AD-754 215

EXPERIMENT SIMULATION

Charles W. Simon

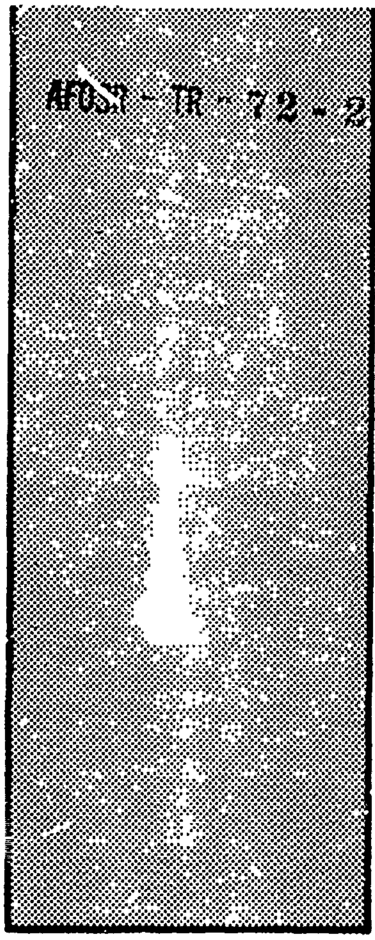
Illinois University
Savoy, Illinois

April 1972

DISTRIBUTED BY:

NTIS
National Technical Information Service
U. S. DEPARTMENT OF COMMERCE
5285 Port Royal Road, Springfield Va. 22151

AD754215



AFOSR - TR - 72 - 2468

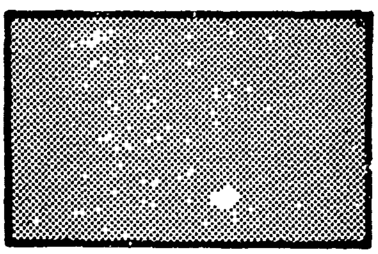
TECHNICAL REPORT NUMBER
ARL-72-7/AFOSR-72-3

Prepared under subcontract with the Aviation Research Laboratory, Institute of Aviation, University of Illinois at Urbana-Champaign, for the Air Force Office of Scientific Research (AFSC), United States Air Force.

F44620-70-C-0105

DDC
JAN 22 1973

EXPERIMENT SIMULATION



Charles W. Simon

AEROSPACE GROUP
HUGHES
DISPLAY SYSTEMS AND
HUMAN FACTORS DEPT
HUGHES AIRCRAFT COMPANY
CULVER CITY, CALIFORNIA

Reproduced by
NATIONAL TECHNICAL
INFORMATION SERVICE
U S Department of Commerce
Springfield VA 22151

Approved for public release;
distribution unlimited.

60

APRIL 1972

UNCLASSIFIED

DOCUMENT CONTROL DATA - R & D

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

1. ORIGINATING ACTIVITY (Corporate author) Aviation Research Laboratory, Institute of Aviation University of Illinois Urbana, Illinois 61801	2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED 2b. GROUP

3. REPORT TITLE

EXPERIMENT SIMULATION

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

Scientific Interim

5. AUTHOR(S) (First name, middle initial, last name)

Charles W. Simon

6. REPORT DATE April 1972	7a. TOTAL NO. OF PAGES ~~48~~ 60	7b. NO. OF REFS 9

8a. CONTRACT OR GRANT NO F44620-70-C-0105 b. PROJECT NO. 9778 c. 61102F d. 681313	9a. ORIGINATOR'S REPORT NUMBER(S) 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) AFOSR - TR - 72 - 2468

10. DISTRIBUTION STATEMENT

Approved for public release; distribution unlimited.

11. SUPPLEMENTARY NOTES TECH, OTHER	12. SPONSORING MILITARY ACTIVITY Air Force Office of Scientific Research 1400 Wilson Boulevard (NL) Arlington, Virginia 22209

13. ABSTRACT

Although the traditional experimental techniques employed by behavioral scientists are considered inadequate for applied studies of humans operating within a man-machine system, researchers have been reluctant to adopt improved methodologies. This reluctance is attributed to inadequate means of evaluating those methodologies in current use and to investigators' lack of experience with new methodologies. It is proposed that a computer program which simulates data generated by laboratory experiments can resolve both these problems quickly and economically. The primary purpose of the current paper is to establish that such a model for experiment simulation can be developed. This report outlines the basic characteristics of the simulation model, which assumes the form of a polynomial regression equation. Next it identifies and discusses many of the factors that usually operate in human factors engineering experiments.

The author emphasizes that the model should provide for both relevant and irrelevant sources of performance variance. Procedures are presented and illustrated for assigning weights to the various factors in the model including those factors that are qualitative and ones that are related nonlinearly to performance. The paper concludes with some considerations for making the technique of experiment simulation as useful and usable as possible, for both one who is designing the model and one who is "experimenting" within the realms of the modeled world.

ia

DD FORM 1 NOV 65 1473 UNCLASSIFIED

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Experiment Simulation						
Simulation Model						
Polynomial Regression Equation						
Coefficients						
Factors						
Methodologies						
Man-Computer Interface Considerations						

EXPERIMENT SIMULATION

A Tool to Exercise, Explore, Evaluate, and Teach Methodologies
for Applied Human Factors Engineering Experiments

Charles W. Simon

April 1972

Technical Report No. ARL-72-7/AFOSR-72-3

Prepared under subcontract with the Aviation Research
Laboratory, Institute of Aviation, University of Illinois
at Urbana-Champaign, for the Air Force Office of
Scientific Research (AFSC), United States Air Force.

F44620-70-C-0105

Equipment Engineering Divisions

AEROSPACE GROUP

Hughes Aircraft Company, Culver City, California

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

FOREWORD

After several abortive efforts to promote improved methods of conducting human factors engineering experiments, it was realized that a major deterrent was the latent fear that most people have of the unknown. Like the "banker's loan syndrome" -- the only people who can borrow money are those who don't need it -- the only persons willing to use new and advanced experimental techniques are those who have already had experience using them. The acceptance of this apparent contradiction as fact lead to the search for the means of resolving the problem.

If we could simulate the response characteristics of a laboratory experiment by storing its mathematical model in a computer, the computer rather than a lot of real people operating equipment in a laboratory could supply the performance data likely to occur under different operating conditions. Instead of finding out what a man would do when certain equipment parameters were manipulated and the data collected according to some systematic plan, one could query the computer data base. If the model of an experiment were properly prepared, experiment simulation would provide a rapid and inexpensive means for investigating, evaluating, experiencing, and teaching data collection techniques for human factors experiments.

In trying to get across the idea to others of what an experiment simulation is and what its contribution to applied human behavioral research would be, three misconceptions commonly occurred. These are explained here in order to help clarify the remainder of the paper.

1. Misconception: The mathematical model of an experiment is intended to substitute for experiments which will provide valid performance data for equipment design.

Fact: The simulation model is of the generic experiment and is not intended as a substitute for a particular experiment to understand man-machine relationships. It is to provide data which will contain sources of error characteristically present in actual laboratory experiments. While the model would naturally provide for the inclusion of equipment factors which could affect performance, it would also include those factors associated with the subjects, apparatus, and environment which tend to interfere with the collection of good data. As a simulation of an experiment (and not merely an engineering system), it could serve as a testing ground for experimental methodologies employed to reduce irrelevant effects so that the relevant can be specified as accurately and as economically as possible.

Validity of results in this form of experiment simulation depends only on the degree to which the results of the system-relevant data of a simulated experiment agree with the parameters in the computerized model. This is a measure of the effectiveness of the methodologies employed.

2. Misconception: Experiment simulation would be used primarily to compare one experimental design against another.

Fact: In the case of statistically correct experimental designs, the relative merits of each can be determined analytically. Knowing the assumptions on which the designs are based, one can specify what information is being lost in order to use a design that is more economical in its distribution of data collection points than a design that includes all factorial combinations. Therefore, no empirical comparison of two designs is necessary, because the goodness of the outcome will depend entirely on whether the information which was sacrificed was needed. If it were, the design should never have been used, and no empirical evaluation will tell anything more about it.

A more important use of experiment simulation is to identify the conditions under which certain designs are more useful than others. As stated earlier, one of the most practical values of experiment simulation is to enable a lot of experience to be gained quickly and inexpensively with certain experimental plans under a wide range of experimental conditions. In a similar vein, the technique can be employed as a tool for training experimentalists.

3. Misconception: To write an equation which would accurately simulate an experiment with human subjects would be too complex and difficult to be of any practical value.

   Fact: An approach will be described in this document which simplifies the development of experiment simulation models so that most persons experienced with the vagaries of laboratory experimentation and an appreciation of the purpose of the simulation can develop useful models in a matter of a few hours. Naming the factors which affect experimental results, both the desired and undesired ones, should be easy for the experienced investigator; these are the terms of the model. The real challenge lies in the ability to assign the proper weights to 30 or more terms for any single experiment in a way that is meaningful and purposeful. This paper shows that this is not only feasible but relatively simple. Whether or not the technique of experiment simulation could ever be developed into a "cookbook" procedure is uncertain. I hope it could not.

The paper does not attempt to document every detail or derive every relationship used in it, for these have been derived many times in statistical textbooks. It does demonstrate one way of getting the job done. Rather than more documentation of how one might simulate an experiment, the next step is to simulate one and gain experience in using this valuable tool.

Charles W. Simon
April 1972

ACKNOWLEDGMENTS

This paper was prepared in the Display Systems and Human Factors Department of Hughes Aircraft Company under Subcontract 2 with the Aviation Research Laboratory, Institute of Aviation, University of Illinois at Urbana-Champaign. The research was supported by the Air Force Office of Scientific Research, Air Force Systems Command, United States Air Force, under prime contract No. F44620-70-C-105 with the University of Illinois. Dr. Glen Finch of AFOSR is technical monitor of the program. Dr. Stanley N. Roscoe, Associate Director for Research of the Institute of Aviation, provided the opportunity and encouragement needed to prepare the paper.

The help of E. J. Dragavon, M. N. Ernstoff, and C. A. Greene of Hughes Aircraft Company, who freely gave their time to discuss many of the points in this paper, is acknowledged with appreciation. However, none of the persons or organizations mentioned above is in any way accountable for the material presented in this paper; that is the sole responsibility of the author.

I..rRODUCTION

Although human factors engineering concepts have yielded significant improvements in equipment design, the contributions of formal human factors engineering experiments have been relatively modest. The quality of experimentally obtained data, depending as it does on the methods employed to obtain the data, has failed to meet the needs of the data users. The traditional techniques used by behavioral scientists have not provided the necessary quantitative, generalizable information needed to design complex equipment and systems.

Although many useful experimental designs have been developed during the past 30 years, the behavioral scientist has tenaciously avoided using them, continuing instead to employ uncritically a narrow set of designs and stylized data collection procedures which are cumbersome, inefficient, uneconomical, and in some cases could not possibly provide the very information for which the experiment is to be performed.* These limitations become particularly acute in applied studies of humans operating as a part of a man-machine system. In spite of this, the search for better methods has remained relatively stagnant and uninspired.

Conditions Working against Improved Methodologies

The failure to improve the methodology employed in behavioral experimentation can be traced to at least two conditions which prevail when

*Simon, C.W. Considerations for the Proper Design and Interpretation of Human Factors Engineering Experiments. Equipment Engineering Division, Hughes Aircraft Company, Culver City, Calif. Tech. Report # ARL-71-27/AFOSR-71-11, December 1971 (In press) applied human factors engineering research is conducted. These are:

1. Inadequate means of evaluating the methodologies currently in use.
2. Inadequate experiences with new methodologies.

Evaluation. There is at this time no truly satisfactory way of evaluating the methods employed in human factors engineering research, particularly when these do not readily lend themselves to mathematical analysis. Experimental designs are a part of the methods in that they indicate a particular patterning of the data collection points. But this is only a small part of the whole process of designing an experiment which can be complicated by practical limitations placed on the data collection and by the differences between human behavior and chemical, agricultural, or engineering processes.

The acceptability of most applied human factors engineering experiments today is based more on whether reports meet contractual obligations of time and certain antiquated and sometimes irrelevant academic standards than on the validity and potential applicability of their results. By the time a report has reached its final form, the trials and tribulations encountered in carrying out a study, which might introduce bias or random error into the results, have long been glossed over or ignored. Although we may do sloppy experiments, we seldom report sloppy experiments. Although it is sometimes possible to detect a poorly conducted study from the data supplied in a finished report, many persons who administer research or who use research results do not have the necessary training to do so. Studies are seldom repeated, so even empirical reliability of the data remains untested. The validity of the data is an even more difficult criterion to evaluate than the methods employed in collection because there are often years between the time the results are used to design a real system and the time the system is tested operationally. By then, the original inputs have been so modified by tradeoffs with inputs relative to nonhuman components that also affect system performance that the validity of the human factors data alone is indeterminable. As a result, because feedback is seldom provided on data quality, much less on the means employed to collect the data, there is little motivation for an experimenter to improve his experimental results through improved methodology.

Experience. Even when it is recognized that a new experimental design could markedly improve the quality of human factors engineering data, it may not be used. For one thing, it is generally necessary to modify techniques developed for other fields before they can be successfully applied to studies involving the collection of human performance data. For another, the better techniques allow for numerous variations which trade data collection economy against information quality. In practice, time and money constraints discourage a period of exploration to determine whether such variations have a practical effect on results. As a result, most human factors engineering experimenters are reluctant to depart from familiar and conventional approaches, however limited, to try designs with which they have had no experience.

A Different Approach

The rapid acceptance of promising new experimental methodologies will occur only when there are ways of quickly and economically evaluating their effectiveness under conditions likely to be found in the conduct of laboratory experiments. One such way would be to develop a computer program which literally simulates the effects which operate on performance data as it is being collected in the course of a laboratory experiment, and use this program to disco[illegible] plans most accurately and inexpensively discover the true characteristics of an artificially-created data base.

The construction and use of such a simulation is conceptually quite simple. A data base of operator performance as a function of all relevant sources of variation that occur in human factors engineering experiments would be stored in a computer. This data base could be queried according to some experimental plan and the resulting sample data analyzed. The accuracy with which the sample data represent the information in the original data base provides a measure of the effectiveness of the experimental plan. The effectiveness of this simulation, therefore, depends on how well it creates the conditions an experimenter must face in the conduct of a laboratory study; it is not intended that this simulation would predict operator performance in some engineering system configuration for a particular mission.

Advantages and Applications

Properly developed and employed, such a simulation could provide an inexpensive basis for comparing experimental designs and methodologies, for evaluating experimenters and proposed data collection plans against absolute standards, and for gaining experience with new designs across a broad spectrum of operational conditions.

Some specific applications in which an experiment simulation could be effectively applied are:

1. To provide a quick and inexpensive way of gaining experience using new experimental designs and data collection procedures to determine under what conditions they are or are not suitable.
2. To determine empirically the effect of deviations from the experimental designs prescribed by the statistician.
3. To develop optimum procedures for collecting human factors engineering data in the presence of different irrelevant sources, types, and amounts of performance variance.
4. To standardized problems to select or quantitatively evaluate the ability of experimental designs and experimenters to handle complexities characteristic of applied behavioral research.
5. To use simulation as a training device to show students the intricacies of experimental design, including the decisions required to collect clean data economically.

Considerations for the development of a model for experiment simulation are discussed in the sections which follow. The primary purpose of this paper is to show that it can be done.

SIMULATING AN EXPERIMENT

Experiment simulation is actually a method by which two people, or two groups of people, or one person wearing two hats, can play a game. One of the two will be called the "creator;" the other will be called the "experimenter." The game consists of the creator building a "world" inside a computer and of the experimenter, through empirical sampling of the stored data, attempting to describe this world.

There would be no game, however, unless there were in this simulation – in the real world – sources of "irrelevant" variance which distort the relationships described in the sampled data from those occurring in the original data base. Both the type and magnitude of these sources of error can be manipulated by the creator, thereby varying the difficulty of the experimenter's task. The experimenter's weapon against this unwanted variance is the astuteness of his experimental sampling plan. The score of the game is determined by the accuracy of the experimenter's description of the stored world and the economy of his effort to derive it.

General Approach

The technique of experiment simulation enables the creator to build a computer program and a data base which will provide an experimenter, using a particular data-collection plan, with performance scores which are functions of the factors relevant to the experiment as well as unwanted conditions which would ordinarily distort the data in the course of an actual experiment. The presence of unwanted conditions in an experiment simulation forces an experimenter -- if he is to do a good job -- to plan for the same contingencies and make the same decisions in selecting an appropriate experimental plan that he would have done were he to have collected his data using real subjects. The steps in the preparation and use of such an experiment simulation are cited below.

First, a polynomial is used to represent the basic model for experiment simulation. Standard terms and frequency distributions are stored in a computer data base. These terms represent those factors affecting performance data in any experiment, i.e. those related to all relevant variables to be included in the experiment, along with other unwanted sources of variance which might artificially develop out of the experimental situation, generally associated with irrelevant (to the experiment) subject, laboratory environment, and experimental methodology factors. The creator assigns identities and measurement scales to the terms of the equation, depending on the particular experimental conditions he wishes to simulate.

Second, the creator assigns coefficients which specify the relative effect each term has on performance. For any particular model, the creator has considerable leeway in how complex in degree and order the model will be and in how much and what type of irrelevant variance will be included.

Third, the creator prepares a scenario for the experimenter. This scenario would provide the experimenter with the same background information that he would have available in an actual study. He would know something about the problems and variables being investigated as well as the limits on time and personnel available for the experiment.

Fourth, the experimenter is given the task of planning his experiment, solving logistic problems, and collecting and analyzing the data. If the simulation is a good one, he will be forced -- step by step -- to make the same decisions he would ordinarily be forced to make, and where his decisions are deficient, so are his experimental results. Instead of measuring the performance of an actual subject, the computer solves the equation introduced by the creator for the data selection points introduced by the experimenter and outputs "experimental" data samples accordingly.

Fifth, the description of the world as determined by the experimenter is compared quantitatively to the actual characteristics of the simulated world. Evaluation criteria include the size and type of the discrepancy between the experimental results and the true simulated world as well as the amount of time, number of subjects, and the number of observations required to arrive at the results.

Implementation

In the sections that follow, a preliminary description of how this simulation can be achieved is presented. First, some basic characteristics of the simulation model will be noted. Then many of the factors that ordinarily play a role in human factors engineering experiments will be identified. These are used to specify the terms of the polynomial equation. Next some practical procedures for assigning coefficients by the creator will be given. Finally, some considerations for building an interface between the computer and both the creator and experimenter will be covered.

THE SIMULATION MODEL

A polynomial regression equation will be the model used in experiment simulation. To prepare such a model, the creator must recognize that factors needed to simulate an experiment can be classified into four general groups, the characteristics of which determine in part how they are to be handled when included in the total equation. These classes and characteristics are shown in Table 1 and discussed on the following pages.

Table 1. Classes and characteristics of factors

CLASS	CHARACTERISTICS	
	Input (Term Value)	Effect (Coefficient Value)
Quantitative	Specific	Specific
Qualitative, Assigned	Artificial	Specific
Qualitative, Random	Artificial	Statistical
Events	Artificial	Statistical or Specific

Quantitative Factors

Quantitative factors are those which can be related by a mathematical function to performance. These factors can be measured on either an ordinal, interval, or ratio scale. "Resolution," "display size," and "rate of movement" are examples of quantitative factors. Certain complex, multicomponent factors which are scaled subjectively on a single dimension are also quantitative; for example, the backgrounds of pictorial imagery used in experiments on target acquisition can be ordered on a single scale called "complexity," although this quality is actually a composite of many conditions of the imagery. "Trials," i.e., the discrete units of time a subject is exposed to experimental conditions, is an example of a quasi-continuous quantitative factor. Although each trial is a discrete entity, a continuum of equal intervals is implied by a function relating trials to performance as in the case of a learning curve.

The relationship between a quantitative, continuous factor, x_i, and performance, y_i, in this simulation can be expressed by a polynomial of the form:

$$y_i = \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 \cdots + \beta_n x_i^n$$

where n is the highest order the creator decides is needed to express the relationship. The Beta coefficients represent the relative weight of the linear, quadratic, cubic, etc., components of the relationship and may be either positive or negative. For quantitative, continuous factors, the X terms can take on any value within the boundaries of the world being simulated.

Two or more factors can be included in this model by simply adding the weighted terms together. If, however, there are interactions, then entirely new terms must be included in the model along with their coefficients which reflect the degree and nature of the interaction. Thus, the terms required to fully represent a two-factor interaction would be

$$y_{ij} = \beta_{11} x_i x_j + \beta_{21} x_i^2 x_j + \beta_{12} x_i x_j^2 + \beta_{22} x_i^2 x_j^2 \cdots \cdots + \beta_{st} x_i^s x_j^t$$

where s may equal t and s + t equals the highest degree term the creator decides to use in the equation. Neither s nor t need equal the n (order) of the terms used in the single factor function, and some Betas can take on the value of zero.

Thus for every quantitative factor to be included in the model, the creator must do the following:

1. Identify the factor (or interaction) by name:

   This name is identified by the subscript.

   e.g. $\overline{X}_i$*

2. Decide which factors will interact:

   e.g. $\overline{X}_i \overline{X}_j$, $\overline{X}_i \overline{X}_m$, $\overline{X}_j \overline{X}_k \overline{X}_l$

3. Decide on the highest order relationship which will be used for each factor:

   This establishes how many terms will be required in the model for a factor.

   e.g. X_i , X_i^2 , X_i^3 (3rd order max.)

4. Decide on the highest degree term for each interaction:

   This establishes how many terms will be required in the model for an interaction.

   e.g. $X_i X_j$, $X_i^2 X_j$, $X_i X_j^2$, $X_j X_k X_m$ (3rd degree max.)

---

*In this paper, the following symbology is used:

Factor = letter with bar
Term = letter without bar

In practice, the effect on performance of any fourth degree (or larger) regression component of an equipment parameter would be expected to be trivial. In fact, by selecting the appropriate measurement scale, the effects of many second degree components of the equipment factors can often be reduced to an insignificant size. In Table 2, the forms of the regression terms required to construct equations of up to and including fourth degree are listed. For example, to write a third degree equation for three factors, all of the forms of the terms shown inside the dotted line in Table 2 would be required plus the comparable terms for Factors $\overline{B}$ and $\overline{C}$ and interactions $\overline{AC}$ and $\overline{BC}$, which parallel those shown for Factor $\overline{A}$ and interaction $\overline{AB}$. Thus the terms for a

TABLE 2. Forms of regression terms needed in multi-factor equations of specified degrees.

REPRESENTATIVE FACTORS AND INTERACTION	DEGREE OF THE EQUATION			
	1st	2nd	3rd	4th
(1) $\overline{A}$	A	A^2	A^3	A^4
(2) $\overline{AB}$		AB	AB^2 A^2B	AB^3 A^2B^2 A^3B
(3) $\overline{ABC}$			ABC	ABC^2 AB^2C A^2BC
(4) $\overline{ABCD}$				ABCD
(5) $\overline{ABCDE}$	(Cannot be represented by a fourth-degree equation.)			

three factor, third degree equation would be:

$$A,\ A^2,\ A^3,\ B,\ B^2,\ B^3,\ C,\ C^2,\ C^3,\ AB,\ AB^2,\ AC,\ AC^2,\ BC,\ BC^2,\ ABC$$

each with one degree of freedom. Whether or not all of these terms would actually affect performance would depend on the coefficients assigned to each by the creator. In the real world of equipment design, it would be realistic to expect many of these terms to have a coefficient of zero.

Qualitative Factors, Assigned

Qualitative factors are those made up of two or more discrete levels, categories, or conditions which cannot be ordered on a numerical scale and are identified only by name. "Manual control devices" is an example of a qualitative factor that might be composed of the following types of devices: joy stick, rolling ball, pencil stick, rotary knob, and so on. Although one can determine the effect each type individually has on performance, no meaningful function can be drawn relating "Manual control devices" to performance.

This qualitative factor is classified "Assigned" because in constructing the simulation equation, the creator will decide exactly how much of an effect each category (or control type) is to have on performance. He does this by assigning the mean change in performance which would occur when each control is being tested by the experimenter in the experiment simulation; the computer would combine this value with those of the other effects which are operating at the time to calculate the performance score for that trial.

Qualitative Factors, Random

Certain qualitative factors are more appropriately simulated by having the creator specify only the mean and standard deviation of the effects of all of the factor, rather than by assigning specific amounts representing the effects of each condition as was done for the previous qualitative factor. Given the distribution data, the computer would select from it a value at random to represent the effect of each condition of the factor. Once a value were selected in this manner and associated with a particular experimental condition, it would remain so for the remainder of the experiment simulation.

"Subject skill" can represent a variable of this type, as in the case when an experimenter selects his subjects directly from a larger population without a priori knowledge of their skill on the experimental task. To simulate this, rather than have the creator assign a specific skill-level value to each subject of the total population, he would instead specify only the mean and standard deviation of a distribution of skill-levels from which the computer would randomly select and assign values for each subject the experimenter decides to use in his experiment simulation. Each time a particular subject is tested under a set of experimental conditions, the computer would adjust the mean performance level for that condition to reflect the skill-level of the subject as determined by the randomly selected value. In this case, the experimenter would have no direct indication of subject's skill-level, except as it is reflected in the performance data.

A second situation can exist in practice which in the simulation would determine the effect of each condition of a qualitative factor by a random selection process. This would be the case when the experimenter selects his subjects on the basis of a task-related pretest. In the simulation, the experimenter would indicate to the computer the number of subjects he intends to use and that he intends to give them a pretest selected from those provided by the creator's scenario. The computer would then provide the experimenter with a printout relating a subject identification number with each pretest score that had been selected at random from a distribution of scores supplied by the creator. The experimenter could then select from the population those subjects which he intends to use in the experiment simulation, on the basis of the pretest scores. The computer in turn would calculate the task performance effect for each subject from the amount of correlation (r) between pretest scores (t) and performance (p), i.e., r_{tp}, a value which had been supplied by the creator. The actual calculations for determining the mean performance adjustment for each subject would be: 1) take the randomly selected pretest score and multiply it by the correlation between the pretest and task performance; 2) include some uncertainty by adding a value selected randomly by the computer from a distribution of the task performance scores multiplied by the coefficient of alienation $(1 - r^2_{tp})$. The lack of a perfect correlation between pretest scores and performance on the task represents another source of irrelevant variance in the results of the experiment simulation, a problem that an experimenter is forced to contend with when he uses a pretest for selecting his subject sample or assigning them to experimental conditions in an actual experiment. The approximate size of this correlation may or may not be revealed to the experimenter through the scenario supplied by the creator.

Events

Events might be treated as another form of qualitative factor. An event is classified separately, however, because unlike the other factors, the experimenter has no control over its introduction into the simulation results. By including such a factor in the experiment simulation, it provides the creator with an option of playing havoc with the experiment and simulating what might happen if some unusual occurrence produced a sudden change in performance without warning, precedence, or repetition. For example, to simulate the effect of a momentary distraction of a subject from the task, a correction might be made to a single trial which yields an abnormally low score. The extent of this effect and its location in time could be specifically assigned if it were to happen only a very few times during an experiment, or the assignment of magnitude and location could be left to the computer which selects a value to random from a distribution supplied by the creator.

Summary

The following summarizes through examples the treatments of the four classes of factors:

Quantitative Factors:

Stored in computer: $y = .45\ x - .12\ x^2$

Equation specified by creator who assigns the coefficients to indicate a certain change in y for a unit change in x and x^2 within finite boundaries.

Experimenter can select any value of x within these boundaries and computer will indicate value of y.

Qualitative Factors, Assigned:

Stored in computer:

	x	y
Pencil stick	1	+.23
Joy stick	2	+.11
Rolling ball	3	-.08
Rotary knob	4	-.26

Creator indicates change in mean performance (y) when experimenter uses each type of manual control in his study.

Experimenter addresses computer and identifies type of control to be used by the arbitrarily assigned numbers, 1 to 4.

Qualitative Factors, Random:

Stored in computer:

Subjects: N = 100
Skill level: Mean = .72, Sigma = .14, Transform = 0.
Task-related test: Mean = 50, Sigma = 10, Transform = 0.
Task-related test: (correlates with performance) .78

Creator specifies above values. Experimenter may or may not decide to use task-related tests which are available.

Experimenter indicates he will use m subjects from population. Subject identification is its order number from 1 to m. Computer assigns the effect each has on mean performance based on values selected at random from the distribution provided by creator. Experimenter may reach these values by first "giving" subjects a task-related test which he uses as a selection device. This test's correlation with performance is involved in estimating the effect each subject will have on performance.

Event, Random and Systematic:

Information stored in computer is same as for two types of qualitative factors.

Creator specifies that data, plus the point in time during an experiment when these will be included in the performance estimate.

Experimenter is unaware that additional inputs have been made, except as he may or may not detect them from the response.

Implications for Computer Programming

The classes of factors described above are all found in actual laboratory experiments Therefore, the basic simulation model must be prepared to include them in case the creator decides they are necessary for the particular world he plans to create. Because factors do break down into these few classes, the requirement for diversity of computer programming is not very great. No matter how many factors are included in a particular model, the problems of programming are fairly limited and, in fact, a generalizable program could be written in modular fashion so that the creator could decide the class of factor, its order, its coefficients, and its sign in some preplanned order, and the computer could construct the complete model.

FACTORS TO BE INCLUDED IN AN EXPERIMENT SIMULATION

In an actual experiment, the performance level on each trial is the result of factors other than those directly related to the equipment and system parameters which the experimenter is studying. For example, performance varies due to differences among the subjects. Further perturbations are added depending on the procedures used to collect the data. Because irrelevant performance variance can distort the relevant measures of performance, the experimenter, in planning his experiment, ordinarily tries to control, compensate, or correct for them.

In a simulated experiment, the experimenter obtains his data from a computer rather than from the actual testing of real subjects under laboratory conditions. The simulation, to be complete, must not only provide an estimate of the performance of an imaginary group of subjects operating equipment but also provide the experimenter with the opportunity and the requirement to make the same kinds of decisions he would need to make were he dealing with the intransigencies of an actual laboratory experiment.

Identities of Typical Factors

In this section, the most common sources of variance that affect the results of a typical human factors engineering experiment will be discussed. These factors can be conveniently divided into the following groups:

1. System (S) Factors. These are factors associated with the equipment, the environment, the task, and other aspects of the system that the experimenter is (or should be) interested in

relating to operator performance. In the design of an experiment, these factors are often referred to as the "independent variables." It is to understand these factors that the experiment is being conducted.

2. Subject (O) Factors. These are the characteristics of a subject -- operator or observer -- which affect or relate to his basic proficiency on the experimental task at the beginning of a study.

3. Temporal (T) Factors. These refer to conditions associated with the subjects, the equipment, or the environment which result in an average change in performance as a function of time. These changes may take place during work or rest periods over the course of the total experimental time.

System (S) Factors

In the simulation it is necessary to characterize the engineering system, which the experimenter, through the collection of data, hopes to describe. System factors are needed to identify equipment, environment, task, and other variables related to the performance of the system being investigated. The experimenter would be expected to include some of these factors in his experimental design.

By way of illustration, some system factors that might be included in the simulation of an experiment on the design of a reconnaissance display are:

Equipment	Environment	Task	Other
Resolution	Ambient illumination	Target recognition	Target size
Dynamic range	Vibration	Response-time limit	Background complexity
Display size			Type of imagery
Image motion rate			

The creator has the option of adding others. Interactions among these factors are commonly found in the real world and should be included in the simulation model as well.

In modelling the experiment simulation, the creator would presumably include more system factors than the experimenter would ever be expected to study in a single experiment. This is characteristic of an actual experiment in which the selection of variables is limited by practical considerations of the size and cost of an effort and the experimenter's astuteness or ambition in ferreting out all of the relevant variables. For all practical purposes, it is believed that fewer than 15 factors and selected interactions should account for nearly all of the total variance attributable to system considerations in a particular task. The creator can use the scenario to keep the experimenter's choice of system factors within simulation boundaries.

Subject (O) Factors

A part of experiment simulation must include the characteristics of the population(s) from which the subjects are to be drawn. These factors should correspond as much as possible to those that would be considered were the subjects real and the experimenter required to select them according to chance or specific task-related criteria. As in the real world, the creator can supply the experimenter with unlimited numbers and types of subjects or can place severe restrictions on their characteristics and availability.

Among the types of information derivable from the simulation that should be available on each subject are:

1. Task Aptitude. This is the basic performance capability for each subject that can be represented by a symmetrical S-shaped (ogive) growth curve and described by two quantities: mean and sigma. The mean is proportional to the maximum growth potential, and the sigma is proportional to the time required to achieve that maximum.
2. Experience Level. This is the level of performance at which each subject begins the experiment. Performance at the beginning of the experiment need not begin at zero.

3. Related Performance. Scores on performance-related tasks may be provided which might be used in the selection of subjects. Performance on these tasks can correlate between ±1 with the actual task.

4. Nonperformance Characteristics. These permit the classification and selection of subjects according to such conditions as sex or age or type of training, etc. They too can be correlated with performance.

Temporal (T) Factors

Time elapses during the conduct of an experiment. During this time changes are taking place in the subjects, the environment, and the equipment. Some of these changes can be controlled, some can be compensated for, while others may only be monitored. In many cases, these changes are irrelevant to the basic purposes of the experiment and can only be expected to introduce disturbances into the pertinent experimental data. Because of their prominence in most human factors engineering experiments for equipment design, however, provisions must be made in the simulation model so that these sources of variance will emerge if the experimenter creates situations which encourage them or fails to employ methodologies which will minimize them.

In this simulation, subject-related temporal factors will be treated somewhat differently from equipment/environment-related temporal factors. In the discussions that follow, these temporal factors are treated as "main" effects; that is, they refer to changes in average performance over time, ones representative of the performance for each subject or system condition. There are also in actual experiments temporal changes that affect individual subjects or system conditions differently; these will be discussed in the section on interactions among S, O, and T factors.

Subject-related temporal factors. Whether or not these factors are operative in a simulation depends upon the experimental design selected by the experimenter. If he has a sufficient number of subjects available for his experiment, he may test each subject only once and avoid the special problems imposed by subject-related performance changes as a function of time. However, in many applied human factors engineering studes, subject availability is limited, the number of experimental conditions is large, and subjects must be used more than once. In some studies, where learning is of primary interest, certain subject-related changes over time would not be considered irrelevant.

For subjects who are repeatedly tested in an experiment, the time continuum can be divided into two major periods: work and rest. These in turn can be further subdivided into intervals that have different but typical effects on subject performance over time (Figure 1). The more common subject-related effects over time that should be included in a simulation model are described below.

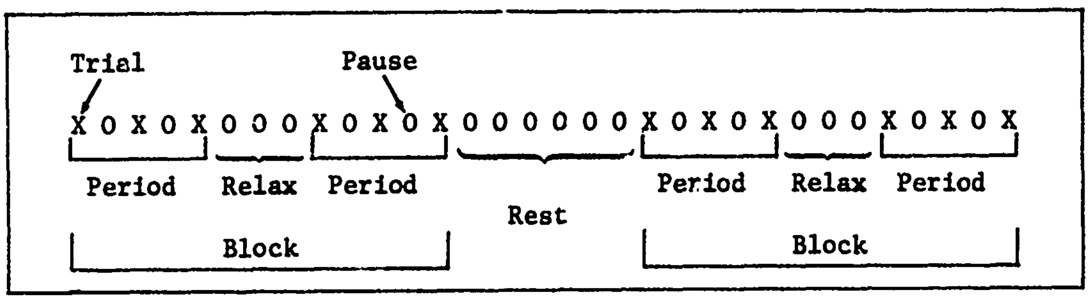


Figure 1. Stylized work-rest patterns

<u>Trials</u>. These are the intervals in which performance data on the task are collected from each subject. For this model, one trial yields one performance score.

There are two distinct types of trial factors: intracondition and intercondition trials. <u>Intracondition trials</u> represent repeated measures of the same experimental condition. <u>Intercondition trials</u> represent the discrete number of measurements made across the same or different experimental conditions.

Performance would vary randomly on the intracondition trials. Performance would be expected to show a rising function over intercondition trials, indicative of learning. The slope of the function depends on the portion of the learning curve that is operative at the time.

<u>Pauses</u>. These are the intervals between two trials, usually just long enough to change or reset the equipment for the next trial or to record the data of the previous trial. During these intervals, performance will remain essentially constant. Any effects of pauses will show up in the period factor.

<u>Periods</u>. These are the total intervals covered by a series of trials and pauses. They are separated from one another by relax intervals. A series of trials and pauses might represent replications of a single experimental condition run sequentially or a series of different equipment settings with only the setup pauses interpolated. During one period, fatigue factors can operate on performance as a function of the length of the period.

<u>Relax</u>. These intervals occur between two periods and are significantly longer than pauses. In an actual experiment, these are generally introduced after a run of trials to allow the subject time to "take a break" and may be used when more complex or time-consuming equipment changes are required.

During these intervals, negative effects that may have built up during a period might dissipate or begin to dissipate.

<u>Blocks</u>. This title is used to designate a clump of periods separated by rather extensive rest intervals from other clumps of periods. Commonly, in actual experiments, when the number

of experimental conditions is so great that subjects cannot be studied on them all during a relatively contiguous time period, the experimental runs are broken into major blocks of times, such as from day to day or morning to afternoon. Average performance changes from block to block, if any, will be stepwise in nature and seldom systematic.

Rests. This is the interval between blocks which tends to eliminate fatigue effects that may have built up within a block. In addition, some performance decrement as a function of forgetting is possible if the period is too long.

Equipment/Environment-Related Temporal Factors. These factors usually fall into three classes:

- Continuous, random
- Continuous, systematic
- Discrete, random

Continuous, random effects are exemplified by a slow drift in equipment voltage over extended periods of time without an obvious pattern being exhibited. Continuous, systematic effects are typified by a voltage drift that occurs during an equipment warmup period or changes in ambient illumination throughout the day. Both of these show definable and repeatable patterns of change. Discrete, random effects can occur when equipment is set at an incorrect value during a particular trial or period. They can also occur from day to day (i.e., block to block) if whatever continuity might actually exist is not readily apparent. These temporal effects may or may not correspond to the beginnings and ends of the intervals considered in the discussion of subject-related temporal effects.

Interactions

In addition to interactions that might exist among factors within a group (e.g., the systems group), terms must be provided in the model for those interactions between groups (S, O, T) that are commonly found in actual experiments. Typical situations that create interactions are:

1. Subjects seldom begin an experiment at the same point on a learning curve or have the same skill levels. This could result in O x T interactions.
2. A piece of equipment may be adjusted incorrectly for several trials, a condition that may not be discovered until later. If the effect on performance were sizeable, this could result in an S x T interaction.
3. As in the case of any unusual event, if the error in Item 2 did not affect the performance of all subjects (perhaps some were tested on a different day when the equipment was set up properly), this could result in an S x O x T interaction.
4. If subjects are not homogeneous and have distinctly different experiences with the equipment conditions being tested, an S x O interaction could occur.
5. The use of counterbalancing as an experimental method could result in S x O x T interactions, or any two-factor interaction.
6. Relative performance levels obtained when operating two equipment conditions can differ and even invert, depending on which condition is tested first. This could result in an S x T interaction.

In an actual experiment, many interaction effects may remain hidden and impossible to isolate because of the experimental design employed. However, because they do occur in the real world, the creator must make provisions for them in his simulation model. Whether they will appear in the results of a simulated experiment depends on the experimental plan employed by the experimenter. If he fails to use a design in which the same subjects run all or some of the experimental conditions, then subject-related temporal factors and the interactions associated with them will not be revealed. The simulation, however, must be capable of creating a realistic effect no matter what route the experimenter ultimately decides to take. When these intergroup interactions do exist, they more often than not represent the "noise" in the experiment and are a major component of the measure of experiment quality.

SPECIFYING THE COEFFICIENTS

Once factors have been selected for an experiment simulation, the creator must then specify the coefficients of the terms representing these factors in the simulation model. These coefficients indicate how much each factor affects performance. Because the game requires the experimenter to use his investigatory skills to ferret out a valid description of the simulated world, whatever it may be, in theory the creator could assign any coefficients he pleased to the model, and the competent experimenter would ultimately be able to discover the characteristics of the simulated world. In practice, however, assigning coefficients in this manner would not be satisfactory. Instead, two limiting principles should be followed if experiment simulation is to be used effectively. These principles are:

1. Experiment simulation should be rational. Assigning coefficients without attempting to produce a realistic simulation creates artificial problems that an experimenter would never expect to find in an actual human factors engineering experiment. The creator, to make his simulation effective, must provide an element of realism in his simulated world. Experimenters seldom approach experimental problems without some knowledge of the world to be investigated, and that world is not a helter-skelter one. This fact is used by the skilled experimenter to design his experiment properly. He makes use of his knowledge of the regularities and relationships that normally exist in the psychophysical world, of certain irrelevant effects that tend to be present from study to study, and of conditions characteristic of his particular problem.

2. Experiment simulation should test the experimenter. Coefficients should be assigned so as to create a simulated world that tests the experimenter's ability to handle situations that might affect the validity of his data. Coefficients can be increased or decreased in value, distorting certain irrelevant effects in order to provide a simulation against which methods for reducing or eliminating these effects can be explored. Coefficients can be assigned which, when combined with restrictions placed upon the experimenter, make data collection a more difficult task; this enables exploration of ways to ease the problems.

For any experiment simulation, the creator will need to assign coefficients to 30 or more terms. To do this knowledgeably so as to satisfy both principles and at the same time be assured that experiment simulation remains a practical tool that will be used frequently, the task of coefficient assignment must be made reasonably uncomplicated. A set of principles, or relationships, makes this possible.

Weighing the Factors

Principle I. The sum of the proportion of total variance accounted for by each statistically indendent factor equals 1.00, and the sum of all subsets combined will equal 1.00.

$$p_t = p_1 + p_2 + P_3 + P_4 = 1.00$$

$$p_2 = p_5 + p_6$$

$$p_t = p_1 + p_5 + p_6 + p_3 + p_4 = 1.00$$

$$p_t = (p_1 + p_5) + (p_6 + p_3 + p_4) = 1.00$$

Rather than think in terms of the coefficients of the model, the creator can begin by thinking of the strength of the relationship he wishes each factor to have on performance. Because "strength of relationship" is expressed by the proportion of total variance accounted for by the factor, the creator's decisions are simplified by the above principle which permits him to consider the problem in parts. Thus, if all items in the list of factors cited in the previous section affect performance, the creator can break these into meaningful subsets related to the type of world he is trying to build. Then he can consider the smaller number of factors in each subset separately and make simple comparative judgments of their relative effect on performance. Each decision the creator makes in these cases reflects his particular goals in constructing the simulation. In the text that follows, the reader will be lead through some considerations and imaginary judgments that a creator might make to illustrate the steps that will end eventually in coefficients describing the world intended by the creator.

To begin, the creator must first decide whether he wants to make the overall difficulty of the experimenter's task great or small. By classifying the many factors into one of two groups -- relevant or irrelevant -- he can influence the task difficulty by the proportion of performance variability he attributes to each group. Relevant factors would be the ones the experimenter is interested in relating to performance. Irrelevant factors would be those that introduce chance errors and bias into these estimations. If the creator starts with a world in which the irrelevant variables contribute relatively little, then no matter how poorly the experimenter may sample that world in order to discover the properties of the relevant variable, his information will be of relatively high quality. If, on the other hand, the creator makes the contribution of the irrelevant factors large, then the experimenter -- in order to obtain untainted data about the relevant factors -- must carefully plan his data collection techniques and use any available means of reducing the unwanted sources of variance.

To reiterate, relevance as used in this paper is defined from the point of view of the experimenter. A factor is relevant when it is the reason the experimenter performs his experiment. Anything that interferes with his obtaining good, clean, and valid information about the relevant factors is called irrelevant. It is not irrelevant in the sense that it creates error in the data, only that it was irrelevant to the experimenter's interest and represents an unwanted source of variance.

Let us conclude, for this example, that the creator decides to distribute the proportional contribution of relevant and irrelevant factors as:

Source of Variance		
Relevant	Irrelevant	
0.85	0.15	= 1.00

What is considered relevant can change from simulation to simulation and depends on the problem posed to the experimenter by the creator in the scenario. Ordinarily all equipment and system (S) factors would be considered relevant, even though information about them all is not requested in the scenario. Thus, an experimenter may be asked in the scenario to find the relationship between display resolution and target detection. However, in the experiment simulation, as in the real world, other factors such as contrast, noise, and so forth, also affect performance. A good experimenter will either add these as extra factors in his study of resolution or specify the values at which these are held constant.

Subject (O) factors may or may not be considered relevant. If the experimenter is expected to determine whether subjects of low, medium, and high skill perform differently with different devices, then the degree of variability among skill levels would be relevant. If, however, different subjects were to be used individually in each cell of the experimental matrix, any variance contributed by them to the measure of performance on the equipment might be considered irrelevant depending on whether or not the subject sample were truly representative of the population. Let us assume for this example that the creator decides to make the skill level (Sk) a relevant factor. At the same time, he would place a "within-skill-level subject variability" factor among the irrelevant ones.

To continue the illustration, we shall let the creator decide to have five equipment or system (S) factors, A, B, C, D and E, along with two sets of two-factor interactions, C x E and D x E. Let us also assume that he decides to include a ninth source of relevant variance, the interaction between subject's skill (Sk) and equipment factor, D. His next step then is to assign relative importance levels to these nine sources of relevant variance. This decision may be based on conditions in the real world, on the distribution typical of the relative effects of factors in an experiment (i.e., they often distribute themselves exponentially), and on how he wishes to test the experimenter. In this example, the following distribution might represent the results when the creator ranked his relevant sources of variance and then assigned the proportion he wished each to contribute to this quality:

Relevant Sources

	1	2	3	4	5	6	7	8	9	
Greatest Effect on Performance	Sk	D	B	E	DxE	SkxD	A	C	ExC	Least Effect on Performance
	.30	.24	.12	.09	.08	.06	.05	.04	.02	= 1.00

The above numbers could be changed in numerous ways, depending on the creator's goal. As written, these numbers show that in this simulation the creator has decided to deemphasize the effect of interactions of performance. Also if, in the scenario, he only mentioned the first four factors, then he has made the experimenter's task easier by not making the remaining relevant factors have too large an effect in case they are overlooked. The experimenter can be "squeezed" by the creator in this kind of simulation in the same manner that he may be squeezed when he has many relevant factors to study but is limited in data collection time and money.

As another example of how the creator can structure the simulation, let us examine some of the decisions the creator might make concerning the irrelevant sources of variance. Considering that these are "irrelevant" sources of performance variability, the creator would first decide whether he would want subject (O) factors or temporal (T) factors to be the major impedance to the experimenter's ability to obtain clean results. If it is the former (that is, if he makes the within-skill-level subject variability larger) it may be because the creator wants to test the experimenter's subject selection techniques. If instead the creator makes the temporal factors larger, he may do so in order to examine methods of controlling certain transfer effects so typical in experiments with humans or to discover ways of minimizing the effects of uncontrollable equipment fluctuations on the relevent information. Let us imagine that for this example the creator decides that, of the irrelevant sources of variance, he wants to place approximately four times as much emphasis on temporal as on subject factors, and he wants to emphasize subject-related temporal factors the most. He might translate this into the following proportions:

Irrelevant Sources

Subject Factors	Temporal Factors		
Within-skill-level	Subject-related	Equipment-related	
.20	.50	.30	= 1.00

What these numbers will mean to the experimenter (if he knew them, and of course he cannot know them except through his sampling of the data base during the simulated experiment) is that he probably would be able to collect cleaner data if he were to use different subjects under each experimental condition than if he were to use the same subject across many experimental conditions. This leads the creator to a number of other decisions that can make the experimenter's task easier or more difficult. If in the scenario he limits the size of the subject population, it may force the experimenter to try to use the same subject over a number of conditions. If he provides some secondary task, which the experimenter could use to select or match his subjects, he may make this secondary task correlate highly with the primary task and thereby serve its purpose well or correlate slightly or not at all, thereby introducing unwanted sources of variance into the study.

In deciding how to weight the subject-related temporal factors, the creator must consider which one should affect performance the most, just in case the experimenter decides to use the same subject more than once during the experiment. By assigning a large proportion of the variance to the intracondition trial factor, the creator will cause the experimenter's data to be less reliable if the same subject is tested repeatedly under the same condition. A large positive coefficient for the intercondition trial factor could simulate subjects who show considerable improvement from " on-the-job practice," whereas a small one would simulate subjects at some learning plateau. The size of the proportion assigned to the periods factor determines whether subjects behave as if they were fatigued, and the size of the proportion for the relax factor determines how long an experimenter must make his simulated subjects wait between sessions to overcome the fatigue effects.

In certain cases, to maintain the proper degree of reality, the assignment of proportions is slightly more complicated; for example, the positive effect of rest should not exceed the negative effect of fatigue. However, as is the case with the entire equation, values can be assigned relatively quickly, and finer adjustments can be made later to bring out qualities that were missed in the original assignment. For this illustration, let us assume that the creator proportions subject-related temporal effects among the following factors as shown here:

Subject-Related Temporal Effects

Intracondition Trials	Intercondition Trials	Periods	Relax	
.40	.20	.20	.20	= 1.00

All temporal factors represent a relation between performance and a unit of time. Because the creator cannot know how much time the experimenter plans to schedule for the various work and rest periods, he must create relationships that cover sufficient units of time to handle all reasonable situations. The creator can control the experimenter through the scenario by limiting the total time available to do the experiment -- a common and realistic situation.

Looking next at the equipment-related temporal factors, the creator might decide to simulate an effect representing a slow, low-amplitude drift in the equipment. The amplitude of the drift would be related directly to the performance loss. In this illustration, he will make this sole factor account for all (1.00) of the variance associated with the equipment-related temporal factors.

Summarizing the Apportioning-of-Variances Examples. These decision processes by the creator would continue until he had succeeded in formulating a world that both resembled reality and stressed those aspects of experimental data collection technology in which he was most interested. In Table 3, the values of the proportions that the creator decided upon in the previous paragraphs have been summarized. In Column 1, the factors to be included in the model are listed. In Column 6, the products of all probabilities in Columns 2 through 5 affecting a particular factor are recorded, and the relative strengths of association for each factor with performance (i.e., the proportion of total performance variance accounted for by each factor) are shown. For example, the proportion of variance accounted for by intracon ition trials is the product of 0.15 x 0.50 x 0.20 equals 0.015.

At this point in the development of the simulation, the creator could examine the combined probability values (Column 6, Table 3) to see whether the values represent the type of world he is trying to simulate. If not, he can make the necessary adjustments with the only restriction that the sum of the proportion must total one.

Table 3. Summarizing the apportioning of variances
and
the coefficients of the standard regression equation

	(1) FACTORS	(2) Relevant/ Irrelevant	(3) System & Temporal Factors	(4) Subject & Temporal Factors	(5) Subject-Related Temporal Factors	(6) Proportion of Variance Accounted For	(7) Coefficients of Standard Regression Equation
RELEVANT	Skill		.30			.255	+ .505
	D		.24			.204	+ .452
	B		.12			f (.102)	+ .316 B - .254 B^2
	E		.09			.076	- .276
	D x E	.85	.08			.068	- .261
	Sk x D		.06			.051	+ .226
	A		.05			q (.042)	+ .154 Aa + .032 Ab - .093 Ac - .093 Ad
	C		.04			.034	+ .184
	E x C		.02			.017	- .130
IRRELEVANT	Subjects w/i Skill			.20		.030	+ .173
	Intercondition Trials				.40	.030	+ .173
	Intracondition Trials	.15		.50	.20	.015	+ .122
	Periods				.20	.015	- .122
	Relax				.20	.015	+ .122
	Drift			.30		.045	- .212
	Total Proportions	1.00	1.00	1.00	1.00	1.00	

f = functional factor

q = qualitative, assigned factor

Finding the Coefficients

Principle II. The magnitude of a coefficient (β) of a standard regression equation equals the square root of the proportion (P) of total performance variance ($\Sigma_y{}^2$) accounted for by each independent factor ($\Sigma_x{}^2$).

$$\beta_i = \sqrt{P_i} = \sqrt{\frac{\Sigma_{x_i}{}^2}{\Sigma_y{}^2}}$$

A standard regression equation is one in which the values of X and Y are given in terms of standard measures, or Z-scores, where Z equals the raw score minus the mean of that set of scores divided by the standard deviation of that set of scores. The square roots of the proportions provide the Beta coefficients of the equation which relate z_y to the z_{x_i} values of the i factors. As long as the different factors are independent of one another, Beta is equal to the linear correlation between that factor and performance.

While the square root of each proportion provides the magnitude of Beta, the creator must specify the sign of these coefficients. This is simple enough to do. For each factor, he must decide whether performance will increase or decrease as the value of a factor increases. If the change is in the same direction, the sign for that coefficient will be plus; if the change is in opposite directions, the sign for that coefficient will be minus.

Thus, starting with the proportions of Table 3, Column 6, the creator, by following the steps cited above, would arrive at the following standard regression equation:

$$z_y \text{ (performance)} = 0.505 \text{ Skill} + 0.442 \text{ D} + 0.316 \text{ B} \cdots \quad \cdots -0.212 \text{ Drift}$$

where every independent factor (e.g., skill, D, B, etc.) is actually a Z-score. The coefficients for all of the terms are shown in Column 7.

Principle III. The coefficient of the raw-score regression equation (b) equals the coefficient of the standard regression equation (β) multiplied by the ratio of the standard deviation of the performance measure (σy) over the standard deviation of the independent variable (σx_i).

$$b_i = \beta_i \frac{\sigma y}{\sigma x_i}$$

Up to this time the creator has worked primarily with bodiless numbers. Next he must relate the standard regression equation to the real world. To do this, he must specify for each term the unit and scale of measurement, the shape of the frequency distribution, and its mean and standard deviation. The standard deviations of all factors or terms are needed to transform Beta coefficients into raw-score coefficients.

The units by which a factor is measured are determined in part by reality. Slant range, for example, would be measured in feet, yards, miles, or meters. Luminance would be measured in foot lamberts. Skill would be measured by the score on some test of skill. The creator has the option of selecting the particular scale along which to express these types of measurements, for example, in meters directly or in logarithms, reciprocals, square roots or some other transformation of the meter values. Transformations would be exercised primarily to simplify the simulation mode, where a linear relationship between a factor and performance is ordinarily preferred.

In order to convert from a standard equation to a regression equation, the creator must be able to specify the means and standard deviations of the different factors. This can be simplified by assuming that all factors are distributed normally, that the standard deviation is one-sixth of the total range, and that the mean is at the center of the range. The creator can fairly easily select the range of values for each factor, for it is he who determines the limits of his simulated world. Although these limits should conform to some extent to reality, the creator can control to a great extent the range of values that the experimenter might decide to investigate by the information supplied in the scenario. If there are reasons to include non-normal distributions, and there can be, these could be handled in the same way that scale transformation are handled.

Once the raw score regression equation has been obtained, the coefficients should be examined directly, for they show the linear relationship between the predictor factor and performance. Thus, a coefficient of + 0.452 for factor D, which, let us imagine, represents the range to a target in meters, would mean that for every meter the observer is away from the target, there would be 0.452 increase in performance units, for example, the number of targets likely to be detected. The ultimate test of the acceptability of the equation can best be made by trying it out before using it with an experimenter. If it fails to reflect the type of world the creator intended to simulate, modifications at this stage should be relatively simple to make.

Multiterm Factors

Were all factors related to performance by a simple linear relationship, the basic approach to assigning coefficients to a large complex model would have been adequately described. Unfortunately, the task is not this simple. The process is complicated when the factors involved are related nonlinearly to performance (and when rescaling the measurements cannot linearize the relationship) and when a factor is qualitative. Some methods for handling these cases are discussed below.

Qualitative factors. A qualitative factor is one in which the levels, or categories within the factor, cannot be ordered or have any quantitative relationship. A qualitative factor might by "types of aircraft" in which the levels or categories are: helicopters, fixed-wing jet, and fixed-wing propeller. To handle these in the simulation model, the creator must decide how much each category affects performance, above or below the mean performance level for that factor.

These judgments are simplified by requiring the creator to decide how much difference in performance each category will make relative to another. For example, if Factor A were a qualitative variable containing four categories, a, b, c, and d, the creator might assign the following relative weights to each:

a, 5; b, 3; c, 1; d, 1.

Presumably, these would characterize the relative effects each condition had on performance in the real world. To fit these into the equation, the numbers must first be converted into values above and below the mean for that factor. In this case, the mean is 2.5 which when subtracted from each weight would leave the following values for each category:

a, 2.5; b, 0.5; c, -1.5; d, -1.5.

The next step is to scale these weights to fit the magnitude of the overall effect of factor A, which was shown in Table 3, Column 6 to be 0.042 parts of the total variance. Mathematically, this means that the sum of squares for factor A was 0.042 part of the total sum of squares. The sum of squares for the four weights (2.5, 0.5, -1.5, and -1.5) assigned to the four conditions equals 11. Therefore, to yield the lower sum of squares, the individual weights must be scaled down.

Principle IV. The square root of the adjustment value (I) which changes one sum of squares (Σx^2) to another sum of squares equals the value required to change all individual scores (X) to values that would yield the second sum of squares.

$$I \; \Sigma x_1^2 : \Sigma x_2^2 :: \sqrt{I}\, X_1 : X_2$$

To make this correction in our example, the weights 2.5, 0.5, -1.5, and -1.5 must be multiplied by $(0.042/11.)^{-2}$, or 0.0618. The results,

$$(0.154\ Z_{Aa} + 0.032\ Z_{Ab} - 0.093\ Z_{Ac} - 0.093\ Z_{Ad})$$

represents the coefficients of a set of dummy terms for factor A in the standard regression equation. Note that the four coefficients have a mean of zero and a sum of squares of 0.042. Insert this multitermed equation in lieu of the square root of the proportion for factor A in the standard regression equation (Table 3, Column 7).

Because the categories a, b, c, and d represent levels of a nominal factor having no quantitative values of their own, the value for each Z_{Ai} term will be either 0 or 1, and only one of the four terms can be designated 1 at any one time. Thus, when the condition is Category A_c, only Z_{Ac} takes on the value of 1, which subtracts 0.093 from the mean performance of the entire equation. The other Z_{A1} terms, being zero, contribute nothing to the performance estimate.

Nonlinear functions. For quantitative variables which relate nonlinearly to performance, the creator will wish to specify the function and then translate it into coefficients that fit into the relationship defined for the total equation. In deciding on the function, he must consider:

- What would be realistic.
- The degree of complexity he wishes to introduce into the simulation.

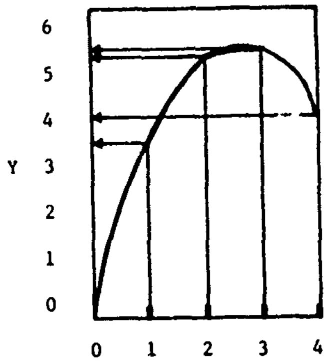
As higher-order terms are introduced into the equation, the experimenter would have to take more data to describe the simulated world accurately. When the data collection effort is limited, the experimenter's data collection plans will have to become more sophisticated.

The following steps will enable the creator to specify the coefficients of a particular function that will satisfy the requirements. These steps are illustrated in Figure 2.

1. Draw on a piece of graph paper the function believed to exist between performance (Y) and the boundaries of the real world for the particular independent factor (X). Emphasis here is on shape rather than numerical accuracy insofar as the values of the two axes are concerned.
2. Divide the X axis into four equal parts. This will allow a relationship up to and including a fourth-order one to be defined. It is assumed that for the purpose of experiment simulation, no function higher than fourth-order would be needed.
3. Label from 0 to 4 the five points at the ends and partitions between spaces on the abscissa (X axis). Label the ordinate (Y axis) with any set of values.
4. For each of the designated X points, find the equivalent Y value.
5. Use the "difference" method to obtain a least square fit for the data.

Step Numbers

(1, 2, 3)



(4)

X	X^2	Y
0	0	0
1	1	3.5
2	4	5.4
3	9	5.5
4	16	4.0

(5) Difference method (least squares):

Y	dY	d^2Y	d^3Y
0			
	+3.535		
3.535		-1.690	
	+1.845		0 ← Yo
5.380		-1.690	
	+0.155		0
5.535		-1.690	
	-1.535		
4.000			

$Y = (1 + d)^k Y_o$
when
$k = 1, 2, \ldots$

$$Y = 0 + 3.535\ X - 1.690\ (X^2 - X)/2 = 4.380\ X - 0.845\ X^2$$

(6)

	X	X^2	Y
Mean	2	6	3.69
Sigma	1.414	5.9	2.0
$\sigma x/\sigma y$	0.707	2.95	

$$\beta_x i = b\left[\frac{\sigma_x}{\sigma_y}\right]$$

$$Z_y = 4.380\ (0.707)\ Z_x - 0.845\ (2.95)\ Z_{x^2} = 3.10\ Z_x - 2.49\ Z_{x^2}$$

(7) Desired proportion for Factor $\beta = 0.102$

$$\therefore\ Z_{y'} = 0.316\ Z_{x'} - 0.254\ Z_{x^2}{'}$$

Figure 2. Example of steps to develop a nonlinear function for the simulation.

6. Convert the coefficients of this raw score regression equation (b) to those of a standard equation (β) by using a variation of Principle III, i.e.,

$$\beta_i = \frac{\sigma_{xi}}{\sigma_y} b_{x_i}$$

7. Take the proportion of each coefficient equal to the proportion of variance desired for that factor. This last step is valid because of the following principle:

<u>Principle V:</u>

$$R^2 = \beta_{12.3}\, r_{12} + \beta_{13.2}\, r_{13}$$

where R^2, the coefficient of determination, is 1.00 because there was no error, and

$$p(R^2) = p\,(\beta_{12.3}\, r_{12} + \beta_{13.2}\, r_{13})$$

where p is the proportion of the total variance to be contributed by this factor in the overall equation.

[Illustration: 0.102 = 0.316 (0.707) - 0.254 (0.479)]

8. Insert the multitermed equation representing the function of the single factor in lieu of the square root of the proportion for Factor B in the complete standard regression equation (Table 3, Column 7).

The one difficulty with this method is that it is possible to get fourth-order effects (since five data points of the function are being specified) when the creator may not wish to have more than a quadratic equation. This can be overcome by selecting only (n + 1) data points spaced equally across the extremes of the X axis to limit the order of the equation to n. This means that a second order polynomial would perfectly describe three values of X and Y, or a third-order would be perfectly described by four values of X and Y. The problem with using this method is that if the experimenter believed that his original higher-order drawing approximated the real world, representing it by a lower-order equation would destroy reality. The creator would have to decide on the merits of the tradeoff.

Eventually, if many equations were developed for experiment simulation, a supply of standard functions could be prepared from which the creator could draw to fit particular situations. Until these have been constructed, however, it is a simple matter to have the computer do the few computations required for applying these techniques.

MAN-COMPUTER INTERFACE CONSIDERATIONS

To make the technique of experiment simulation usable, useful, and used, it must be made easy to apply by both creator and experimenter. "Easy" in this case means letting the computer do most of the calculations and letting the man supply the guidance and basic inputs. Optimizing this relationship, which involves problems of computer programming, will not be considered in detail at this time, except to indicate some considerations that should be emphasized.

Creator

Ideally, the program would enable the creator and computer to engage in a dialogue in which the computer asks questions and the creator answers. The questions would follow the pattern illustrated earlier, taking bits of the problem small enough to permit purposeful answers to be given. Furthermore, while the creator would supply numbers, the computations, even simple ones, would be done within the computer.

An effective program from the standpoint of the creator would be one that would permit corrections or changes to be made easily, both in terms of individual numbers and blocks of numbers. This would allow a variety of "worlds" to be built and modified without having to go through the entire procedure each time. There should be provisions for storing worlds that they could be reused.

Once the basic techniques for simulating experiments have been adequately developed, the process could be made more sophisticated.

Instead of simulating a single world, the creator should be allowed to input multiple worlds by specifying those dimensions he wishes to change systematically. For example, he may describe a series of worlds in which the proportion of relevant to irrelevant variance is shifted systematically, or he may shift the proportion of variance originally accounted for by a subject-related time factor to an equipment-related time factor. If the multiple worlds are constructed by changing certain parameters not evident to the experimenter after the experimenter has collected his data, it would be possible to see how effective one particular experimental plan would have been across the whole range of conditions.

Experimenter

For the creator, the computer helps to construct a world; for the experimenter, the computer stores the world as a data base from which the experimenter must sample. The interface between computer and experimenter must make the experimenter's task of withdrawing data as easy as possible. As in an actual experiment, he should be able to indicate the coordinates of the space he wishes to sample and the computer should provide him with an estimate of performance at each coordinate by solving the equation for the values specified. Furthermore, as a matter of convenience, the experimenter should have the option of requesting data from one or a series of data points at one time, and the results should be fed back to him immediately.

The experimenter should be provided with a scenario by the creator that gives him the problem he is expected to solve, some background material, and the limits to be placed on his data collection strategy. The scenario should provide the experimenter with all of the supporting information about the problem that he would probably have were he to do the experiment in an actual laboratory. The limiting conditions -- the availability and type of subjects, the amount of time allowed for the study (which translates into data collection units), and the characteristics of the apparatus and the environment -- may be manipulated by the scenario along a continuum of greater to smaller restrictions which, when combined with the characteristics of the simulated world, produce a range of difficulties that the experimenter must be prepared to master to obtain the information at a desired quality level.

In addition, in order for the experimenter to be able to communicate with the computer, he must be told which measuring units are being used in the simulation and some information about the boundaries of the simulated world. Provisions must be made in the computer program so that when an improper instruction is given, it will be rejected and the experimenter notified.

From the point of view of the experimenter, the time continuum is considerably less complex than for the creator. The experimenter need think only in units of time, the length of his runs, his rest periods, and the time between sessions. He literally must schedule his simulated experiment as he would a real one, only he must convey this information to the computer in a more stylized manner, i.e., in units of time associated with every operating and nonoperating period carefully designated.

The same question-answer, dialogue style would be used between the computer and experimenter as with the creator. The computer would ask for information and provide the experimenter with the format he must use. Whatever program is prepared, it should -- if possible -- be sufficiently modular that it may be corrected or upgraded with as little rewriting as possible.

CONCLUSIONS

A method of simulating the characteristics of experiments in which human subjects are used has been proposed. Admittedly brief, presumably sufficient detail has been provided to show the reader that a model for this purpose is feasible.

After such an effort has been completed once, subsequent efforts will be easier. As part of a total program, the process could be simplified considerably by developing and using standardized aids. These aids would be applicable to those parts of the experiment model that are similar from experiment to experiment. These aids would consist of prespecified curves and relationships of varying magnitudes from which the creator could select the ones that best meet his needs.

The most immediate application for experiment simulation is to allow for experiences with new advanced methodologies. If the math model for experiment simulation is rough in regard to realism in the beginning, this still should not reduce its value for examining methodologies. As the model is improved and methods for establishing standards developed, the technique can be used to evaluate experimenters.

Experiment simulation as a tool can be expected to improve the quality of applied behaviorial research materially.

BIBLIOGRAPHY

While many statistical textbooks contain the basic information required to understand and apply the techniques proposed in this report, the following books represent one set which the author found useful for this prupose.

Chew, V. (ed.) Experimental designs in industry. New York: Wiley, 1958.

Davies, O.L. (ed.) Desi~ and analysis of industrial experiments (2nd edit.) New York: Hafner, 1967.

Draper, N. R. and Smith, H. Applied regression analysis. New York: Wiley, 1966.

Ezekial, M. and Fox, K. A. Methods of correlation and regression analysis. (3rd ed.) New York: Wiley, 1967.

Goulden, C. H. Methods of statistical analysis. (2nd ed.) New York: Wiley, 1952

Hays, W. L. Statistics. New York: Holt, Rinehart and Winston, 1963.

Kempthorne, O. The design and analysis of experiments. New York: Wiley, 1952.

Peters, C. C., and Van Voorhis, W. R. Statistical procedures and their mathematical bases. New York: McGraw-Hill, 1940.

Sokolnikoff, I. S., and Redheffer, R. M. Mathematics of physics and modern engineering. (2nd ed.) New York: McGraw-Hill, 1966, pp. 663-665.