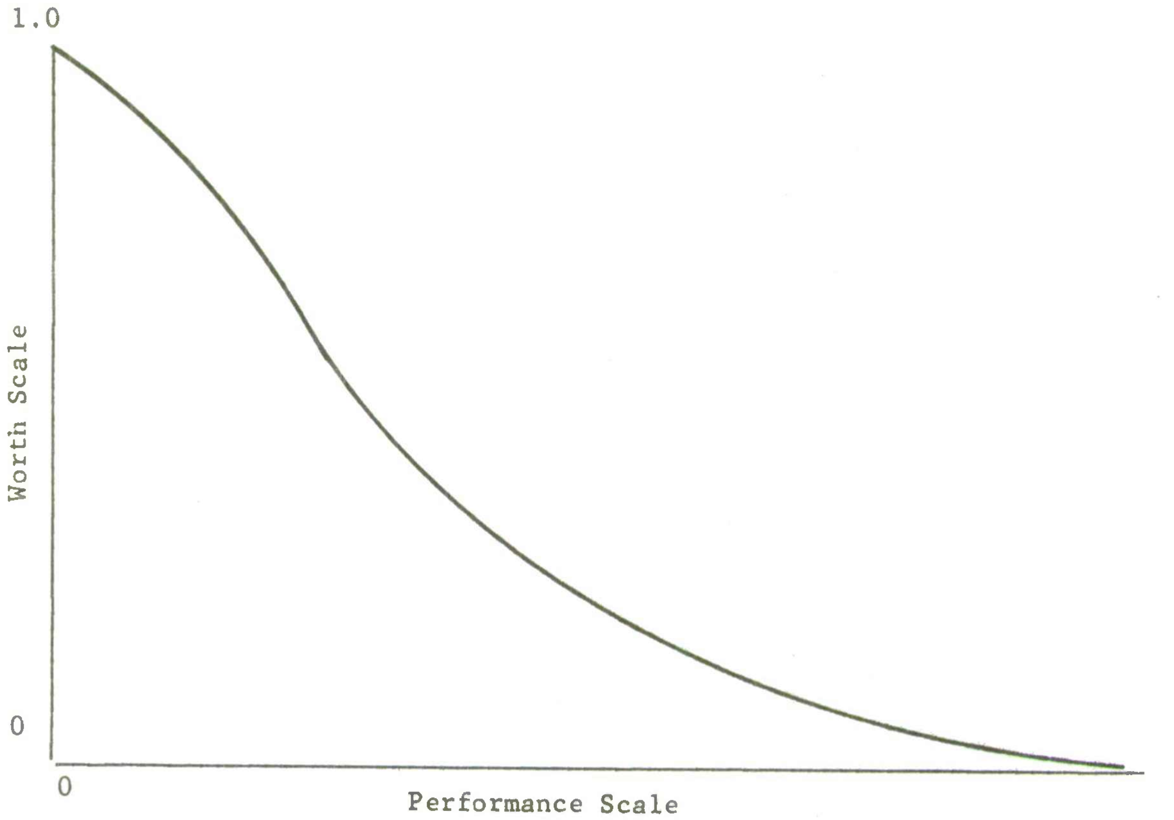


FIGURE 14

APPENDIX XX

SCORING PROCEDURE 20

References in text: Section 7.5.1.3 and Appendixes VII through XIX.

This procedure constitutes a continuation of each of the previous procedures listed below:

1. procedure 7
2. procedure 8
3. procedure 9
4. procedure 10
5. procedure 12
6. procedure 13
7. procedure 14
8. procedure 15
9. procedure 16
10. procedure 17
11. procedure 18
12. procedure 19.

The general shape of the scoring function to be formulated has already been determined and inspected visually in one of these previous procedures. The purposes of this procedure is to select a specific curve of the general shape already determined.

Step 1. Assemble one or more decision makers.

Step 2. Prepare a standard sheet of graph paper for each decision maker layed out and marked off in the following manner.

1. Lay the worth scale along the vertical axis of a cartesian coordinate plane.

2. Mark off zero worth points at the origin of the graph and one worth point on the vertical axis near the top of the graph.

3. Mark off tenths of a point at equally-spaced intervals along the vertical axis between zero and one worth point.

4. Lay the performance scale along the horizontal axis.

5. Mark off zero performance at the origin of the graph and either the logical upper bound (if one exists) or some amount of performance substantially in excess of (say 50 percent greater than) the anticipated maximum produced performance on the horizontal axis near the right-hand edge of the graph.

6. Establish convenient, equally-spaced performance subdivisions along the horizontal axis, and mark these off.

Step 3. Each decision maker will then ask himself the following question. "What level of performance, if promised by an alternative, should be considered ten percent successful in satisfying the related lowest-level performance criterion?" Indicate this level of performance by placing an "x" in the interior of the graph at the position corresponding to that estimated level of performance along the horizontal performance scale and the ten percent or one-tenth worth point level along the vertical worth scale.

Step 4. Repeat Step 3 for the twenty percent, thirty percent, forty percent, fifty percent, sixty percent, seventy percent, eighty percent, and ninety percent worth levels, respectively.

Step 5. Each decision maker should now have on his sheet of graph paper nine "x" marks. If the performance measure for which a scoring function is being formulated possesses a logical upper bound, two additional "x" marks may be placed on the graph - one at zero performance, and the other at the logical upper bound. If the performance measure possess no logical upper bound, only one additional "x" mark may be placed on the graph corresponding to zero performance. Place the additional "x" mark(s) on the graph.

Step 6. Collect the graphs from each separate decision maker. Compute the (possibly weighted) arithmetic mean (averaged over separate decision makers) for each of the nine percentage levels along the worth scale.

Step 7. Prepare a new sheet of graph paper identical to the sheets prepared in Step 2.

Step 8. Plot the nine average points computed in Step 6 on this new sheet prepared in Step 7.

Step 9. With the aid of a French curve, draw a smooth curve of the predetermined general shape through the average points plotted in Step 8. The result is a scoring function in graphical form.

Step 10. To use this graphical scoring function, note the actual amount of performance promised by an alternative, and read the corresponding point score directly off the graph.

This completes the procedure.

BIBLIOGRAPHY

Churchman, C. W., et al, Introduction to Operations Research, New York, John Wiley and Sons, 1957.

Churchman, C. W., Prediction and Optimal Decision: Philosophical Issues of a Science of Values, Englewood Cliffs, N. J., Prentice Hall, 1961.

Cronbach, L. J., and G. C. Gleser, Psychological Tests and Personnel Decisions, Urbana, Illinois, University of Illinois Press, 1965.

Eckenrode, R. T., "Weighting Multiple Objectives", Management Science, 12, 3 (1965), 180-192.

Grayson, C. J., Jr., Decisions Under Uncertainty: Drilling Decisions by Oil and Gas Operators, Boston, Division of Research, Graduate School of Business Administration, Harvard University, 1960.

Miller, J. R., The Assessment of Worth: A Systematic Procedure and its Experimental Validation, Cambridge, Unpublished Doctoral Dissertation, Massachusetts Institute of Technology, 1966.

Raiffa, H., and R. Schlaifer, Applied Statistical Decision Theory, Boston, Division of Research, Graduate School of Business Administration, Harvard University, 1961.

Pound, W. H., "Research Project Selection: Testing a Model in the Field," IEEE Transactions in Engineering Management, 11 (1964), 16-22.

Rosenthal, S., "Analytical Technique for Automatic Data Processing Equipment Acquisition", 1964 Spring Joint Computer Conference, 1964.

Schlaifer, R., Probability and Statistics for Business Decisions, New York, McGraw-Hill Book Company, Inc., 1959.

Smith, N. M., Jr., "A Calculus for Ethics: A Theory of the Structure of Value", Behavioral Science, 1 (1956), 111-142 and 186-211.

Von Neumann, J., and O. Morgenstern, Theory of Games and Economic Behavior, 1st ed., Princeton, N. J., Princeton University Press, 1944.

Yntema, D. B., and W. S. Torgerson, "Man-Computer Cooperation in Decisions Requiring Common Sense," IRE Transactions on Human Factors in Electronics, HFE-2, 1 (1961), 20-26.

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)* The MITRE Corporation Bedford, Massachusetts	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE A SYSTEMATIC PROCEDURE FOR ASSESSING THE WORTH OF COMPLEX ALTERNATIVES		
4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)* N/A		
5. AUTHOR(S) *(First name, middle initial, last name)* Miller, James R. III		
6. REPORT DATE November 1967	7a. TOTAL NO. OF PAGES 198	7b. NO. OF REFS 13
8a. CONTRACT OR GRANT NO. AF 19(628)-5160 b. PROJECT NO. 8510 c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) ESD-TR-67-90 9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)* MTR-260	
10. DISTRIBUTION STATEMENT This document has been approved for public release and sale; its distribution is unlimited.		
11. SUPPLEMENTARY NOTES N/A	12. SPONSORING MILITARY ACTIVITY EDP Equipment Office, Electronic Systems Division, L. G. Hanscom Field, Bedford, Massachusetts	

13. ABSTRACT

This paper addresses itself to the problem of assessing worth. It is assumed that a decision context has been specified and that a fixed set of discrete alternatives has been produced. It then remains to assess the worth of each alternative, to estimate the resource drains required by each, and to combine these considerations, along with considerations of risk/uncertainty, so as to arrive at a final decision. The bulk of this paper is directed toward worth assessment. To aid in the assessment process, a detailed procedure has been devised. The purpose of this procedure is set forth, and step-by-step instructions for its actual implementation are presented. A live instance of its complete application is also provided for illustrative purposes. Major conclusions are that the procedure can be carried out successfully--at least by professional decision makers in a laboratory setting--and that all phases of it exert an important impact upon the decision making process. Critical to its overall impact, and particularly to its success, are its quantitative aspects.

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
RESOURCE ALLOCATION						
Systematic Procedure for						
Complex Alternatives, Assessment of Worth						
Problem Definition						
Constraints						