R-2719-AF

Subjective Transfer Function Approach to Complex System Analysis

Clairice T. Veit and Monti Callero

March 1981

Rand

PROJECT AIR FORCE

The research reported here was sponsored by the Directorate of Operational Requirements, Deputy Chief of Staff/Research, Development, and Acquisition, Hq USAF, under Contract F49620-77-C-0023. The United States Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation hereon.

The Rand Publications Series: The Report is the principal publication documenting and transmitting Rand's major research findings and final research results. The Rand Note reports other outputs of sponsored research for general distribution. Publications of The Rand Corporation do not necessarily reflect the opinions or policies of the sponsors of Rand research.

Published by The Rand Corporation

R-2719-AF

Subjective Transfer Function Approach to Complex System Analysis

Clairice T. Veit and Monti Callero

March 1981

A Project AIR FORCE report
Prepared for the United States Air Force

Rand
SANTA MONICA, CA. 90406

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED

PREFACE

This report summarizes major problems with commonly used approaches to subjective measurement and describes recently developed techniques for resolving some of the difficulties. The resolutions discussed involve using experimental designs to obtain judgment data that allow *tests* of underlying judgment models and thus provide the constraints needed to induce metric scale values from ordinal information.

The report describes a new approach to complex system analysis—the subjective transfer function approach—which incorporates these experimental designs. Tactical air command and control is the complex system used to illustrate major features of the analytic technique. The key points should be of interest to those involved in the Air Force tactical air command and control and force employment system as well as to other agencies responsible for formulating and using subjective measurement techniques.

The work was performed under the Project AIR FORCE research project "Tactical Air Command and Control."

SUMMARY

This report describes the subjective transfer function approach to complex system analysis. This approach resolves major problems encountered with approaches to subjective measurement currently being used to evaluate complex systems.

Commonly used approaches to complex system analysis include "direct" scaling techniques and typical applications of multiple regression and decision analyses. In these approaches, major premises underlying conclusions about subjective processes (those that cannot be observed directly but are inferred from observed judgments) cannot, in principle, be tested within the framework.

The resolution of the testability problem lies in the major features of the algebraic modeling approach: (a) factorial experimental designs, (b) tests of proposed subjective algebraic judgment models, and (c) derivation of subjective measures from appropriate models. The basic idea in this approach to measurement is that factorial designs allow *tests* of the predictions of the proposed judgment model and provide the constraints needed to induce metric scale values from ordinal information. The model describes how components of a complex system affect judgments of an outcome. A proposed model is accepted as the appropriate description of judgments when the judgments obey the predictions of the model. Subjective measures of stimuli and responses are derived from an *appropriate* model and have meaning with respect to that model.

These basic ideas are incorporated in the subjective transfer function approach to complex system analysis. The subjective transfer function approach has additional features especially developed for complex system analysis in which causes and effects of numerous variables on judged outcomes have to be explained. In this report we use the Air Force tactical air command and control and force employment system to illustrate measurement problems, resolutions, and features of the subjective transfer function approach.

ACKNOWLEDGMENTS

The authors wish to thank Barbara Rose, Michael Birnbaum, and Ted Donaldson for helpful comments on earlier drafts of this report.

CONTENTS

PREFACE iii

SUMMARY v

ACKNOWLEDGMENTS vii

Section

I. INTRODUCTION 1
- Outline of the Report 1
- Tactical Air Command and Control Force Employment Representation 2

II. MEASUREMENT APPROACHES COMMONLY USED IN EVALUATING COMPLEX SYSTEMS 4
- Direct Scaling Framework 5
- Multiple Regression Analyses 6
- Decision Analysis 6
- Comments 7

III. THE ALGEBRAIC MODELING APPROACH TO SUBJECTIVE MEASUREMENT 8

IV. SUBJECTIVE TRANSFER FUNCTION APPROACH 13
- Defining a Complex System Representation 13
- Obtaining Transfer Functions 17
- Complex System Comparison 21
- Comments 26

Appendix

A. TACTICAL AIR COMMAND AND CONTROL AND FORCE EMPLOYMENT 27

B. "DIRECT" SCALING FRAMEWORK 30

REFERENCES 37

I. INTRODUCTION

A system is termed "complex" when numerous different components are needed to adequately describe it. Cause and effect relationships among many of the components of a "complex" system must be understood before one can understand what affects the important outcomes of the system as a whole. A system's effectiveness is assessed according to these outcomes.

System effectiveness is often evaluated using subjective measurement techniques. Typically, "experts" are asked to use a numerical scale to respond to questions about particular aspects of a system. Measurement techniques are termed "subjective" when they require interpretation of experts' responses in terms of processes that cannot be directly observed. Response interpretations usually concern (a) the "subjective scale values" associated with specified characteristics of the system and/or (b) how subjective values of the system's characteristics affect judgments of a system's outcome (which requires specifying the expert's underlying judgment model). Response interpretations are usually used as input to operational and management decisions.

Subjective measurement evaluations of complex systems are encountered in both civilian and military sectors. Within the Air Force, subjective measurement is the primary method used by Mission Area Analysis in support of the planning, programming and budgeting process, and it is being considered for application to long range planning as well as for evaluating tactical air command and control. It is therefore important to develop sound measurement techniques that allow *tests* of causal theories about judgments of system effectiveness.

The purpose of this report is to describe a new approach to complex system analysis—the *subjective transfer function* approach. This approach allows *tests* of judgment theories of complex systems, thereby resolving major measurement problems encountered with commonly used approaches.

The subjective transfer function approach to complex system analysis was developed during research on evaluating the command, control, and employment system for the United States Air Force tactical air forces. Consequently, that complex system will be used to discuss subjective measurement issues and describe the subjective transfer function approach.

OUTLINE OF THE REPORT

In the remaining portion of this section, we present a simplified version of the components and component interrelationships that make up a tactical air command and control and force employment representation. The variables in this representation will be used for illustrative purposes throughout the report.

In Section II we draw on a body of literature to discuss flaws in measurement approaches commonly used to evaluate complex systems, including "direct" scaling and typical applications of multiple regression and decision analysis. The basic measurement problem with these methods is that conclusions concerning subjective scale values and models rest on premises that are untested in practice and, more important, untestable in principle.

In Section III, we describe and illustrate the algebraic modeling approach to subjective measurement, which provides a framework for resolving the testability problem identified in Section II in fairly simple situations where only a few components are involved.

In Section IV, we describe our subjective transfer function approach to complex system analysis. This measurement approach resolves the problems identified in Section II for a system composed of numerous components, or a "complex" system. The measurement problems are resolved by incorporating the experimental designs that characterize the algebraic modeling approach and by adding design features necessary to functionally interrelate components of a complex system. There are three parts to the subjective transfer function approach: (1) defining a complex system representation, (2) obtaining subjective transfer functions, and (3) using the transfer functions for comparative system analysis.

TACTICAL AIR COMMAND AND CONTROL AND FORCE EMPLOYMENT REPRESENTATION

Tactical air command and control is the means by which an air commander brings tactical air forces to bear against an enemy in war. As such, it directly and vitally influences the force employment, or the effectiveness of the tactical air forces in accomplishing military objectives and thereby substantially influences the overall conflict outcome. Hence, it is important to determine how well our tactical air command and control processes and systems can meet wartime requirements and how changes in those processes and systems would affect command and control effectiveness. Evaluating command and control was the primary goal in developing the subjective transfer function approach.

A simplified conceptualization of tactical air command and control and force employment is shown in Fig. 1. This representation depicts a network of hypothesized system components and their interrelationships. The components of this graphic representation will be used throughout the paper to illustrate and discuss measurement issues. For those unfamiliar with the nature of Air Force tactical air command and control and force employment, a detailed description of each of the components shown in Fig. 1 is presented in Appendix A.

The system has been structured into four tiers. At the highest tier is a particular land battle in which tactical air forces would be employed to help gain a favorable outcome. The next tier down represents the specific Tactical Air Operations (TAOs) the tactical air forces would perform: Close Air Support (CAS), where tactical aircraft attack enemy ground forces that are in contact with friendly ground forces; Interdiction, where tactical aircraft attack enemy forces and resources well behind the main battle line; and Airlift, where tactical transports deliver men, munitions, weapons, and equipment to the forces involved in the battle.

The next two tiers—Functions and Elements—characterize tactical air command and control. Command and control affects the TAOs by Planning each operation ahead of time, Directing specific units to perform each operation, and Controlling (monitoring and adjusting) each operation during execution of the plan. These Functions—Planning, Directing, and Controlling—must each meet different specific requirements for each different TAO.

The bottom tier represents the Elements used to perform the functions. For this simplified illustration we selected the following: the Friendly Information and Enemy Information coming into the command and control system; the Processes by which information is made available for use in the system; and the Communications (COMM) used to give directions to the tactical forces.[1]

[1]Components in the Function and Element tiers in Fig. 1 that are labeled the same are to be considered different. For example, Planning Interdiction operations and Planning Airlift operations address greatly different environments, goals, tactics, etc., and hence require different considerations, techniques and procedures. Thus, this hierarchy has 24 different Element components that impact through the intermediate Function and TAO components on the Land Battle.

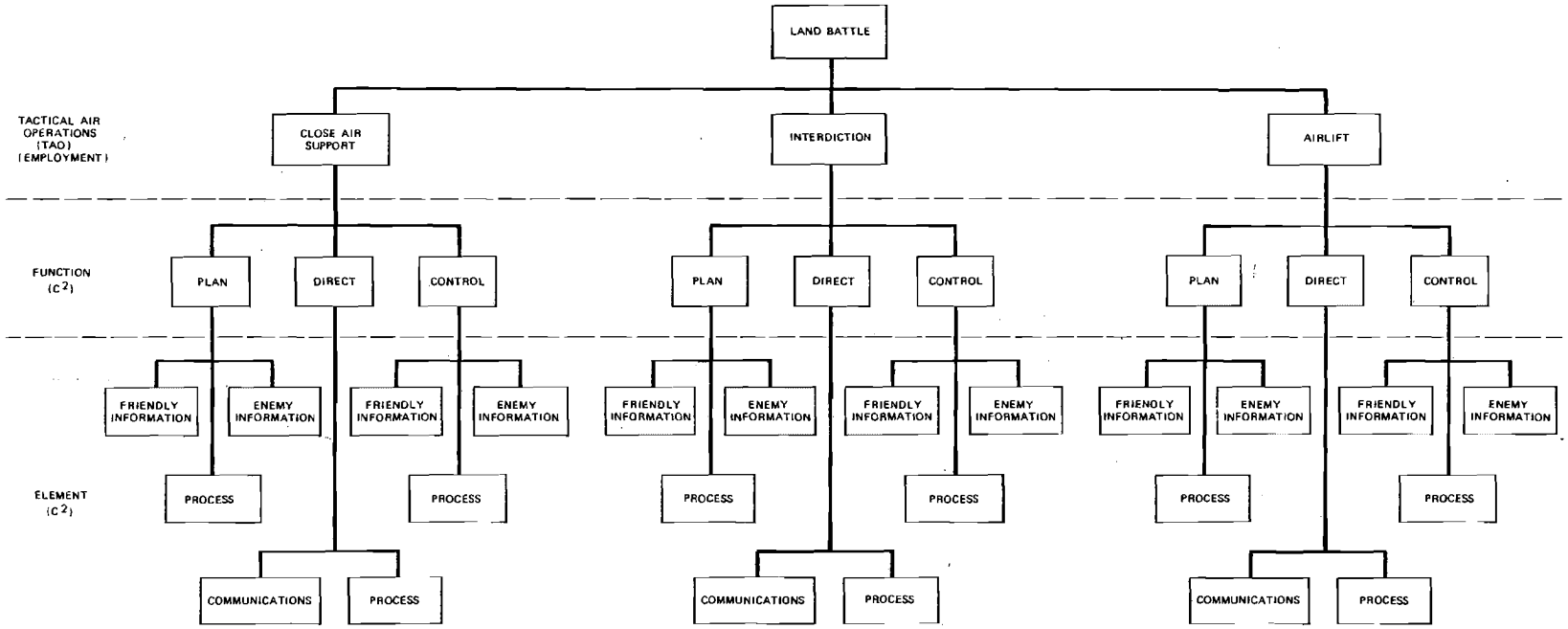
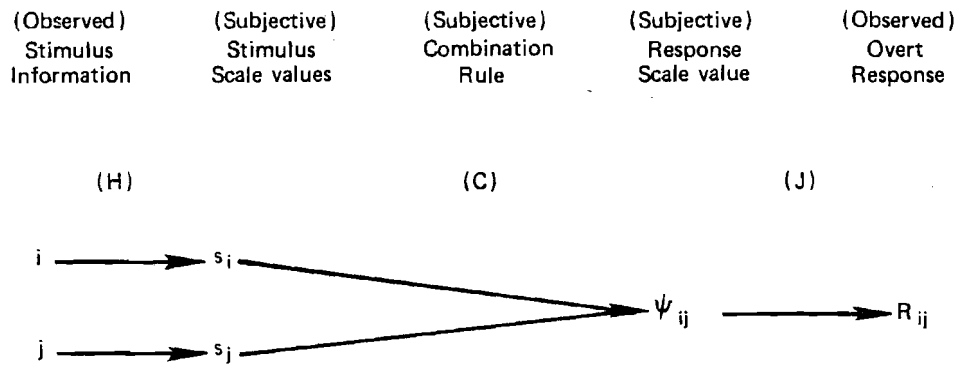


Fig. 1—Simplified conceptualization of a possibleAir Force command and control (C^2) and force employment representation

II. MEASUREMENT APPROACHES COMMONLY USED IN EVALUATING COMPLEX SYSTEMS

Some of the more commonly used subjective measurement techniques in complex systems analysis include so-called "direct" scaling techniques, multiple regression analyses, decision analyses, and various combinations of these approaches. Problems with these techniques have been discussed in detail elsewhere (Anderson, 1974; Birnbaum, 1973, 1974; Birnbaum and Veit, 1974; Krantz, 1972; Shepard, 1976; Veit, 1978). The problems have to do with the testability of conclusions about subjective scale values and judgment models, and can be described in terms of the outline shown in Fig. 2.



The function H transforms each stimulus (i and j) into a subjective value (s_i and s_j); the function C is the algebraic model respondents use to combine the scale values into a subjective response, ψ_{ij}; J transforms the subjective response into an observed response, R_{ij}.

Fig. 2—Outline of subjective processes

Figure 2 proposes that for two pieces of information (stimuli), i and j,[1] presented to a respondent (for example, in a questionnaire item), three subjective processes occur within the stimulus-response interval. First, the two pieces of stimulus information describing characteristics of the system are transformed by the function H to subjective scale values, s_i and s_j. Second, the scale values are combined by the function C to form an overall subjective response, Ψ. This function is the model that specifies how the scale values affect the subjective response. Third, the subjective response is transformed into the observed response, R, by the function J (the judgment function). All three of these functions are subjective in the sense that

[1]The outline can easily be extended to include a number of stimuli.

they can only be inferred from what is observed—the stimulus information (i and j) and the response (R).

Any complete theory of judgment must specify all three subjective processes. These specifications are credible if they result from empirically verified hypotheses. It is important that they be credible since the purpose of using human judgments for evaluating complex systems is to provide information for making operational and management decisions. The next sections discuss testability problems resulting from the methods currently used to evaluate complex systems.

DIRECT SCALING FRAMEWORK

In the "direct" scaling framework proposed and developed by S. S. Stevens (1946, 1957, 1971), the questions posed are generated from experimental designs that manipulate a single factor (single-factor designs).[2] For example, in Fig. 1, Close Air Support might be the single factor. To be manipulatable, a factor must have several levels (i.e., values or categories along the factor continuum). The levels of Close Air Support could be performance descriptions—good, fair, and poor. This example of a single-factor design would thus have three factor levels while other components in the system described in the questionnaire scenario would be held constant at one level. The task posed to respondents (e.g., "experts") might be to judge the value of each Close Air Support level in gaining a favorable outcome in a specified land battle using a given numerical scale.[3]

Manipulation of a single factor allows assessment of its effect on judgments at the constant level of the other factors included in the questionnaire scenario. However, this information is rarely of interest, and, in fact, levels of factors are usually selected to ensure that main effects will occur. The interest is usually in obtaining the subjective scale values associated with each level of the manipulated factor. These scale values are assumed to be "directly" related to the numbers given as responses. It is also assumed that the function used by respondents to combine these subjective values (C in Fig. 2) follows the form dictated by task instructions. (Typical instructions are to report the "ratio" of two factor levels or the "interval" between two factor levels.) From this assumption, it is further concluded that the scale properties of the numbers ("scale values") are what might be expected under that instructional model; ratio properties[4] are usually assumed for numbers resulting from a ratio task and interval properties for numbers resulting from an interval task. The major problem with these conclusions is that they are, in principle, untestable in this single-factor design framework. The framework does not provide the design constraints necessary for determining the subjective stimulus scales (s_i, s_j), the subjective response scales (Ψ_{ij}), and the model (C) from the observed responses (Krantz and Tversky, 1971; Anderson, 1974; Birnbaum and Veit, 1974; Shepard, 1976; Veit, 1978).

For example, to test a ratio model for "ratio" judgments, it should be possible to determine if

$$R^R_{ij} = \frac{R^R_{ik}}{R^R_{jk}} \quad ,$$

[2]For a detailed discussion of this approach, see Appendix B.

[3]Many single-factor designs could be extracted from the representation shown in Fig. 1. Each component could be treated as a factor with qualitative descriptions as factor levels (as described above for Close Air Support); or each tier could be treated as a factor with the components defining the tier as the factor levels. Decisions on how to define the factors or variables of a representation depend on the hypotheses under consideration.

[4]It has commonly been held that a ratio model yields scale values to a ratio scale. However, scale values derived from a ratio model yield numbers with log-interval scale properties (Krantz, Luce, Suppes, and Tversky, 1971).

where R^R represents the "ratio" response. A subtractive model for "difference" (interval) judgments predicts that

$$R^D_{ij} = R^D_{ik} - R^D_{jk} ,$$

where R^D represents the "difference" response. A test of these simple predictions requires the additional constraint of a second factor, k, to be manipulated in the design; that is, it requires at least a two-factor experimental design. Section III, which discusses the algebraic modeling approach to measurement, describes how hypothesized models (C in Fig. 2) can be tested with appropriate designs and how, once the model is known, subjective stimulus (s_i, s_j), and response (Ψ_{ij}) scale values can be derived separately from the model.

MULTIPLE REGRESSION ANALYSES

Problems with the typical use of multiple regression to explain judgment data have been discussed extensively by Birnbaum (1973, 1974a). When multiple regression is the data analytic technique, the subjective combination function (C) is usually assumed to be some form of the linear multiple regression model. "Subjective" values of predictor variables are sometimes obtained from "direct" scaling techniques; physical values are often used when stimuli are measured on the physical continuum.

The basic problem is that both the subjective combination model and the subjective scale values are unknown. The experimental designs typically used in this research do not provide the necessary constraints to *test* the hypothesized form of the linear regression model; nor do they provide a means for verifying the "correctness" of the "direct" scales or physical values used as input to "test" the model.[5] Indices of goodness-of-fit (e.g., R^2_i) are usually used to assess the model. But, such goodness-of-fit indices may be misleading since they can be high when *deviations* from model predictions are significant and systematic (Anderson, 1971), and higher for an incorrect than a correct model (Birnbaum, 1973, 1974a).

DECISION ANALYSIS

The typical application of decision analysis uses "direct" scales in the subjective expected utility (SEU) model. The SEU model proposes that choices should be or are (depending on whether the model is thought of as a prescriptive or descriptive theory) made by maximizing the sum of the products of utility and probability associated with the outcomes; that is, given a choice between *m* alternatives, it is proposed that people choose (or should choose) the one that maximizes

$$SEU = \sum_{i=1}^{m} w_i u_i , \qquad (1)$$

where w_i and u_i correspond, respectively, to the subjective weight and subjective utility (scale value) of the ith outcome, and $\Sigma w_i = 1$. Multiattribute utility (MAU) theory extends the SEU model to choices between probabilities of outcomes, each of which has multiple attributes.

[5]When input values are physical measures, the untested assumption is that H in Fig. 2 is an identity function.

Decision analysts interested in complex system evaluation usually use Eq.(1) as a prescriptive theory. The model serves to link the components throughout a hierarchical representation (e.g., Fig. 1) to a final outcome (e.g., the Land Battle). Values used as input to the model are usually "direct" scales (see, for example, recommendations presented by Gardiner and Edwards (1975)) and physical values (e.g., probabilities) associated with the stimulus outcomes. Both input values lack validity. The same problem of validating that we discussed in relation to "direct" scales exists with physical values such as probabilities. In this latter case, the untested assumption is that the physical values are the *same* as their subjective counterparts; that is, that H in Fig. 2 is an identity function. Since the procedures used with this approach do not provide a way to validate values (the weights (w) and utilities (u) in Eq. (1)) used as input to the model, there is no way of knowing whether prescribed choices are those that would be "prescribed" by the model.

COMMENTS

All three of the methods described above are commonly used in complex system evaluation. None of the methods employs experimental designs that provide the constraints needed to *test* hypotheses about subjective scale values and/or combination functions. Thus, conclusions about these subjective events are based on untested assumptions.

In the next section we use illustrations to demonstrate why a single-factor design is not sufficient for deriving scales or testing combination functions. We also show how questionnaires can be generated from experimental designs that allow tests of the hypothesized combination function (model). Subjective scale values are derived from the model when the data obey the model's predictions. The model validates the scales.

III. THE ALGEBRAIC MODELING APPROACH TO SUBJECTIVE MEASUREMENT

The algebraic modeling approach to subjective measurement resolves the major problem of testability encountered with the approaches described above. However, this approach is not practical for complex systems involving many variables (factors) and interlinking causal hypotheses. In this section, we describe the basic ideas and experimental designs that characterize the algebraic modeling approach. In the next section, we describe how these ideas and experimental designs are incorporated (along with additional design features) into the subjective transfer function approach for complex system analyses.

The basic idea behind the algebraic modeling approach to subjective measurement is to use experimental designs to generate questionnaires that allow tests of the hypothesized combination model (C in Fig. 2). When judgment data satisfy the predictions of the model, subjective scale values (s_i, s_j, and Ψ in Fig. 2) are derived from the model. Thus, the model that specifies *how* the stimulus scale values affect judgment is the empirical validation base for those values.

Factorial combinations of stimuli[1] are a key design feature in model testing. When questionnaires are generated from factorial designs, crucial predictions of hypothesized combination models can be tested. The following example illustrates the main ideas of the approach.

Suppose you wanted to know how performance of different tactical air operations affected the "expert's" judgment of their value in bringing about a favorable outcome to a specified land battle. Suppose further that the level of performance for each TAO could be good, fair, or poor.

It might initially be hypothesized that the subjective response (Ψ in Fig. 2) was the simple sum of the separate values placed on each TAO performance—a simple additive model. For two TAOs, Close Air Support and Interdiction, the additive model may be written:

$$\Psi_{CAS_i,INT_j} = s_{CAS_i} + s_{INT_j}, \tag{2}$$

where s_{CAS_i} and s_{INT_j} are the respective scale values for the ith and jth performance levels of Close Air Support and Interdiction, and Ψ_{CAS_i,INT_j} is the subjective response scale value. Figure 3 shows a factorial design that would allow a test of this hypothesis. Close Air Support and Interdiction are the two factors, and their possible performances (good, fair, or poor) are the three factor levels for both factors. A questionnaire generated from this design would consist of nine questions (stimulus items). Each item would describe the performance of Close Air Support and Interdiction for a specified Land Battle. For each item, experts might be instructed to judge the overall value of the performance of these two TAOs in bringing about a favorable outcome to the Land Battle using a 9-point category rating scale. A one would be used if the two TAO performances seemed to be not at all valuable in effecting a favorable outcome, a nine would be used if they seemed very valuable, and the other numbers in the scale would be used for judgments falling between the two extremes.

Hypothetical data (individual, mean, or median responses) for this task are shown in panel A of Fig. 4. Panel B of Fig. 4 shows a plot of the data shown in panel A as a function of Interdiction performance level, with a separate curve for each Close Air Support performance

[1]In factorial designs, every level of each factor is combined with every level of every other factor.

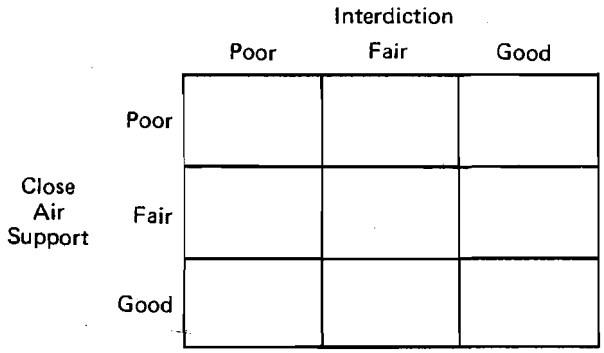


Fig. 3—Factorial design of Close Air Support and Interdiction

level. The slopes of the curves represent the effect of Interdiction performance on judged value; separations between the curves represent the effect of Close Air Support performance. When the function relating subjective and observed responses (J in Fig. 2) is assumed to be linear,[2] the additive model of Eq. (2) predicts that the curves in Panel B should be parallel; that is, the vertical distance between any two points of any two Close Air Support curves should be the same, independent of the level of Interdiction (value on the x-axis). These data are perfectly consistent with this parallelism prediction.[3] If obtained data plotted as in Panel B of Fig. 4 revealed systematic deviations from parallelism, the additive model would be rejected.[4] Note that if only one factor were used in the design as in the "direct" scaling framework, only one of the curves shown in panel B of Fig. 4 would be obtained. It is not possible to test the parallelism prediction with only one curve (one factor).

When data are consistent with the predictions of the hypothesized model, the subjective stimulus scale values are least squares estimates under the model. For the additive model, these are the row and column marginal means for the row and column stimuli, respectively (see Fig. 4A).[5] The subjective responses (Ψ) are the cell values.

If responses to the task described above turned out as in Fig. 5A, the additive model would be rejected as an explanation of the expert's combination model. A plot of these data (Fig. 5B) reveals a systematic divergent interaction. Thus, for these data, neither the marginal means nor the cell values would contain any special meaning.

Upon close inspection of the data shown in Fig. 5B, it would be discovered that the divergent interaction followed the particular form predicted from a range model (Birnbaum and Stegner,

[2]This assumption can be tested using additional experimental design features to the simple factorial design shown in Fig. 3. (For details see Birnbaum (1974b), Birnbaum and Veit (1974b), and Veit (1978).)

[3]The apparent convergence of the curves toward high Interdiction values on the x-axis is illusory as can be seen by measuring the distances between the curves.

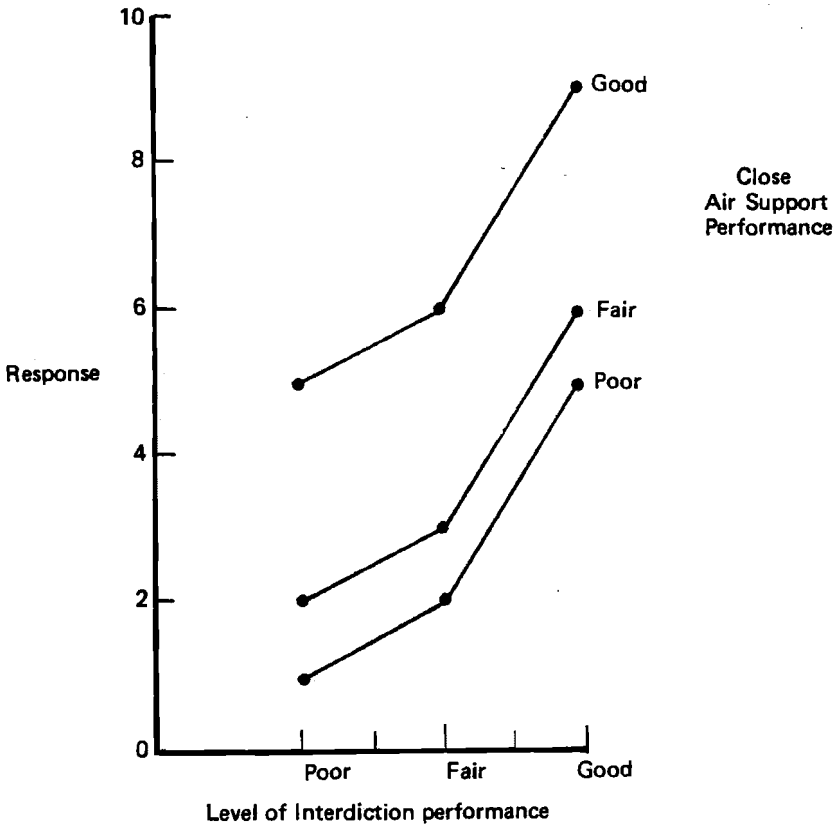
[4]The hypothetical data graphed in panel B of Fig. 4 are ideal. In reality, data would contain error. Graphic analyses aid in determining if error is random or contains systematic deviations from model predictions. An Analysis of Variance test of the interaction provides a statistical test of the additive model.

[5]Under an additive model, the row marginal means in the factorial design are linearly related to the scale values of the row stimuli, and the column marginal means are linearly related to the scale values of the column stimuli.

PANEL A

Close Air Support	Interdiction: Poor	Fair	Good	Row Marginal Means
Poor	1.0	2.0	5.0	2.67
Fair	2.0	3.0	6.0	3.67
Good	5.0	6.0	9.0	6.67
Column Marginal Means	2.67	3.67	6.67	

PANEL B



PANEL A Cell values could be mean, median, or an individual subject's responses.

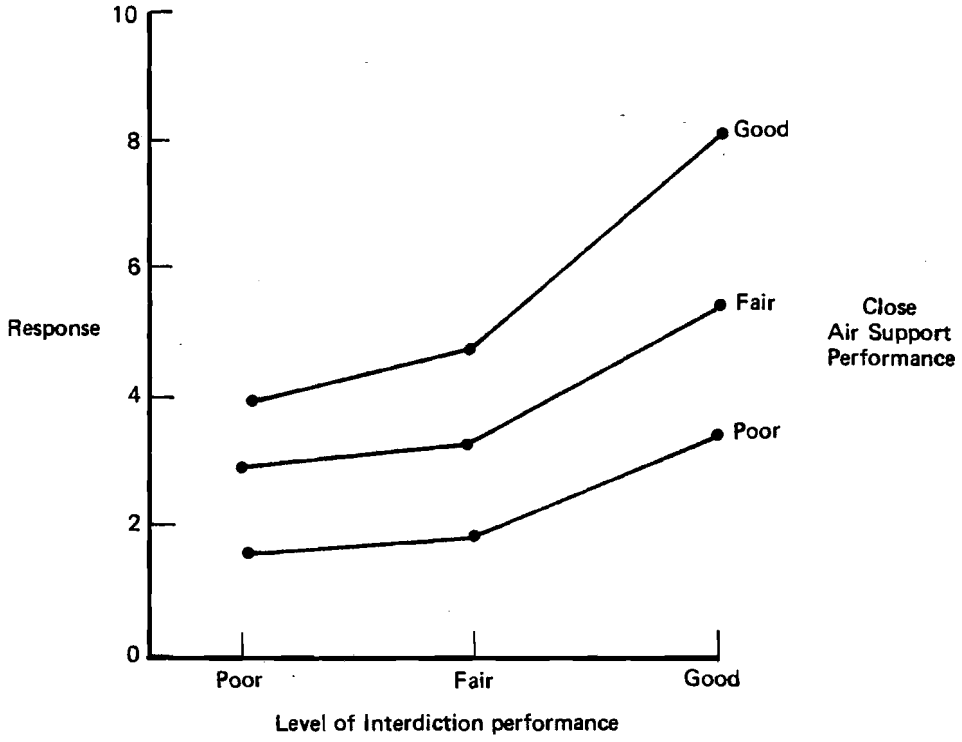
PANEL B Responses shown in Fig. 4, are plotted as a function of Interdiction level with a separate curve for each Close Air Support level; the parallel curves are predicted by the additive model.

Fig. 4—Hypothetical data consistent with the additive model

PANEL A

Close Air Support \ Interdiction	Poor	Fair	Good
Poor	1.6	1.9	3.5
Fair	3.0	3.4	5.5
Good	4.0	4.9	8.1

PANEL B



PANEL A Cell values could be mean, median, or an individual subject's responses.

PANEL B Graph of hypothetical data shown in Fig. 6; the divergent interaction infirms the additive model.

Fig. 5—Hypothetical data that violate an additive model

1979, 1980). The range model predicts that on a particular trial the response results from the following subjective process:

$$\Psi_{CAS_i,INT_j} = \frac{w_0 s_0 + w_{CAS} s_{CAS_i} + w_{INT} s_{INT_j}}{w_0 + w_{CAS} + w_{INT}} + \omega(s_{max} - s_{min}) , \quad (3)$$

where $w_0 s_0$ is the weight and scale value associated with the initial impression (what the judgment would be in the absence of specific information); w_{CAS_i} and w_{INT_j} are the subjective weights associated with Close Air Support and Interdiction, respectively; s_{CAS_i} and s_{INT_j} are the subjective scale values associated with the ith and jth levels of Close Air Support and Interdiction, respectively; s_{max} and s_{min} are the highest and lowest valued stimuli, respectively, in the ijth set; and ω is an empirical constant that represents the magnitude of the range effect.

The range model predicts that for each item presented for judgment, respondents take a weighted average of the stimuli but alter this average by taking into account the subjective range of the information contained in the item. Thus, this model proposes that the extremity of the information affects the judgment. Once the model is known, scale values (least squares estimates) can be derived from it.

When human judgments are used to evaluate complex systems, it is vital to require that conclusions regarding their meaning be based on tested premises. The algebraic modeling approach to subjective measurement resolves the testability problem encountered with other approaches presently used to evaluate complex systems. This resolution lies in generating questionnaires from experimental designs that allow tests of the respondent's combination model.

The key design feature to testing model predictions is the factorial design.[6] One problem with factorial designs is that the burden on the respondent increases rapidly as the number of factors and factor levels are increased. For example, suppose that the Air Force wanted to know how changes in the Element components shown in Fig. 1 affect the expert's perceived outcome of the Land Battle. Since the interest is on *causal and perceptual* links among the variables, the answer requires experimental manipulation of the Elements in designs that allow tests of hypothesized judgment models. If each of the 24 Elements were treated as a factor and each factor had three factor levels (for example, enemy information could be 9, 5, or 2 hours old), a questionnaire generated from a completely crossed design would contain 3^{24} (3 levels of each of the 24 factors) items for each respondent to answer. Further, each item would contain 24 pieces of information. This would be an impossible task to pose to respondents because of questionnaire length and the amount of information contained in each item. It is possible to reduce both questionnaire length and item size using variations on the completely crossed design shown in Fig. 3 (see Birnbaum and Stegner (1979, 1980) for details). However, these kinds of reductions would not be sufficient to allow tests of hypotheses among the numerous variables usually needed to adequately define a complex system. The subjective transfer function approach was designed to handle this problem.

[6]Additional features to the simple crossed design (e.g., Fig. 3) are necessary to adequately test the predictions of some models. For example, a more extended design would be necessary to test the major predictions of the range model (Eq. (3)) and independently derive weight and scale value parameters. Discussions of these additional designs are presented in Birnbaum and Stegner (1979, 1980) and Norman (1976).

IV. SUBJECTIVE TRANSFER FUNCTION APPROACH

In this section, we describe how the subjective transfer function approach resolves the measurement problem of testability (verifiability) described in Section II for complex systems that involve numerous variables that interlink causally throughout the system. Basically, this is accomplished by (a) incorporating the experimental designs that characterize the algebraic modeling approach, and (b) providing additional design features necessary for linking all the components of a system to its overall outcome (e.g., the Land Battle in Fig. 1).

The discussion is divided into three areas. First, we discuss procedures for defining components of a complex system representation and formulating testable hypotheses about their effects on system outcomes. Second, we describe how to obtain combination models (C in Fig. 2) (referred to as transfer functions for reasons described in this section) that link components of the system to one another and to the highest tier in the representation (e.g., Land Battle in Fig. 1). Third, we discuss the application of transfer functions to a comparison of two or more complex systems.

DEFINING A COMPLEX SYSTEM REPRESENTATION

The first step in defining the components of a complex system is to gather information from "experts" about the important system outcomes. For example, effecting a favorable outcome to a particular Land Battle would be important to those involved in tactical air command and control and force employment. The next step is to gather information from experts about what system components might affect the battle outcome. From the pool of possible components, a system's hierarchical structure is hypothesized. Some of these components are hypothesized to be influenced by other components in the pool and thus serve an intermediary role in their effects on the final outcome.

Preliminary Hypotheses

In Fig. 1, Air Force professionals would have hypothesized that the Land Battle is affected by Tactical Air Operations, which are affected by the Functions, which, in turn, are affected by the Elements. Specifically, experts might have hypothesized that Tactical Air Operation *performance* affects the Land Battle; tactical air operation performance is affected by the *ability to perform* the Functions of Planning, Directing and Controlling, which are affected by some *features* of the Elements.

These relationships are stated as preliminary hypotheses in Table 1. These hypotheses would be considered preliminary because they precede hypotheses that specify factor levels (e.g., levels of performance of Tactical Air Operations) and combination models that explain how these levels affect judgment. For discussion purposes, independent variables (the factors to be manipulated have been underlined; dependent variables (response dimensions) have been set in italics. As can be seen from Table 1 and Fig. 1, intermediary components are hypothesized as both being affected by components lower in the representation and having

Table 1

PRELIMINARY HYPOTHESES ASSOCIATED WITH COMPONENTS SHOWN IN FIGS. 1 AND 6

1. TAO Performance[a] affects perceived *chances of bringing about a favorable outcome to the Land Battle.*[b]

2. Ability to perform the Function (Plan, Direct, or Control) affects perceived *Close Air Support performance.*

3. Ability to perform the Function (Plan, Direct, or Control) affects perceived *Interdiction performance.*

4. Ability to perform the Function (Plan, Direct, or Control) affects perceived *Airlift performance.*

5. Features of the Elements affect perceived *ability to perform Planning for Close Air Support.*

6. Features of the Elements affect perceived *ability to perform Directing of Close Air Support.*

7. Features of the Elements affect perceived *ability to perform Controlling of Close Air Support.*

8. Features of the Elements affect perceived *ability to perform Planning for Interdiction.*

9. Features of the Elements affect perceived *ability to perform Directing of Interdiction.*

10. Features of the Elements affect perceived *ability to perform Controlling of Interdiction.*

11. Features of the Elements affect perceived *ability to perform Planning for Airlift.*

12. Features of the Elements affect perceived *ability to perform Directing of Airlift.*

13. Features of the Elements affect perceived *ability to perform Controlling of Airlift.*

[a]Independent variables (the factors to be manipulated) are underlined.
[b]Dependent variables (the response dimensions) are in italics.

effects on components higher in the representation. Thus, development of the hierarchical representation is based upon a series of hypotheses concerning causes and effects within the system that ultimately affect the final outcome.

Experimental Units

After hypotheses are formed, the complex system is divided into experimental units that correspond to these hypotheses. In Fig. 6, the command and control and force employment representation shown in Fig. 1 is labeled with experimental units that correspond to the hypotheses listed in Table 1. Each unit contains the components that make up the independent and dependent variables needed to test its hypothesis. These experimental units make it possible to use factorial designs to generate questionnaires of reasonable length; questionnaire items would contain a maximum of about five pieces of information. Combination models (C in Fig. 2) are sought to explain the judged relationship among the variables within each unit separately. Thirteen judgment experiments would be conducted to test hypotheses about the representation shown in Fig. 6.

An additional advantage of representing a complex system in terms of its experimental units is that different experts might be required for different units. For example, in Fig. 6, one group of Air Force professionals might be expected to know about Planning Close Air Support missions at the Function tier of the hierarchy (experimental unit 2) but not about the Process Support for Directing Interdiction (experimental unit 9). Conversely, a group that knew about the Process Support for Directing Interdiction would not know about Planning Close Air Support missions.

Procedure

Once an initial set of hypotheses is formed, preliminary experiments must be performed on the respondent population to find out (a) if the tasks make sense (that is, whether components (factors), component descriptions (factor levels), and dependent variables are understandable in terms of the judgment task); (b) if selected components statistically affect judged outcomes; and (c) what combination model (transfer function) might appropriately explain component effects in the different experimental units.

These assessments can be made by performing judgment experiments like the one described in the last section within each experimental unit separately. The sense of the tasks can be assessed by examining each respondent's data. If a respondent's data exhibit numerous violations of fundamental algebraic axioms such as commutativity and transitivity, it is concluded that the respondent did not understand the task. Tests of component effects are made using simple statistical analyses (e.g., analysis of variance). Tests of initially hypothesized combination functions are made using statistical and graphical (Figs. 4B and 5B) analyses. If tests indicate that a selected component does not affect judgments of the designated outcome, that component is omitted from the representation and a new one may be sought. Iterations of judgment experiments within each unit continue until appropriate components and an appropriate combination model (transfer function) have been found. The *final* definition of the complex system representation (i.e., specification of the components and a diagram of their interrelationships) emerges when judgment experiments are completed for every experimental unit.

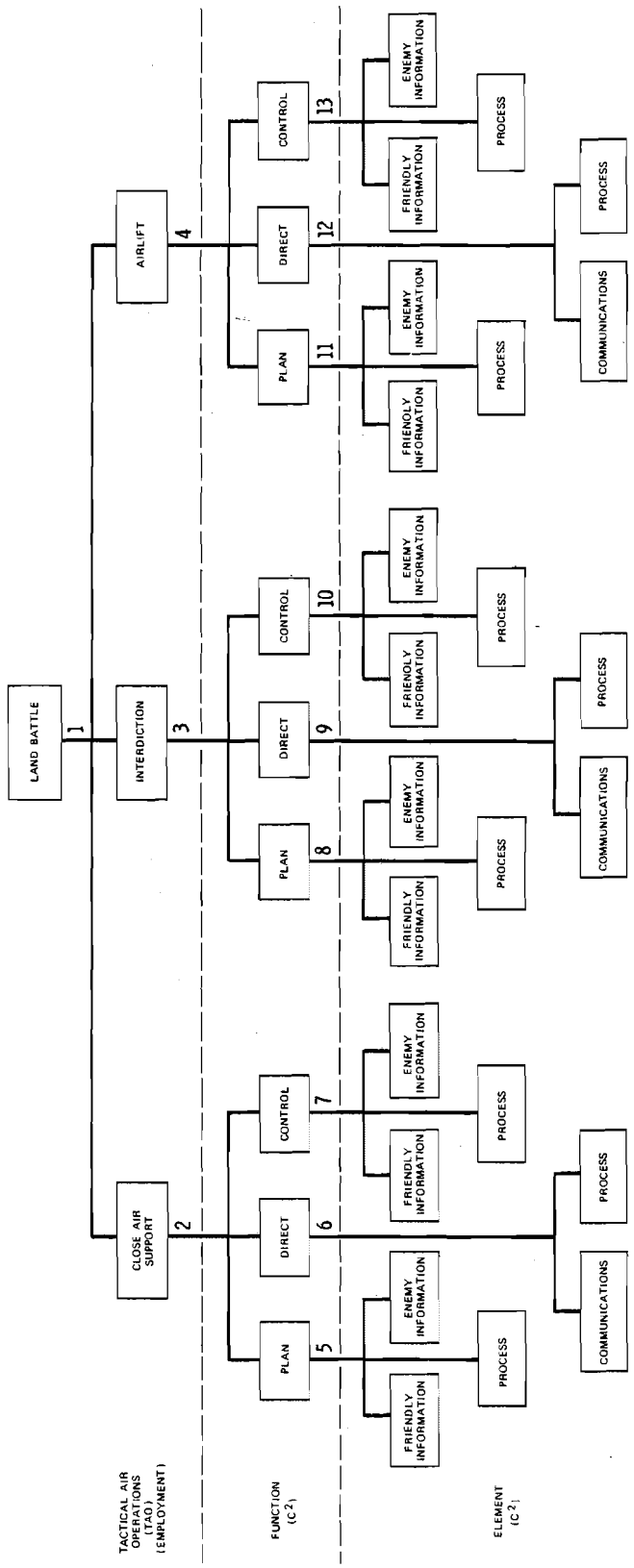


Fig. 6—Air Force command and control (C^2) and force employment representation divided into experimental units (Corresponding hypotheses are given in Table 1)

OBTAINING TRANSFER FUNCTIONS

We describe below two features of the subjective transfer function approach. The first is the construction of experimental designs within each experimental unit. The second relates to definitions of independent and dependent variables.

Experimental Designs

Within each experimental unit, questionnaires would be generated from factorial designs of the independent variables. Designs that allowed tests of hypothesized combination models would be selected. An initial basis for selecting a model when little is known about the variables under consideration could be its success in other domains. Statistical "badness-of-fit" tests provide the researcher with information about the data's deviations from model predictions. Graphic tests of fit (in which data are plotted as in Figs. 4B and 5B) aid in diagnosing the magnitude and direction of model deviations. When an appropriate model is determined, stimulus and response scale values are derived from it. The goal would be to diagnose an appropriate combination model for each experimental unit in the hierarchy.

If the initial tactical air command and control and force employment system turned out to be like Fig. 6, a specified Land Battle would set the scene for all the experiments. Table 2 outlines a single experiment that might be performed at the TAO tier of Fig. 6. Each TAO has been operationally defined in terms of its performance—good, fair, or poor. A fully crossed factorial design of the three factors at this tier would produce 27 questionnaire items to present

Table 2

OUTLINE OF POSSIBLE JUDGMENT EXPERIMENT AT THE TAO LEVEL

A.	Factors	Close Air Support Performance	Interdiction Performance	Airlift Performance
B.	Factor Levels	Good Fair Poor	Good Fair Poor	Good Fair Poor
C.	Item Descriptions	1. Good 2. Good 3. Good 4. . . . . 27. Poor	Good Good Good . . . . Poor	Good Fair Poor . . . . Poor
D.	Sample Item[a] (3 above)	If you knew that Close Air Support performance was good, Interdiction performance was good, and Airlift performance was poor, what would you judge your chances to be of effecting a favorable outcome in the Land Battle?		

[a]All 27 items would be randomly ordered within the questionnaire.

to the expert for judgment. (Variations on a completely crossed factorial design might be necessary to adequately test the models under investigation. Different hypothesized models may require different design variations.) An outline of the 27 different item descriptions is presented in Panel C of Table 2. The respondent (an Air Force professional) would judge the chances of effecting a favorable outcome in the Land Battle, given the information in each item.

Table 3 outlines an experiment for unit three at the Function tier of Fig. 6. This experiment would be designed to test a model that specified the effects described in the third preliminary hypothesis of Table 1. In this example, the factors, ability to Plan, Direct, and Control, could be described as good, fair, or poor. Again, 27 questionnaire items would be generated from a simple factorial design of all three factors. A combination function would be sought that specified the relationship between ability to perform the functions and perceived Interdiction performance in the specified Land Battle.

Experiments for each of the nine units at the Element tier would follow the same outline. Each of the components would be described along a certain dimension of interest. For example, at unit 8 in the hierarchy, *currency* (in terms of how frequently the battle field is observed and the time it takes to get the information to the command and control system) might be the dimension selected to define Friendly and Enemy Information, and *time* to process incoming information might be the dimension selected for the Process component. Levels of each of these factors would have to be specified and factorially combined to produce questionnaire items for the respondent to answer. For each item, respondents would be asked to judge the ability to Plan Interdiction. Other experimental units at the Element level would use independent and dependent variables corresponding to their hypotheses of concern (see Table 1).

Table 3

OUTLINE OF A POSSIBLE JUDGMENT EXPERIMENT FOR UNIT 6 AT THE FUNCTION LEVEL

A.	Factors	Ability to Plan Interdiction	Ability to Direct Interdiction	Ability to Control Interdiction
B.	Factor Levels	Good Fair Poor	Good Fair Poor	Good Fair Poor
C.	Item Descriptions			
	1.	Good	Good	Good
	2.	Good	Good	Fair
	3.	Good	Good	Poor
	4.	.	.	.
	.	.	.	.
	.	.	.	.
	.	.	.	.
	27.	Poor	Poor	Poor
D.	Sample Item[a] (2 above)	If you knew that the ability to Plan Interdiction was good, Direct Interdiction was good, and Control Interdiction was fair, what would you judge the Performance of Interdiction to be?		

[a]All 27 items would be randomly ordered within the questionnaire.

Operational Definitions of Independent and Dependent Variables

Careful construction of operational definitions of independent and dependent variables (components) provides the "transfer" features of the models and thus the functional link among experimental units throughout a representation.

Note in Fig. 6 that every component except those at the lowest and highest tier serve as independent variables in one experimental unit and dependent variables in another experimental unit. For example, in experimental unit 5, Plan is the *dependent* variable for the Friendly Information, Process, and Enemy Information independent variables. However, in experimental unit 2, Plan is an *independent* variable along with Direct and Control; the dependent variable for unit 2 is Close Air Support. Similarly, for experimental unit 1, Close Air Support is an independent variable along with Interdiction and Airlift for the Land Battle dependent variable. Transfer functions are obtained by operationally defining the components that serve as both independent and dependent variables in the *same terms* for both uses (i.e., in both experimental units). Thus, the operational definition of these components as dependent variables is the same as their operational definition as independent variables. These matching operational definitions are illustrated in Table 1. For the fifth hypothesis (corresponding to experimental unit 5), the *dependent* variable definition for Plan (*ability to perform* Planning for CAS), coincides with the definition of Planning when this component serves as an independent variable (*ability to perform* the function (second hypothesis in Table 1)). Similarly, the dependent variable definition of Close Air Support—CAS performance—coincides with the definition of Close Air Support when it is used as an independent variable (hypothesis number one). This matching of dependent and independent variable definitions occurs throughout the representation for all components serving as both independent and dependent variables. Thus, when combination models are determined for all experimental units in the representation, scale values of a dependent variable (response scale values, Ψ) in one experimental unit are on the same definitional continuum as the scale values of its associated independent variable (stimulus scale values at the next highest tier in the hierarchy). These "matching" scale values provide the rationale for using obtained models as *transfer* functions in complex system comparison. When the models are used as transfer functions, an *output* (Ψ) model value obtained by computing a function at one hierarchical tier is transferred for use as an *input* value for its associated model at the next highest tier in the representation.

For an example, take models that might be obtained for experimental units two and one. The model for the variables shown in experimental unit two would be some known function of the values of Planning, Directing and Controlling ability. Computing these known values according to the dictates of the model would yield the model's output—the value of *Close Air Support performance.* The model at unit one would be a known function of the values placed on Close Air Support, Interdiction, and Airlift performance. These values are needed in order to calculate this model's output. One of these input values—Close Air Support performance—would be obtained by computing the output to the model at unit two. Models at experimental units three and four would provide the remaining values of Interdiction and Close Air Support performances, respectively, needed for calculating an outcome to the model at unit one. Because of this transfer feature, combination models (C in Fig. 2) are referred to as transfer functions (T). A transfer function is sought for each experimental unit in the hierarchy, as illustrated in Fig. 7. We discuss next the usefulness of the transfer functions in comparing complex system outcomes.

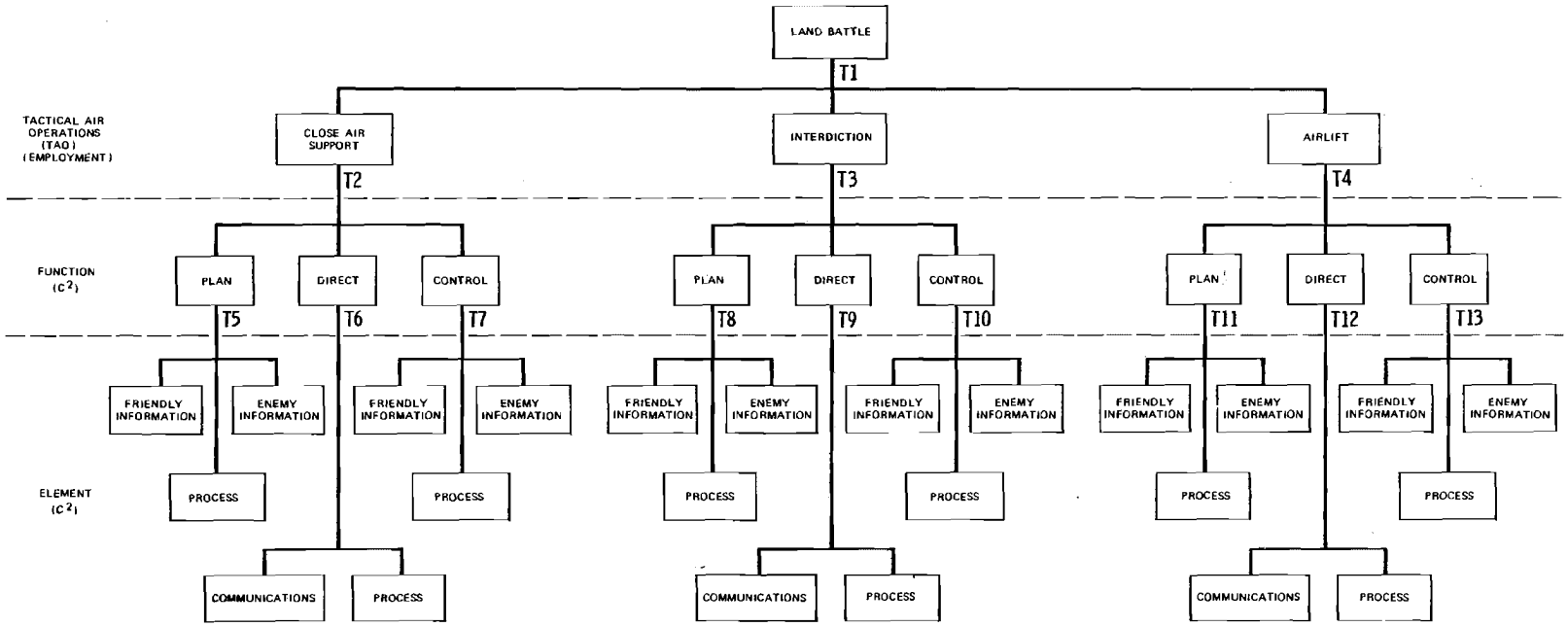


Fig. 7—Transfer functions associated with experimental units

Each transfer function specifies the subjective relationship between the independent and dependent variables within its experimental unit

COMPLEX SYSTEM COMPARISON

Once the transfer functions are known for all experimental units comprising a representation, it is possible to compare systems having the components and component definitions making up that representation. Comparisons can be made for every outcome (there is one outcome for each experimental unit) in the representation. Two procedures can be used for system comparison. One procedure is to compute the subjective transfer functions. An additional or alternative procedure that can be used for smaller subsets of experimental units is to analyze graphic displays of the subjective responses (Ψ values in Fig. 2) derived from the transfer functions. These procedures are described below.

Computing Subjective Transfer Functions

Three kinds of information have to be known before the subjective transfer functions can be used to compute outcomes: (1) the systems to be compared have to be defined; (2) the scale values needed as input to the models at the lowest hierarchical tier need to be determined (these models would not have other model's outcomes to use as input, as is the case with all models in the hierarchy above the lowest tier); (3) these scale values need to be calibrated. In this section we briefly discuss how to obtain this information and then we demonstrate how the subjective transfer functions are calculated and used for complex system comparison.

Defining a Particular Complex System. A given representation defines numerous complex systems. A particular system characterized by the representation is identified by the component levels at the lowest hierarchical tier. For two systems to be *different*, they must differ in *at least* one component level at the lowest tier. For example, for all systems defined by the command and control and force employment representation shown in Fig. 7, a particular system would be identified by its Element levels. Two different systems would have to vary in at least one Element level.

For example, in Fig. 7, *Communications* (for Directing Interdiction (T9 in Fig. 6)) reflects an actual communications capability used in command and control systems. Systems would be different if they had different communications capabilities for Directing Interdiction. (Specific differences in communication capabilities hypothesized to be important would have been factor levels manipulated in the experiment.) Also, systems would be different if they had different qualities of friendly and/or enemy information (provided by different information gathering and reporting networks). The question is, how do these different systems vary according to internal outcomes (at each experimental unit) and according to their overall outcomes (e.g., the Land Battle in Fig. 1).

Determining Initial Input Scale Values. Before subjective transfer functions can be used to compute outputs, subjective input values must be provided to the functions at the lowest tier in the hierarchy. Once these are provided, model output values serve as inputs to all models at higher tiers.

Subjective input values are obtained in one of three different ways:

1. If the particular component levels defining the system were used in the experiment, then the subjective scale values are known; they are part of the experimental data;
2. If the component levels were not used in the experiments but are physical values, then the functions relating these physical values to their subjective counterparts (H in Fig.

2) are known and are part of the experimental data, and can be used to transform the new physical measures to the subjective values[1] needed as input to the model;

3. If the component levels were not used in the experiments and are qualitative descriptions rather than physical values, pre-evaluation experiments, similar to the original experiments, would have to be performed for the experimental units involved in order to determine the subjective values of those component levels.

Once all the subjective values associated with the factor levels at the bottom of the hierarchy are known and calibrated (described next), the transfer functions can be used to compare outcomes at all levels in the hierarchy.

Calibrating Initial Input Scale Values. The researcher can usually claim to know scale values of component levels (at the lowest hierarchical tier) at least to a linear transformation.[2] When different component levels are scaled in different experiments (as the separate experimental units suggest), resulting linear transformations of the values vary across experimental units. Therefore, use of the subjective transfer function requires calibrating scale values of component levels at the lowest (e.g., Element) hierarchical tier. This can be accomplished by using experimental designs at the lowest tier that cut across experimental units. For example, some of the component levels selected to define friendly information for unit eight might also be employed in the experimental design for units nine and ten at the Element tier. The idea is that by pinning down the relationships among factors that are repeated across experimental units, it is possible to convert all scale values in those experimental units to the same unit of measure. Different situations would require different solutions to the calibration problem. Solutions would vary with the components selected to describe the system, with "expertise" differences in respondent populations among experimental units, and with the form of the transfer functions for the experimental units at the lowest tier.

Using Subjective Transfer Functions To Compare Systems. As described in the last section, transfer function analyses require transferring the subjective response value (Ψ in Fig. 2) obtained from a model at one tier in the hierarchy to the model at the next highest tier in the hierarchy along the same path. There are nine paths in Fig. 7, one corresponding to each element group.

Transfer function analysis is illustrated in Fig. 8. First, input values to T5, T6, and T7 would yield a subjective response scale value (Ψ) for each function. Next, the response scale value obtained from computing T5 would be used as the scale value associated with Planning (p) needed to compute T2. Similarly, the response scale value obtained from computing T6 would be used as the scale value associated with Directing (d), also needed to compute T2, and so forth until all input scale values for T2, T3, and T4 are computed. In the same manner, the response scale values obtained by computing T2, T3, and T4 would be the input scale values for T1. Finally, the output obtained by computing T1 is the overall subjective effectiveness index.

To illustrate how transfer functions are computed, consider the following numerical example. Suppose T5 in Fig. 8 is an additive model; that is,

$$T5(\text{fr in},\text{en in},\text{proc}) = (\text{fr in}) + (\text{en in}) + (\text{proc}),$$

[1]H in Fig. 2 relates physical values (stimuli) to subjective values. When subjective stimulus values are derived from a known model, the plot of these values as a function of the physical values yields the form of the H function. Thus, when stimuli are physical measures and the model and scales are known, the form of the H function is known.

[2]The entire class of additive models and a number of interactive models (e.g., Eq. 3) yield interval scales of the stimuli and responses when enough constraints are built into the test of the model by the design. Multiplicative models, however, yield values known only to a power transformation (Krantz et al., 1971).

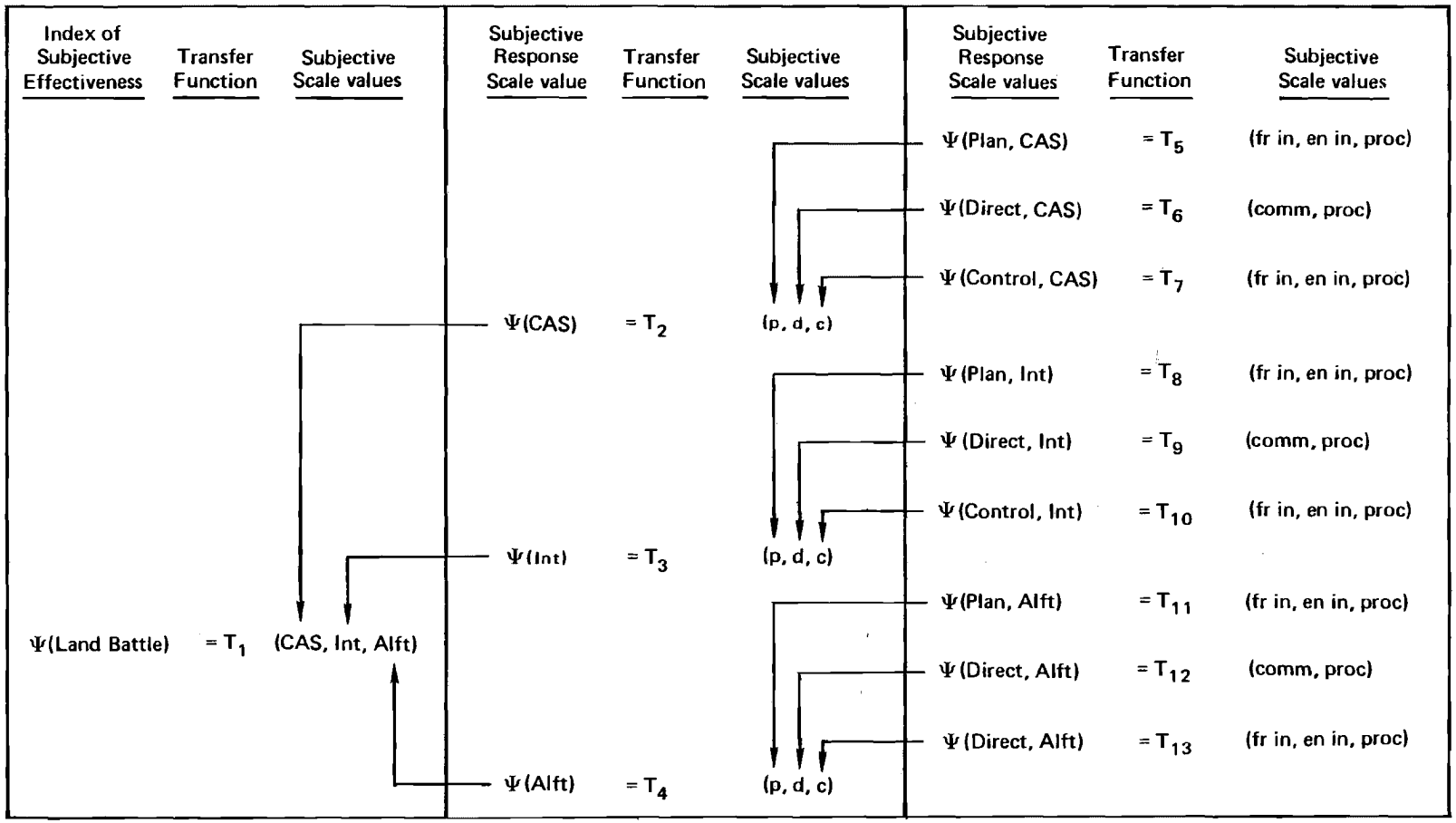


Fig. 8—System evaluation

Subjective response scale values from transfer functions at one level in the hierarchy serve as subjective scale values (Inputs) to the transfer function at the next highest level in the hierarchy on the same path. Friendly Information (fr in), Enemy Information (en in), Process (proc), Communication (comm), Plan (p), Direct (d), Control (c), Close Air Support (CAS), Interdiction (Int), Airlift (Alft).

and the subjective input scale values for Friendly Information (fr in), Enemy Information (en in), and Process (proc) are 7, 2, and 5, respectively. Then the output for T5 would be obtained as follows:

$$\begin{aligned}\Psi_{\text{Plan,CAS}} &= (\text{fr in}) + (\text{en in}) + (\text{proc}) \\ &= 7 + 2 + 5 \\ &= 14 .\end{aligned}$$

This output value of 14 obtained from T5 is the input value for Planning (p) in T2.

Suppose T6 is a range model of the form shown in Eq. 3, and the weights of the initial impression (w_0), Communication (comm) factor, and Process (proc) factor are 1, 3 and 5, respectively, and the weight of the range term, ω, is -0.8. Then the model for T6 would be

$$\Psi_{\text{Direct,CAS}} = \frac{(1)s_0 + 3(\text{comm}) + 5(\text{proc})}{1 + 3 + 5} + [-0.8(s_{\max} - s_{\min})] .$$

If the scales values are 3, 6 and 8 for the initial impression (s_0), Communication (comm) and Process (proc), respectively, substituting these values into the model would yield

$$\begin{aligned}\Psi_{\text{Direct,CAS}} &= \frac{3 + (3)(6) + (5)(8)}{1 + 3 + 5} - 0.8(8-6) \\ &= 5.18 ,\end{aligned}$$

the output value for T6. This output value of 5.18 is the input value for Directing (d) in T2.

Suppose T7 is also a range model and the weights for the initial impression (w_0), Friendly Information (fr in) factor, Enemy Information (en in) factor, and Process (proc) factor are 1, 4, 7, and 3, respectively, and ω is -0.7. Then the model for T7 would be

$$\Psi_{\text{Control,CAS}} = \frac{(1)(s_0) + 4(\text{fr in}) + 7(\text{en in}) + 3(\text{proc})}{1 + 4 + 7 + 3} - [0.7(s_{\max} - s_{\min})] .$$

If the subjective scale values are 2, 8, 2, and 6 for the initial impression (s_0), Friendly Information (fr in), Enemy Information (en in), and Process (proc), respectively, substituting these values in this model would yield

$$\begin{aligned}\Psi_{\text{Control,CAS}} &= \frac{(1)(2) + (4)(8) + (7)(2) + (3)(6)}{1 + 4 + 7 + 3} - 0.7(8-2) \\ &= 0.2 ,\end{aligned}$$

the output value for T7. This output value of 0.2 is the input value for Controlling (c) in T2.

Once the three Ψ values are obtained from T5, T6, and T7, T2 can be calculated.

Suppose T2 is an additive model,

$$T2(p,d,c) = (p + d + c) .$$

Substituting the Ψ values obtained from T5, T6, and T7 into this model yields

$$\begin{aligned}\Psi_{\text{CAS}} &= 14 + 5.18 + 0.2 \\ &= 19.38 ,\end{aligned}$$

the output for T2. This output value of 19.38 is the input value for Close Air Support (CAS) in T1. Calculating T1 also requires an Interdiction (Int) and an Airlift (Alft) value. In order to obtain the Interdiction (Int) value for T1, it would be necessary to compute the transfer functions T8, T9, and T10 to get the input values to T3. Calculation of T3 provides the Interdiction value for T1. Similarly, in order to obtain the Airlift value for T1, it would be necessary to compute the transfer functions T11, T12, and T13 to get the inputs to T4. Calculation of T4 provides the Airlift value for T1. Finally, calculating T1 yields the subjective effectiveness index.

Usually, the concern in complex system analysis is on what *changes* the subjective effectiveness index; that is, why and where systems differ in the representation. An important feature of the subjective transfer function approach is that system comparisons can be made among all outcomes within the system (at each experimental unit) and the overall system outcome.

Figure 8 can be used to illustrate system comparison. If two or more systems differed in their communication and/or process capabilities for Directing Close Air Support (unit 6), they could be compared at three different outcome points in the hierarchy—their abilities to Direct (T6), their Close Air Support performances (T2) and their relative influences on the Land Battle (T1). Another example would be two systems that differed in their Process support capabilities for Controlling both Close Air Support and Interdiction missions. These two systems could be compared in their T7 and T10 outcomes (the ability to Control Close Air Support and Interdiction operations, respectively); their T2 and T3 outcomes (the relative abilities of the two systems to perform Close Air Support and Interdiction missions); and finally, their T1 outcomes (their relative influences on the Land Battle). Thus, the transfer functions can be used to compare outcomes among systems at all units in the hierarchy.

Graphic Analyses

Graphs provide a useful mode for simultaneous comparison of all systems defined by the manipulated factors and, through extrapolation, other systems with Element levels that lie within the manipulated range. Graphic displays within each experimental unit would resemble that shown in Fig. 5B except that *subjective* responses (Ψ values derived from the model) would be plotted on the y-axis and subjective scale values (s in Fig. 2) would be plotted on the x-axis.[3] Such graphic displays would allow visual inspection of subjective tradeoffs in values of the independent variables that produce various outcomes. For example, Fig. 5B could be thought of as representing the outcomes from nine systems that differ on levels of Interdiction and Close Air Support performance at the TAO hierarchical tier in the representation.[4] If the values shown in this graph were derived from theory (the subjective transfer function), the data would indicate that poor Interdiction and fair Close Air Support performance are valued about the same in effecting a favorable outcome in the Land Battle as good Interdiction and poor Close Air Support performance. Comparisons among other pairs or groups of data points allow similar evaluative interpretations. The next step would be to examine similar theoretic plots at the Function tier to examine how the ability to perform the different Functions affects Interdiction and Close Air Support Performance, and so forth until it is determined how one or more of the

[3]This would be a plot of the model's predictions.

[4]Figure 5B depicts data for a two factor experiment. If three factors were used as suggested in Tables 2 and 3, a two-dimensional graph would be displayed for each level of the third factor.

Elements can be changed to alter the ability to perform the Functions and hence the perceived outcome of the Land Battle.

This type of analysis gets complicated when comparisons include many experimental units and tradeoffs are between a number of variables as the analysis proceeds from the top down to lower tiers in the hierarchy. In these cases, use of the subjective transfer functions is a more practical evaluative tool.

COMMENTS

The subjective transfer function approach is a valuable tool for complex system analysis because it provides a framework for *testing* cause and effect hypotheses. The framework further allows the testing of judgment models that specify the nature of these effects. Thus, information resulting from the analysis provides guidance for changing aspects of the system to achieve desired outcomes. The approach is being demonstrated and refined at The Rand Corporation for application to command and control and force employment evaluation.

Appendix A

TACTICAL AIR COMMAND AND CONTROL AND FORCE EMPLOYMENT

In wartime, the tactical Air Force contains fighter aircraft, reconnaissance aircraft and transport aircraft organized by tactical "wings"—each wing having 36 to 72 of one type of aircraft and the men, equipment, supplies and facilities needed to maintain and operate those aircraft in combat. The tactical Air Force also contains a command and control system leading downward from the overall commander of the tactical air force to the wings. This system, called the Tactical Air Control System (TACS), manages the employment of the forces—determines which enemy targets to destroy, which information to collect, and where and what to airlift, and directs specific wings to perform specific tasks at specific times.

The TACS includes a network of operations centers, communications systems, and ground and airborne radars. It maintains as complete a picture as possible on the unfolding air and land battles and of the posture of unengaged friendly and enemy forces by processing friendly information and enemy information provided to it. From this picture and consideration of national and military plans and objectives, senior officers in the TACS make the force employment decisions and direct the wings accordingly.

The force employment decisions are made in two different contexts: future and present operational time periods. Deciding how to employ the force in a future operational time period (historically, the next day), is called Planning. In each period, the employment of the entire tactical force expected to be available is planned for the following period. When the plan is being executed, decisions are required on adjustments to the planned employment in response to currently perceived situations that differ from those projected at the time the plan was made. This employment decisionmaking is called Controlling. In both cases, the decisions take the form of specifying operational missions to be flown by the tactical aircraft, and the wings are "directed" to do so. Hence, Tactical Air Command and Control performs three main functions—Planning, Directing and Controlling.

Tactical air forces affect the course of military events by flying (or having the potential to fly) combat missions. These missions are categorized into Tactical Air Operations (TAOs) indicating primary mission objectives. The TAOs include Air Defense, Reconnaissance, Search and Rescue, and Offensive Counter Air, and, of course, the three selected for illustration in the main body of this report—Close Air Support, Interdiction, and Airlift. Hence, tactical Air Force employment in general can be thought of as the performance of the Tactical Air Operations, and the effectiveness of force employment can be thought of as the effectiveness of appropriate TAOs in affecting the course of military events.

A land battle can be defined as a single military event in which tactical air forces play an important role. In large-scale conventional warfare, opposing forces engage in many battles on the ground in order to achieve military objectives (such as occupying territory or destroying opponents' forces) that are expected to contribute to the ultimate attainment of national goals. In these battles, while the army engages opposing army forces on the ground, tactical air forces conduct tactical air operations to influence the outcome of the battle. They attack and destroy enemy army forces in direct contact with our army forces (Close Air Support); fly in reinforce-

ments and resupplies to our army forces (Airlift); attack and destroy enemy forces and equipments, close roads and other lines of communication in the enemy's rear to keep new enemy forces from joining the battle (Interdiction); attack and destroy enemy aircraft attempting to attack our army forces (Air Defense); and carry out other TAOs having less direct influence on the course of the battle.[1]

Evaluation of the effectiveness of Tactical Air Command and Control (the fundamental research issue which led to the development of the subjective measurement technique) must be in terms of how it can affect the course of military events in wartime. For our discussion in the main body of the report we have chosen to use a land battle as the military event against which to measure. Command and control affects military events only through its effect on the performance of tactical air operations. Hence, the representation (Figs. 1, 6 and 7) shows the Land Battle influenced at the top tier and the TAOs directly influencing it, which follows from the above discussion. Command and Control is brought in at the third tier, reflecting that the effectiveness of the TAOs depends in large part on how well they can be planned, directed and controlled. And finally, the elements which go into Planning, Directing and Controlling form the bottom tier.

TACTICAL AIR OPERATION DEFINITIONS

Close Air Support

Air attack against hostile targets which are in close proximity to friendly forces and which require detailed integration of each air mission with the fire and movement of those forces.

Interdiction

The attack of specific objectives by fighter, bomber, or attack aircraft on an offensive mission. It includes air operations conducted to destroy, neutralize, or delay the enemy's military potential before it can be brought to bear effectively against friendly forces. These air operations are conducted against categories of targets at such distances from friendly forces that detailed integration of each air mission with the fire and movement of friendly forces is not required.

Tactical Airlift

The carriage of passengers and cargo within a theatre in the context of airborne operations, air logistic support, special missions and aeromedical evacuation missions.

---

[1]Because of the heavy involvement of tactical air in these battles they are now considered in a composite sense as "the air/land battle."

FUNCTION DEFINITIONS

Planning

The activities and decisionmaking that determine how the tactical air resources are to be used in a future operational time period (usually the next day). It encompasses establishment of strategy, selection of air missions to be flown and targets to be attacked, specification of aircraft to fly the missions, and development of detailed tactics to be used in accomplishing each mission. It is based on knowledge and perception of both friendly and enemy force dispositions, capabilities, and intentions.

Controlling

The monitoring and evaluation of the current military situation and the adjusting of plans and ongoing operations as necessary to achieve military objectives.

Directing

The issuance of orders to all units involved in the execution of plans generated by the planning function and plan adjustments generated by the controlling function. It encompasses both the preparation and the transmittal of orders and instructions, which must be timely and comprehensive to enable forces to perform assigned tasks and accomplish planned missions.

ELEMENT DEFINITIONS

Friendly Information

Information on friendly events, resources and capabilities. It is measured in terms of its currency, accuracy, and content.

Enemy Information

Information on enemy events, resources, and capabilities. It is measured in terms of its currency, accuracy, and content.

Process Support

The means by which information within the command and control system is processed, displayed, and communicated internally.

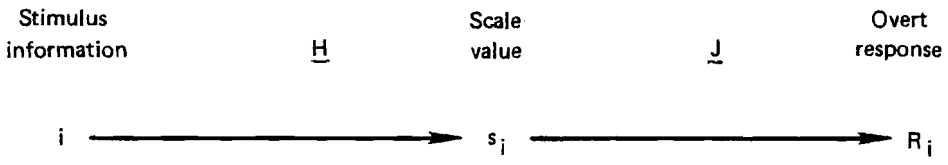
Communications

The capacity of the system used to communicate with the operational wings to direct them to perform tactical air missions.

Appendix B

"DIRECT" SCALING FRAMEWORK

The "direct" scaling approach (Stevens 1946, 1957, 1971) has been used widely in psychology and has been adopted readily by researchers in other areas. The appeal of the approach is its simplicity and the belief that subjective scale values are obtained by having people assign numbers to stimulus objects according to a set of rules (S. S. Stevens proposed this in 1946). An outline for discussing this approach is presented below in Fig. B.1. Note that only two events in the outline are directly observable: the stimulus information, i, and the overt response (R_i). The first subjective process, *H*, transforms the stimulus information into a corresponding subjective scale value, s_i. The second subjective process, *J*, transforms the scale value into an overt response, R_i. Thus, the outline postulates two subjective transformations that occur between the presentation of the stimulus information and the occurrence of the response. Stimulus information could consist of descriptive statements (e.g., a sentence that describes the use of Interdiction in a particular Land Battle) or dimensions that have associated physical measures (e.g., distance, time, number of messages coming into a system, number of sorties).[1]



H represents the function that transforms the stimulus information i to its subjective counterpart, s_i; J represents the judgment function that transforms the subjective value to an overt response, R_i.

Fig. B.1—Outline of "direct" scaling

Scaling Examples

The main features of the "direct" framework are (a) the emphasis on obtaining scale values (the s_i values in Fig. B.1) and (b) the use of single-factor experimental designs (described below) to generate questions posed to the respondent. As will be seen, a single-factor design does not allow tests (verification) of basic assumptions underlying conclusions about subjective events.

[1]Researchers using "direct" scaling usually use stimuli that can be measured on the physical continuum. Graphs of responses as a function the physical values are assumed to yield the form of H in Fig. B.1 (called the psychophysical function).

Two illustrations are presented below that describe how "subjective" scales might be obtained in this framework. For both examples, consider the stimuli to be levels of Close Air Support performance; performance could be good, fair, or poor. Think of the respondents as Air Force professionals who were asked to perform the specified task. For each task, written category labels operationally define the numbers used by the Air Force professional in making a judgment concerning the *value* of each level of Close Air Support performance in effecting a favorable outcome in a given land battle. In the hypothetical examples, however, numbers are deliberately disassociated with particular performance levels since the examples are presented solely for purposes of illustrating the "direct" scaling approach; specific scaling interpretations are not intended. Data could be obtained on an individual basis or be the result of a group decision. Data that result from a group decision can present special interpretive problems which require understanding how characteristics of the group differentially influence the decision; but, these are not of concern here since the main points of discussion are independent of such issues.

Example 1: Category Ratings of Close Air Support Value. Figure B.2 below illustrates a single-factor design where Close Air Support from the third tier in Fig. 1 serves as the single factor. The factor (often referred to as the variable stimulus) has three levels that correspond to Close Air Support performance capability. The three performance levels have been arbitrarily labeled a_1 through a_3 in order to disassociate them from the hypothetical numbers. On a given trial, an Air Force professional might be asked to judge the value of a particular Close Air Support performance level in producing a favorable outcome in a specified Land Battle. The task might require respondents to use a nine-point category rating scale to make their judgments; a nine would represent a "very valuable" performance level and a one would represent a performance level that appeared "not at all valuable;" numbers in between would represent gradations of these extremes. Typically, stimuli (performance levels) would be presented in a random fashion *after* respondents were familiar with all possible choices so that they knew when to use a one, a nine, and all of the other possible numbers in the response scale.

Hypothetical data for this task have been inserted in the cells of the single-factor matrix shown in Fig. B.2.[2] Numbers in the cells could represent either judgments obtained from a

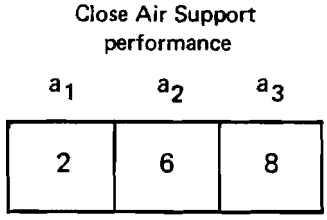


Fig. B.2—Hypothetical data matrix for category ratings of Close Air Support value

[2]The numbers (and hence the corresponding Close Air Support performance levels) in the hypothetical data matrix can be considered to have been ordered *post hoc* in terms of increasing magnitude.

single individual, mean, or median judgments of a number of individuals. For these hypothetical data, performance level a_1 got a very low "value" judgment while performance level a_3 got a high "value" judgment.

In the "direct" scaling framework, numerical responses are assumed to be linearly related to the underlying subjective scale values (i.e., J in Fig. B.2 is assumed to be linear).

Example 2: Magnitude Estimations of "Ratios" of Close Air Support Performance Level Values. The second example is a task that requires respondents to make ratio judgments of the relative value of each piece of information to a selected standard by using 100 if the variable performance level appears equally as valuable as the standard in the success of the given land battle, 50 if it appears half as valuable, 200 if it appears twice as valuable, etc. This is an example of a magnitude estimation response scale with a modulus ("ratio of 1") equal to 100. When magnitude estimations are used, respondents are typically told to use any number they wish that follows the described pattern in making their ratio judgments. For this task, the three variable levels of Close Air Support performance would be paired randomly with a selected standard performance level for judgment. Again, the goal would be to get the subjective scale values (s_i in Fig. B.1) associated with each performance level. Hypothetical magnitude estimations (means, medians, or an individual respondent's data) for this task are presented in Fig. B.3. For these data, performance level a_1 appeared half as valuable as the standard, level a_2, while performance level a_3 appeared four times as valuable as level a_2. In this framework it would be concluded that the scale values associated with TAOs a_1, a_2, and a_3 are linearly (actually linear with a zero intercept for a ratio task) related to 50, 100, and 400, respectively.

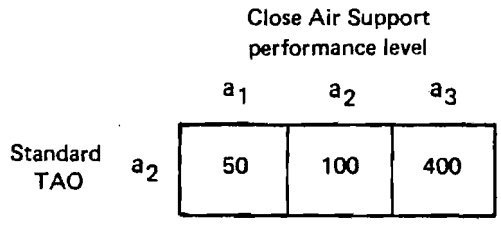


Fig. B.3—Hypothetical data matrix for magnitude estimation of Close Air Support value

Problems with "Direct" Scaling

After responses are collected, the numbers require interpretation. As mentioned above, researchers using "direct" scaling usually interpret responses as linearly related to the underlying subjective values of the stimuli along the operationally defined response continuum; the numbers given to the Close Air Support levels in Figs. B.2 and B.3 would be interpreted as the subjective scale values of those levels in effecting a favorable outcome in the specified Land Battle.

Scrutiny of "direct" scaling interpretations has led to criticisms of the approach. One major criticism is that interpretations based on operational definitions are simply tautologies; they do not *explain* what the numbers mean. In the above examples, nothing is known about what

affected the Air Force professional's response, or under what conditions other numbers might have been given as judgments. For an interpretation to be useful, it has to be based on criteria that permit its falsification; that is, there has to be a way to determine whether it is "wrong." The major problem with the "direct" approach to scaling is that this is not possible. This can be seen by examining the two major assumptions on which data interpretation is based. The first assumption is that the judgment function (J in Fig. B.1) is linear; the second assumption is that respondents combine subjective stimulus values according to the rule dictated by instructions. For example, when the instructions are to judge "intervals" or "ratios," responses result from actual subjective interval or ratio computations. In this single-factor framework, these assumptions are, in principle, untestable. This is because when only one factor is used in the design, responses are a confounded composition of the two subjective transformations, *H* and *J* in Fig B.1. It is not possible to separate stimulus scaling (H) from judgmental processes (J), or test theories of the respondent's combination rule. These assumptions are discussed next.

Assumption of a Linear Judgment Transformation

The assumption that responses are a "direct" scale of subjective value implies that the judgment transformation (*J* in the outline of Fig. B.1) is linear for "interval" estimates (example 1) and linear with a zero intercept for "ratio" estimates (example 2). This implies that "scales" obtained from the two types of tasks should be *linearly* related.[3] *Empirically*, however, magnitude estimations of "ratios" are typically a *positively accelerating* function[4] of category ratings of "intervals" for a wide variety of psychophysical and social judgment dimensions (Stevens, 1968; Stevens and Galanter, 1957). This typical nonlinear relationship found between different operations for "measuring" the same stimuli in the "direct" framework is illustrated in Fig. B.4 for the hypothetical data. The graph in Fig. B.4 is a plot of magnitude estimations of "ratios" as a function of category ratings of "intervals." (Both sets of numbers would be interpreted by the "direct" scaling researcher as representing the subjective scale values of the *same* Close Air Support levels.)

Failures of "scale" convergence have also been found *within* a given task. For magnitude estimations of "ratios," responses to a given set of stimuli have been demonstrated to change with changes in contextual features of the experiment. For example, responses depend on the magnitude of the standard stimulus and the modulus (the number selected to represent a "ratio of 1") (Poulton, 1968), as well as the range of the magnitude estimation response-scale examples and distributional features such as spacing and frequency of the stimuli presented for judgment (Birnbaum, 1980). Category ratings of a given set of stimuli also depend upon features of the stimulus distribution (see for example, Parducci and Perrett, 1971; Birnbaum, 1974c; Birnbaum, 1980).

The outline shown in Fig. B.1 makes a clear distinction between the response, R, and the underlying subjective scale value, s, associated with the stimulus. The empirical nonlinear relationship between response values assigned to the same stimuli shown in Fig. B.4 implies either that the judgment function, J, is not linear or that *subjective scale values* associated with particular stimuli change in different ways, depending on contextual features of the experi-

---

[3]This can be seen from the following reasoning. If $R_i = as_i + b$ for "intervals" and $R'_i = cs_i$ for "ratios," then $R_i = a(R'_i/c) + b = a'R'_i + b$, where $a' = (a/c)$, and R_i and R'_i represent the ith "interval" and "ratio" response, respectively.

[4]A positively accelerating function is one that increases at an increasing rate, that is, the second derivative of the function is greater than zero. Examples are power functions with exponents greater than one, or exponential functions.

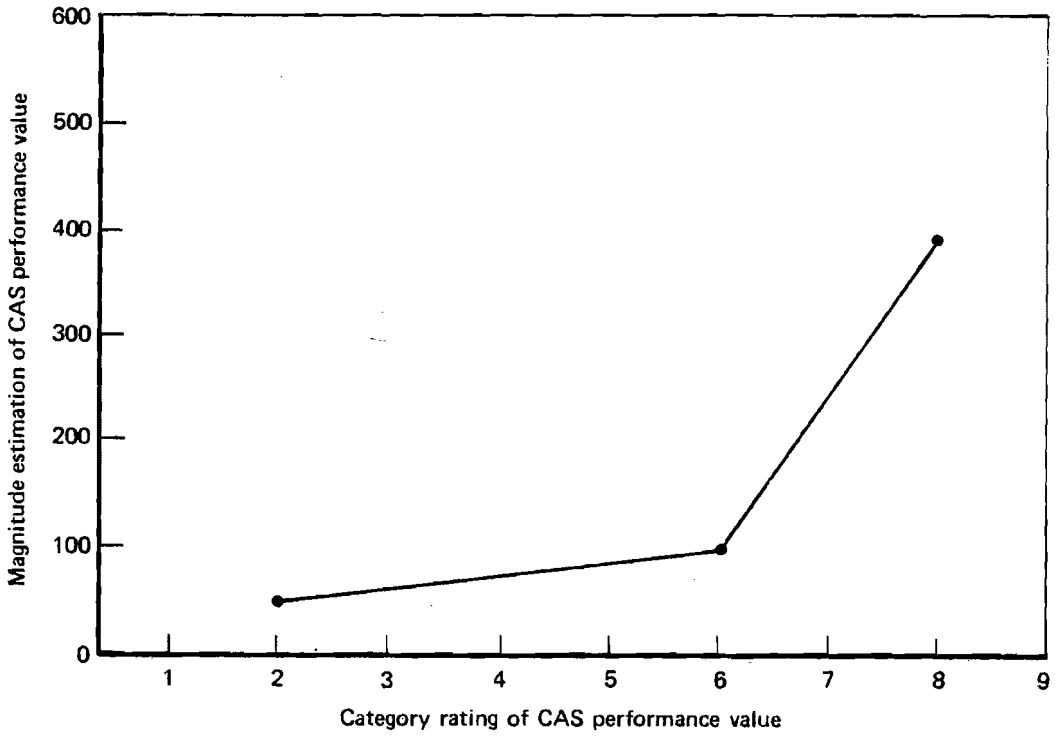


Fig. B.4—Test of "scale" convergence

ment (e.g., response, scale, task). More than one factor is needed in the stimulus design to test between these two possible interpretations.

What if response values to the same stimuli agree? This is the case for some stimulus continua (Stevens and Galanter, 1957). Linear agreement between "subjective scales" of the same stimuli can be considered a necessary but not sufficient criterion for determining scale validity (Birnbaum and Veit, 1974a). A nonlinear function between operational definitions of sensations of the same stimuli suggests that at least one set of scales is "wrong." A linear function, however, does not imply that either "scale" is "right"; both scales could be wrong with respect to some validity criterion.

Assumption that Respondents Obey Task Instructions

The second major assumption of researchers using the "direct" scaling approach is that respondents follow task instructions; that is, their mental computations are as prescribed by the task. Thus, when instructions are to judge "ratios," the respondent's subjective combination process is assumed to be a ratio rule. From this assumption, it is further assumed that resulting numbers (responses) represent a ratio scale of sensation of the stimulus information. When respondents are instructed to estimate "intervals," it is assumed that their psychological process corresponds to task instructions and thus the resulting numbers represent an interval scale of subjective value. To *test* the hypothesis that the respondent's subjective combination

rule corresponds to the dictates of the task, at least *two* factors are needed in the stimulus design, as we demonstrated in the section on the algebraic modeling approach to measurement. Determining scale properties is a separate issue and depends on the constraints placed on the test of the model by the experimental design.

SUMMARY REMARKS

"Direct" scaling contains measurement problems that cannot be resolved without further constraints. Because the basic assumptions of the framework are, in principle, untestable in the framework, many psychologists interested in determining these subjective events have questioned the usefulness and meaningfulness of "scales" obtained with "direct" scaling methods (Birnbaum and Veit, 1974a; Krantz, 1972; Savage, 1966; Shepard, 1976; Treisman, 1964; Veit, 1978).

REFERENCES

Anderson, N. H., "Integration Theory and Attitude Change," *Psychological Review,* Vol. 78, 1971, pp. 171-206.

Anderson, N. H., "Information Integration Theory: A Brief Survey," in D. H. Krantz, R. C. Atkinson, R. D. Luce, and P. Suppes (eds.), *Contemporary Developments in Mathematical Psychology,* Vol. 2, Academic Press, New York, 1974.

Birnbaum, M. H., "The Devil Rides Again: Correlation as an Index of Fit," *Psychological Bulletin,* Vol. 79, 1973, pp. 239-242.

Birnbaum, M. H., "Reply to the Devil's Advocates: Don't Confound Model Testing and Measurement," *Psychological Bulletin,* Vol. 81, 1974a, pp. 854-859.

Birnbaum, M. H., "The Nonadditivity of Personality Impressions," *Journal of Experimental Psychology,* Vol. 102, 1974b, pp. 543-561 (monograph).

Birnbaum, M. H., "Using Contextual Effects To Derive Psychophysical Scales," *Perception and Psychophysics,* Vol. 15, 1974c, pp. 89-96.

Birnbaum, M. H., "Controversies in Psychological Measurement," in B. Wegener (ed.), *Social Attitudes and Psychophysical Measurement,* Erlbaum, Hillsdale, N.J., 1981 (in press).

Birnbaum, M. H., and S. E. Stegner, "Source Credibility in Social Judgment: Bias, Expertise, and the Judge's Point of View," *Journal of Personality and Social Psychology,* Vol. 37, 1979, pp. 48-74.

Birnbaum, M. H., and S. E. Stegner, "Measuring the Importance of Cues in Judgment for Individuals: Subjective Theories of IQ as a Function of Heredity and Environment," *Journal of Experimental Social Psychology,* Vol. 17, 1981.

Birnbaum, M. H., and C. T. Veit, "Scale Convergence as a Criterion for Rescaling: Information Integration with Difference, Ratio, and Averaging Tasks," *Perception and Psychophysics,* Vol. 15, 1974a, pp. 7-15.

Birnbaum, M. H., and C. T. Veit, "Scale-Free Tests of an Additive Model for the Size-Weight Illusion," *Perception and Psychophysics,* Vol. 16, 1974b, pp. 276-282.

Gardiner, P. C., and W. Edwards, "Public Values: Multiattribute-Utility Measurement for Social Decision Making," in M. F. Kaplan and S. Schwartz (eds.), *Human Judgment and Decision Processes,* Academic Press, New York, 1975.

Krantz, D. H., "Measurement Structure and Psychological Laws," *Science,* Vol. 175, 1972, pp. 1427-1435.

Krantz, D. H., R. D. Luce, P. Suppes, and A. Tversky, *Foundations of Measurement,* Academic Press, New York, 1971.

Krantz, D. H., and A. Tversky, "Conjoint-Measurement Analysis of Composition Rules in Psychology," *Psychological Review,* Vol. 78, 1971, pp. 151-169.

Norman, K., "A Solution for Weights and Scale Values in Functional Measurement," *Psychological Review,* Vol. 83, No. 1, 1976, pp. 80-84.

Parducci, A., and L. Perrett, "Category Rating Scales: Effects of Relative Spacing and Frequency of Stimulus Values," *Journal of Experimental Psychology,* Vol. 89, 1971, pp. 427-452 (monograph).

Poulton, E. C., "The New Psychophysics: Six Models for Magnitude Estimation," *Psychological Bulletin,* Vol. 69, 1968, pp. 1-19.

Savage, C. W., "Introspectionist and Behaviorist Interpretations of Ratio Scales of Perceptual Magnitudes," *Psychological Monograph,* Vol. 80, 1966, pp. 1-32.

Shepard, R. N., "On the Status of 'Direct' Psychological Measurement," in Savage (ed.), *Minnesota Studies in the Philosophy of Science,* Vol. IX, University of Minnesota Press, Minneapolis, 1976.

Stevens, S. S., "On the Theory of Scales and Measurement," *Science,* Vol. 103, 1946, pp. 677-680.

Stevens, S. S., "On the Psychophysical Law," *Psychological Review,* Vol. 64, 1957, pp. 153-181.

Stevens, S. S., "Ratio Scales of Opinion," in D. K. Whitla (ed.), *Handbook of Measurement and Assessment in the Behavioral Sciences,* Addison-Wesley, Menlo Park, California, 1968.

Stevens, S. S., "Issues in Psychophysical Measurement," *Psychological Review,* Vol. 78, 1971, pp. 426-450.

Stevens, S. S., and E. H. Galanter, "Ratio Scales and Category Scales for a Dozen Perceptual Continua," *Journal of Experimental Psychology,* Vol. 54, 1957, pp. 377-411.

Treisman, M., "Sensory Scaling and the Psychophysical Law," *Quarterly Journal of Experimental Psychology,* Vol. 16, 1964, pp. 11-22.

Veit, C. T., "Ratio and Subtractive Processes in Psychophysical Judgment," *Journal of Experimental Psychology: General,* Vol. 107, No. 1, 1978, pp. 81-107.