

AD-A035 108

HUGHES AIRCRAFT CO CULVER CITY CALIF ENGINEERING EQU--ETC F/6 5/5
ECONOMICAL MULTIFACTOR DESIGNS FOR HUMAN FACTORS ENGINEERING EX--ETC(U)
JUN 73 C W SIMON F44620-72-C-0086
HAC-P73-326A NL

UNCLASSIFIED

1 OF 3
AD
A035108



ADA 035108

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) Hughes Aircraft Company Culver City, California		2a. REPORT SECURITY CLASSIFICATION Unclassified	
		2b. GROUP	
3. REPORT TITLE Economical multifactor designs for human factors engineering experiments.			
4. DESCRIPTIVE NOTES (Type of report and inclusive dates) 9) Final report			
5. AUTHOR(S) (First name, middle initial, last name) 10) Charles W. / Simon		12) 190 p.	
6. REPORT DATE 11) Jun 73	7. TOTAL NO. OF PAGES 172	7b. NO. OF REFS 52	
8. CONTRACT OR GRANT NO. 15) F44620-72-C-0086	9a. ORIGINATOR'S REPORT NUMBER(S) HAC-P73-326A		
9. PROJECT NO.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)		

D D C
 REPRINTED
 FEB 1 1977
 RELEASED

10. DISTRIBUTION STATEMENT
Revised edition: Distribution list included at end of report.

11. SUPPLEMENTARY NOTES Original report, F73-326, June 1973 Revised report, F73-326A, August 1976 <i>Superseded 767739 per the mc</i>	12. SPONSORING MILITARY ACTIVITY Air Force Office of Scientific Research
--	---

13. ABSTRACT
Experimental data collection plans are described that permit the study of from five to thirty experimental factors. These plans were selected from those employed in physical science research and were suitable for human factors engineering research. The maximum number of data collection points required for each basic design never exceeds three hundred and often is less. The method of employing these designs is a two-phase one. In the first phase, a large number of potentially critical factors are systematically screened quickly and economically in a way that identifies the more important ones. In the second phase, functions are obtained that relate the more important quantitative factors to operator performance. The need for this type of approach is justified on the basis of data obtained from an analysis of human factors engineering experiments published between 1958 and 1972 in the journal, Human Factors. Five principles that enable economical multifactor human factors experiments to be successfully conducted are stated and when assumptions are made concerning the application of these principles, empirical data is provided in support.

DISTRIBUTION STATEMENT A
 Approved for public release;
 Distribution Unlimited

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Experimental Design Screening Experiments Response Surface Designs Fractional Factorial Designs Human Factors Research Evaluation Economical Multifactor Experimental Designs						

Assignment for

NTIS Write Section
D.C. Draft Section

UNANNOUNCED

JUSTIFICATION

BY

DISTRIBUTION/AVAILABILITY CODES

Dist. AVAIL. and/or SPECIAL

A

ECONOMICAL MULTIFACTOR DESIGNS
FOR
HUMAN FACTORS ENGINEERING EXPERIMENTS

Charles W. Simon
Hughes Aircraft Company

Technical Report No. P73-326A

Prepared under contract with the
Air Force Office of Scientific
Research, (AFSC), United States
Air Force.

June 1973

Equipment Engineering Division
AEROSPACE GROUP
Hughes Aircraft Company • Culver City, California

"It is always pedantic to try to make forced use of statistical devices borrowed from another field when they only poorly fit. Statistical procedures are tools to be drawn upon only as needed for definite and well-understood purposes, and those tools are best which are not only most natural for the worker but also most readily understood by the reader to whom the findings of the research are to be addressed. The great historical contributions to statistics did not come about by the intention of the author to make a statistical formula; on the contrary, they were inventions devised for interpreting certain baffling research problems with which the investigator was confronted in some concrete setting. It is such natural emergence of procedures from the needs of the situation, rather than the imitative use of statistics, that should be the ideal toward which we work."

C. C. Peters and W. R. VanVoorhis,
"Statistical Procedures and
their Mathematical Bases",
1940

ABSTRACT

Economical multifactor designs already employed in other scientific vocations were selected for their applicability to human factors engineering problems. With the designs in this report, the effects of between five and thirty factors can be investigated with fewer than 300 observations. As presented here, these are not merely a conglomerate of experimental designs, but an approach which, if followed, should provide laboratory data from which more precise prediction of field performance is possible.

To make certain that the most important factors are being investigated, an initial experiment should contain the 15 or 30 factors -- or more if necessary -- that the investigator suspects have critical effects on performance. Using screening designs, a long list of factors can be ordered approximately in terms of their relative effects on performance and the factors which account for most of the performance variability can be identified. Generally, only a few factors -- probably less than 10 -- will be responsible for most of the variance. If these factors are quantitative, as most of them are in human factors engineering research, the function relating them to performance can be approximated by a polynomial using response surface designs.

Human factors engineering experiments, as results from analyses of 14 years of published research reveal, have generally studied too few factors and have taken many more observations than should have been required for the job. The need for new approaches is attested to by the failure of the conventional methods to provide experimental data that will account for most of the performance variations within the experiment or will predict field performance.

To take less data and obtain more information than is ordinarily done requires a shift in experimental philosophy and in experimental method. Economical multifactor designs are viable research tools in human factors engineering research because:

1. Replicating basic designs is often unnecessary.
2. Higher-order effects are usually negligible.
3. Analyzing the data as the experiment progresses often permits an early termination of the study.
4. Experimental logic may be substituted at times for actual data collection.
5. A conscientious effort on the part of the experimenter can usually make a small amount of data both reliable and accurate.

Justifications for these principles are discussed.

FOREWORD

In a 1956 edition of a physics book, the author discussed the theory of space flight. He concluded with the prognosis that although such an adventure was theoretically possible, man would never leave the earth's atmosphere until he had developed a more powerful fuel, one capable of creating the required thrust calculated by the author. A year later, however, the Russian's sent up their first satellite using the same fuel considered inadequate in the physics book. Instead of using only one rocket -- the basis for the calculations in the book -- they used two, one to boost the other, and by this different methodology were able to accomplish the "impossible".

Investigators concerned with the design of man-machine systems usually concede that the real world is more complex than their experimental simulation. Some are satisfied with this, believing that the simplicity facilitates the interpretation of the data and that eventually the results from many experiments can be combined into a multifactor, cohesive data base. Others are more skeptical of this simple approach, having observed how numerous efforts to quantitatively synthesize the results from human factors experiments have not been very successful. Instead, they would prefer that their experiments represent the real world more completely (i. e. include a much larger number of variables). This objective has been hindered in the past in part because when conventional experimental data-collection plans have been used, they quickly exceed practical limits imposed by the available time and money. The typical human factors experimenter, were he asked to obtain empirical data necessary to relate fifteen or more factors to operator performance, might acknowledge that it was theoretically possible, but would probably abstain, considering the request impractical and irresponsible, if he were to consider it seriously at all.

This report provides an alternative approach for the investigator working on applied human factors engineering problems who is not satisfied with how well his experimental results solve real world problems. It is a follow-on to the report, "Considerations for the design and interpretation of human factors engineering experiments", written last year. In that report, a number of misconceptions and inappropriate methods commonly employed

in human factors experiments were discussed. Suggestions on how to improve the quality of experimental data were made. This report provides the rationale and the tools for accomplishing this, particularly in regard to collecting multifactor data economically. While these tools are basically the same as those traditionally employed in human factors engineering research, the way that they are used is changed. This change makes the seemingly impossible possible.

Much of the philosophy for experimental design in this report was adapted from that described in numerous papers by G. E. P. Box, who has developed many ingenious experimental plans for physical science research. The designs described in this paper have been employed for many years in chemical and agricultural research; they are included here because they are suitable for research involving human behavior, particularly when the independent variables are physical system parameters.

This report emphasizes method, design, and interpretation as they apply to the practical conduct of formal experimentation. Enough statistics are supplied as tools -- generally in an oversimplified form -- to help an investigator use the designs immediately, but with little concern for either statistical theory or methods of analysis. These must be obtained from the original papers. This report should be read in order from beginning to end. The chapters are sequenced so that each provide a class of designs that match a corresponding phase of an entire experimental program. Knowledge of subsequent chapters depends on the understanding of the previous chapters.

The time available for this report did not permit certain important aspects of conducting economical multifactor experiments to be discussed. A noteworthy omission is a discussion of the conduct of economical multifactor "undesigned" experiments. These are the ones in which the independent variables are not under the experimenter's control, as is the case in many field studies. Nor does this report provide information on how to control the bias that the order in which different experimental conditions are presented to

the same subject may have on the effects of primary interest. Nor is the cost involved in building a simulator suitable for truly multifactor experimentation considered in this report, although it represents a practical problem equal in magnitude to that of collecting data economically. Some of the above omissions will be treated later in separate papers.

I am convinced that the approach proposed in this report -- if used properly -- can make a material improvement in the quality of human factors engineering research. It must be put to use before the more subtle details can be worked out and the conditions under which it will be optimally effective identified. I welcome hearing from readers who care to discuss the content of this report, who may need clarification on any point, or who disagree with its content.

Charles W. Simon
1973

ACKNOWLEDGMENTS

This paper is the final report of a research project conducted at the Hughes Aircraft Company, Culver City, California, under Contract No. F44620-72-C-0086 with the Air Force Office of Scientific Research, Air Force Systems Command, United States Air Force. Dr. Glen Finch, Program Manager, Life Sciences Directorate, was the technical monitor for AFOSR.

Numerous persons contributed to the preparation of this report. Marilyn A. Wilson collected, organized, and performed the basic analysis of the experiments published in the Human Factors journal. Flynard E. Roberts prepared the computer programs used for additional analyses of that basic data. Linda L. Young made many valuable suggestions toward the editing of the final report. Claire Rosen coordinated the publication of this report. Their help in these matters -- often above and beyond the call of duty -- is gratefully acknowledged. Special appreciation is also expressed for the support given this program by Robert L. Herbelin, J. William Weber, and John G. Bean.

TABLE OF CONTENTS

	Page
CHAPTER I: REQUIREMENTS FOR UPGRADING HUMAN FACTORS	
EXPERIMENTS	1
IMPORTANCE OF EMDS FOR HUMAN FACTORS	
ENGINEERING RESEARCH	1
An analysis of some Human Factors Experiments	1
Experimental Design Characteristics	3
Quality of Experimental Results	6
Establishing EMD requirements	9
More Factors in the Experiment	10
Avoid Excessive Data Collection	13
APPLICABILITY OF EMD TO HUMAN FACTORS RESEARCH	13
ECONOMICAL MULTIFACTOR DESIGNS -- A PREVIEW OF THINGS TO COME	14
 CHAPTER II: BASIC PRINCIPLES OF ECONOMICAL MULTIFACTOR DESIGNS	
SOURCES AND METHODS OF ECONOMY	16
Sources	16
Methods	17
EMD PRINCIPLE I. DON'T REPLICATE A BASIC EXPERIMENTAL DESIGN UNNECESSARILY	19
Why Replicate?	19
Replicating to Measure Performance More Precisely	20
Situations in which Replication for Precision is Generally Unwarranted	20
Situations in which Precision can be Obtained without Replication	21
Replicating to Obtain an Error Estimate for a Significance Test	23
Situations in which Error Estimates can be Obtained without Replications	23
Situations in which Error Estimates are Unwarranted ...	24

TABLE OF CONTENTS (continued)

	Page
When Subjects and Trials are Treated as Experimental	
Factors	25
Situations in which Multiple Subjects and Trials	
are not True Replications	25
Situations in which Measures of Individual	
Differences are Irrelevant	26
Special Situations	28
When Replications are Desirable	29
Investigating Transfer Effects	29
Developing the Power of the Test in Comparison	
Studies	30
Estimating Error Variances for Significance Tests	
from Partial Replications	30
Minimizing Experimental Artifacts	31
EMD PRINCIPLE II. IF HIGHER-ORDER EFFECTS ARE ASSUMED	
NEGLIGIBLE, THE DATA REQUIRED TO ISOLATE THESE	
EFFECTS NEED NOT BE COLLECTED UNTIL THERE IS	
EVIDENCE THAT THE ASSUMPTION IS INVALID	32
What is "Negligible"?	32
When Psychologists have used this Assumption	33
Two Types of Higher-Order Effects	36
Arguments and Evidence that Higher-Order Effects are	
Negligible	38
Mathematical and Intuitive Arguments	38
Higher-Order Interactions	39
Three-Factor Interactions	41
Two-Factor Interactions	43
Higher-Order Terms of the Polynomial	43
Methods of Minimizing Higher-Order Effects	46
EMD PRINCIPLE III. COLLECT AND EVALUATE DATA IN A	
SEQUENCE OF PROGRESSIVE ITERATIONS	47

TABLE OF CONTENTS (continued)

	Page
EMD PRINCIPLE IV. SUBSTITUTE EXPERIMENTER'S KNOWLEDGE AND ANALYTIC SKILLS FOR DATA COLLECTION	55
Selecting the proper Measurement Scale	55
Identifying which Confounded Effects are Important	56
EMD PRINCIPLE V. MINIMIZE BIAS EFFECTS ON EACH INDIVIDUAL MEASUREMENT	57
Sources that Bias Experimental Measurements	59
Design	60
Equipment	60
Subjects	60
Procedures	61
Analysis	61
 CHAPTER III: ECONOMICAL DESIGNS FOR QUALITATIVE FACTORS (FRACTIONAL FACTORIALS)	 63
SOME UNDERLYING CONCEPTS AND NOTATIONS	63
Developing a Sign Matrix for Two-Level Factorial Designs	64
Estimating the Effects (Mean Differences)	67
Calculating Sums of Squares and Mean Squares	68
Orthogonality	69
CONSTRUCTING FRACTIONAL FACTORIALS FOR FACTORS AT TWO LEVELS	69
Blocking and Confounding	69
Fractioning and Aliasing	73
The Resolution of a Fractional Factorial	78
The Other Block	79
CREATING SMALLER 2^{k-p} FRACTIONAL FACTORIALS	81
SOME 2^{k-p} FRACTIONAL FACTORIAL DESIGNS	84
FRACTIONAL FACTORIALS FOR FACTORS WITH MORE THAN TWO LEVELS	85
Symmetrical Fractional Factorial Designs with Three or Four Levels	85
3^{k-p} Designs	85
4^{k-p} Designs	87

TABLE OF CONTENTS (continued)

	Page
Non-Symmetrical Fractional Factorials	87
USING FRACTIONAL FACTORIAL DESIGNS WITH QUANTITATIVE AND QUALITATIVE FACTORS	87
 CHAPTER IV: ECONOMICAL DESIGNS FOR SCREENING A LARGE	
NUMBER OF FACTORS	89
GENERAL APPROACH	90
STAGE ONE OF THE SCREENING PROCESS: SATURATED	
DESIGNS	91
Constructing Saturated Designs when the Number of Conditions Equals a Power of Two	91
Variations of the Basic Saturated Designs	98
When There Are Fewer Than N-1 Factors	98
Interaction Effects	99
Estimating Error	99
When the Basic Design is Blocked	99
When Unplanned-for Information is "Discovered"	100
"Discovering" Error Estimates	101
"Discovering" a Factorial Design	101
Saturated Designs when the Number of Conditions is a Multiple of Four	102
Confounding	103
Precision	103
Selecting a P-B Design	104
STAGE TWO OF THE SCREENING PROCESS: AUGMENTATION	
DESIGNS	105
A. D. 1. To isolate a single main effect and all its two-factor interactions from the remaining effects, unbiased by any other main effects or two-factor interactions	105
Isolating Aliased Effects	107
A. D. 2. To isolate all main affects from all two-factor interactions, leaving the two-factor interactions still aliased among themselves	109

TABLE OF CONTENTS (continued)

	Page
Investigator Logic	111
A. D. 3. To help the investigator analytically identify critical main and two-factor interaction effects	112
Using the (I) column to Measure Block Effects	115
A. D. 4. To add a new factor to the study	115
A. D. 5. To obtain unbiased estimates of all main and interaction effects among any three factors if the remaining factors are of no importance	115
STAGE THREE OF THE SCREENING PROCESS: ISOLATION	
DESIGNS	116
I. D. 1. To separate a single pair of two-factor interactions with one extra condition	117
Obtaining other conditions	119
I. D. 2. To separate four members of a single string of two-factor interactions with three extra experimental conditions	120
I. D. 3. To separate members of a string of three-factor interactions	123
I. D. 4. To isolate the second-order coefficients of a response surface	124
CHAPTER V: ECONOMICAL DESIGNS FOR QUANTITATIVE FACTORS	126
CHARACTERISTICS OF RESPONSE SURFACE DESIGNS	127
Economy	128
Applications	128
Types of Designs	130
CENTRAL-COMPOSITE SECOND-ORDER DESIGNS	131
Construction	131
Features of Central-Composite Designs	132
Design Parameters	136
Partial Replication of Central-Composite Designs	139

TABLE OF CONTENTS (continued)

	Page
SECOND-ORDER RESPONSE SURFACE DESIGNS WITH THREE-LEVELS PER FACTOR	141
Incomplete Block Designs	141
Construction	143
THIRD-ORDER RESPONSE SURFACE DESIGNS	146
NON-SYMMETRICAL SECOND-ORDER RESPONSE SURFACE DESIGNS	146
RESPONSE SURFACE DESIGNS FOR "MESSY" EXPERIMENTAL SPACES	148
Construction	149
Practical Considerations	151
 CHAPTER VI: CONCLUSIONS	 152
 REFERENCES	 154
 APPENDIX I: AN ANALYSIS OF THE METHODOLOGY AND EFFECTIVENESS OF SOME REPRESENTATIVE HUMAN FACTORS EXPERIMENTS	 159
APPENDIX II: FRACTIONAL FACTORIAL DESIGNS AT TWO LEVELS ...	164
APPENDIX III: PLACKETT AND BURMAN DESIGNS	169
APPENDIX IV: THREE-LEVEL RESPONSE-SURFACE DESIGNS	171

FIGURES

<u>Figure</u>	<u>Title</u>	<u>Page</u>
I-1	Distribution of proportions of variance accounted for by experimental factors in 239 experiments	5
II-1	Latin square experimental design for three factors	33
II-2	Exploration strategy in the development of a response surface	51
V-1	Spatial arrangement of the coordinates of a central-composite design for three factors	133
V-2	Spherical characteristic of the space covered by a central-composite design	133
V-3	Information contours of experimental designs	137
V-4	"Messy" experimental designs	150

TABLES

<u>Table</u>	<u>Title</u>	<u>Page</u>
I-1	Percentage of 239 Experiments Studying Different Numbers of Factors	2
I-2	Analyses of 236 Experiments Published in <u>Human Factors</u>	4
I-3	Median Proportions of Variance Accounted for as a Function of the Number of Experimental Factors being Studied in the Experiment	7
I-4	Some Relative Measures of Experimental Results	8
II-1	Two Methods of Partitioning Sources of Variance	37
II-2	Analyses of the Proportion of Variance Explained by Equipment Interaction Effects	40
II-3	Analyses of Three-Factor Interaction Effects Accounting for More Than .05 of the Total Variance	42
II-4	Proportion of Variances of Main Effects Accounted for as a Function of the Order of the Polynomial	45
III-1	Experimental Conditions, Sign Matrix, and Scores	67

TABLES (continued)

<u>Table</u>	<u>Title</u>	<u>Page</u>
III-2	Blocking Alternatives for a 2^2 Factorial	71
III-3	Blocked 2^2 Factorial	72
III-4	Sign Matrix for a 2^4 Factorial Design	74
III-5	Sign Matrix for a 2^{4-1} Fractional Factorial Design (Principle block) (I = ABCD)	75
III-6	Sign Matrix for a 2^{4-1} Fractional Factorial (I = -ABCD)	80
III-7	Sign Matrix for a Quarter Replicate of a 2^4 Factorial (I = ABCD = ABD = C)	82
III-8	Fractional Factorials with Three Levels Found in Conner and Zelen	86
IV-1	Sign Matrix for a 2^3 Design - Design I	92
IV-2	Sign Matrix for a 2^{4-1} Design - Design II	93
IV-3	Sign Matrix for a 2^{5-2} Design - Design III	95
IV-4	Basic Design (2^{7-4})	96
IV-5	Basic Design and A. D. 1	106
IV-6	Basic Design and A. D. 2	110
IV-7	Two Sets of Performance Data for Seven Effects	111
IV-8	Performance Data for the Basic and A. D. 3 Designs	113
IV-9	Other Experimental Conditions that Might be Used to Isolate the Effects of (AF+BE)	120
V-1	Parameters for Designing Orthogonally Blocked, Second-Order Central-Composite Designs	138
V-2	Plans For Partially Replicating Central-Composite Designs . . .	140
V-3	Second Order Response Surface Designs with Three Levels per Factor	145

CHAPTER I.
REQUIREMENTS FOR UPGRADING HUMAN FACTORS EXPERIMENTS

"Economical multifactor designs" (EMDs), as the term is used here, refer to data collection plans that enable a large number of factors to be investigated in a single experiment while keeping the total number of observations to a reasonable size. For this paper, "large" refers to at least five factors and at times as many as 15 or 30. "Reasonable size" refers to a basic experimental design that contains no more than 300 observation points and usually a great many less. The designs described here were selected because they are suitable for most human factors engineering experiments concerned with the problem of equipment design. Whether or not they are suitable for other problems in which human performance is involved will not be considered in this report.

IMPORTANCE OF EMDS FOR HUMAN FACTORS ENGINEERING RESEARCH

The importance of these designs to human factors engineering research cannot be fully appreciated unless one has examined critically the information produced from traditional methods of studying these problems. While the practical value of the results of formal human factors experiments has been questioned in general (1), little effort has been made to evaluate the productivity and effectiveness of the methods employed in this experimentation. Simon (44) compared the methods used in human factors engineering research with the types of questions the research was intended to answer. He concluded that the methods most commonly employed were often misapplied or inadequate for obtaining the desired information.*

An Analysis of Some Human Factors Experiments

To provide a quantitative evaluation of the quality of data produced in human factors engineering experiments and the methods employed to obtain this data, an

* Dunnette (26) has made similar criticisms about psychological research in general. Campbell and Stanley (15) have questioned some of the techniques employed in educational psychology.

analysis was made of the experiments published in the journal, Human Factors, between 1958 and 1972.* Their design characteristics and effectiveness in accounting for the variability in operator performance in the experiments were determined. The results of the analysis showed clearly that many of these formal experiments were little more than extravagant exercises, examining factors that explained little of the results of the particular experiment and less when related to performance in the real world. A reanalysis of 239 analysis-of-variance tables reported in this journal during the fourteen year period showed that in approximately 60 percent of the experiments, the experimental factors that were purpose-ly varied in order to measure their effect on performance accounted for less than half of the total performance variance within the experiment. Since the median number of factors studied in these experiments was two, the chances that this data would predict performance in a complex operational situation with any degree of accuracy are slim.

The 239 experiments were grouped for many analyses according to the number of factors studied in the experiment. The percent of experiments having from none to seven equipment (and system and environment) factors are shown in Table [I-1]. In this same table, the percent is shown when the experiments were regrouped

Table [I-1] Percentage of 239 Experiments Studying Different Numbers of Factors

<u>Number of Factors</u>	<u>Equipment Factors</u>	<u>Equipment, Subject, and Temporal Factors</u>
0	0.8	-
1	29.7	20.5
2	38.9	43.5
3	23.0	25.5
4	5.4	7.5
5	1.7	2.5
6	0	0
7	0.4	0.4

} 7.5 } 10.4

*Reference will be made throughout this report to this analysis of the experiments in the journal, Human Factors. The conditions and scope of this analysis are described briefly in Appendix I.

according to the number of equipment, subject, and temporal factors in an experiment. Sources of variance due to subjects and trials were defined as "experimental factors" only when they were examined in the experiment for meaningful effects as opposed to being treated merely as forms of replication. Only ten of the 239 experiments examined specific subject characteristics and only 36 looked for systematic effects of performance over trials. In the remaining experiments, while subjects or trials would usually be removed as a source of variance in an analysis-of-variance table, these effects were never examined or interpreted further. Over half of the experiments in which subjects or trials were treated as an interpretable factor occurred in an experiment studying a single equipment factor. This accounted for the largest shift in the two distributions seen in Table [I-1] when there was a drop in the number of one-factor experiments and an increase in the number of two-factor experiments when subjects and trials were considered to be factors.

The median proportion of performance variance in an experiment accounted for by the experimental factors when only equipment factors were considered versus when equipment, subject, and temporal factors were considered differed by only three percent overall. Histograms for the two analyses are shown in Figure [I-1]. It is apparent that whether or not subject and temporal factors were included in the analyses of these human factors engineering experiments (with the emphasis on equipment parameters), it made only a marginal difference in how much of the performance variability was explained by the particular set of experimental factors.

Experimental Design Characteristics. Characteristics of the 236 experiments with from one to five equipment factors in an experiment are shown in Table [I-2]. The contents of this table should be self-explanatory; the conclusions drawn from this table are summarized and interpreted as follows:

- 1) The median number of equipment factors studied in all of the experiments were two. Less than eight percent of the experiments studied four or more equipment factors in an experiment (Column 2). Even

Table [I-2]. Analyses of 236 Experiments Published in Human Factors^a

(1) Number of Equipment Factors in a Single Experiment	(2) Number of Experiments in this Category	(3) Median Number of Levels per Experiment	(4) Median and Maximum Number of Replications Based on		(5) Total Number of Observations in Experiments
			Subjects (Max)	Trials (Max)	(Median) (Range)
1	71	4*	6 (30)	1 (70)	72 18-1120 (7680) ^c
2	93	3	10 (64)	1 (10)	180 15-1944 (2016)
3	55	3	6 (36)	1 (12)	192 24-5184 (9600)
4	13	3	6 (18)	1 (2)	768 48-3888
5	4	2	6 (10)	1.5 (5)	1200 192-3000

^a There were also three other experiments not included here: 2 zero-factor and 1 seven-factor study.

^b "Equipment" factors are those associated with the equipment, system, and environment.

^c Numbers in parentheses refer to the upper limit of the total number of observations when the effects of trials represented a meaningful factor such as learning, fatigue, etc. The upper limit of the total number of observations not in parentheses refers to those experiments when trials were not a factor but treated simply as a form of replication.

* This value is distorted upward since two factor experiments with only two levels were excluded from the count.

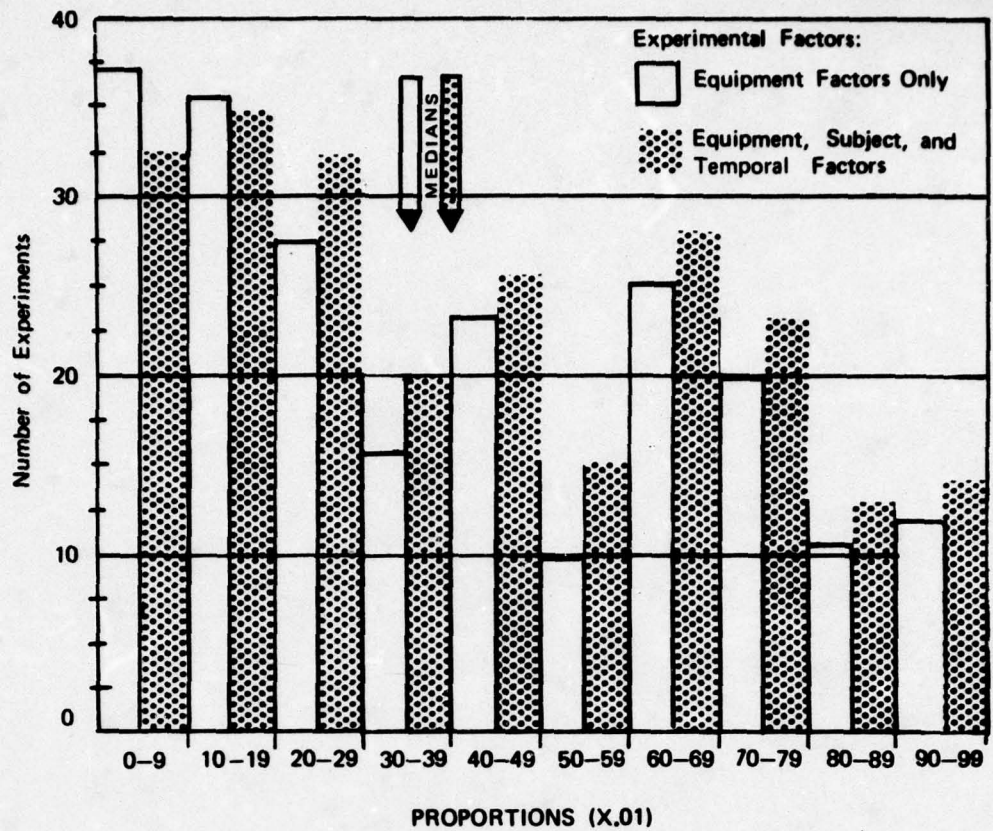


Figure [I-1]. Distribution of proportions of variance accounted for by experimental factors in 239 experiments.

when subject and temporal factors were counted, in only ten percent of the experiments were four or more factors studied. [Table I-1]

- 2) The median number of levels per equipment factor for all experiments was three (Column 3). For the five-factor studies, the median number of levels per factor reduced to only two, which meant that non-linear relationships between main effects and performance could not be estimated.

Considering the normal complexity of the real world, these experiments of few factors would appear to be examining only a small part of it. One can reflect on the problems that the blind men from Istanbul had in defining the elephant in order to understand why the results from human factors engineering experiments have had only marginal success in quantitatively predicting performance in the field.

- 3) Some experiments used more than a thousand observations to study the effects of from one to five factors. Most of this effort — several times more than that needed to collect information on the basic design — was expended making repeated measurements on the same experimental conditions. (Column 5)
- 4) Subjects rather than trials were the primary method of replicating; while the median values per groups appear reasonable, the maximum number of replications for all groups were quite high. As the number of factors in the experiments increased, the size of the experimental effort seemed to deter the amount of replication to some extent. (Column 4)

It seems that a great deal of effort which might have been expended collecting new information over a larger experimental space was used only to make repeated measurements on the same conditions.

Quality of Experimental Results. In Table [I-3], the median proportion of the total performance variance accounted for by the experimental factors (and their interactions) is shown for experiments grouped according to the number of factors in the experiment. Two sets of data are shown — when the number of factors are based only on equipment variables and when the number is based on equipment, subject, and temporal variables. The conclusions drawn from Table [I-3] are:

- 5) On the average, the more equipment factors in an experiment, the greater the proportion of the performance variability in the experiment that will be accounted for by those factors. This is essentially the same when equipment, subject, and temporal factors are all considered. (Column 2)
- 6) When the equipment plus temporal (or subject) factor studies are removed from the one-factor group and added to the two-factor group, the proportion accounted for by the remainder of the one-factor group increases and for the newly formed two-factor group decreases (Column 2). This shift suggests that the presence of subject or temporal factors (as opposed to subject and trial replications) tends to depress the proportion of the total variance accounted for by the equipment factors alone, but that the effects of a subject or temporal factor is on the average not as great as that of an equipment factor in these human factors experiments.

Table [I-3]. Median Proportions of Variance Accounted for as a Function of the Number of Experimental Factors being Studied in the Experiment.

(1) Number of factors in a single experiment	(2) Proportion of variance accounted for by the experimental* factors where "factors" include		(3) Proportion of experiments in which equipment factors account for at least half the total variance
	Equipment only (N)**	Equipment, Subject, Temporal (N) **	
1	0.16 (71)	0.29 (49)	0.30
2	0.31 (93)	0.27 (104)	0.34
3	0.45 (55)	0.42 (62)	0.42
4	0.61 (13)	0.58 (17)	0.62
5	0.65 (4)	0.70 (6)	1.00

* This also includes the proportions of variance from all interactions among factors. Subject and Temporal sources of variance are considered to be an experimental factor when they were seriously interpreted and not when these categories were merely forms of replications.

** N refers to the number of experiments in each group. For the "equipment only" column, two other experiments with no equipment factors and one with seven were not included. For the other column, only one seven factor experiment is not included.

COPY AVAILABLE TO DDC DOES NOT PERMIT FULLY LEGIBLE PRODUCTION

- 7) The probability that at least half of the performance variability will be accounted for by equipment factors increases as the number of equipment factors increases. "Probability" is based on the percent of experiments in the particular groups that equaled or exceeded the amount shown. (Column 3)

These results show two important trends: One, with less than four independent variables in an experiment, the factors that were purposefully varied accounted for less of the variability in performance on the average than other conditions which were supposedly replicated or inconsequential in the experiment. Two, even when five factors were studied, there is still an uncomfortably large proportion of the variance not accounted for. Outside the laboratory, other factors not included in the experiment can also affect performance: this would serve to increase the proportion of variance not accounted for by the experimental data were it applied to the operational situation.

The values related to Equipment factors only in Tables [I-2] and [I-3] can be combined to provide a measure of the quality of the experimental results and the redundancy and effectiveness of the experimental designs. These are referred to as "indexes" in Table [I-4] and are interpreted and defined as follows:

- 1) The quality of the data improves as the number of equipment factors increases. "Quality" is defined here as the ratio of the proportion of

Table [I-4]. Some Relative Measures of Experimental Results.

<u>Number of Equipment Factors in Single Experiment</u>	<u>(1) Quality Index</u>	<u>(2) Redundancy Index</u>	<u>(3) Effectiveness Index</u>
1	0.19	24.	2.2
2	0.45	30.	1.7
3	0.82	19.	2.3
4	1.56	51.	0.8
5	1.86	57.	0.5

variance associated with the equipment factors to the proportion associated with irrelevant sources of variance. (Column 1)

- 2) The redundancy in data collection more than doubles when more than three factors are studied. "Redundancy" is defined here as the number of observations in an experiment over the minimum number required to obtain the coefficients of a polynomial describing a second degree response surface. This calculation was based on the assumption that there were three levels per factor. (Column 2)
- 3) The effectiveness of the experimental design decreases markedly when more than three factors are studied. "Effectiveness" is defined here as the ratio of the proportion of variance accounted for by the equipment factors over the total number of observations required to obtain it. (Column 3)

These three indexes are based on median values of the proportions and total observations for each group of experiments. Had the same measures been made for individual experiments, the range of values would be quite large; the use of average here only helps to identify a trend. The measures of quality, redundancy, and effectiveness should not be taken too seriously as absolute indexes; however, as a crude indication of relative merit they can be useful in the comparison of human factors engineering experiments.

Establishing EMD Requirements

To the extent that the experiments published over the past fourteen years in the journal, Human Factors, are representative of human factors engineering research in general, the numbers in Tables [I-3] and [I-4] would seem to indicate that a great deal of time and effort has gone into obtaining information of limited practical value. Particularly noteworthy is the increase in quality of the data at about the same place - number of factors - that experimental efficiency drops off.

The results of the analysis clearly indicate the characteristics that future human factors experimental designs must have if the goal of predicting field performance from laboratory data is ever to be achieved. Specifically, experiments must include a great many more factors than are currently included in a single

experiment and the number of observations must be held to a minimum in order to make multifactor experiments economically feasible.

More Factors in the Experiment. The critical question is: Approximately how many equipment factors should be included in an experiment if one hopes to predict real world performance from laboratory data? A first approximation to answer this question can be obtained from the data based on the equipment factors only in Table [I-3], Column 2.

If in the one-factor experiments, one equipment factor accounts for 0.16 of the total variance (on the average) how many factors would be needed to account for "all" of the variance? The arithmetic answer would be determined by dividing one by 0.16 to obtain 6.3 factors. Based on the same calculation using the proportions of variance accounted for by two, three, four, and five factors, we would need 6.5, 6.6, 6.5, and 7.7 factors respectively to account for "all" of the variance in an experiment.

Considering the degree of independence among these groups of data, the answers — centering around seven factors — are remarkably consistent. Perhaps the extra factor which seemed necessary in the case of the five-factor experiments was needed to compensate for an inability to estimate the contribution of the quadratic component of the main effects since for this group, there was a median of only two levels per factor. On the other hand, this variation may have been merely a quirk resulting from the small amount of available data. In any case, these numbers suggest that had approximately seven factors and three levels per factor been used in these experiments most of the performance would have been accounted for, on the average. The phrase, "on the average," reminds us that the calculations were based on median values, and indicates that the hypothesized number of factors would be sufficient only fifty percent of the time. To account for most of the variance ninety percent of the time, that number of factors would have to be increased. A suitable correction would suggest that to account for most of the

performance variance in an experiment most of the time, approximately ten factors, each at three levels, should be included in a single study.*

This recommendation, based as it is on empirically derived data, ignores the theoretical principle that if an experiment were performed properly, then most of the performance variance within the experiment should be accounted for by the experimental variables whether the number of factors were ten or one. Supposedly in the experiment, nothing else has been changed to cause the performance to vary. However, in our sample, among experiments studying the same number of factors, the proportion of variance accounted for ranged from 0.10 to 0.90. It is obvious that other conditions must also be operating that have not been taken into account yet.

Two rather obvious situations can exist when the experimental factors fail to account for most of the performance variance within the experiment: either the relative effects of "random" (chance) variability — although small — are overwhelming the effects of experimental factors that have only small, albeit reliable, effects, or there are major sources of uncontrolled or irrelevant variances running rampant in the experiment that distort the data so as to cause even important effects to appear relatively small.

Therefore, the recommendation based on the empirical data that approximately ten factors are needed to account for most of the experimental variance in an experiment must be accompanied by the implicit assumptions that 1) there will be a well-conducted experiment, and 2) the most important factors affecting performance on the particular task are included in the experiment. How to conduct a "clean" experiment is an art that will not be treated to any great extent in this

* It might be argued that to consider including ten factors in an experiment intended to compare the effects of three hand-controls, for example, on tracking performance would be meaningless. This might be so if an experimenter could know that all other conditions of the experiment were identical to those that would be experienced under operational conditions and that there would never be a wish to generalize beyond these specific conditions. Since it is highly unlikely that this would ever be so, many more factors than "type of control" could be added to the experiment relevant to the characteristics of the task, the environment, and the personnel.

paper, although how this is done is neither intuitively obvious nor dealt with to any great extent in institutions from which human factors researchers receive their training. How to include the most important factors in an experiment will be a major technique discussed later on in this report, and it has a bearing on the topic at hand, i. e., number of factors in an experiment. One method of finishing a program with some assurance that the most critical factors have been studied is by starting the program with more factors than will be eventually needed and allow the empirical data to identify the most important. While it might be difficult for an investigator to name the exact ten most important factors affecting performance on a given task, even if he is quite familiar with that task, he can probably select fifteen or twenty factors within which he believes the ten most important ones will eventually be found. Thus, it would be safer to begin a research program with approximately fifteen or so factors in the experiments. This not only increases the chances of having the more important factors included in the experiment but also increases the chances of including the factors needed to accurately predict performance in the operational situation.

If the initial fifteen or so factors are carefully selected by a knowledgeable experimenter and the experiment is performed with reasonable care, the set of factors that will ultimately describe and predict performance in the real world might be smaller than the hypothetical ten. Budne's (14) comments are relevant on this point. He wrote:

"Experience in a large number of screening experiments in industrial situations has consistently shown that there are only a few critical variables and a large number of unimportant variables associated with each specific problem. There is limited practical value in attributing 'statistical significance' to any number of the 'unimportant' variables while one or more of the 'critical few' variables escape consideration. In the real world it becomes useful to assume that total variation and total effect can be broken into all of their components and that each component may be attributed to a particular variable or cause. In the light of experience, it is both practical and useful to make the assumption that a very few of these many variables or causes contribute a major portion of the total variation or effect." (p. 140)

Avoid Excessive Data Collection. The data in Table [I-2], Column 5, reveals that in some cases thousands of observations were made to study the effects of one or two factors. If truly multifactor experiments are to be conducted, there must be some way to reduce the magnitude of the effort. An examination of the 239 experiments that were analyzed revealed a considerable amount of redundancy in the data collection. For example, in 44 percent, the same subject was tested more than once on the same experimental conditions. In 93 percent of the experiments more than one subject was tested on the same experimental conditions. These replications add to the magnitude of the data collection process and tend to reduce the number of factors an experimenter is willing to study. The question therefore is: how many observations are a reasonable number to consider when selecting economical multifactor designs. Until more experience has been acquired, the following logic was employed to answer this. If it is only necessary to determine the second-order relationship between fifteen factors and operator performance, then a minimum of 136 observations would be required to estimate the 135 coefficients of a polynomial approximating that relationship. Since many experiments will be considering fewer factors, somewhat arbitrarily, it would seem that any experimental design initially requiring more than 300 observations would be wasteful. Many should require less.

APPLICABILITY OF EMD TO HUMAN FACTORS RESEARCH

In the discussions that follow, methods of economically collecting multifactor data will be described. No designs are included in this report that handle less than five factors, and some techniques are described that will permit an examination of from fifteen to thirty factors.

An effort was made to include only those techniques and designs that were particularly suitable for experiments to arrive at design parameters for equipment used by the human operator. While EMDs may not be applicable to all human factors engineering problems, they will be to a great many. There is so much room for improvement in our research methods that even when certain designs are not directly applicable, the principles behind these designs can still be useful.

Human factors engineering experimentation lends itself markedly to the application of economical multifactor designs. First of all, because human factors engineering research involves relating physical equipment and environmental factors to operator performance, a majority of the factors to be studied can be measured on quantitative and continuous scales. Second, human factors engineering research must ultimately find solutions that are applicable to real world problems. Any success in this regard will not be arrived at using procedures of the past, performing multitudes of small independently planned experiments with the aim of ultimately consolidating their results. The failure of the approach has emphasized the need to look at a bigger picture in a single experiment, even if some precision is sacrificed initially. Third, the majority of these experiments are conducted under circumstances in which time and money are limited. These designs will enable the most information to be obtained at the least cost. Finally, the designs, when used properly, encourage an investigator to seek solutions to problems rather than to merely do experiments. In general, they provide a method of arriving at the best answer with a minimum of elegance and in some instances provide a means of evaluating their own effectiveness.

ECONOMICAL MULTIFACTOR DESIGNS – A PREVIEW OF THINGS TO COME

The chapters that follow are arranged to be read consecutively. Unless a reader is thoroughly familiar with each chapter in turn, he will not understand a subsequent chapter. Although the designs described in each chapter have been developed and even used in other disciplines as independent entities, the approach proposed here for human factors research considers selected designs from each chapter as steps in a sequence of designs which the experimenter would employ as he progresses through the program. Thus the discussion in Chapter III on fractional factorials – suitable for qualitative and quantitative factors – is included here primarily to familiarize the reader with the technique to be employed in Chapter IV. Chapter IV employs fractional factorials for purposes of quickly screening large numbers of variables to discover which are the most important ones. Chapter V describes the techniques whereby the effects of these more important variables can be measured and related quantitatively to operator performance. Presumably if out of a great many candidate factors the most important ones are chosen, the final equation of seven to ten factors should permit more precise predictions of field performance than has been possible up to now. Before the

reader can be comfortable with the methods proposed in any of these three chapters, however, he must understand, accept, and use the principles on which economical multifactor designs are based. These are discussed next in Chapter II.

CHAPTER II.
BASIC PRINCIPLES OF ECONOMICAL MULTIFACTOR DESIGNS

The economical multifactor designs (EMDs) discussed in this paper do not differ markedly from designs traditionally employed in human factors research. They are all branches of the same general family, stemming from the theory of multiple regression and its specialized form of the analysis of variance. Emphasis is given to special features of the factorial designs, particularly the 2^n series including the principles of single degrees of freedom, confounding, and fractional replicates. The experimenter who is familiar with these topics of mathematics and statistics will have little trouble understanding the underlying structure of economical designs.

The difficulties that may arise in the use of EMDs will come from the shift in philosophy, the difference in the experimental approach, and the degree of control and involvement the investigator must have as compared to the way much human factors engineering research has been traditionally handled.

SOURCES AND METHODS OF ECONOMY

EMDs can best be understood through an understanding of the principles on which they are based. Since the only practical means of including more factors within the same experimental plan is to reduce the amount of data that must be collected to an absolute minimum, the principles for EMDs revolve about the sources from which and the methods by which the economy can be obtained. The following sections outline the contents of Chapter II.

Sources

Which experimental conditions should be eliminated from replicated factorial designs in order to reduce the size of an experiment without a material loss of information? In most experiments, some economy can be achieved by:

- 1) Minimizing repeated measurements of the same experimental conditions. Given the choice between including more factors in an experiment or replicating a smaller experimental space, the former will generally provide the most unbiased estimates of the effects of interest. In fact, there are many instances in which there is little to be gained from replication.
- 2) Not measuring experimental conditions in order to estimate effects that are likely to be non-existent or relatively unimportant. To plan to isolate the effect of a fifth-order interaction, for example, is to combine unwarranted optimism with an inordinate waste of time and effort.

These principles seem so intuitively obvious. How wasteful it is to study the same thing over and over again or to study unimportant aspects of the problem when the same effort might have been used to examine a larger portion of the critical space. Yet from its inception, human factors research has tended to emphasize the former approach.

Methods

"It ain't what you do but the way you do it," an old song informs us. Economy can be achieved in the collection of experimental data by changing the way data has been typically collected in the past. Three more principles of EMDs are:

- 3) Use a more flexible approach to data collection. Begin with an experimental plan that can be modified as the experiment progresses, changing direction if necessary or terminating the data collection when it is apparent no more information can be obtained.
- 4) Substitute investigators' knowledge and analytical skills for actual data collection. Experimenter objectivity is a desirable goal, but not to the point that known data is discarded and time and effort is wasted trying to rediscover it.

- 5) Take extra precautions to minimize irrelevant sources of performance variance that creep in to bias the data collection phase. This can reduce the need to take extra data that serves primarily as a cover-up for poor procedures.

Each of these principles is based on continued investigator involvement from the pre-experimental planning stage, through the data-collection, to the analysis and interpretation of the data. How different this is from the usual approach where the principle investigator plans an experiment and turns it over to his lesser-trained assistant to run and analyze. In using EMDs, the investigator will consider more carefully why he is collecting his data and what he really wants to get out of it. He will be more interested in finding answers than in doing "experiments." He will rely more on himself and less on the experimental design to guide his data collection and analysis. He will become willing to accept flexible plans and probabilistic guesses as important tools of the research process. He will find himself more involved in the total experimental process than ever before.

The rationale on which these principles are based will be discussed in the sections that follow. The reader who finds some of these principles difficult to accept because of his previous experiences should be reminded that those experiences have come almost totally from experiments in which fewer than five equipment factors have been investigated in a single study. It will be shown how some uneconomical methods employed in human factors research in the past were needed to cover the limitations which result from studying only two or three factors in an experiment, and how the reasons for many of these methods disappear when five or more factors are to be studied. To really feel comfortable employing economical multifactor designs, the investigator must get used to "thinking big." The very conditions that must be present to safely study many factors economically are the ones that exist (for the careful experimenter) when many factors are included in a single design. The principles themselves tend to support one another.

The five basic principles underlying EMDs are described in detail below, including the rationale, supporting data, and the conditions under which they are and are not applicable.

EMD PRINCIPLE I. DON'T REPLICATE A BASIC EXPERIMENTAL DESIGN UNNECESSARILY.

If the number of replications of an experimental design is held to a minimum, the savings that result from not making repeated measurements on the same conditions can be used to make measurements on different experimental conditions. Some replication may be necessary if for no other reason than it may give the experimenter more subjective confidence in his data and when no economical limits are placed on the data collection and analysis, it need not be discouraged. In the past, however, replicating has been too often the tail that wagged the dog; in order to be able to replicate, fewer factors had to be studied in the experiment. This proves to be the wrong choice in most cases, for a precise study of a small portion of the experimental space has little predictive power in an operational situation in which performance is affected by a great many factors. The results of a great many small experiments have never been satisfactorily consolidated (44) When there are limits on time and money and it is necessary to choose between making repeated measurements on the same experimental conditions or taking new data over an expanded experimental space, the latter alternative will generally result in more and better information, particularly when a large number of factors is involved.

Why Replicate?

Most experimenters tend to replicate automatically whether they need to or not. In this section, the reasons that investigators replicate will be discussed and a distinction will be made between those circumstances when it is and is not necessary. By attending to this distinction, data collection can be reduced and the savings redirected toward studying more factors at nominal costs.

Replication of a basic design can be achieved by testing more than one subject on the same experimental conditions or testing the same subject more than once on the same experimental conditions. There is an implicit assumption that the subjects have been drawn at random from a homogeneous population, and that on retesting the same subject, performance on subsequent trials can be measured independently

of performance on preceding trials. There are several reasons why repeated measurements are made in these experiments:

- 1) To increase the precision with which mean performance can be estimated.
- 2) To obtain an estimate of error variance to test for statistical significance.
- 3) To measure the effects of individual differences or changes in performance over time by treating the dimension being replicated as an experimental factor.

As the discussion which follows will show, some of these goals may be achieved without replication and some are irrelevant to the original purpose of conducting the experiment. If economy in experimentation is important, an experimenter must be able to distinguish among the different circumstances.

Replicating to Measure Performance More Precisely

Investigators have often replicated their basic experimental plan to obtain a more precise measure of mean performance. In some human factors experiments, the use of multiple replications has been justified on the basis of improved precision when in fact:

- the use of replication for precision is unwarranted
- an alternative to replicating for precision exists.

Situations in which Replication for Precision is Generally Unwarranted. There are certain tasks in which wide, unexplained fluctuations in performance occur from trial to trial which totally obscure the "true" measures of mean performance. When this occurs, many investigators will make repeated measurements on the same conditions, using either extra trials or many subjects, and use only an average measure in subsequent analyses and discussions to smooth the effects of the unwanted fluctuations. Quite often this situation arises when experiments are being conducted in the field, where the need for economy is often greater than in the laboratory. To solve this "problem" by replicating many times is both inappropriate and unwarranted.

What the experimenter has done by smoothing the data in this way is to obtain a clearer picture of the mean effects of the conditions that were included in his experiment. What he has failed to do is to understand the reason for the large fluctuations that did occur and will probably reoccur under operational conditions. Thus averaging can give a precise estimate of a trivial effect and an inflated sense of the importance of the experimental conditions in the experiment, but may cause the investigator to overlook far more important sources of variance that must be understood if generalization from the laboratory data to the operational situation is to be of practical value. Replicating an experimental design for this purpose is not only uneconomical but becomes the means by which the investigator avoids his research responsibilities. It allows him to be lazy in the planning and the conduct of the experiment, hiding rather than identifying, controlling, and isolating effects. If the investigator, instead of replicating, had used the same data collection effort to identify the causes of the wide fluctuations, the quality of his information would have been markedly improved. In those cases where it is not possible to control factors that are suspected of accounting for the fluctuations, both economy and understanding can be achieved if suspected parameters are measured and their effects evaluated using a regression model. Replication should not be used to hide sources of variances which instead should be and could be identified and measured.

Situations in which Precision can be Obtained without Replication. There will always remain some fluctuations in performance that cannot be readily identified; replication can be used in these cases to obtain a more precise estimate of mean performance. The standard error of the mean — the measure of its precision — is inversely related to the square root of the number of observations used to obtain the mean. Therefore the more observations that are involved, the narrower the range within which the true mean can be expected to lie.

It should be noted that to replicate for precision may have grown out of an experience involving only experiments in which a few factors were studied. Ninety-two percent of the experiments published in Human Factors between 1958 and 1972 included only three or fewer equipment factors. With designs of that small size, some replication may have been necessary to obtain a sufficient number of degrees of freedom and a comfortable degree of precision. However when experiments

8 with five or more factors are studied, as considered in this report, replication for precision will be unwarranted.

Making repeated measurements of the same experimental condition is not the only way to increase the number of observations used to estimate a mean. If the number of factors in an experiment is large enough, there is sufficient hidden replication (30, p. 103) within the basic design to provide a reasonable precision without replication. For example, if a factorial experiment were conducted on eight factors, each at two levels, the total number of observations (or experimental conditions) in the experiment would be 256. Therefore each mean of a main effect would be calculated from one-half of the observations, or 128 in this case. Each effect, whether main or any order of interaction in these 2^n designs, being merely mean differences between two halves of the experiment, will all be based on 128 measures.

8 Hidden replications can have certain advantages over true replications. Hidden replications of the different levels of a single factor, since they are actually taken in combination with many different levels of the other factors, provide a more representative measure of mean performance within the experimental space than would be the case if the replications were repeated measures of the same condition. This is desired when the purpose of the experiment is to obtain a general multifactor function relating the equipment variables to performance across a great many conditions. If, of course, the interest rests in a specific task, the precision of the equation— an average across the total multifactor space — may not describe as precisely some points in that limited space representating the particular task under consideration. This then, in a particular mission and task situation, might be the one, situation-specific case where replication for precision might be justified.

8 The general rule however is not to replicate when hidden replications provide enough observations to make reasonably precise estimates of the main, two-factor interaction, and possibly three-factor interaction effects.

Replicating to Obtain an Error Estimate for a Significance Test

Replication can provide an estimate of experimental error, which is used in tests of statistical significance as the standard against which observed differences among experimental conditions are tested. Chew (17, p. 5) defines experimental error as "the failure of two identically treated plots, or experimental units, to give identical yields, or responses." This error is assumed to be distributed normally. In practice, where economy is a viable criterion in the design of an experiment,

- there are acceptable alternatives to replicating to obtain an estimate of the error variance
- there are circumstances when no error variance is needed.

Situations in which Error Estimates can be Obtained without Replications.

Behavioral scientists have tested the statistical significance of effects without replicating to obtain an estimate of the error variance. When factorial designs have been too big to replicate, the highest-order interaction has been used in lieu of an error term. In doing this, the experimenter implicitly made the assumption that the effect of this interaction on performance is negligible. By definition, for an effect to be negligible, the variability in performance would be no greater than could be expected by chance. In practice, this assumption is almost never checked. Circumstances under which this assumption is likely or not likely to be valid will be discussed later.

Error estimates can also be obtained without replicating the basic design by using what might be called — employing a chess-term analogy — a "discovered" replication. If a large number of factors is studied in a single experiment, it is highly probable that the effects of some will be negligible. If data is collected originally with the unreplicated design, that data collected on factors with negligible effects (and we need not know which these will be ahead of time) can be used to obtain an estimate of error. This is so, of course, because by definition if an effect is negligible, there are no meaningful differences among levels and they therefore represent a replication.

If the number of factors in the study is large enough, there will be little trouble in deciding whether or not effects are negligible. Mere consideration of the

practical value of observed differences should suffice. In addition, if the proportion of variance attributable to the particular factor is small, its relative importance within the experiment is established.

Situations in which Error Estimates are Unwarranted. Traditionally tests of statistical significance have been used to identify critical factors in human factors engineering research. An investigator would select a group of factors that he believed to be important, would collect some performance data on the conditions representing combinations of these factors, and would apply a significance test to decide whether his "hypothesis" (that these were important factors) was correct. Since a majority of human factors experiments have been relatively small, replicating the basic design has been the only way the error estimate used in the significance test could be obtained.

But tests of statistical significance only measure the reliability of an effect, not how much of an effect there is. The results of such tests can be influenced by any number of decisions on the part of an investigator. There have been serious questions raised as to its suitability for factor identification (3)(4)(31)(37)(38)(41)(44). With many replications, the biggest danger is the identification of statistically significant (i. e., reliable) factors that have only trivial effects.*

The identification of critical factors, therefore should be based on whether the effect on performance is important rather than merely reliable. Ordinarily in a multifactor experiment, if the former is true, the latter will follow.

*In the analysis of the experiments in Human Factors, out of 494 main effects that were examined, 194 accounted for 0.04 or less of the total variance in their particular experiment. Of these, the investigators concluded that 11.7 percent of them were "statistically significant" effects. In one three-factor study involving 3024 observations (49), one of the factors was statistically significant at $p < 0.01$. However, all of the factors including the significant one and their interactions combined accounted for less than one-half of one percent of the total variance, while the error variance (experimenter's categorization) accounted for 0.92 of the total variance. The remaining 0.07 was almost equally distributed between subject and subject-by-experimental factor interactions. Much discussion was generated by the identification of the "significant" but trivial affect.

For factor identification, or screening experiments as they are referred to in this report, tests of significance need not be made. It is assumed instead that with reasonable care and effort, if enough factors are included in the basic experiment, any random variance associated with replication would be inconsequential relative to the other effects. As Budne (14, p. 140) stated: "the existence of high residual variation in an experiment merely indicates that the most important variables were not included in the experimental design. [Note: Other possible explanations for high residual variations are considered in the discussion of EMD Principle V.] Statistical significance alone is a function of sample size and this residual variation, and is thus not a good measure of what is or is not important in the real world. The absolute magnitude of residual variation must be considered. When estimating error variances for significance tests is of secondary interest, then replicating designs for this purpose is unwarranted. As Davies (23, P. 20) points out, obtaining error estimates for screening experiments is unjustified since these are not the types of experiments in which irrevocable decisions must be made.

When Subjects and Trials are Treated as Experimental Factors

Psychologists have always had some concern for individual differences. To study individual differences, more than one subject must be tested under the same set of conditions. Psychologists have also had a long involvement in problems of learning, forgetting, and other phenomena of changes in performance over a period of time. To study these temporal factors, the same individual must make repeated measurements on the same task. In some experiments, however, knowledge of the effects of trials or subjects on performance is of primary interest; in others, this knowledge is treated merely as an irrelevant fact.

Situations in which Multiple Subjects and Trials are not True Replications.

Some human factors experiments include among their experimental factors - in addition to the equipment factors - subject and temporal factors. For example, an experiment may be designed to answer questions such as: Do pilots perform differently with a new display-control configuration than non-pilots? Should equipment be designed differently to compensate for age or sex differences among operators? How much difference is there in operator performance using different devices after a great deal of practice? What effect does the design of a piece of equipment have on the ability to perform a monotonous task? Answers to these questions can only

8 be obtained from experiments in which several subjects perform the same tasks or single subjects repeat the same task several times.

In these cases, however, multiple measures on the same condition of an equipment factor are not really replications. In these quasi-replications, the special characteristics of the subjects or the positions of the trials in a series of trials are intended to represent (in conjunction with the equipment characteristic) a unique experimental condition. When this is so, the subject characteristics and temporal changes over trials are treated as meaningful factors in the same sense that the equipment factors are. In these cases, the interactions among equipment, subject, and temporal factors are also meaningful. This is not the case when subjects and trials are intended to be only replications.

8 When multiple subjects and trials are employed to estimate a meaningful effect, therefore, the discussion on minimizing replications does not apply. However, the analysis of the 239 human factors experiments revealed that in only 5 percent of those in which multiple subjects were used was there a concern about the effect some subject characteristic such as their sex, experience, or handedness had on the ability to perform using the equipment. Only 35 percent of the experiments in which the same subjects were tested repeatedly on the same experimental conditions was there any interest in the effects of such temporal characteristics as learning or the effects of performing a task over extended periods of time.

In view of the limited number of times when subject and temporal changes were actually meaningful factors, it is important to distinguish when that is the case and when it is not. The distinction affects critical decisions for the design of the experiment.

0 Situations in which Measures of Individual Differences are Irrelevant. Relatively few human factors engineering experiments actually include subject characteristics as factors of their design. This probably has its historical origin in the fact that human factors experiments are conducted to find ways of optimizing performance by improving the equipment rather than by selecting or training the operator. Whether or not the separation of these various effects on performance is wise cannot be discussed here. Obviously where a disordinal interaction can be

expected between subject and equipment characteristics, both sources of variance must be included in the same experiment.

In the majority of the cases, however, when repeated measures are made on the same experimental conditions using different subjects, the variance associated with subjects will sometimes be isolated and presented in an analysis-of-variance table, and thereafter ignored. Neither the discussion of the results nor the conclusions drawn concerning the data will mention anything about this subject variance. Under these circumstances, this data must be considered irrelevant.

It is sometimes argued, however, that knowledge of subject variance is a measure of individual differences and will be important when the results of the experiment is to be applied to the real world. In practice, unfortunately, this is seldom the case; the variance attributable to subjects is seldom a useful piece of information. To be useful,

- the subjects employed in the experiment must be truly representative of the population to which the data is to be extracted
- representativeness must be based on multiple characteristics
- the values of the characteristics for the sample and population must be known.

These conditions seldom exist for the typical human factors experiment. When multiple subjects are run as replications, the chances that they are representative of the population is slight for the following reasons:

- 1) The average number of subjects in human factors experiments run around ten.
- 2) In many cases, no systematic sampling of subjects is or can be made and the ones that are used are those that are available.
- 3) When subjects have been selected, it is often on the basis of a single label (for example, Air Force pilot). It is seldom that additional considerations (such as amount of flying time, types of aircraft flown, etc.) that account for wide variations in performance are taken into account.

- 4) Quantitative descriptions of population and samples are seldom available making it impossible to adequately identify to what sub-portion of the population experimental results refer.

In addition, the artificiality of the experimental situation also influences the performance of individual subjects. A part of the variability between subjects' performance reflects the basic instability of a mean score for subjects that are often still learning how to handle the experimental situation as the study progresses and do so at different rates. In summary, it is highly presumptive to believe that the variance associated with the performance of a small group of subjects used to replicate an experimental design has much permanency or practical validity insofar as the experimental results may be applied to the real world. Under these conditions, replicating experiments for this purpose is not justified.

Special Situations. There are some common situations in which it is easy to confuse quasi-replication with the irrelevant replication and vice-versa. One experimental design not readily recognized as an example of quasi-replication is that in which several subjects are tested on the same series of experimental conditions but the order of presentation is varied to compensate for any biased effects that sequential position might have on the measure of the particular condition. Although the basic experimental design is being repeated by each subject, it is not true replication since the subjects are confounded with an additional variable - order of presentation.

In a second case, when a quantitative variable is being studied, collecting too many levels of that variable is not only uneconomical but also an unwarranted form of replication. As a general principle, to take more than $N + 1$ levels of a quantitative variable that can be related to performance by an equation of degree N is wasteful. A simple example of this would be when the relationship between an independent and dependent variable is known to be a straight line within the limits of interest (i. e., can be approximated by a first degree equation). Under these circumstances to include more than two levels of the independent variable in the experimental design is to obtain redundant information and would be equivalent to replicating. In practice, of course, there is always some uncertainty of the order of the relationship to expect and there are certain economical experimental designs

which require a large number of levels per variable. However, where economy is a consideration, more than the necessary number levels of a quantitative variable must be considered to be another form of unwarranted replication.

When Replications are Desirable

Although replicating has frequently been used unnecessarily, it would be foolish to claim that replications are never necessary. There are circumstances when multiple measurements of the same experimental condition can provide additional information toward an understanding of the problem under investigation.

Investigating Transfer Effects. In human factors engineering studies, the experimenter must be careful to minimize the effects of the order in which a number of experimental conditions are presented to the same subject. This has already been discussed and the point was made that running more than one subject for the purpose of reducing this order of presentation effect is not considered to be ordinary replication. Instead, the requirement that each subject to be tested on experimental conditions in a different order adds a new factor, another source of variance, to the experiment and should be considered in the analysis.

In the experiments analyzed in the journal, Human Factors, many investigators verbalized concern about this order-of-presentation effect; however, when systematic counterbalancing was employed to compensate for such effects, relatively few statistically removed the variance associated with the effort. This could inflate the error variance for the tests of significance of the differences among the equipment variables. What's more, in no experiment was an effort made to isolate the transfer effects from the direct effects of the mean performance for each experimental condition. A transfer effect is the residual effect that carries over from one experimental condition to affect the results of the experimental condition tested next. Residual effects may carry over to more than one subsequent condition. In some experiments, transfer effects are considered to be a nuisance; in others — particularly in training problems — interest in the residual effects may be as great or greater than in the direct effect.

To obtain the best information about transfer effects, replication of experimental conditions is required. In some designs, each subject repeats the first one of a series of experimental conditions in a row of a latin square design used to counterbalance order. The methods of handling transfer effects is too large a topic to be treated here, but may be found in other references. (2)(16)(39)

Developing the Power of the Test in Comparison Studies. Simon (44) distinguishes between human factors engineering experiments which try to identify the factors having the most important effects on performance and those which compare the relative effectiveness of several experimental conditions. In the former case, replications can increase the degrees of freedom to a point where tests of significance begin to spotlight trivial effects. In the latter case, particularly when the interest in the comparison study is to establish that there is no difference in operator performance on two or more equipment conditions (for example, in order to justify the use of the least expensive piece of equipment), then some way is needed to increase the power of the F-test. To insure that the F-test will not lead to an erroneous conclusion that a difference of a certain magnitude does not exist when in fact it does, the degrees of freedom of the error estimate should be large. Replication may be the only way to achieve the required power.

Estimating Error Variances for Significance Tests from Partial Replications. Although the importance and value of tests of significance have been more or less downgraded in this report, they cannot be totally discarded. Tests of significance can be helpful when the experimental data was sloppily collected so that there is a relatively large amount of variability outside the effects of interest. If the numbers that make up means are highly variable, then differences between means must be viewed with some degree of skepticism. Tests of significance can be applied to check the overenthusiasm of an experimenter who might otherwise accept as reliable differences between means that were actually attributable to irrelevant sources of variance.

To keep the amount of data collection within reasonable bounds for this purpose, economy can be maintained by replicating only a portion of the total experimental

design. There have been a number of plans for partial replication for specific purposes. These include:

- 1) Making repeated measures at the center of the experimental space. This technique has been used to obtain an error term in central-composite (response surface) designs with which to test how well first or second order regression models fit the experimental data. (9)
- 2) Repeating a fraction of the complete design. It is sometimes feared that variability may change away from the center of an experimental design. Dykstra (27) suggests a technique of partially replicating over the entire experimental area to obtain more precise estimates of the coefficients of a second order polynomial, more degrees of freedom to estimate experimental error, and a more powerful test of the adequacy of the second order model. Box (6) also discusses this alternative.

Minimizing Experimental Artifacts. Pragmatically, there will be times when running a subject repeatedly on the same experimental condition may be justified in the name of economy. If the number of observations in a basic design is large and there is a need to be concerned with the order of presentation, then it may be desirable to run several trials sequentially on the same condition to offset transfer effects without the need for counterbalancing. In many transfer situations, with reasonable planning the residual effect will usually subside after a trial or two. In practice, this assumption should be tested. However, if it is so, and if care is taken to avoid learning or fatigue effects, several trials may be run (for example, four). Then the first two, which may be biased from the residuals effects from the previous experimental condition, can be eliminated from any calculation. This situation differs from the one in which replications are used to escape from the responsibility of explaining real but unidentified sources of variance. On the contrary, in the present example, replication is used to overcome an artificial and identified variance which came from an experimental procedure. In the end, as an alternative to counterbalancing (which can require a great many additional observations), the above technique could prove more economical. Its validity could be tested by running two subjects on all conditions in opposite orders to see how well the means correlate. If order effects have been eliminated, and if there is no reason to believe there are disordinal equipment-by-subject interactions, then an almost perfect positive rank-order correlation should be obtained.

A second situation in which the experimental method might introduce a bias is the habituation and expectation found in psychophysical measures of sensory threshold in "methods of limits" studies. By running each subject at least two trials on the same condition, first approaching the threshold from above and the second time approaching it from below, these artificial biases should be averaged out.

In general, if an experimenter must replicate — if it helps him feel more comfortable about his results — he can always do it after he has run through a study once and examined his data. That would be the time to decide just how much replication is required, rather than before the experiment has been planned. Even if three or four replications are ultimately made, this is still more economical than the number that have typically been employed.

EMD PRINCIPLE II. IF HIGHER-ORDER EFFECTS ARE ASSUMED NEGLIGIBLE, THE DATA REQUIRED TO ISOLATE THESE EFFECTS NEED NOT BE COLLECTED UNTIL THERE IS EVIDENCE THAT THE ASSUMPTION IS INVALID.

Some EMDs reduce the amount of data to be collected by not isolating certain higher-order effects which are assumed to be negligible. Higher-order effects generally refer to three-factor (or higher) interactions, and any third degree (or higher) component of a function relating operator performance to an experimental variable. If the assumption of negligible higher-order effects is valid, then this reduction in the data collection effort is obtained with essentially no loss of critical information. Preferably, the experimenter begins with the assumption of negligible higher-order effects, which he continues to check as the experiment progresses.

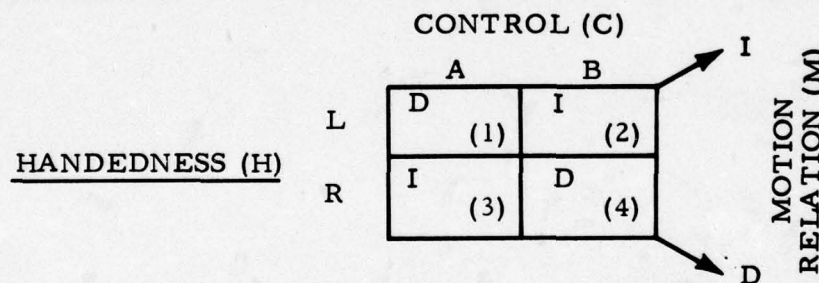
What is "Negligible"?

At least two criteria are useful for deciding whether an effect is negligible or not. The first is whether the absolute size of the effect in the experimenter's judgment would have any practical effect on performance. The second is how the size of the effect compares with the size of the other effects in the same study. Proportion of performance variance (η^2) is a useful measure (31), and within any experiment, effects accounting for one or two percent of the total variance ordinarily can be considered small.

Whether an effect is negligible or not has nothing to do directly with whether or not it is statistically significant. It may be statistically significant yet negligible or not significant and not negligible. *

When Psychologists have used this Assumption

There are a number of situations in which psychologists have traditionally employed the assumption of negligible interactions. Perhaps the most common has been their use of a Latin (or Graeco-Latin) square design. A Latin Square design provides an economical way of estimating the effects of three factors on performance provided that no interactions exist among the factors. For example, to study the effects of two types of controls (A or B), handedness of the operator (Left or Right), and direction of movement relationship between display and control (Direct or Indirect), a single replicate of a complete factorial design would require eight experimental conditions. On the assumption that no interactions exist among the factors, a Latin square design with only four experimental conditions can be used to estimate the effects of the three factors. The experimental plan would look like Figure [II-1].



- Cell (1) = Control A, Left handed operator, and Direct motion relation
- Cell (2) = Control B, Left handed operator, and Indirect motion relation
- Cell (3) = Control A, Right handed operator, and Indirect motion relation
- Cell (4) = Control B, Right handed operator, and Direct motion relation

Figure [II-1]. Latin square experimental design for three factors.

*In the analysis of the experiments in the journal, Human Factors, from 1958 to 1972, 16 percent of the main and interaction equipment effects that accounted for less than 5 percent of the variance in the experiment were considered statistically significant ($p \leq 0.05$) by the investigator.

If twenty operators were randomly distributed equally among the four cells, the total experimental design would consist of 19 degrees of freedom partitioned as follows:

<u>Source</u>	<u>d. f.</u>
Control (C)	1
Handedness (H)	1
Motion Relation (M)	1
Residual	16

After performance data were collected, the effect of Controls would be determined by the differences between means of the A and B columns, the effect of Handedness would be determined by the differences between means of the L and R rows, and the effect of the Direction of Motion Relation would be determined by the differences between means of the I and D diagonals. These three estimates are orthogonal (i. e., independent of one another). No further reduction of the residual variance is possible.

When a Latin square design such as this is analyzed, the traditional analysis of variance table would appear as if no estimate of the CxH, CxM, HxM, or CxHxM interactions were ever made. In fact, the estimates attributed to the three main effects (C, H, and M) are actually the results of main effects confounded with interactions. Specifically, the indicated effect of C, H, and M are not due to these main effects alone, but are actually the combined effects:

$$C + (H \times M)$$

$$M + (C \times H)$$

$$H + (C \times M)$$

The reader can check this himself by noting that exactly the same combinations of cell values would have been used, for example, to calculate main effect M and the interaction C x H, namely (Cell 1 + Cell 4 - Cell 3 - Cell 2). The resulting value is divided by two. In this design, the effects of these two sources of variance cannot be estimated independently. Therefore, in the case of M + (C x H), we will obtain an unbiased estimate of the effect of M only if the effect of C x H is negligible.

Behavioral scientists using Latin square designs make the implicit assumption that interactions are negligible. However, there is a high probability that the

assumption is invalid where two-factor interactions are concerned and that many of the main effects estimated from Latin square designs are distorted. Two-factor interactions are not to be considered "higher-order" effects.

A second situation in which psychologists have assumed that an interaction effect is negligible is when the highest-order interaction is used as a substitute error term in the analysis of an unreplicated multivariate experiment. No test of the assumption is ever made. However, the chances are good that the assumption will be valid for all practical purposes, provided four or more factors are being studied.

Except in the two situations cited above, psychologists have generally failed to collect their data to take advantage of the possibility that higher-order effects are negligible, even when they believe it. One experimenter (28, p. 120) tested twelve subjects under every combination of a 2^7 (seven factors at two levels each) factorial design, a total of 1536 measurements. In his analysis, he calculated separately all of the main, two-factor, three-factor, and four-factor interactions but pooled the effects of five-factor, six-factor, and seven-factor interactions, which he had determined accounted for less than 2.5 percent of the treatment variance. An examination of his data revealed that had he also included the four-factor interactions with the others, the pooled portion would still have accounted for only four percent of the treatment variance and less than 1 percent of the total variance within the experiment. Since the pooled portion encompassed the effects of 64 sources of variance, for any practical purpose, these effects are negligible. Had the investigator anticipated or been willing to assume that four-factor interactions or higher would have negligible effects, and had not taken the measurements required to isolate them from the lower-order effects, he could have reduced his data collection by half with practically no loss of information. If he then wished to maintain his original level of effort, the 768 observations that were not needed to complete the original design could have been more effectively employed to study the effects of additional variables or to determine if any of the original variables related non-linearly to performance.

Two Types of Higher-Order Effects

EMDs can be conveniently categorized by the type of higher-order effects which they assume negligible. In a general sense, certain designs make no provisions for collecting the data required to isolate selected higher-order interaction effects; others partition interaction effects and are designed to ignore third-degree or higher terms of an equation relating the independent variables to performance. Those that assume higher-order interactions are negligible are the fractional factorials and screening designs discussed in Chapters III and IV. These are based on an analysis of variance model and are most suited for the study of qualitative factors or any two-level factors. Those that assume higher-degree terms are negligible are based on a regression model and are most suited for quantitative variables. This breakdown is employed in the designs discussed in Chapter V. While there are exceptions to this method of partitioning, the distinction is useful for understanding EMDs based on the principle of negligible higher-order effects.

To illustrate the distinction between the two types let us imagine an experiment to study the effects of three variables, A, B, and C, at 3, 3, and 2 levels respectively. Eighteen observations will be needed to complete the basic factorial design. The total 17 degrees of freedom could be partitioned in two ways as shown in Table [II-1], depending on whether the ANOVA or the regression model is to be used.

In Table [II-1], Column I, the 17 degrees of freedom are partitioned into main and interaction effects. In Column II, however, the partitioning is even greater, each effect being associated with a single degree of freedom representing a term in a polynomial. The sources of variance adjacent to one another are wholes or parts of the same. Thus, the four degrees of freedom of the $A \times B$ Interaction effect can be partitioned into sources of one degree of freedom each: $A \times B$, a linear-by-linear portion of the interaction; $A^2 \times B$, a quadratic-by-linear portion; $A \times B^2$, a linear-by-quadratic portion; and $A^2 \times B^2$, a quadratic-by-quadratic portion. If most of the variance associated with the interaction can be accounted for by the linear-by-linear portion, $A \times B$, then it may not be necessary to isolate the remaining three higher-order effects. In later chapters, how the assumption of negligible higher-order effects can be employed to reduce the amount of data taking will be explained in more detail.

Table[II-1]. Two Methods of Partitioning Sources of Variance

I		II	
Sources of Variance and Degrees of Freedom in Analysis of Variance Model		Terms of the Polynomial and Degree of Terms in Regression Model (1 d.f. each)	
<u>Main Effects</u>	Factor \bar{A} 2 d.f.	A_2	1st
		A	2nd
	Factor \bar{B} 2 d.f.	B_2	1st
		B	2nd
	Factor \bar{C} 1 d.f.	C	1st
<u>Two Factor Interactions</u>	Int. $\overline{A \times B}$ 4 d.f.	$A_2 \times B$	2nd
		$A^2 \times B_2$	3rd
		$A_2 \times B_2$	3rd
		$A^2 \times B^2$	4th
	Int. $\overline{A \times C}$ 2 d.f.	$A_2 \times C$	2nd
		$A^2 \times C$	3rd
Int. $\overline{B \times C}$ 2 d.f.	$B_2 \times C$	2nd	
	$B^2 \times C$		
<u>Three-Factor Interactions</u>	Int. $\overline{A \times B \times C}$ 4 d.f.	$A_2 \times B \times C$	3rd
		$A^2 \times B_2 \times C$	4th
		$A_2 \times B_2 \times C$	4th
		$A^2 \times B^2 \times C$	5th

*The line over the letters indicates that an effect is being referred to; without the line it is a term of a polynomial with one degree of freedom.

Arguments and Evidence that Higher-Order Effects are Negligible

While it is easy to assume that higher-order effects are negligible, whether they are in fact is another matter. Even when the assumption is made tentatively, to be checked as the experiment progresses, it would lose much of its practical value in the use of economical designs if it were valid only infrequently. There is evidence, however, that this is not the case.

Mathematical and Intuitive Arguments. The assumption of negligible higher-order effects is made continually in the statistical literature on experimental design. Plackett and Burman (40, p. 306) wrote that "if main effects are regarded as being first order of small quantities and if the function relating them to performance may be differentiated (i. e., is a smooth relationship), then when p variables are measured on a continuous scale we may validly neglect all the interactions above a certain order, for a $(p - 1)^{\text{th}}$ order interaction is of the p^{th} order of smallness." They further stated that when some variables are qualitative rather than quantitative, "the justification for the assumption must be found in considerations outside the data which the experiment provides in commonsense or philosophical grounds." Box and Hunter (10, p. 213), as justification for designs that do not supply coefficients for higher-degree terms of a polynomial approximating a response surface, allude to the expectation that higher-order effects are negligible "assuming the properties of similarity and smoothness."

The economical designs proposed by these statisticians were devised originally for research in the physical sciences where variables are quantitative. Will the assumption hold, and will the designs be useful for behavioral science research? In human factors engineering research and other areas of applied experimental psychology, because of the interest in equipment and system parameters, many of the independent variables can be ordered quantitatively, sometimes on a continuous and occasionally on a discrete but ordered scale that can be treated as if it were continuous (e. g., 1, 2, 3, 4, 5, etc. targets). However until recently there has been no empirical evidence to support or reject this assumption insofar as human factors research is concerned.

The analysis of all experiments published in the journal, Human Factors, determined empirically how important higher-order effects were in that population

of studies. Some of the results of this analysis are reported here; a description of the study and the measure employed is found in Appendix I.

Higher-Order Interactions

For each experiment, the proportion of variance accounted for by each main and interaction effect of the equipment factors was calculated. After separating the data, in Table [II-2], by the order of the interaction being examined (Column 1) and by the number of factors in the experiments from which the data was taken (Column 2), this data was analyzed in two ways. In one case, the sum of the proportion of variance accounted for by all of the interactions of the same order in an experiment (Column 3) was the basic unit for the analysis; in these cases, the term "combined" was used (Columns 5 through 9). In the other case, the proportion of variance for the individual interactions were analyzed (Columns 10 and 11).

For example, in half of the experiments studying four factors at a time, the sum of the proportion of variance accounted for by four three-factor interactions was 0.03 or less. The maximum proportion accounted for by the sum of the four three-factor interactions in any of these four factor experiments was 0.11. Of the 13 combined proportions accounted for by the sum of the four three-factor interactions in each experiment, 13.2 percent (or two combined interactions) accounted for more than 0.05 of the total variance. When individual interactions were examined, only 1.9 percent (or one interaction out of the 52 of that category) accounted for more than 0.05 of the total variance.

Since all interaction effects of the same order in a single experiment seldom accounted for approximately the same proportion of the variance, it would probably be misleading to divide a combined proportion by the number of proportions that were summed to obtain it. For example, the combined proportion of variance accounted for by all four of the two-factor interactions in a four-factor experiment was 0.11. On the average, each two-factor interaction would account for 0.0275 parts of the total variance. This is not recommended; it would be better to think the combined values of these effects as representing how much of the total variance would not have been accounted for if all two-factor interactions had never been calculated, or would have been confounded with other effects had these effects never been collected.

Table II-2. Analyses of the Proportion of Variance Explained by Equipment Interaction Effects

Order of Interaction Effect (Number of Factors in the Interaction)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Number of Factors Studied in the Experiment	Number of Interactions of each Order in an Experiment	Number of Experiments in this Category	Combined* Proportion of Total Variance Accounted for by All Interactions of the Same Order Within Each Experiment	50%ile	75%ile	100%ile	Percent of Interactions (Combined or Individual) in which the Proportion of Variance Accounted for Exceeds the Indicated Proportion	Combined Interactions of Same Order (n=col. 4)	Individual Interactions (n=cols. 3x4)	
					>.05	>.10	>.10	>.05	>.10	>.05	>.10
5th (5 FI)	5	1	4	.00	.01	.01	.01	0	0	0	0
4th (4 FI)	5	5	4	.00	.01	.01	.01	0	0	0	0
3rd (3 FI)	4	1	13	.00	.01	.02	.02	0	0	0	0
	5	10	4	.02	.02	.02	.02	0	0	0	0
	4	4	13	.03	.03	.11	.11	13.2	7.6	1.9	0
	3	1	55	.01	.02	.19	.19	10.9	5.4	10.9	5.4
2nd (2 FI)	5	10	4	.08	.16	.16	.16	50.0	50.0	5.0	2.5
	4	6	13	.09	.10	.24	.24	76.9	7.6	9.0	0
	3	3	55	.05	.10	.36	.36	43.6	20.0	13.9	4.8
	2	1	93	.02	.04	.65	.65	24.0	11.3	24.0	11.3

* The term "combined" as used throughout this table means that the proportion of variance accounted for by all of the interactions of the same order within each experiment was summed, and it is the summed value that is being analyzed.

From the data in Table [II-2], the following generalizations can be made:

- 1) The more factors studied in a single experiment, the smaller the *proportion of variance* accounted for by individual interactions.
- 2) The higher the order of interaction, the lower the proportion of variance accounted for by that order.
- 3) Four-factor interactions and higher are for all practical purposes negligible.
- 4) In over 75 percent of the experiments, three-factor interaction effects can be considered to be negligible. However, as the number of variables studied in an experiment decreased, some three-way interactions effects were large enough to require further examination.

Three-Factor Interactions. From Table [II-2], it can be seen that when five factors were studied in an experiment, the three-factor interaction effects were negligible. However, this is based on the results from only four experiments. Three-factor interaction effects also appear to be negligible for all practical purposes in the four-factor studies. The maximum combined value of four interactions accounted for only 0.11 of total variance. Of the four interactions that were summed to make that amount, only one accounted for more than 0.05 of total variance; it accounted for 0.06.

All of the experiments in which the combined three-factor interactions accounted for more than 0.05 of the total variance are listed along with some descriptive data in Table [II-3]. This was the case in only eight of the 72 experiments which could be analyzed for three-factor interaction effects. Six of these eight were the effects of individual three-factor interactions; two were the combined value of four effects. Only four of the eight accounted for more than ten percent of the total variance. Two (No. 4 and No. 8) were the combined value of four individual three-way interaction effects of which only one of the six individual interactions ones accounted for 0.06 of the total variance. Two (No. 2 and No. 3), although accounting for 0.18 and 0.16 of the total variance in each experiment, were used in lieu of an error term. That means that the experimenter treated these effects as if they were due to pure chance, i. e., were negligible. One case (No. 7) was not reliable, i. e., statistically significant. The factors making up this group of three-factor

Table [II-3]. Analyses of Three-Factor Interaction Effects Accounting for More Than .05 of the Total Variance

	Number of Factors in the Experiment	Proportion of Total Variance Accounted for by Combined 3FIs	Number of Interaction Effects Summed	Proportion Accounted for by Individual 3FIs	Number of Levels	Type of Variable*	Type of Interaction
1	3	.19	1	.19	2, 2, 2	LLL	Disordinal
2	3	.18	1	.18	3, 2, 30	LLL	**
3	3	.16	1	.16	3, 4, 2	LLL	**
4	4	.11	4	.06, .04, .00, .01	3, 2, 2	NNN	Ordinal
5	3	.10	1	.10	2, 3, 2	LLL	***
6	3	.09	1	.09	3, 3, 5	LLN	Ordinal
7	3	.08	1	.08	3, 3, 5	LLN	Not Significant
8	4	.06	4	.04, .01, .01, .00	20, 3, 2	LNN	Ordinal

*L=qualitative; N=quantitative; LLN=2 qualitative, 1 quantitative; LNN=1 qualitative, 2 quantitative

**Used as error term

***Insufficient data to decide

interactions were primarily qualitative variables; there was only one exception (No. 4). Only one (No. 1) of these three-factor interactions (among those for which it could be determined) was of the disordinal type (X-type). A disordinal interaction is one in which the performance at different levels of a factor will be ordered differently depending on the level of a second factor which is operating when the performance is measured. The others were the ordinal type of interaction (V-type) which could probably have been eliminated had a different measurement scale been used or if the performance scores had been appropriately transformed. It is of interest to note that in the worst case, that is the case in which the three-factor interaction accounted for 0.19 of the total variance, the absolute difference between the worst and the best of the eight experimental conditions in that experiment was 1.44 bits/second of transmitted information from display to control. In reaction time alone, the difference amounted to 0.78 parts of a second.

It is apparent that a tentative assumption that three-factor interactions are negligible is the most parsimonious one to make. In a very few cases, it may be wrong. However if the measurement scales are selected from the beginning to linearize the data as much as possible, the number of critical three-factor interactions will be reduced. Non-negative effects are more likely with qualitative factors.

Cochran and Cox (16, p. 219) suggest watching the two-factor interactions for clues that three factor interactions might be important. They suggest that if the main effects and two-factor interactions of a set of factors are large, then it is likely that some three-factor interactions might also be large. If the two-factor interactions are small, it is less likely (but not impossible) that the three-factor interactions are large.

Two-Factor Interactions. While most economical multifactor designs are constructed so as not to ignore two-factor interactions, it still is of interest to obtain quantitative information on how important these effects are likely to be. From the data in Table [II-2], the following generalizations can be made about the two-factor interaction effects:

- 1) The more factors studied in an experiment, the more likely an individual two-factor interaction will be negligible.
- 2) If all of the data from experiments with three or more factors were combined, only 36 out of 72 experiments had the combined effects of the two-factor interactions in the studies accounting for more than 0.05 of the total variance. Only 11.3 percent of the individual two-factor interactions in the studies involving three or more factors accounted for more than 0.05 of the total variance. Only 3.2 percent of the individual two-factor interactions in the studies involving three or more factors accounted for more than 0.10 of the total variance.
- 3) Two-factor interactions, in general, cannot a priori be assumed negligible.

In general, interaction effects tended to be somewhat higher when qualitative factors were involved than quantitative.

Higher-Order Terms of the Polynomial

The functions relating quantitative factors to performance can be approximated by a graduated polynomial. Each term of the polynomial will represent a single degree of freedom. Thus the main effect of a three-level factor with two degrees of freedom in the analysis of variance, will be represented by two

terms in the equation — a linear and a quadratic term. The interaction of two three-level variables with four degrees of freedom in the analysis of variance would be represented by the following four terms, each with a single degree of freedom, in the polynomial:

$x_i x_j$	(linear-by-linear interaction)	2nd degree term
$x_i^2 x_j$	(quadratic-by-linear interaction)	3rd degree term
$x_i x_j^2$	(linear-by-quadratic interaction)	3rd degree term
$x_i^2 x_j^2$	(quadratic-by-quadratic interaction)	4th degree term

The degree of the term is equal to the sum of the exponents in the term; the order of the equation is equal to the highest degree of any term in the equation. The majority of economical multifactor designs that can be used with quantitative factors limit the data collection to that required for a first or second degree models. In the above example of the two-factor interaction, this would mean that only the linear-by-linear component of the interaction would be estimated and the other three components would be assumed negligible.

Similarly, if a factor contained five experimental levels, its relation to performance could be represented by four terms:

$$x_i, x_i^2, x_i^3, x_i^4$$

of which the cubic and quartic terms would be assumed negligible. The question is: How likely is it that these higher-order effects are really negligible?

Because the analysis of variance model dominated the analyses of the experiments published in the journal, Human Factors, between 1958 and 1972, there was less data available for checking this assumption. However, whenever the means of every level of a quantitative main effect were published, it was possible to determine how well equations containing from first to fifth-order terms would fit these main effects. An analysis was performed on all quantitative main effects with three, four, five, or six levels that had accounted for 0.25 or more of the total performance variance in the experiment. The results are shown in Table [II-4].

Table [II-4]. Proportion of Variances of Main Effects Accounted for as a Function of the Order of the Polynomial

Number of Levels Involved	Order of the Polynomial														
	1st			2nd			3rd			4th			5th		
	Percentile Ranks**														
	1	50	100	1	50	100	1	50	100	1	50	100	1	50	100
3 (20)*	.71	.96	1.0	--	1.0	—									
4 (10)	.55	.76	1.0	.92	.98	1.0	--	1.0	—						
5 (4)	.80	.97	1.0	.95	.99	1.0	.99	1.0	1.0	—	1.0	—			
6 (2)	-	.60	-	-	.98	-	-	1.0	-	-	1.0	-	—	1.0	—

*Numbers in parentheses indicate the number of main effects included in the analysis. Only main effects that accounted for .25 or more of the total variance were included.

**Percentile rank is interpreted to mean: 1 is the smallest proportion of variance of any main effect explained by that order polynomial; 50 is the median proportion explained; 100 is the highest proportion.

Table [II-4] shows the proportion of the variance of quantitative main effects that is accounted for when represented by polynomials of different orders. Obviously an equation of order ($d - 1$) will account for all of the variance of any main effect with d levels. For each group of data, the lowest, median, and highest proportions accounted for are presented as 1, 50, and 100 percentile ranks. One can conclude from the data in this table that for the sample involved, the inclusion of higher-than-second order terms in the polynomial will account for a negligible proportion of the main effects.

When this is so for main effects, then Plackett and Burman's (40) statement regarding critical order of the interaction of quantitative variables is likely to be applicable, and the importance of third-degree and higher effects should be slight. In the few cases when the assumption is not valid, a fact that can be detected if the proper experimental design is employed, more data may have to be collected. The chances for large higher-order effects however may be minimized by the techniques described next.

Methods of Minimizing Higher-Order Effects

The degree to which higher-order effects may be negligible is not totally dependent on characteristics of the factors themselves. Instead the manner in which the experimenter designs his experiment and collects his data can do much to influence the validity of the principle of negligible higher-order effects, as it affects the use of economical multifactor designs. There are a number of steps that can be taken to increase the probability that higher-order effects will be negligible. These are:

- 1) Keep the range of values over which a factor is varied relatively small.

This procedure simply recognizes the fact that sufficiently small sections of any curve can be approximated by a straight line. The investigator should know enough about his factors from preliminary studies to be able to set his boundaries so as to encompass most of the space of interest without exceeding second-order relationships.

- 2) Employ a scale that will linearize the relationship between independent and dependent variables whenever possible.

In order to simplify relationships, transformations of the data are often employed. This should be done beforehand by selecting the values of the levels of the independent variable at proper intervals on a scale that linearizes the function between it and performance.

- 3) Exert proper administrative control during the data collection phase to minimize disruptive events.

When interactions are detected, there is of course no way to distinguish why they occurred by merely examining the data. With

two-factor interactions, the reasonableness of their presence might be determined rationally. With higher-order interactions, this is less likely and it is not impossible that these may have occurred as a result of a subject fouling-up several times or changing his strategy mid-stream in an experiment. None of these conditions are associated with the experiment but are actually artifacts of the experimental situation. Many experimenters attempt to meet these problems by running many subjects or many trials on the same experimental conditions and averaging out these effects. However this is not conducive to data collection economy. The other alternative is to give maximum attention to see that as each piece of data is collected the chances for contamination from irrelevant sources be minimized.

4) Exercise proper controls to eliminate systematic but irrelevant sources of variance.

Many interaction effects of the ordinal variety in experiments in which the same subject is tested under more than one experimental condition come from systematic changes in operator performance, such as learning or fatigue. Other systematic but irrelevant sources of variance can be attributed to such factors as equipment drift. The commonly employed counterbalancing techniques do not always reduce these effects and in fact at times may enhance them. Techniques such as "blocking" (42), using practiced subjects, and monitoring equipment which can't be controlled are all ways in which these systematic sources of irrelevant variance (generally appearing as interactions) can be reduced.

EMD PRINCIPLE III. COLLECT AND EVALUATE DATA IN A SEQUENCE OF PROGRESSIVE ITERATIONS.

Most psychological experiments are completely planned and all the data is collected before the results are formally analyzed. Implicit in this approach has been the attitude that it's not quite cricket to change one's mind once the design has been devised or the data collection is on its way. As a result, the cost of obtaining information has usually been inflated unnecessarily since data collection

generally continues long after the desired information has been obtained. In addition to the higher cost of doing experiments in this way, the information is often of marginal quality because the investigator failed to anticipate disruptive conditions or stop the study when such conditions became apparent as the program progressed.

Had the experiments been planned in such a way that the data required for a complete design be collected and analyze a little at a time before the entire design was completed, the knowledge gained from the first blocks of data could be used to decide what to do next. This knowledge may lead to the decision to alter the course of data collection into more profitable directions or to stop the experiment if there are signs that additional data collection would have contributed little additional information. This principle of progressive iteration is fundamental to most economical designs and provides the safety feature when minimizing replications and assuming negligible higher-order effects.

Box and Hunter (8) have noted that "the only time an experiment can be properly designed is after it has been completed." They state: "It might be possible to devise some rigid system of experimentation which proceeded in accordance with some set of unalterable rules; but this, since it would have to sacrifice the experimenter's basic knowledge, would be extremely inefficient and would commend itself to no one who had any exposure to the realities of experimentation. In practice, what one can do is proceed sequentially and have available at each stage a variety of useful techniques which will help the experimenter to decide what to do next. The aim should be to apply a process which, when properly handled, will converge to the required solution" (p. 139). This principle - so successful in chemical engineering research - can be equally so in human factors experiments for equipment design.

Whereas the two previous principles of design economy were concerned with what measures might be omitted, this principle deals with the way data should be collected. By using this process of progressive iteration, the amount of data

which must be collected to obtain a certain level of information can generally be reduced. This economy is achieved:

- 1) By first obtaining a less precise overview of the effects of a great many factors in order to select the most important to study more precisely later

Too many human factors experiments have expended effort studying factors which after an elaborate experiment was completed was found to have only trivial effects on performance. If a sequential study has been planned, a relatively small amount of data could have been collected first on a great many factors, enough to decide which had the greatest effect on the performance under investigation. Any loss in precision could be compensated for later when only a few truly critical factors are being studied.

- 2) By avoiding the exploration of parts of an experimental space that are uninteresting, uninformative, or unimportant

Instead of collecting data according to a regular pre-arranged pattern which samples at regular intervals throughout an experimental space, an investigator may skirt selectively through the space by collecting a little data at a time, analyzing it, and using it to guide him to the regions of greatest importance. If in the earlier stages of the study the effects of certain factors are found to be negligible, they may be dropped from later data collection efforts. If a first order polynomial doesn't adequately fit the data, the experimental space can be expanded to obtain an estimate of non-linear relationships. If the boundaries of the experimental space don't encompass the coordinates of the optimum response, the foci of the experimental space can be shifted.

Box and Hunter (8) use an iterative approach to find the coordinates of a multifactor space where the chemical yield is maximum. Describing an imaginary system, they illustrate how they would search a two-factor space composed of temperature and percent chemical concentration to find the combination of values which give the optimum chemical response. They point out the extravagance of mapping the entire space since this would include a great many conditions where the response level would be of little

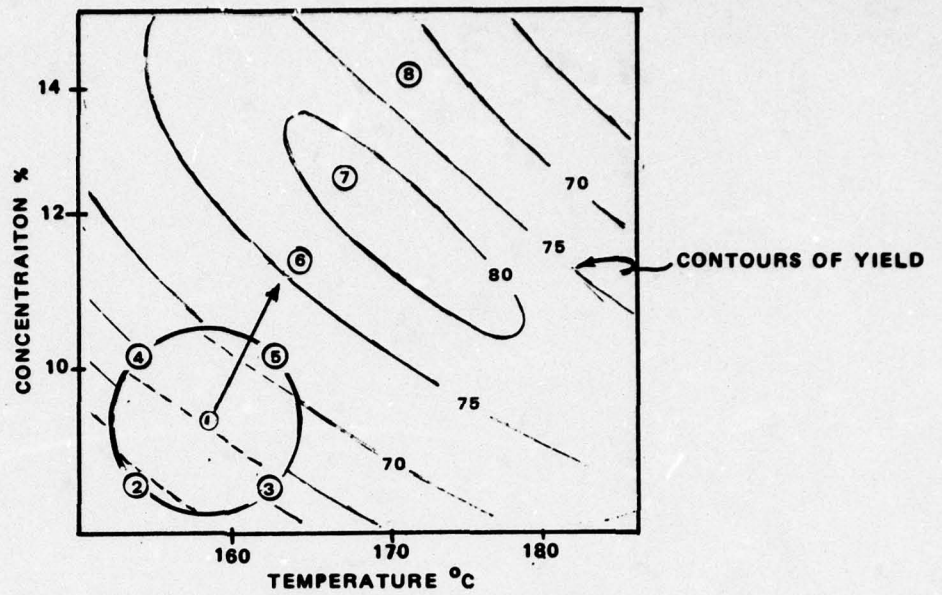
interest. Instead they arrive at the optimum through a series of small iterative steps as follows:

- a) By starting at the "best guess" location, take enough data points to fit by the method of least square, a polynomial of sufficient order to provide a local approximation of the surface (Figure [II-2, A] measures 1-5).
- b) From this information, take additional measures in the region at which higher responses were likely to occur (Figure [II-2, A] measures 6, 7, and 8).
- c) Continue to repeat this until the region of optimum response can be identified (Figure [II-2, B] measures 9, 10, etc).
- d) At the final stage of this progression, before making a complete map of the region, transform the variables and conduct the final mapping experiment in the coordinates of the new scales (Figure [II-2, C]). This eliminates interaction effects, makes the response surface more symmetrical, and simplifies locating the optimum position fairly accurately.

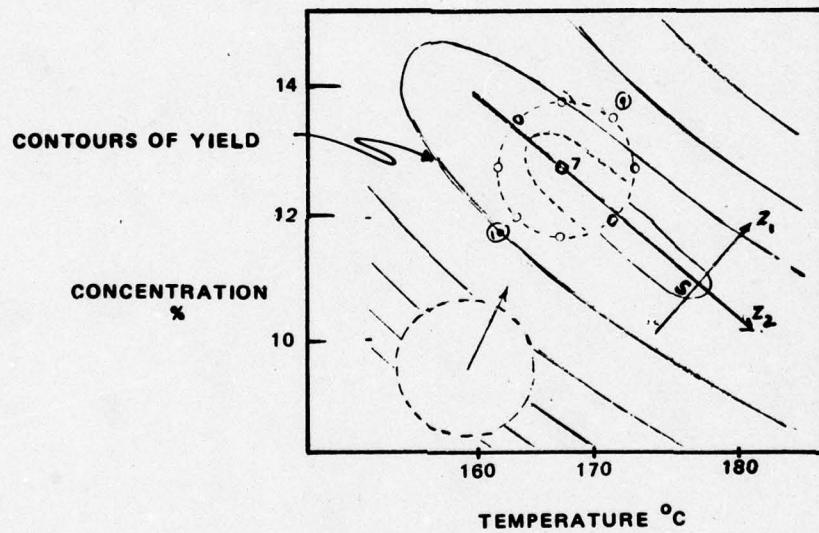
Response surface designs of this type will be discussed in Chapter V.

- 3) By terminating the experiment as soon as all of the desired information has been obtained or when the data already collected explains most of the observed variance

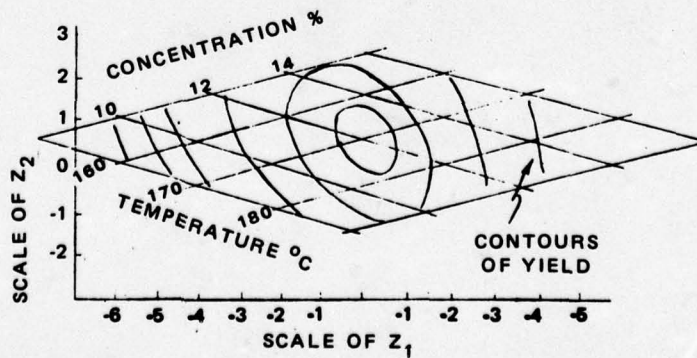
Since one or both of these events generally occur long before the data for a factorial design has been collected, the savings is substantial. A case in point was the seven-factor experiment described earlier (28) to illustrate the savings by assuming higher-order effects are negligible. Its purpose was to assess the effects of seven factors germane to establishing the optimum design of a peripheral vision display. The factors were line width, black-to-white ratio, display area, display shape, visual fixation point, rate of movement, and angle of lines; there were two levels of each factor. The experimenter designed and ran 12 subjects on the full factorial design consisting of the $2^7 = 128$ treatments.



(A) Contour representation of a response surface and first order experimental design.



(B) Contour representation of a response surface and second order experimental design.



(C) Response surface plotted in terms of the transformed variables z_1 and z_2 .

Figure [I-2]. Exploration strategy in the development of a response surface.

[Adapted from Box and Hunter (8)]

The experimenter might have made a considerable savings in the amount of data he had to collect with no loss of information, and in some cases with a gain, by employing one of several possible progressive iteration approaches.

The first approach involves the previously discussed assumption that fourth-order or higher interaction effects will be negligible. Here the investigator would collect enough data to complete a half-replicate (64 conditions) of a 2^7 factorial design. This type of design is discussed in Chapter III. With this much information he could determine whether third-order interaction effects were sizeable and if they were not, could feel reasonably secure that still higher-order effects would be even smaller, and terminate the experiment. If the third-order effects were large, however, he could then complete the other half of the complete factorial to learn more of the higher-order effects. The chances are against the latter case and by using the progressive iteration approach, he exercises an option to reduce the data collection if the facts warrant it.

A still more economical application of collecting data sequentially in progressive iterations can be found using the screening technique described Chapter IV. With this approach, the investigator would begin by collecting data on 24 of the 128 experimental conditions in the complete factorial. However, after each block of eight condition, he would stop and analyze his results in order to determine which new conditions to study in the next block. From the 24 conditions, 3/16th fraction of the total factorial, he would have been able to determine with reasonable confidence which factors and their two-factor interactions were the most important. Although he may decide to add another block to increase the precision of his estimates or resolve some uncertainty that might still exist, he still could terminate his experiment early for he would have learned just about everything he eventually did learn when he completed his entire design.

In still a third approach, described in Chapter V on response surface methodologies, the investigation could run the half-replicate of the 2^7 factorial in blocks of eight conditions, and add in each block an additional experimental point located at the center of the experimental space. The

eight extra center points would provide enough additional data to make a crude test to see if a linear model fit the data already collected. If the linear model was adequate, the experiment could end; if it were not, with eighteen additional data points added to the basic design a second order polynomial could be written for seven variables. This ability to estimate quadratic effects would provide more information than the investigator could have obtained with the complete factorial and at less cost.

Still greater economy might have been achieved by combining the screening approach with the response surface approach. However, with only seven variables, the potential savings would be very small. When the number of variables reaches 15 or more, the potential savings would allow multifactor experiments to be conducted that otherwise might never have been possible.

- 4) By being able to evaluate some data before the decision to replicate is made

The arguments against replication have already been presented. When there are indications that a half-replicate of a factorial has given all of the information that a full replicate would give, the experiment should stop. To repeat the same half-replicate cannot be justified. If one is going to waste measurements in that manner, possibly to increase precision, it would be better to complete the factorial without repetition. The same would be true if only a $1/16^{\text{th}}$ fraction of the total factorial had been run. Rather than repeat the same experimental conditions again, the same precision and more information would be obtained if a different $1/16^{\text{th}}$ fraction were run.

An iterative approach provides little justification for replication. If in an early analysis, the data seems uncontrolled, then the solution is to find the source of the unexplained variance rather than a means of hiding it. Suspected variables might be added to subsequent stages in the build-up of the design, or at least controlled or measured.

If an analysis of an incomplete factorial with one observer does appear to provide all of the information needed, replicating the

same set of experimental conditions using a second subject might serve one purpose which has not been dealt with up to now. It may provide a quick check of the reliability of the first set of estimates. While this is exactly what a test of statistical significance is intended to do when two sets of data such as this are combined and analyzed, if the economy of data collection has minimized the number of degrees of freedom to the point where a test of significance will have little power, more subjective confidence (rather than statistical reliability) may be acquired for the data if a visual check of a scatter-plot of the means of the experimental conditions from two subjects yields a relatively straight diagonal with a slope of one. This type of plot is also useful for quickly detecting the conditions on which two subjects deviate significantly: this permits explanations to be sought before continuing the data collection.

Severe differences between the ranks of the experimental conditions ordered on the mean performance will appear in an analysis of variance as a subject-by-factor interaction. Unless a specific subject characteristic is being investigated, such an interaction has little information value. Commons reasons for a disagreement among rank orders of experimental conditions between presumably homogeneous subjects are: 1) poor experimental control, 2) large order-of-presentation effects, 3) inadequate training and/or practice, and 4) momentary distractions, either internal or external, to the subjects.

EMD PRINCIPLE IV. SUBSTITUTE EXPERIMENTER'S KNOWLEDGE AND ANALYTIC SKILLS FOR DATA COLLECTION.

Cookbook experimental designs and mechanical data collection procedures are inordinately wasteful. Considerably greater economy can often be achieved when the experimenter becomes more personally involved. As a result of their "behavioristic" background, experimental psychologists have frowned upon this approach. Reacting to "arm-chair" psychology, many psychologists have tried to emulate the "scientific approach" by eliminating all subjective considerations from the data collection, both on the part of the subject and the experimenter. As a result, many of them have stopped doing research and began doing rigid experiments of meager depth and limited breadth. To reverse this trend, more investigator involvement is needed.

The use of the experimenter's judgment to modify the course of an experiment has already been discussed in EMD Principle III. As the experiment progresses, the investigator can decide whether or not he needs to continue to collect more data, whether to add or drop factors, or to shift the experimental space, or to replicate or not. But these judgments are made in order to avoid collecting data unnecessarily, that is, to avoid collecting data that will add essentially nothing new to the information already obtained. There are, however, applications of experimenter judgment wherein this knowledge and skill can be used to obtain information in lieu of actual data collection.

Selecting the Proper Measurement Scale

In many experiments, the investigator's experience with the independent and dependent variables is sufficient to enable him to anticipate the shape of their functional relationships. If he puts this information to proper use, he can usually reduce the amount of data he must collect without any material loss of information. For example, there is an abundance of psychophysical data to show that when the intensity of light (in the middle brightness range) is increased in equal physical increments, the change in brightness will be perceived by the observer as a curvilinear function, i. e. monotonic and negatively accelerated. To approximate this curve, at least three points would have to be plotted. However, by knowing that this is the approximate function relating physical and

psychological brightness, the experimenter could plan his data collection by selecting levels of light intensity distributed at equal intervals on a logarithmic scale. In this case, brightness as perceived by an observer would be essentially linearly related to the physical change (on a log-footlambert scale) and a minimum of only two levels would be required to approximate it. In this hypothetical example where only the minimum possible data points are considered, the difference between two or three levels for this single factor may appear small and of little practical consequence. However, if there were seven factors in an experiment with a similar amount of savings, then for a complete factorial design the number of data collection points would drop from $3^7 = 2187$ to $2^7 = 128$, and for a fractional factorial that keeps all main and two-factor interactions from being confounded with one another, the reduction would be from $3^{7-2} = 243$ to $2^{7-1} = 64$. Furthermore, by preplanning so that the experimental factors are scaled so as to approximately linearize their individual relationships to performance as much as possible, not only is the amount of data to be collected reduced, but also the chances that higher-order interaction effects will be negligible is increased (EMD Principle No. 2).

When relationships are not known beforehand, the experimenter can often obtain sufficient experience quickly and cheaply by making a few preliminary measurements. An informal exploration of factors and their parameters before any serious planning begins is probably the quickest and safest way to provide an observant and reasonably sophisticated investigator with the clues needed to select the best candidate experimental factors and their measurement scales, as well as to forwarn of potential problems that might arise during the data collection. This preliminary effort will almost always enhance the quality of final experimental results, and materially reduce the effort required to collect good data.

Identifying which Confounded Effects are Important

The economy achieved by not isolating confounded effects that are assumed to be negligible was discussed in EMD Principle II. For example, if Factor A and Interaction ABCD are confounded, there would be no need to isolate the two effects if it could be assumed that the four-factor interaction effect were negligible. Any

measured effect in this case — actually the sum of $A + ABCD$ — must be assumed due to Factor A.

There are circumstances however when a number of effects are confounded and the chance that all of them are negligible is low. A typical case in a sequential screening design is the confounding of a string of two-factor interactions. The ordinary approach would be to collect a complete block of additional data to separate the effects of the different two-factor interactions in the string; this would permit the important ones to be identified. A more economical approach, described in Chapter V, would be to collect a little extra data in such a way that the experimenter can determine analytically from the existing data which two-factor interactions are most probably important.

The situation is analogous to an electronic technician who must troubleshoot a complex piece of equipment in order to determine the cause of a malfunction. He may follow a highly proceduralized job aid that takes him a step at a time through a standard sequence of checks, looking for the signals that will indicate where the trouble lies. Or he may know from the combination of observable symptoms the approximate location of the trouble and start his testing near there, rather than go through the entire, more elaborate sequence. If he is correct, he has reduced the number of steps needed to find the trouble.

As a general principle, whenever the experimenter's judgment can be used in place of data collection, it should be as long as provisions are made to have this judgment eventually checked.

EMD PRINCIPLE V. MINIMIZE BIAS EFFECTS ON EACH INDIVIDUAL MEASUREMENT.

If one wishes to collect less data while obtaining essentially the same information, the data that is collected must be as accurate as possible. All of the principles of economical multifactor designs depend on this being the case. Yet if one examined the experimental literature, the size of some error variances seem to negate precision and accuracy for much of the human factors data.

In half of the 239 experiments analyzed in Human Factors, more than 25 percent of the total performance variance within the experiments could not be explained by the equipment factors and their interactions, subject factors, and temporal factors combined. In a quarter of the experiments analyzed, 44 percent of the performance variance was "unexplained"* by those factors. Among individual experiments, there were some in which the unexplained variance was less than 10 percent and some in which it was more than 90 percent. Since these percentages describe only the amount that was not explained within the experiment and since experiments ordinarily include only some of the conditions operating in the real world, the experimental results could be expected, on the average, to describe very little of what would happen under operational conditions.

There is a prevailing attitude — implied if not actually expressed — that a large residual variance in so far as human performance is concerned is natural, i. e., it is a normal phenomenon to be deplored but accepted. As a result, even when half of the variance in an experiment is not accounted for by the factors that were intentionally varied (or those, like subjects and trials, that might be expected to vary), the quality of the data is seldom questioned. Instead, experimenters (anticipating that such a condition might exist) rely upon massive, redundant data collection programs and a mystical faith in the ability of a statistically elegant experimental design to purify badly conceived and poorly executed experiments. With economical multifactor designs, such laxity can no longer be tolerated.

EMD Principle V reverses this trend by emphasizing the importance of being concerned with the purity of the measurement of each individual data point. It is based on the premise that much of what has been considered to be error or residual variance, that is, the large unexplained variance within an experiment, is not an inherent and inescapable characteristic of human performance, but the result of inadequate experimental planning, improper data analysis, and poorly managed data collection techniques. If the more common sources that frequently bias experimental measurements were reduced, eliminated, or measured as each piece of data is collected, then:

- 1) That which is called residual error variance within an experiment will shrink to an inconsequential size;

*See page 163, in Appendix I, for the specific definition of the term "unexplained", as used in this report.

- 2) Field performance will be predicted more accurately from laboratory data.

The goal of EMD Principle V, of course, is to make each individual measurement so bias-free that it can stand alone as a valid representation of performance under analogous conditions in the real world. Some examples commonly found in human factors engineering experiments that may bias the results are discussed below.

Sources that Bias Experimental Measurements

Most experienced experimenters will acknowledge that financial pressures, time limitations, political considerations, and other sources not directly related to the experiment can create an environment in which biased data is likely to occur. In this environment, the least-experienced personnel are assigned the tasks of collecting and analyzing the data; these are the ones least prepared to recognize surreptitious sources of bias or to know how to handle them if they are recognized. As a result, conditions that bias experimental measurements are quite commonly found. In general, these conditions fall into two major classes:

- 1) Those that affect individual experimental conditions differentially.
- 2) Those that affect the experimental conditions uniformly.

In the first class, uncontrolled and/or unidentified factors vary throughout the experiment and become confounded with estimates of the means, interactions, and residual variances. Isolated incidents and events that appear and disappear at random throughout the experiment also have effects on performance. These confounded effects result in mean distortions that remain hidden (since there are no standards against which to compare them); they are revealed however by the large residual of unaccounted-for variance and a failure to predict outside the laboratory.

In the second class, the conditions of the experiment are non-representative of the conditions found in the real world. These distortions cannot be recognized from an examination of the experimental data; they are revealed when experiments with little internal residual variance fail to predict performance in the operational situation. To predict should be the ultimate criterion of experimental quality.

A few of the more common circumstances that can distort human factors experimental data and account for a high residual error variance are:

Design. Too few factors and too few levels per factor are used because it was believed (incorrectly, as this report will show) that to include more would make the size unmanageable. Some of the factors that are studied are not the important ones because the customer is not sophisticated enough to ask the right questions and the experimenter is not sufficiently motivated to educate the customer. Certain nominal factors (e. g., airfield) are in fact a composite of several factors (e. g., object size, object-to-background brightness contrast, object pattern); among different airfields, these critical visual factors are allowed to vary indiscriminately. Insufficient time is allotted to a pre-experimental period in which a fruitful range of values for the experimental factors can be established and the procedures tested so that the experiment can be run smoothly.

Equipment. Left-over equipment from a previous study is used in spite of the fact that it was not designed to simulate the new task properly. The parameters of many equipment factors held constant are unknown. Complex stimuli in the real world are represented unrealistically in the experiment to simplify the task of defining and controlling them. A technique for simulation is selected because it is cheaper rather than because it is representative. Equipment is built with little regard for the problems of running an experiment; as a result changing experimental conditions becomes so complicated and time consuming that mistakes are made and subjects grow weary. Environmental parameters that affect performance but cannot be controlled are not measured as they vary so that their effect can be removed statistically after the fact. Experimenters fail to understand how components of a physical system interact, so that when one condition is set, others are unknowingly changed. Equipment is not properly debugged before the experiment is begun.

Subjects. Subjects are selected from "the guys in the lab" or student "volunteers" from the Psychology 100 class. "Image interpreters" are borrowed from the military for a target recognition study; the fact

that they have been trained to interpret photographic imagery while the study involves imagery from an advanced radar system is considered irrelevant. Subjects are improperly motivated or instructed; they become bored with the proceedings or modify their procedures part way through the experiment. Limitations on the use of subjects are arbitrarily imposed by such things as union rules in industry or military protocol. Inadequate monitoring of the subject during the actual data collection can result in a failure to note that he is not following instructions, has become tired, or was not paying attention at the appropriate time. Subjects are distracted or disturbed by conditions of the environment when the laboratory is not properly shielded.

Procedures. So much time and money are used to construct the equipment that the data collection phase must be hurried. Long experiments are divided into blocks of time without regard for the advantages of orthogonal blocking. Concern with possible order of presentation effects but without the knowledge of how to properly handle them causes an experimenter to randomize the order. No effort is made to determine at the time of occurrence why an extreme performance score occurred — was it an artifact or just an extreme of a normal distribution? The experimenter has insufficient experience to know what to do during the experimental run when data on a particular condition is lost.

Analysis. Although counterbalancing for order of presentation effects, these sources of variance are not isolated during the analysis of the data. The use of particular designs such as a Latin square makes it impossible to estimate certain interaction effects (e. g., equipment X subject X trials) which are almost certainly going to have an effect. Error variances are actually what's left over after the experimenter has removed what he may be interested in rather than what he should have. Small experimental designs leave too few degrees of freedom to make powerful enough test of significance. Data is analyzed automatically by computer and is not studied for peculiarities by the experimenter. The experimenter does not know how to handle outliers or missing data.

In summary, as each measurement is made the investigator must constantly assess whether or not the critical parameters associated with the equipment, the subjects, the environment, and the task (including those artificially introduced by the experimental procedures) at that moment are representative of the conditions in the field to which the experimental results are to be eventually extrapolated. If they are not, the data at that point is distorted and the results of the experiment will be distorted. This same type of assessment must be made to guide the experimenter who must decide how to correct a detected distortion. Biasing circumstances can be eliminated with a little care, should be if the quality of the experimental data is to be maintained, and must be if economical multifactor designs are to be viable.

CHAPTER III.
ECONOMICAL DESIGNS FOR QUALITATIVE FACTORS
(FRACTIONAL FACTORIALS)

A factorial design is made up of experimental conditions in which every level of every factor is combined once with every level of every other factor. A fractional factorial design, or fractional replication, is made up of only a portion of the experimental conditions of the complete factorial selected in such a way that higher-order effects are not isolated from lower-order effects. Thus the economy from fractional factorials is based on the assumption that higher-order interaction effects are negligible and need not be independently estimated.

Fractional factorials have been used infrequently in human factors engineering research, appearing primarily in the form of a Latin square. These designs are presented here because they do represent one form of economical design, but more important, because their characteristics and methods of construction are basic to the designs discussed in later chapters. While fractional factorials at two levels are suited for both qualitative and quantitative variables, those which can handle three or more treatments of a variable will probably be more useful in the study of qualitative variables.

Because so much excellent material has been written about the construction and characteristics of fractional factorials (16)(17)(18)(19)(23)(29)(34)(45)(51) only enough information will be presented here to familiarize the reader with some of the fundamental concepts, notations, and techniques of forming fractional replicates so that he will understand their applications in subsequent chapters. For a more complete treatment, supplemental reading is urged.

SOME UNDERLYING CONCEPTS AND NOTATIONS

There are a number of ways of conceptualizing the conditions of an experimental design. Of these, a sign matrix is a particularly useful form for understanding two-level factorial and fractional factorial designs. This discussion will show the

relationship between the symbology conventionally employed by the psychologist to describe his experimental design and the sign matrix. The development here may seem slow to some; it has been purposefully oversimplified to be sure to get the ideas across.

Developing a Sign Matrix for Two-Level Factorial Designs

Psychologists often design experiments by drawing cells to represent the experimental conditions. For example, in a two-factor, two-level design, the following would represent the experimental plan.

		Factor A	
		Low (l)	High (a)
Factor B	Low (l)	I 4	II 2
	High (b)	III 6	IV 4

The Roman numerals in each cell of the design serve to identify the cells. The Arabic numbers in the lower-right corner of each cell are fictitious performance scores assigned to each condition. The alphanumeric, (l) and (a) or (b), beside the levels Low and High, respectively, of each factor are abbreviated notations used to represent those levels. In an experiment, each experimental condition is formed by combining the levels of the two factors in each cell, as follows:

<u>Cell Number</u>	<u>Factors</u>		<u>Fictitious Performance Scores</u>
	<u>A</u>	<u>B</u>	
I	low(l)	low(l)	4
II	high(a)	low(l)	2
III	low(l)	high(b)	6
IV	high(a)	high(b)	4

This can be more simply expressed by using only the alphanumeric designations for the low and high levels:

<u>Cell Number</u>	<u>Factors</u>		<u>Fictitious Performance Scores</u>
	<u>A</u>	<u>B</u>	
I	(1)	(1)	4
II	a	(1)	2
III	(1)	b	6
IV	a	b	4

If only the letters of the factors in which the higher level is being used are written down, this matrix can be shortened still more. Thus:

<u>Cell Number</u>	<u>Experimental Condition</u>	<u>Fictitious Performance Scores</u>
I	(1)	4
II	a	2
III	b	6
IV	ab	4

Every factor contributes to each experimental condition; therefore, where no letter is shown in the notation, the low level for the corresponding factor is in fact being used. For example, Cell Number III is a combination of the low level of factor A (since a is missing) and the high level of factor B (since b is present). A condition in which all levels are low is designated by (1).*

Experimental conditions of a 2^k design (where k equals the number of factors) can also be described by denoting the low level of a factor by a minus sign (-) and the high level by a plus sign (+), thus:

*The concepts of low and high can only be applied to quantitative factors. When qualitative factors are being studied, no such distinction can be made. If one level of the qualitative factor can be considered the standard from which deviations are to be measured, that is usually designated the low level.

<u>Cell Number</u>	<u>Experimental Condition</u>	<u>Factors</u>	
		<u>A</u>	<u>B</u>
I	(1)	-	-
II	a	+	-
III	b	-	+
IV	ab	+	+

With four observations of a 2^2 design, three independent effects can be estimated. One is the effect of factor A, another is factor B, and as in any factorial design, the third is the interaction of A and B. The signs of the AB Interaction can be determined by "multiplying" the signs for A and B according to conventional arithmetic rules—multiplying two of the same signs gives a plus and two different signs gives a minus.*

Thus for signs for the AB interactions would be

<u>A</u>	<u>B</u>	<u>AB</u>
-	-	= +
+	-	= -
-	+	= -
+	+	= +

These can be combined into a sign matrix along with a fourth column, referred to as the Identity (I) column, which can be used to calculate the mean of the data. A column of the fictitious performance scores for each experimental condition is also added. The completed matrix is shown in Table [III-1].

*If there had been three factors, A, B, and C, and if for a particular experimental condition the signs were -, -, and + respectively, then the signs for the four possible interactions would have been: AB, +; AC, -; BC, -; and ABC, +. Actually, a + represents +1 and a - represents -1, and it is the ones that are actually being multiplied. These are eliminated in the notation and discussion for the sake of simplicity.

Table [III-1]. Experimental Conditions, Sign Matrix, and Scores

Cell Number	Experimental Conditions	Sign Matrix			Fictitious Performance Scores	
		Primary A	B	Derived AB		Identity (I)
I	(1)	-	-	+	+	4
II	a	+	-	-	+	2
III	b	-	+	-	+	6
IV	ab	+	+	+	+	4

Estimating the Effects (Mean Differences)

The primary section of the sign matrix in Table [III-1] shows the combinations of levels of each factor that define each experimental condition. Thus condition a in Cell II would be made up of the high value of factor A and the low level of factor B.

The entire matrix can be used to estimate the effect of each factor and its interactions. For example, to estimate the effect of factor A, the signs in the A column would be attached to the corresponding performance values, thus:

-	4
+	2
-	6
+	4

Summing these, we would get -4. This sum must then be divided by 2^{k-1} , where k is the number of factors in the experiment. In this case, $2^{k-1} = 2^{2-1} = 2^1 = 2$. When -4 is divided by 2, we get -2, the effect of factor A. It is also the mean difference between the performances in the high and the low conditions of factor A.

Using the sign matrix [III-1] to estimate the AB interaction effect, we would assign the signs in column AB to the experimental conditions and obtain the following:

$$\text{Effect AB} = [+1 - a - b + ab]/2$$

and when the performance scores are substituted,

$$\text{Effect AB} = (+4 - 2 - 6 + 4)/2 = 0/2 = 0$$

Similarly we could estimate the effect of factor B, thus:

$$\text{Effect B} = (-4 - 2 + 6 + 4)/2 = +4/2 = +2$$

In calculating the mean using the signs of the Identity column, one must divide by the total number of experimental conditions, thus:

$$\text{Mean} = (+4 + 2 + 6 + 4)/4 = +16/4 = +4$$

Calculating Sums of Squares and Mean Squares

For a 2^k factorial design, the sum of squares can be obtained directly from the estimated effect since

$$\text{Sum of squares} = 2^{k-2}(\text{effect})^2$$

where k is the number of factors in the experiment. In the above example, with two factors, then $2^{k-2} = 1$, and the sum of squares for each source of variance would be:

$$\text{Sum of squares for A} = (-2)^2 = 4$$

$$\text{Sum of squares for B} = (+2)^2 = 4$$

$$\text{Sum of squares for AB} = 0 = 0$$

The total sum of squares would be 8. This can be checked by the conventional method of summing the squares of the deviations of the performance scores in each experimental condition from the grand mean, or

$$(4 - 4)^2 + (2 - 4)^2 + (6 - 4)^2 + (4 - 4)^2 = 8$$

Since each sum of squares is associated with a single degree of freedom, the sum of squares for each effect equals the mean square, or variance. With this design, there is no estimate of error.

Orthogonality

The property of orthogonality can be illustrated with the sign matrix in Table [III-1]. When two independent factors are orthogonal, they are uncorrelated, unconfounded, and their effects can be independently estimated.

Orthogonality is said to exist between any two factors if their cross products sum to zero, or in the case of a sign matrix, where the cross products of their corresponding signs contain an equal number of plus and minus signs. Although we talk of + and - signs, we are, in reality, dealing with +1 and -1 but for convenience have ignored the numbers. Thus if we multiplied the signs of columns A and B of the sign matrix in Table [III-1], we would get (from top to bottom) +, -, -, +. Effects of A and AB, B and AB, A and I, B and I, and AB and I are also orthogonal.

CONSTRUCTING FRACTIONAL FACTORIALS FOR FACTORS AT TWO LEVELS

In this section, the complete factorial will be divided into smaller blocks and only some of these blocks will be used — a fraction of the total design. Of course when less data is taken, some information is lost. The construction of fractional factorials depends on the selection of what will be saved and what will be lost.

Blocking and Confounding

Blocking refers to a technique of dividing the experimental conditions of a complete factorial design into smaller units, or blocks. When the correct set of

experimental conditions are assigned to each block, an average performance change between blocks will not bias the estimates of the effects of greatest interest. Blocking is useful when, for example, it is not possible to run an entire factorial design on a single day. Instead of dividing up the conditions in some random fashion to do half on one day and half on another, more systematic blocking techniques should be employed. Then even if something happens between days to cause the performance on all second day conditions to be higher, blocking can prevent the effects of greatest interest from being biased by this shift. But there is a price. Each time an experiment is blocked to preserve certain effects, the estimates of some other effects will be lost by being confounded with any effects due to differences between blocks.

Confounding means that the effects of two or more sources of variability are not independent, i. e., orthogonal. When effects are confounded, it is not possible to determine which effect is responsible for observed differences in performance. For this reason, when blocking, the investigator tries to select those experimental effects to be confounded in which he is least interested or which he believes to be unimportant in the first place.

Although too simple a situation to be of any practical value, let us continue to use the 2^2 factorial design to illustrate how blocking and confounding occur. In the original sign matrix of Table [III-1], the performance scores associated with the four experimental conditions were: (1) = 4; a = 2; b = 6; and ab = 4. However, before the experiment these "true" values would be unknown to the experimenter. Let us imagine that he wishes to determine the effects of factors A and B and their interaction, AB, but must run half the experiment on each of two days. He suspects that there are uncontrollable changes in his equipment from day to day, and is concerned how he should divide the four experimental conditions into two sets of two. He has three alternatives, as shown in Table [III-2].

If there is an average change in performance from day to day which the investigator cannot measure, and he divides the experimental conditions according to the first alternative, he will obtain erroneous information on the effect of factor B which is confounded with the effect of differences between days (blocks).

Table [III-2]. Blocking Alternatives for a 2^2 Factorial

	Alternatives		
	1	2	3
ONE DAY*	(1) a	(1) b	(1) ab
ANOTHER DAY	b ab	a ab	a b

*No distinction is made here as to which is the first and second day, a consideration which would produce three more alternatives.

This can be seen from [III-2], since a difference between performance on the conditions on the two days is the same as the difference between the high and low levels of factor B. Similarly, should he choose the second alternative, he will obtain erroneous information on the effect of factor A. Should he choose the third alternative, he will receive erroneous information about the AB interaction effect.

In larger studies, the number of alternatives would be equal to $N(N - 1)/4$ where N is the number of experimental conditions to be divided into two days and 4 reflects the fact that there was no effort to distinguish which day a block of conditions will go into. When the number of experimental conditions are larger than in this over-simplified example, many of the possible alternatives will leave the estimates of all of the effects biased — confounded with blocks — if the experimenter does not understand the principles of blocking. That is why in situations such as this, the worst thing to do is to assign the conditions into days according to some random plan. Instead, the investigator should block his experimental conditions so that he will lose the information he cares least about and will preserve the information in which he is most interested. Although the choices are ridiculously limited in this simple example, let us assume that the investigator is least interested in the AB interaction. This means that he should confound the effects of the AB interaction with the effects due to days. This is done by placing in one day all experimental conditions which in the sign matrix in Table [III-1]

are minus on the AB interaction and in the other day, those which are plus. This is shown in Table [III-3]. Performance scores on each condition are the same as those given in the original sign matrix of Table [III-1] except that all scores on the second day were increased by 3 points to represent the additional effect that uncontrolled changes in the equipment had on performance. Since the experimenter can never know what the real (original) values were, he must use the above data to estimate the effects.

If we calculate the effect of A, B, and AB as we did before we would find:

$$\text{Effect A} = [a + ab - (1) - b]/2 = (+5 + 4 - 4 - 9)/2 = -4/2 = -2$$

$$\text{Effect B} = [b + ab - (1) - a]/2 = (+4 + 9 - 4 - 5)/2 = +4/2 = +2$$

$$\text{Effect AB} = [(1) + ab - a - b]/2 = (+4 + 4 - 5 - 9)/2 = -6/2 = -3$$

By comparing these values with the earlier calculations, we can see that in spite of an increase of +3 in the last two conditions, the effects of factors A and B are unaffected. On the other hand, the estimate of the AB interaction effects has changed. While we know from the way the problem was devised that the change came from the increase during the second block, ordinarily an investigator would

Table [III-3]. Blocked 2^2 Factorial

		Effects			Performance
		A	B	AB	
Day 1	(1)	-	-	+	4
	ab	+	+	+	4
Day 2	a	+	-	-	5
	b	-	+	-	9

never know whether the observed effect was due to an AB interaction or a difference in blocks or both. But by sacrificing the estimate of one effect, in this case the AB interaction, the investigator was able to obtain an unbiased estimate of the remaining effects.

Had the investigator blocked by confounding factor A, perhaps because he was interested in obtaining an unbiased estimate of the AB interaction, then conditions a and ab would be in Day 1 and b and (1) would be in Day 2. Days are equivalent to blocks, of course. In this case, the estimates of the B and AB effects would be unaffected by an increment of +3 in performance in the last half of the experiment, but the estimate of the effect of factor A would be totally confounded with the effect of blocks.

In larger experiments, a design could be divided into more than two blocks and in that case more than one effect would be lost. As the number of factors increase, it becomes more probable that some higher-order effect confounded with block will be negligible. In that case, blocking can be accomplished without any practical loss of information.

Fractioning and Aliasing

A fractional factorial design is created by using the experimental conditions of some of the blocks in the total factorial and eliminating the remaining blocks of conditions from the experiment. If certain criteria are met, the information obtained from the fractional replicate will be for all practical purposes, as good as that obtained from the full replicate. It is in creating and selecting a fraction most likely to meet the required criteria that the problems of design arise.

To illustrate the problems, conditions, and techniques associated with the design of fractional factorials, we shall begin with the complete factorial for four variables at two levels each. The complete sign matrix for a 2^4 factorial is shown in Table [III-4].

Table [III-4]. Sign Matrix for a 2^4 Factorial Design

	EFFECTS															
	I	A	B	AB	C	AC	BC	ABC	D	AD	BD	ABD	CD	ACD	BCD	ABCD
(1)	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
a	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-
b	+	+	-	-	+	+	-	-	+	+	-	-	+	+	-	-
ab	+	-	-	+	-	+	+	-	-	+	+	-	-	+	+	-
c	+	+	+	+	-	-	-	-	+	+	+	+	-	-	-	-
ac	+	-	-	-	+	+	-	-	+	-	-	-	+	+	-	-
bc	+	+	-	-	+	-	-	-	+	-	+	+	-	-	-	-
abc	+	-	+	+	+	+	+	-	-	-	-	-	-	-	-	-
d	+	-	-	-	+	+	+	-	+	-	-	+	-	+	+	-
ad	+	+	-	-	-	+	+	-	+	+	+	-	-	-	+	+
bd	+	-	+	-	-	+	-	-	+	-	+	-	-	-	+	+
abd	+	+	+	+	-	-	-	-	+	+	+	+	-	-	-	-
cd	+	-	-	-	+	-	-	+	+	-	-	+	+	-	-	+
acd	+	+	-	-	+	+	-	-	+	-	-	-	+	+	-	-
bcd	+	-	+	-	+	-	-	-	+	-	+	-	-	+	-	-
abcd	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

EXPERT'L
CONDITIONS

We begin by dividing the 16 conditions of the 2^4 factorial design into two blocks, using the ABCD interaction as the basis for the division. Any effect used to block a factorial is referred to as a defining contrast. In this case, there is only one, the ABCD interaction. As with the 2^2 factorial, the experimental conditions are assigned to blocks by putting all of the conditions with a plus sign in the ABCD column into one block and all with a minus sign into the second block.

The size of the experimental design is reduced by eliminating one of the blocks. In this example, the block with the minus signs in the ABCD column was not used. The sign matrix of the remaining block, with only plus signs in the ABCD column, is shown in Table [III-5]. Since this is the block with the (1) condition in it, the one with the lower level of all factors, it is referred to as the "principle" block. The original sixteen experimental conditions of the 2^4 factorial have been reduced to eight conditions, a half-replicate of the complete factorial. This is expressed as a

2^{4-1} design

Table [III-5] Sign Matrix for a 2^{4-1} Fractional Factorial Design
(Principle block) (I = ABCD)

EXPERIMENTAL CONDITIONS	EFFECTS															
	I	A	B	AB	C	AC	BC	ABC	D	AD	BD	ABD	CD	ACD	BCD	ABCD
(i)	+	-	-	+	-	+	+	-	-	+	+	-	+	-	-	+
ab	+	+	+	+	-	-	-	-	-	-	-	-	-	+	+	+
ac	+	+	-	-	+	+	-	-	-	+	+	-	-	+	+	-
bc	+	-	+	-	+	-	+	-	-	+	+	-	-	+	-	+
ad	+	+	-	-	-	+	+	+	+	-	-	-	-	+	+	+
bd	+	-	+	-	-	+	+	+	-	+	-	-	-	+	-	+
cd	+	-	-	+	+	-	-	+	+	-	+	-	+	-	-	+
abcd	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

or 2^{-1} (one-half) of the 2^4 factorial, which of course is composed of $2^{4-1} = 2^3 = 8$ experimental conditions.

When half the data required for the complete factorial is collected, half the information which might have been estimated is lost. This can be understood by studying the sign matrix in Table [III-5]. It can be seen that:

- 1) No estimate of the effect of the defining contrast, the ABCD interaction, can be made. Only the positive conditions of ABCD are in this block making the calculation of the ABCD effect equal to that for the mean, which is calculated from the Identity column, I.
- 2) Every other effect has an equal number of plus and minus conditions. Thus all of these effects can be estimated from the differences between high and low levels within blocks.
- 3) However, certain pairs of effects have an identical sign pattern, for example, effects A and BCD, effects B and ACD, effects BC and AD,

etc. In fact every effect has one other effect with the same sign pattern. That means that when the performance values associated with each experimental condition are summed according to the sign pattern, the effects of these matched or aliased sources will be the same. These aliased effects are totally confounded; no independent estimate of the effects of the aliased pairs are possible. It is impossible to know whether the measured effect of A is due to factor A or interaction BCD or some combination of both.

Instead of constructing a sign matrix and relying on visual inspection to determine which effects are aliased, there is a rather simple way to determine this. First, the defining contrast is specified as

$$I = ABCD$$

where I is referred to as the Identity factor and when multiplied by any effect is treated as unity (one).

To determine the alias of A, the defining contrast, ABCD, is "multiplied" by A, as if by the usual rules of algebra, but dropping all squared terms.* Thus:

Defining Contrast	$I = ABCD$
Multiplied by	$A = A$
Results in	$A = A^2BCD = BCD$

The alias of an interaction is calculated in the same way, e. g.,

Defining Contrast	$I = ABCD$
Multiplied by	$AC = AC$
Results in	$AC = A^2BC^2D = BD$

*This procedure has its mathematical basis in modular arithmetic.

With eight experimental conditions from the 2^{4-1} fractional factorial it is possible to estimate seven independent effects, no matter how many variables are being studied. The entire aliased set would be:

Defining Contrast	$I = ABCD$
Effect 1	$A = BCD$
Effect 2	$B = ACD$
Effect 3	$C = ABD$
Effect 4	$D = ABC$
Effect 5	$AB = CD$
Effect 6	$AC = BD$
Effect 7	$AD = BC$

As with Latin squares, when effects are aliased, e. g., $A = BCD$, the effect that is actually being measured is

$$A + BCD$$

The plus sign does not necessarily mean that the apparent effect of A will always be enhanced if the effect of BCD is not negligible. BCD may have a negative effect, so that when it is aliased with the effect of A,

$$A + (-BCD) = A - BCD$$

the observed effect might appear smaller than the independent effect of A, or the two large effects could conceivably cancel each other out.

An examination of the aliases in this design reveals the importance, when fractional factorials are used, of the assumption that higher-order interaction effects are negligible. With this particular design in Table [III-5], unbiased estimates of the main effect are possible only if the three-factor interactions are negligible, and unbiased effects of any three of the two-factor interactions are possible only if their aliases — another set of two-factor interactions — are negligible. For human factors engineering problems, with two-factor interactions aliased with one another, the usefulness of this 2^{4-1} design would be quite limited.

Suppose that there had originally been a five-factor factorial, a 2^5 design, from which a half-replicate was created with the ABCDE interaction for the defining contrast. Using the multiplication technique just described, it becomes apparent that all main effects will be aliased with only four-factor interactions and all two-factor interactions will be aliased with only three-factor interactions. The possibility of getting unbiased main and two-factor interaction effects has increased considerably with this design.

The Resolution of a Fractional Factorial

The resolution level of a fractional factorial design indicates the degree and nature of its alias pattern. Of particular interest is the alias pattern of the main effects and the two-factor interactions. In this report, designs of Resolutions III, IV and V or higher have the greatest applications. The relationships between some resolution levels and which main and interaction effects are confounded are as follows:

Resolution III: Main effects are unconfounded with one another but aliased with all interaction effects.

Resolution IV: Main effects are unconfounded with one another and two-factor interactions but two-factor interactions are aliased among themselves. Both are aliased with higher-order interactions.

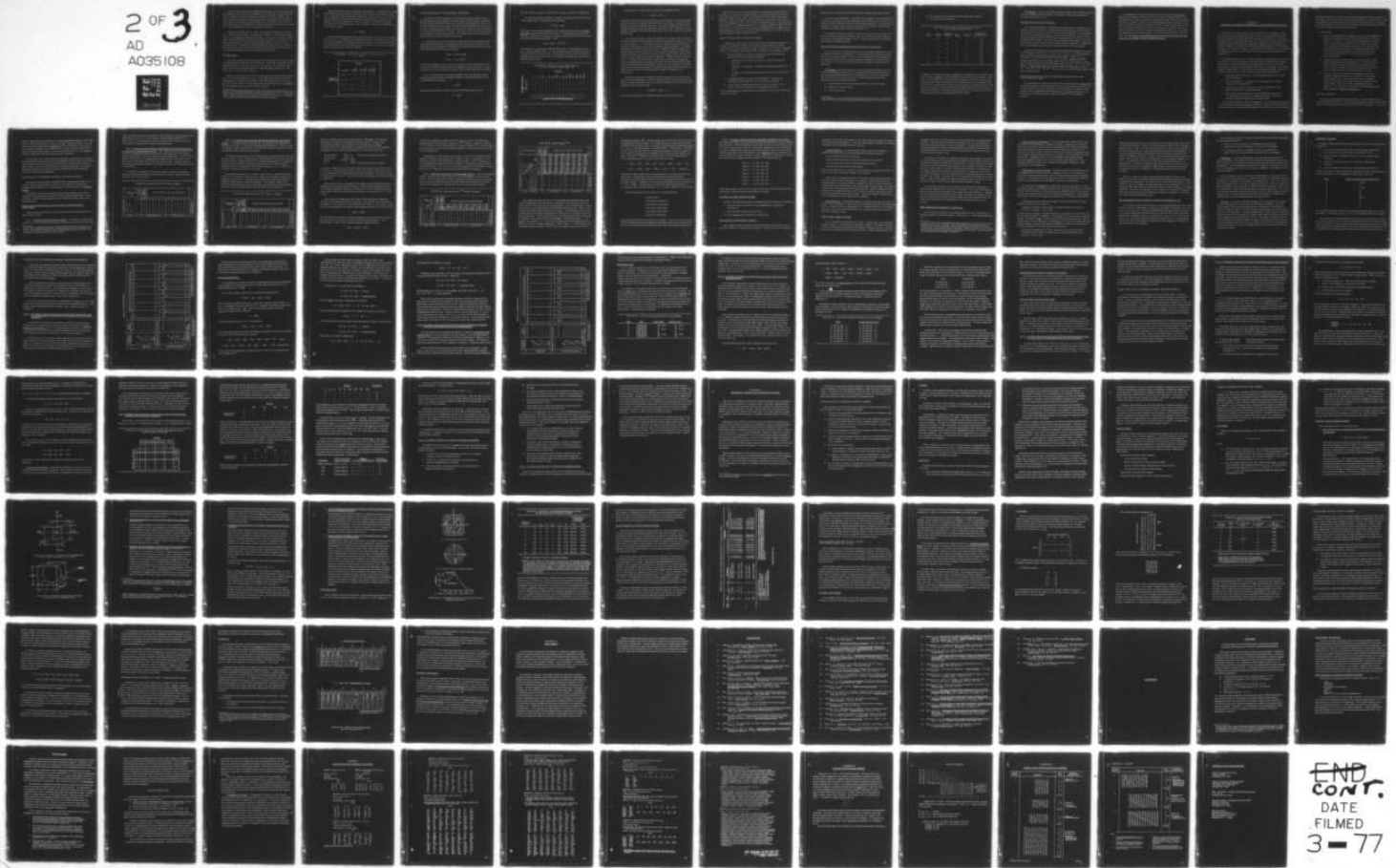
Resolution V: Main effects and two-factor interactions are unconfounded with one another but are aliased with higher-order interactions.

AD-A035 108

HUGHES AIRCRAFT CO CULVER CITY CALIF ENGINEERING EQU--ETC F/6 5/5
ECONOMICAL MULTIFACTOR DESIGNS FOR HUMAN FACTORS ENGINEERING EX--ETC(U)
JUN 73 C W SIMON F44620-72-C-0086
HAC-P73-326A NL

UNCLASSIFIED

2 OF 3
AD
A035108



END
CONT.
DATE
FILMED
3-77

Since for most human factors engineering problems, one cannot assume with any confidence that two-factor interactions are not important, designs of Resolution V or higher are the most interesting. They are the first in which neither main effects nor two-factor interaction effects are confounded within or between one another, being aliased only with higher-order interactions. This does not mean that there are no applications for designs of Resolutions III and IV, for there are. Some important uses will be discussed in Chapter IV.

Designs of Resolutions III, IV, and V are sometimes referred to as three-, four-, and five-letter designs, referring to the number of letters in the smallest "word" in the defining contrast for the design.* It is easy to see how this relates to the degree of aliasing. A Resolution III design with a three-letter word in the defining contrast (e. g., XYZ) must alias a main effect (X) with a two-factor interaction (YZ). A Resolution IV design with a four-letter word (e. g., WXYZ) will alias a main effect (X) only with a three-factor interaction (WYZ) but some two-factor interactions (XY) will be aliased with others (WZ).

The Other Block

In the first example of a fractional factorial, the principal block was selected to represent the half-replicate of the 2^4 factorial. This block included the experimental conditions: (1) ab, ac, ad, bc, bd, cd, and abcd. But what if the other block had been chosen which contained the remaining eight experimental conditions: a, b, c, d, abc, abd, acd, bcd?

Should the selection of one block or the other affect the results of the experiment? Not if the assumptions are met. If the higher-order aliased effects are truly negligible, then lower-order effects will be the same whether one block or the other is used. However, if the higher-order aliased effects are not negligible, then the combined effects in the two blocks will differ.

*Up to now, all defining contrasts have had only a single word since we have considered only half-replicate designs. When smaller fractions are developed, more than one effect will be involved. When these are strung out, e. g., $I = ABCDEF = CDE = ABF$, the effects are referred to as "words." In this case, the smallest word has three letters and it would be a Resolution III design.

How does the block which is used affect the notations? An examination of Table [III-6] shows that all the signs in the ABCD column in the second block are negative. No estimate of this interaction is possible, of course, and it is still aliased with the Identity (I) column representing the mean of the block. However, the signs of these two columns are reversed. The signs in the Identity column are still plus, but in the ABCD column, they are minus. Therefore, the defining contrast would be written

$$I = -ABCD$$

Using the multiplication technique described earlier, the alias of the A effect would be -BCD. An examination of Table [III-6] reveals that these two effects also have identical patterns, but that the signs are reversed. This characteristic will be found with all of the aliased effects in this second block; one of the two aliased pairs will be positive and the other negative.

Table [III-6]. Sign Matrix for a 2^{4-1} Fractional Factorial
($I = -ABCD$)

		EFFECTS															
		I	A	B	AB	C	AC	BC	ABC	D	AD	BD	ABD	CD	ACD	BCD	ABCD
EXPERT'L CONDITIONS	+	+	-	-	-	-	-	-	-	-	-	-	-	+	+	-	-
	+	-	+	-	-	-	+	-	+	-	+	-	+	+	-	+	-
	+	-	-	+	+	-	-	-	+	-	+	+	-	-	+	+	-
	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-
	+	+	-	-	-	+	+	-	-	+	+	-	-	+	+	-	-
	+	-	+	-	-	+	-	+	-	+	-	+	-	+	-	+	-

CREATING SMALLER 2^{k-p} FRACTIONAL FACTORIALS

If instead of the half-replicate, a still smaller design was desired, then a quarter-replicate of the original 2^4 factorial design could be created. This time the half-replicate would be divided into two parts by selecting another effect to be sacrificed and using for the quarter-replicate only the half with either all plus or all minus conditions for that effect. Of course, there would be no practical reason to use a quarter-replicate of a 2^4 design. That would involve only four experimental conditions for the entire experiment. The example is used here only to illustrate how smaller replicates can be constructed.

Let us assume that the experimenter decides that he is not interested in the effect of the ABD interaction, and decides to use it for the next division. If the eight experimental conditions of Table [III-5] are divided on the basis of the signs in the ABD column, the two blocks would be:

+ block = ac, bc, cd, abcd

- block = (1), ab, ad, bd

Note that the conditions of one block (+) are all those with an odd number of the letters in the ABD interaction and that the conditions of the other block (-) has all conditions with an even number (or none) of letters found in the ABD interaction.

If the + block is used as the quarter-replicate, then the signs for quarter-replicate effect ABD would correspond with those of the Identity factor and the relationship would be written

$$I = ABD$$

Had the other block been selected, then the relationship would have been

$$I = -ABD$$

For this example, however, we will use the + block, as shown in Table [III-7].

To create this quarter-replicate, estimates of the effects of ABCD and ABD were purposely lost. We may write the expression

$$I = ABCD = ABD$$

and refer to the two effects associated with the Identity factor (I) as the defining generators rather than the defining contrast. By multiplying these generators together, we generate a third effect, or word, which is also aliased with the Identity factor and the other two effects. Thus,

$$ABCD \times ABD = A^2B^2CD^2 = C$$

An examination of the quarter-replication sign matrix, Table [III-7], will show that no effects of ABCD, ABD, or C can be estimated since only the high level (+) conditions of each are in that block. No contrast with a lower level is possible. Aliasing between main and two-factor interactions is considerable; each effect is aliased with four others.

Table [III-7]. Sign Matrix for a Quarter Replicate of a 2^4 Factorial ($I = ABCD = ABD = C$)

		EFFECTS															
		I	A	B	AB	C	AC	BC	ABC	D	AD	BD	ABD	CD	ACD	BCD	ABCD
EXPERIMENTAL CONDITIONS	ac	+	+	-	-	+	+	-	-	-	-	+	+	-	-	+	+
	bc	+	-	+	-	+	-	+	-	-	+	-	+	-	+	-	+
	cd	+	-	-	+	+	-	-	+	+	-	-	+	+	-	-	+
	abcd	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
		I	A	B	D	I	A	B	D	D	B	A	I	D	A	B	I
ALIASES OF EFFECTS IN CORRESPONDING COLUMN																	

The presence of a main effect as a part of the defining contrast,

$$I = ABCD = ABD = C$$

raises some question as to the desirability of the particular set of defining generators that were used. Ordinarily it is preferable to be able to estimate all main effects. But what alternatives are there? One might try different effects in the defining contrasts, but for this particular example, no other combination would eliminate aliasing at least one main effect and in some, more than one main effect would be lost.

In the early discussion of fractional factorials, the highest-order interaction was used to create the half-replicate. However, when further divisions are made, the "best" design — i. e., the one that permits the highest Resolution possible — may not necessarily be created using the highest-order interaction. The quarter replicate of a 2^5 factorial is a case in point. A higher resolution design can be obtained using two three-factor and one four-factor interactions for the defining contrast than using a five-factor interaction with any other effect. With the 4-3-3 factor interaction selection, a quarter-replicate, 2^5 design, would be of Resolution III. No main effects would be confounded with one another, although they would be confounded with two-factor interactions. Had we used instead the highest, five-factor interaction with either a four-factor, three-factor or two-factor interaction as the other generator, the complete defining contrast would have at least one word of one or two letters and be a design of Resolution I or II respectively.

Another characteristic of defining contrasts containing more than one word has to do with the sign pattern. Had we selected the set of four experimental conditions associated with the - sign for the quarter replicate of the 2^4 factorial, the defining contrast would have been

$$I = ABCD = -ABD = -C$$

Note that the multiplication of signs is retained across these conditions.

While 2^{k-p} fractional factorials represent a considerable economy in data collection over the use of a complete factorial, particularly when the number of factors are eight or more, there are relatively few human factors engineering problems in which the interest is strictly limited to a great many factors having only two conditions or levels. The advantage of understanding the construction and symbology of 2^{k-p} fractional factorials (i. e., fractional replicates, $\frac{1}{2^p}$ of the total, with each of k factors at two levels) will be more apparent in later chapters when these designs are employed as a step in the screening process or a part of a more complex design to obtain response surfaces.

SOME 2^{k-p} FRACTIONAL FACTORIAL DESIGNS

In Appendix II, some two-level fractional factorial designs for from five to 15 factors are provided. These were selected from the document entitled "Fractional Factorial Experiment Designs for Factors at Two Levels," published by the U. S. Department of Commerce (45), according to the following criteria:

- 1) All main effects and two-factor interactions are unconfounded with one another, with the following exceptions:
 - 13 factors: 12 two-factor interactions could not be estimated (out of 78).
 - 14 factors: 2 two-factor interactions could not be estimated (out of 91).
 - 15 factors: 2 two-factor interactions could not be estimated (out of 105).
- 2) All designs required less than 300 observations. (Actually, the maximum number was 256 for 10, 12, 14 and 15 variables.)
- 3) No more than 16 experimental conditions are in any block, and no main or two-factor interactions is confounded with blocks.

In the original document, other designs are available in which some of the above criteria are not met.

FRACTIONAL FACTORIALS FOR FACTORS WITH MORE THAN TWO LEVELS

If the factors are quantitative, somewhere in the progress of the experiment it will often be necessary to look at a minimum of three and as many as five levels in order to determine whether non-linear relationships might exist. If the factors are qualitative, however, there can be occasions where the number of experimental conditions of a single factor might be more than five.*

Fractional factorials for factors with more than three levels are a part of the body of economical multifactor designs. However, they are beyond the scope of this report. This group of designs, however, might be useful in the investigation of the effects of qualitative factors.

Symmetrical Fractional Factorial Designs with Three or Four Levels

The economy of the fractional factorial over the complete factorial becomes a necessity if more than a few factors are to be examined and these factors contain three or four levels per factor. However, these designs will not be discussed in this report since the material receives excellent treatment in a number of other sources and is not critical for understanding other designs considered later in this report.

3^{k-p} Designs. Excellent discussions on the construction of three-level fractional factorial designs can be found in Cochran and Cox (16) and Davies (23). A government publication prepared by Conner and Zelen (19) provides designs for four to ten factors, at three levels each. Of these designs, those listed in Table [III-8] satisfy the following criteria:

- 1) Require less than 300 experimental conditions in the basic design
- 2) Handle five or more factors
- 3) Allow for blocking

*In the analysis of 14 years of human factors engineering research, only eight percent of the factors in 239 experiments looked at more than five levels per factor.

- 4) No main effects are confounded with any other main effects or two-factor interaction effects.

Table III-8. Fractional Factorials with Three Levels Found in Conner and Zelen (19)

Number of Factors	Fractional Replicate	Number of Observations in Replicate	Number of Blocks	Observations per Block	Clear Two-Way Interactions Over Total Number Possible
5	1/3	81	9	9	9/10
			3	27	10/10
6	1/3	243	27	9	13/15
			9	27	15/15
7	1/9	243	27	9	18/21
			9	27	21/21
8	1/27	243	27	9	24/28
			9	27	28/28
9	1/81	243	27	9	30/36
			9	27	36/36
10	1/243	243	9	27	43/45

In addition, most of the two-factor interactions are independent of one another, but in some cases, portions of the interactions are aliased with other portions. Interactions of two three-level factors contain four degrees of freedom, each of which can be isolated. While some of these are the portions that are confounded, other portions of the same interaction may still be isolated, and an effect estimated. Had these few exceptions not been allowed, much larger blocks or more total observations would be needed. As with any fractional factorial, both main and two-factor interaction effects are aliased with higher-order effects.

4^{k-p} Designs. Four-level fractional factorial designs can be constructed from two-level fractional factorial designs. A method for doing this is explained in Cochran and Cox (16, p. 273).

Non-Symmetrical Fractional Factorials

It is not always possible nor desirable for an experimenter to assign an equal number of levels to all factors. Fractional factorial designs for factors with two and three levels have been worked out by Conner and Young (18). In these designs, as in the other fractional factorials noted here, the grand mean, all main effects and all two-factor interaction effects can be estimated, that is, they are not confounded with one another. They are of course confounded with higher-order interaction effects, which tentatively must be assumed to be negligible. An explanation of how these designs are constructed is given in Conner and Young's paper.

Conner and Young provide non-symmetrical fractional designs for each of the $39 \cdot 2^m 3^n$ designs, from $(m + n) = 5$ to $(m + n) = 10$, $(m, n \neq 0)$. Of these designs, ten exceed the 300 observation limit set for this report. These were: $2^7 3^3$, $2^6 3^4$, $2^4 3^5$, $2^5 3^5$, $2^2 3^6$, $2^3 3^6$, $2^4 3^6$, $2^2 3^7$, $2^3 3^7$, and $2^2 3^8$. Available ten factor designs requiring less than 300 observations were $2^1 3^9$, $2^8 3^2$, and $2^9 3^1$. Nine factor designs requiring less than 300 observations were $2^1 3^8$, $2^5 3^4$, $2^6 3^3$, $2^7 3^2$, and $2^8 3^1$. Eight factor designs requiring less than 300 observations were $2^1 3^7$, $2^3 3^5$, $2^4 3^4$, $2^5 3^3$, $2^6 3^2$, and $2^7 3^1$. No seven factor or smaller design required more than 300 observations.

USING FRACTIONAL FACTORIAL DESIGNS WITH QUANTITATIVE AND QUALITATIVE FACTORS

For experiments in which the majority of the critical factors are quantitative, fractional factorials are best employed as a device for achieving economy in conjunction with the screening process and in the development of response surface designs. These applications will be discussed in the subsequent chapters. For this purpose, fractional factorial designs of Resolution V will probably be sufficient.

For experiments in which the majority of critical factors are qualitative, appropriate fractional factorials can prove to be more economical and still provide essentially the same information as complete factorials when a large number of factors are studied. However, as an added safety precaution, designs of Resolution VII should be employed if possible in order to keep the third-order interactions unconfounded among themselves and with lower-order effects. This may mean increasing the number of experimental conditions within a block or employing a slightly larger fractional replicate. To do otherwise, however, would be risky since the chances of getting important third-order interactions with qualitative factors are higher than with quantitative factors.

CHAPTER IV.
ECONOMICAL DESIGNS FOR SCREENING A LARGE NUMBER OF FACTORS

At the start of a human factors research program, the investigator is often aware of fifteen to thirty equipment, system, and/or environmental factors, that could conceivably have an important effect on operator performance. Typically in experiments on equipment design, this list is reduced to from two to four factors usually on the basis of expediency, equipment availability, and experimenter or customer interests, with little regard for their relative importance in the scheme of things. As a result, in the past, considerable time and money has been expended investigating factors which have relatively small effects on the performance of interest (44).

The experimental plans in this chapter provide a method with which the effects of from fifteen to thirty variables can be studied while taking far fewer measurements than have often been made in some experiments of two or three factors. These plans, referred to in the statistical literature as "screening" or "saturated" designs, are all forms of the fractional factorial designs discussed in Chapter III. They are treated here as a special class of economical designs because:

- 1) They should be used early in a research program when less is known about the problem.
- 2) They are intended for "screening" very large number of variables to identify the most important.
- 3) They are not intended to obtain an accurate representation of any particular part of the experimental space.
- 4) They trade any loss in precision for the opportunity of obtaining a comprehensive overview of the experimental space in order to know what should be studied later in greater detail.

The screening process is as much an approach as it is an experimental design. Every principle of economical designs is employed. With fractional factorials as a basis, screening is accomplished by applying the judicious use of the progressive

iteration principle and by using the experimenter's judgments at times in place of more data collection to unravel certain confounded effects. Mathematically, screening designs are the same whether applied to chemical or human factors problems, but methodologically there may be some additional considerations which affect the use of these designs when humans are involved.

These include:

- 1) When screening designs are employed in some chemical engineering and agriculture experiments, only N observations may be used to study $N - 1$ variables. In human factors experiments, while each block of data should be examined as it is obtained in accordance with the principle of progressive iteration, it is unlikely that the screening study would end before $3N$ observations are made. At least $3N$ observations are needed to isolate two-factor interactions from main effects and to identify the important two-factor interactions.
- 2) In human factors screening studies, the order in which experimental conditions are presented serially to an observer is more likely to introduce biased results than in chemical research. This is a general problem found in all studies in which a man is his own control and will not be discussed in this section.
- 3) For many human factors engineering problems, building the apparatus needed to perform a truly multifactor study could become prohibitively costly, particularly since the primary purpose of the study is to eliminate most of the variables from future studies. While this is not an experimental design problem, it can influence the selection of both designs and experimental problems, whether it should or not.

GENERAL APPROACH

Screening studies progress in several stages. The first stage involves a saturated design in which $(N - 1)$ effects will be isolated by using at least N experimental conditions carefully selected from the total factorial design. The effects

that are isolated in saturated designs are usually independent estimates of the main effects*, each confounded with two-factor and higher interaction effects. For this reason, the basic design must be augmented in the second stage of the screening study, usually by adding N more observations, to isolate the main effects from at least the two-factor interactions. Further observations may be added to isolate or at least identify which two factors interactions are important.

By this stage, there should be enough information to grossly order the factors and two-factor interactions in terms of the magnitude of their effects on performance. Still the number of measurements taken will have been relatively few, yet with a large number of factors, the precision of the estimates fairly high. The quality of the data increases as the number of factors increase and so does the savings incurred from using screening designs.

STAGE ONE OF THE SCREENING PROCESS: SATURATED DESIGNS

The number of experimental conditions in these experimental designs must be at least one more than the number of factors to be studied in the experiment. Designs with this high factor-to-condition ratio are often referred to as saturated designs.

The number of experimental conditions in the basic design can be used to identify two types of saturated designs for slightly different applications. In one the number conditions must equal some power of two; in the other, they must be divisible by four.

Constructing Saturated Designs when the Number of Conditions Equals a Power of Two

Using a technique described by Box and Hunter (10) saturated designs can be constructed as follows:

Step 1. Determine the size of the basic design. The number of experimental conditions in this basic (saturated) design should equal the next power of two larger

*In this report, screening studies are conducted with each factor having only two levels. An investigator has the task of selecting these levels to represent the points between which a maximum range of performance is likely to occur. The importance of some exploratory work is evident.

than the number of factors to be studied. With six factors, the next power of two higher than six is $2^3 = 8$. With 25 factors, the next higher would be $2^5 = 32$, and so forth. To illustrate the procedure, a plan for the study of seven factors requiring eight experimental conditions will be developed.

Step 2. Construct a sign matrix of N experimental conditions which permits N-1 effects to be independently isolated. Since eight experimental conditions are needed to independently estimate seven effects, the sign matrix for a factorial design already known to meet the conditions is written down first. The 2^3 factorial permits seven effects to be independently estimated: three main effects, three two-factor interactions, and one three-factor interactions. Using the notations and symbols described in Chapter III, the sign matrix for the 2^3 factorial is shown in Table [IV-1].

The matrix is orthogonal. When any two columns of signs are multiplied together, the product column has an equal number of + and - signs which (being actually +1 and -1) sum to zero.

Table [IV-1]. Sign Matrix for a 2^3 Design - Design I

EXPERIMENTAL CONDITIONS	DESIGN TYPE AND RESOLUTION 2^3_V	GENERATING PRODUCTS	DEFINING CONTRASTS (*GENERATORS)	INDEPENDENT FACTORIAL EFFECTS AND ALIASED INTERACTIONS						PERFORMANCE SCORES	
				I=Mean	A	B	C	ABC	BC		AC
1	(1)		+	-	-	-	-	+	+	+	6
2	a		+	+	-	-	+	+	-	-	4
3	b		+	-	+	-	+	-	+	-	8
4	a b		+	+	+	-	-	-	-	+	10
5	c		+	-	-	+	+	-	-	+	0
6	a c		+	+	-	+	-	-	+	-	2
7	b c		+	-	+	+	-	+	-	-	22
8	a b c		+	+	+	+	+	+	+	+	36
DESIGN I				PRIMARY SIGN MATRIX			DERIVED SIGN MATRIX				

Step 3. Convert the eight treatment, three-factor matrix to a seven-factor matrix. The preceding matrix, Table [IV-1], while enabling seven independent estimates to be made, is suitable for handling only three factors. What is needed is a design of eight treatments which will enable the effects of seven factors to be isolated and estimated with equal precision.

To illustrate the procedure, the 2^3 design is first converted so that it can handle four factors. This is accomplished by substituting the fourth factor for an interaction effect in the original design which is assumed (tentatively) to be negligible. Without any other evidence, the highest order interaction is usually selected. Therefore, factor D is substituted for the ABC interaction, i.e. $D = ABC$, and Design II, shown in Table [IV-2], is formed.

Note that the sign matrix for Design II, Table [IV-2], is identical to that of Design I [IV-1]; the labels, however, have changed, for example, to reflect the addition of another factor. The double line of effects above the sign matrix now indicates which effects are aliased in Design II. For example, Design II was created by making Effect D equal to ABC; this means that D and ABC are aliased. However, in any design where there are four factors at two levels each,

Table [IV-2]. Sign Matrix for a 2^{4-1} Design - Design II

EXPERIMENTAL CONDITIONS	DESIGN TYPE AND RESOLUTION		GENERATING PRODUCTS	DEFINING CONTRASTS ("GENERATORS")	INDEPENDENT FACTORIAL EFFECTS AND ALIASED INTERACTIONS						PERFORMANCE SCORES			
	2 ³ V	(4-1) IV			I	ABCD*	BCD	ACD	ABD	D		AD	BD	CD
					I=Mean	A	B	C	ABC	BC		AC	AB	
1	(1)			+	-	-	-	-	+	+	+	6		
2	a	d		+	+	-	-	+	+	-	-	4		
3	b	d		+	-	+	-	+	-	+	-	8		
4	a b			+	+	+	-	-	-	-	+	10		
5	c	d		+	-	-	+	+	-	-	+	0		
6	a c			+	+	-	+	-	-	+	-	2		
7	b c			+	-	+	+	-	+	-	-	22		
8	a b c	d		+	+	+	+	+	+	+	+	36		
DESIGN I			II	PRIMARY SIGN MATRIX				DERIVED SIGN MATRIX						

there can be fifteen mean and interaction effects. When there are only eight experimental conditions in a design for four factors (a 2^{4-1} design), then it is understood implicitly that effects will be aliased with one another. The Identity factor (I), representing the estimate of the mean, is determined by multiplying the only known aliased effects by one of its own terms, thus:

Aliased effects	$ABC = D$	(This was an arbitrary selection.)
Multiplied by	$ABC = ABC$	
Equals	$A^2B^2C^2 = ABCD$	
Or	$I = ABCD$	(Defining Contrast for Design II.)

In Design II, Table [IV-2], the word in the defining contrast aliased with the Identity factor, I, is written above I in that column. The aliases of the other effects are also written above the effects with which they are aliased. For example, when $I = ABCD$, then $A = BCD$, $B = ACD$, and so forth, according to the rules described in Chapter III.

Since a new factor, D, was added to the design, this must be reflected in the designations for the experimental conditions. This is done by adding the letter d to the original designations whenever the high level of factor D is included in the experimental condition (as indicated by the presence of a plus sign in the $D = ABC$ column).

The procedure of substituting a new factor for the interactions tentatively assumed to have negligible effects can continue. For example, the next substitution might be a fifth variable, E, for interaction BC. Since $E = BC$, then $I = BCE$ (as previously explained). But since I is also aliased with ABCD, we can now write

$$I = ABCD = BCE.$$

BCE and ABCD are referred to as design generators, since their product will produce still another word (effect) which is also aliased with the others. Thus $(BCE)(ABCD) = ADE$, and the complete defining contrast becomes:

$$I = BCE = ABCD = ADE.$$

We have now completed a design in which five factors, from A to E, are being studied with eight experimental conditions, a 2^{5-2} design, or a quarter-replicate of a 2^5 factorial. This is Design III which is shown in Table [IV-3]. In this design, the seven effects can still be independently isolated, but each effect has three other effects aliased with it.

The technique of substituting a new variable for each of the interactions in the original 2^3 design could continue until $D = ABC$, $E = BC$, $F = AC$, and $G = AB$ to form a seven factor, eight experimental condition design. This 2^{7-4} design is a one-sixteenth fractional factorial of the complete $2^7 = 128$ experimental conditions of the full factorial. The seven effects are still independent of one another, but now each effect is a composite of a main effect aliased with 15 interactions.

Step 4. Determining the aliases of the independent effects. This fully saturated design, referred to as the Basic design in future discussions, is shown in Table [IV-4]. While the sign matrix remains unchanged, the labels show each increment in the substitution process along with the new aliased effects and the new designations for the experimental conditions that occur in this seven factor design.

Table [IV-3]. Sign Matrix for a 2^{5-2} Design - Design III

DESIGN TYPE AND RESOLUTION		GENERATING PRODUCTS	DEFINING CONTRASTS (*GENERATORS)	INDEPENDENT FACTORIAL EFFECTS AND ALIASED INTERACTIONS							
				12	DE	ABDE	ACDE	BCDE	ABCDE		CDE
2 ⁵⁻² IV		12	ADE	DE	ABDE	ACDE	BCDE	ABCDE	CDE	BDE	
2 ⁴⁻¹ IV		2	BCE*	ABCE	CE	BE	AE	E	ABE	ACE	
2 ³ V		1	ABCD*	BCD	ACD	ABD	D	AD	BD	CD	
			I=Mean	A	B	C	ABC	BC	AC	AB	
EXPERIMENTAL CONDITIONS	1	(1)									6
	2	a	d	+	-	-	-	+	+	+	4
	3	b	d	+	-	+	+	-	-	-	8
	4	a b	d	+	+	+	-	-	-	+	10
	5	c	d	+	-	-	+	-	-	+	0
	6	a c	e	+	+	-	+	-	+	-	2
	7	b c	e	+	-	+	+	+	-	+	22
	8	a b c	d e	+	+	+	+	+	+	+	36
DESIGN I II III		PRIMARY SIGN MATRIX				DERIVED SIGN MATRIX					

Table [IV-4]. Basic Design (2^{7-4})

DESIGN TYPE AND RESOLUTION		GENERATING PRODUCTS	DEFINING CONTRASTS (GENERATORS)	INDEPENDENT FACTORIAL EFFECTS AND ALIASED INTERACTIONS									
				1	2	3	4	5	6	7			
EXPERIMENTAL CONDITIONS	1 (1) 2 a 3 b 4 a b 5 a c 6 a b c 7 b c 8 a b c	e f g d e f g d e f g d e f g d e f g d e f g d e f g d e f g	1234	BCDEFG	ACDEFG	ABDEFG	DEFG	ADEFG	BDEFG	CDEFG	6 4 8 10 0 2 22 36	PERFORMANCE SCORES	
			234	EFG	AEFG	BEFG	CEFG	ABCEFG	BCEFG	ACEFG			ABEFG
			134	ADFG	DFG	ABDFG	ACDFG	BCDFG	ABCDFG	CDFG			BDFG
			34	BCFG	ABCFG	CFG	BFG	AFG	FG	ABFG			ACFG
			124	BDEG	ABDEG	DEG	BCDEG	ACDEG	CDEG	ABCDEG			ADEG
			24	ACEG	CEG	ABCEG	AEG	BEG	ABEG	EG			BCEG
			14	CDG	ACDG	BCDG	DG	ABDG	BDG	ADG			ABCDG
			4	ABG*	BG	AG	ABCG	CG	ACG	BCG			G
			123	CDEF	ACDEF	BCDEF	DEF	ABDEF	BDEF	ADEF			ABCDEF
			23	ABEF	BEF	AEF	ABCEF	CEF	ACEF	BCEF			EF
			13	BDF	ABDF	DF	BCDF	ACDF	CDF	ABCDF			ADF
			3	ACF*	CF	ABCF	AF	BF	ABF	F			BCF
12	ADE	DE	ABDE	ACDE	BCDE	ABCDE	CDE	BDE					
2	BCE*	ABCE	CE	BE	AE	E	ABE	ACE					
1	ABCD*	BCD	ACD	ABD	D	AD	BD	CD					
	I=Mean	A	B	C	ABC	BC	AC	AB					
DESIGN I	II	III	IV	V	PRIMARY SIGN MATRIX				DERIVED SIGN MATRIX				

Each Design increment (II, III, IV and V) indicates the additional aliases and altered designations of the experimental conditions as each new main effect was substituted for an interaction in the original design. In the column headed: "Defining Contrast", the Defining Generators are marked with an asterisk. These are the original confounding of main and interaction effects. These generators were multiplied together in all possible combinations to determine the remaining words of the Defining Contrasts. How the generators are to be multiplied to produce each word is indicated by the numbers in the column headed: "Generating Products." For example, the numbers 1, 2, 3, and 4 are associated with the Defining Generators ABCD, BCE, ACF, and ABG respectively. When 1 and 2, i. e. (ABCD)(BCE), are multiplied together, as shown in Table [IV-4], the product forms a new word for

the Defining Contrast, ADE. Another example can be found with 234 near the top of the Generating Products column. This means that the corresponding word in the Defining Contrast column, EFG, was formed by multiplying the defining generators, 1, 2, and 3, or $(BCE)(ACF)(ABG) = (ABC^2EF)(ABG) = (ABEF)(ABG) = (A^2B^2EFG) = EFG$. There is a new word for all possible combination of the four generators (four things two at a time plus four things three at a time plus four things four at a time). For this 2^{7-4} design, the Defining Contrast, with an Identity factor, I, four Defining Generators, plus eleven new words, becomes:

$$I = ABCD = BCE = ADE = ACF = BDF = ABEF = CDEF = ABG \\ = CDG = ACEG = BDEG = BCFG = ADFG = EFG = ABCDEFG.$$

From this the other aliased effects can be generated in the manner described in Chapter III. This generation has already been done in the Basic Design, Table [IV-4], and the aliases of each of the seven independent effects are listed in the same columns. It should not be forgotten that although we write the aliases as $A = BCD = ABCE = DE$ and so forth, actually any single estimated effect is the combined effect of all of its aliases, e. g. $A + BCD + ABCE + DE$ and so forth.

In this saturated design, each main effect is associated with

- No main effects
- 3 two-factor interactions
- 4 three-factor interactions
- 4 four-factor interactions
- 3 five-factor interactions
- 1 six-factor interaction.

Each independent effect has its unique set of aliases. The entire matrix of aliases is totally dependent upon the Defining Contrasts which were in turn determined by the particular interaction with which each main effect was originally confounded.

Step 5. Facing the realities about the assumption of negligible interactions.

The likelihood of higher-order interactions having any appreciable effect has already been discussed. Without attempting to draw a fine line at the moment between what is or isn't a higher-order interaction, the evidence suggests that one can generally feel quite comfortable ignoring four-factor interactions or higher and quite uncomfortable ignoring two-factor interactions. If we tentatively assume that we won't be concerned with three-factor interactions either — this will be checked later — then the particular aliases of greatest concern in this 2^{7-4} saturated design are:

$$\text{Effect 1} = A + BG + CF + DE$$

$$\text{Effect 2} = B + AG + CE + DF$$

$$\text{Effect 3} = C + AF + BE + DG$$

$$\text{Effect 4} = D + AE + BF + CG$$

$$\text{Effect 5} = E + AD + BC + FG$$

$$\text{Effect 6} = F + AC + BD + EG$$

$$\text{Effect 7} = G + AB + EF + CD$$

In the Basic Design, Table [IV-4], the main and two-factor interaction effects have been printed in bold-face type to make them more visible.

Variations of the Basic Saturated Designs

The basic saturated design might be modified under the following circumstances:

- 1) When the number of effects to be investigated with N treatments is less than N-1.
- 2) When the Basic design is to be blocked.
- 3) When unplanned-for information is "discovered."

When There Are Fewer Than N-1 Factors

Basic designs were discussed as if there would always be saturation, that is that there would be N-1 factors for N treatments. However since the N number of

treatments are limited to some power of two, there will be times when the number of factors might be less than $N-1$. Since it is still possible to estimate $N-1$ effects, how might the extra available effects be utilized?

Interaction Effects. In the case where N equals eight and the number of factors are less than seven, the seven independent estimates could be:

6 factors and one two-factor interaction

5 factors and the interaction of one factor with each of two others

4 factors and all two-factor interactions between any three

3 factors and all interactions between them.

What has been described, of course, are the situations that existed in the build-up from a 2^3 factorial to the Basic design – but in reverse.

When a particular interaction effect is of interest, then care must be taken to see that the sign matrix fits the need. For example, if there had only been six variables and the investigator wished to estimate the effect of the AC interaction unconfounded with any main effect, then in developing the Basic design, the factor F could not have been substituted for the interaction AC. Instead it might have been substituted for the interaction BC (since no factor G is being used) leaving the AC interaction independent of any main effects. This would change the defining relations accordingly as well as which effects are aliased.

Estimating Error. In saturated designs, no estimate of experimental error is possible when the degrees of freedom are used up estimating the main and interaction effects. If however there are fewer than $N-1$ identifiable effects or if the remaining interaction effects are in fact unimportant, then the extra observations could provide some estimate of error.

When the Basic Design is Blocked

When the number of observations in the Basic design is large, the investigator may wish to block. In human factors engineering research, if the same individual is tested sequentially over a number of conditions, irrelevant sources of variances tend to creep in and distort the estimates of interest. Both the individual and the

equipment can vary as a result of factors created artificially by the experimental situation. When the number of observations exceed ten or so, the possibility of blocking the design should be seriously considered. The principles of blocking and its advantages for reducing irrelevant sources of variance are discussed in considerable detail elsewhere (16)(23), and specifically for human factors engineering research by Simon (42).

The purpose of blocking is to separate the experimental conditions into blocks in such a way that if an average performance difference exists between these blocks the effects of greatest interest will not be distorted. To achieve this more precise measure of the effects within blocks, however, the investigator must sacrifice the precision of those effects confounded with the effects of blocks. Presumably the investigator selects those effects to be sacrificed from among the ones in which he is least interested, or the ones which would be so obviously large that a precise estimate is not needed. Of course, if all of the columns in the design are not used (that is, there are fewer than $N-1$ main and interaction effects of interest), one of the extra columns could be used for blocking within the Basic design.

The blocking of the design into two parts is accomplished by assigning all experimental conditions with a plus sign in the column to be used for blocking into one block and all with a minus sign in that column into the second block. Although the effect of differences between blocks is totally confounded with whatever effect may have been measured by that column (including all of the aliased effects), the effects of blocking will still be independent of the remaining effects in other columns.

When Unplanned-for Information is "Discovered"

As the number of variables in a saturated design increase, the probability of getting negligible effects also increases. When this is so, some special advantages occur.*

*Results of a basic saturated design must be interpreted with caution insofar as the negligible results are concerned. Since main effects in a saturated design are aliased with two-factor interaction effects and the measured value is the sum of all the aliases, the failure to discover an effect could conceivably be due to a large positive main effect combined with a large negative interaction. This will be discovered when the Basic design is augmented.

"Discovering" Error Estimates. As in any experiment, whether involving a saturated or factorial design, if an effect is found to be negligible, the high and low levels of that factor can be treated as replications in the experimental design and the data collected at each level can be used to obtain an estimate of error. Although it may not be known in advance which factors will have a negligible effect, as the number being studied exceed 10 or so, there is a high probability that some of them will. Although no plan for error is included in the original design, discovered negligible effects serve as the source for estimating error. Any bias that might exist in this error estimate will be upward, leading to a more conservative test of significance.

"Discovering" a Factorial Design. When a large number of factors are being investigated and the effects of some are negligible, there are circumstances in which the original saturated design reverts to a factorial design for the important variables. This means that an investigator can have his cake (study N variables) and eat it too (estimate all factorial effects if only two are important).

The concept of the resolution of a design was discussed in Chapter III. Saturated (Basic) designs are designs of Resolution III, since no main effect is confounded with any other main effect, but main effects are confounded with two-factor interactions and two-factor interactions are confounded with each other. The example in [III-4] with seven factors and eight observations is a 2^{7-4} design of Resolution III.

To determine how many factorial designs can be found within the saturated design, the general rule to apply is:

A design of resolution R will provide a complete factorial in any sub-set of the (R-1) variables. (10, p. 342)

For Resolution III designs, complete factorials are possible for any sub-set of two factors out of the total N-1 variables. For example, in the seven-factor saturated design, if the effects of any - we do not need to know which ahead of time - pair of factors prove to be important and the remaining are not, then the original saturated design already provides the data needed to estimate the effects of the two factors and their interaction. A second, more inclusive rule is:

If a design of resolution R is used to screen sub-sets of R factors, then full factorials will result for certain sub-sets and fractional factorials for others. (10, p. 343)

Fractional factorials would occur for any sets of three effects if they were a word in the defining contrasts. Thus, for the Basic Design, Table [IV-4], a defining generator was $I = ABG$. If the sub-set of factors A, B, and G were discovered to be the three important ones, only fractional factorial projections are possible. If the three important factors were not all in one of the words in the defining contrasts, for example, factors A, B, and C, then the complete factorial could be projected. When the Basic design is combined with certain augmented designs (e. g., A. D. 2 discussed later in this section), the combined design becomes one of Resolution IV. Therefore, although we don't know ahead of time which sub-set of factors will turn out to be important, whichever does, the design still provides the capacity to examine a complete or a fractional factorial for four of those variables.

While it is unlikely in human factors experiments that only three or four factors out of a great many being screened will be the only important ones and the remainder will be negligible, the reader should be aware that there will be circumstances in which the original screening design may provide estimates of the error variances and the data needed to estimate the effects of some complete or fractional factorials. However, Box and Hunter (10, p. 343) warn: "Evidence from experiments of this kind should only be regarded as suggestive and subject to confirmation rather than supplying definite proof."

Saturated Designs when the Number of Conditions is a Multiple of Four

The technique used by Box and Hunter to develop saturated designs can be used when the number of experimental conditions is a multiple of some power of 2, e. g., 8, 16, 32, 64, 128, etc. Plackett and Burman (40) developed saturated designs which enable independent estimates of up to $N-1$ factors each of two levels with N experimental conditions where N is a multiple of 4, e. g., 8, 12, 16, 20, 24, 28, 32, 36, etc. up to and including 100 with the exception of 92. For those designs where N is a power of 2, Plackett and Burman's (P-B) designs are the same as Box and Hunter's (B-H) designs.

Two differences between P-B and B-H designs which can affect how the designs are used and interpreted are:

- 1) When it occurs, the degree of confounding between main and interaction effects in P-B designs is less than in B-H designs.
- 2) P-B designs provide an investigator with more opportunities to control the degree of experimental precision than do B-H designs.

Confounding. In B-H saturated designs, main and interaction effects are totally confounded. This means that with fractional factorial designs of two levels, the calculation for the main effect would be identical with that for the aliased interaction effect. This confounding was discussed in Chapter III, and should be obvious in the B-H designs since they were constructed by equating, for example, factor E, with interaction AD.

With P-B designs in which the number of observations are not some power of two, the degree of confounding can be less than 100 percent. Tukey (48) calculated (as indexes to the degree of confounding) such values as 0.11, 0.16, 0.11, and 0.10 for P-B designs with N's of 12, 20, 24, and 25 respectively. Since 1.00 represents total confounding, Tukey concluded: "If simple two factor interactions concern you in an experiment, Plackett-Burman patterns are unusually attractive." (48, p. 171) If one intends to estimate only main effects and run only a Basic design, then those designs in which the main effects and interactions are least correlated would be expected to give the least biased estimate of the main effects.

Precision. Precision in saturated designs (2^{k-p}) of Resolution III is proportional to the square root of the number of experimental conditions. The more experimental conditions measured, the greater the precision. Increasing precision increases the power of the tests of statistical significance and increases the accuracy of the estimation. If an investigator wishes to estimate an effect approximately five times more precisely than a single experimental condition can be measured, he must select a design where the $N=25$. The nearest P-B design requires 24 observations and would be suitable; the nearest B-H design would have been for $N=32$, providing more precision than was needed. Both P-B and B-H designs are constructed so that one is able to estimate the effects of all factors with equal and maximum precision.

Selecting a P-B Design

Plackett and Burman (40) have provided the data to construct their saturated designs for:

- 1) Up to N-1 two-level factors, where N experimental conditions are for all cases of N/4 up to 100 (with the exception of N=92).
- 2) Up to N-1 three-level factors, for N=9, 27, and 81 experimental conditions.
- 3) Up to N-1 five-level factors, for N=25 and 125 experimental conditions.
- 4) Up to N-1 seven-level factors for N=49 experimental conditions.

Of these, the data to construct all designs involving 32 or fewer experimental conditions is provided in full in Appendix III. Specifically, these are:

<u>Level</u>	<u>Number of Experimental Conditions (N)</u>
2	8 (2^3)
2	12
2	16 (2^4)
2	20
2	24
2	28
2	32 (2^5)
3	9
3	27
5	25

The information needed to construct larger P-B saturated designs can be found in their paper. The designs for N=8, 16, or 32 are equivalent to the B-H designs.

Three and five level designs might be useful when the non-linearity of quantitative factors is expected to be large and the opportunity for much additional follow on work is small. More than likely they would be employed with qualitative factors if that many different classes existed. Any results obtained with this design where main effects are confounded with interactions must be interpreted with caution.

STAGE TWO OF THE SCREENING PROCESS: AUGMENTATION DESIGNS

In human factors engineering research there will ordinarily be little reason to plan a Basic (saturated) design without augmenting it with an additional N observations. With this total of $2N$ observations, $2N-1$ effects can now be independently estimated. Properly selected augmentation designs can, for example, isolate the main effects from a string of two-factor interaction effects, which should not be assumed to be negligible a priori.

Just which augmentation design is best depends on what the analysis of the first block of data shows and what the investigator believes needs isolation to complete the screening process. These augmentation designs, a second block of the complete factorial, also satisfy the requirement of economy and follow the principle of progressive iteration in data collection. In addition, this is the stage at which investigator's judgment begins to play a more critical role.

Except where otherwise indicated, the augmentation designs (A. D.) described below were selected primarily from the papers by Box and Hunter (10)(11). Each design, when combined with the Basic design, serve a specific purpose, as indicated.

A. D. 1. To isolate a single main effect and all its two-factor interactions from the remaining effects, unbiased by any other main effects or two-factor interactions.

If a second set of experimental conditions are added to the Basic design, Table [IV-4], such that the two matrices are identical except that the sign of a single factor is reversed, then the combined design will provide an estimate of the main effect of the switched factor and all associated two-factor interaction effects unbiased by any other main effects or two-factor combinations.

To illustrate this, a second set of eight conditions are developed in which the signs of factor E have been reversed, but the signs of all other factors remain the same. This means that in forming this set of eight experimental conditions, the opposite levels of factor E are used. The conditions for A. D. 1 and those in the Basic design are shown in a new sign matrix of Table [IV-5].

TABLE [IV.5]. BASIC DESIGN AND A.D.1.

EXPERIMENTAL CONDITIONS	PRIMARY SIGN MATRIX	INDEPENDENT FACTORIAL EFFECTS AND ALIASED INTERACTIONS*														
		BG CF DE	AG DF EC	AF DG EB	AE CG FB	AD BC FG	AC BD EG	AB CD EF	1	2	3	4	5	6	7	
BASIC DESIGN	A B C D E F G	A	B	C	D	E	F	G								
	1 efg	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
	2 a de	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
	3 bdf	-	+	-	+	-	+	-	+	-	+	-	+	-	+	
	4 abg	+	-	+	-	+	-	+	-	+	-	+	-	+	-	
	5 cdg	-	+	-	+	-	+	-	+	-	+	-	+	-	+	
	6 acf	+	-	+	-	+	-	+	-	+	-	+	-	+	-	
	7 bce	-	+	-	+	-	+	-	+	-	+	-	+	-	+	
8 abcdefg	+	+	+	+	+	+	+	+	+	+	+	+	+	+		
INDEPENDENT EFFECTS OF BASIC DESIGN		1	2	3	4	5	6	7								
AUGMENTATION DESIGN NO. 1	A B C D E F G	A	B	C	D	E	F	G	AB	BC	CD	DE	EF	FG	ABCD	
	9 fg	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
	10 a d	+	-	-	+	-	-	+	-	-	+	-	-	+	-	
	11 b def	-	+	-	-	+	-	-	+	-	-	+	-	-	+	
	12 ab e g	+	-	+	-	+	-	+	-	+	-	+	-	+	-	
	13 cde g	-	+	-	+	-	+	-	+	-	+	-	+	-	+	
	14 a c ef	+	-	+	-	+	-	+	-	+	-	+	-	+	-	
	15 bc	-	+	-	+	-	+	-	+	-	+	-	+	-	+	
16 abcd fg	+	+	+	+	+	+	+	+	+	+	+	+	+	+		
INDEPENDENT EFFECTS OF BASIC AND A.D.1. COMBINED		8	1	2	3	4	5	6	7	14						

*With sixteen observations, identity column can be used for fifteenth effect if signs are reversed in A.D.1.

* Three-factor and higher order interaction aliases not shown. e.g. D=ABC

Inspection of this new sign matrix shows how the additional conditions make only the sign pattern for factor E and all of its two-factor interactions distinctive from those of the previously aliased main and two-factor interaction effects. All other previously aliased effects are unchanged, still aliased to higher-order interaction effects tentatively assumed to be negligible.

Isolating Aliased Effects

Let us digress for a moment in order to relate what occurred in the sign matrix to what happens with the defining generators and aliased effects when A. D. 1. is combined with the Basic design.

For the Basic design, Table [IV-4], the defining generators were

$$I = ABCD = ABG = ACF = BCE$$

since D had been confounded with ABC, G with AB, F with AC, and E with BC. In A. D. 1., the signs in factor E were reversed. Originally, E had been equated with BC in the 2^3 design, and $I = BCE$. Now, with the sign reversal, -E would be equated with BC (or E with -BC) and

$$I = -BCE.$$

The defining generators for A. D. 1. are therefore:

$$I = ABCD = ABG = ACF = -BCE$$

If we write down only those words in the defining contrasts for I which are composed of no more than three letters, we would have

$$I = ABG = ACF = BCE = CDG = BDF = ADE = EFG \quad (\text{Basic})$$

$$I = ABG = ACF = -BCE = CDG = BDF = -ADE = -EFG \quad (\text{Augmentation})$$

Every word in the augmentation design with an E or a BC in it is now associated with a negative sign.

With the data from either half of the design, either the Basic or the augmentation half, there is no way to independently estimate the size of the effects of A, BG, CF, or DE, regardless of sign. But by combining the two halves, two separate estimates can be made. When the two are subtracted, the effects of the factor with the reversed sign can be estimated. When they are added, the effects of the remaining string of aliased effects can be estimated. Let us look at an example of how this works. To shorten the example, we will look at only the three strings of three two-factor interactions each aliased to A, B, and E. To these we will add some new fictitious performance scores.

The aliases for A in each half of the design are

$$(A + BG + CF + DE) = 7 \text{ (Basic)}$$

$$(A + BG + CF - DE) = -3 \text{ (Augmentation)}$$

Now if we add the two sets of equalities, we would get

$$(2 A + 2 BG + 2 CF) = 4 \quad \text{or} \quad (A + BG + CF) = 2$$

and if we subtract them (changing all of the signs in the lower set), we'd get

$$(2 DE) = 10 \quad \text{or} \quad DE = 5$$

To continue with the strings of aliases associated with the B effect, we would have

$$(B + AG + CE + DF) = -5 \text{ (Basic)}$$

$$(B + AG - CE + DF) = 4 \text{ (Augmentation)}$$

When these two strings are added we get

$$(2 B + 2 AG + 2 DF) = -1 \quad \text{or} \quad (B + AG + DF) = -0.5$$

and when they are subtracted, we get:

$$(2 CE) = -9 \quad \text{or} \quad CE = -4.5$$

Taking just one more effect out of the seven, we determine the aliases for the strings associated with the E effect. These are

$$(E + BC + AD + FG) = 10 \quad (\text{Basic})$$

$$(E - BC - AD - FG) = 2 \quad (\text{Augmentation})$$

which yields $(2 E) = 12$ or $E = 6$, when added, and $2 (BC + AD + FG) = 8$ or $(BC + AD + FG) = 4$, when subtracted.

Thus it can be seen that by reversing the signs in only the E column of the basic matrix in Table [IV-4] (making $-E = BC$), we have been able to isolate the main effect E and its interactions AD and BE. Had we continued, the other two-factor interactions for E would also have been isolated from the other main effects and two-factor interactions. With the 16 experimental conditions, 15 effects can be independently isolated. In this example, they are the one main effect, E, each of its six interactions with factors A, B, C, D, F, and G, and the seven strings of still aliased two-factor interactions. All of these, however, still are aliased with higher-order interactions. The fifteenth effect in this case is between blocks.

A.D.2 To isolate all main effects from all two-factor interactions, leaving the two-factor interactions still aliased among themselves.

The sign matrix for this augmentation design is created by reversing the sign of every factor (but not the Identity) in the Basic design. The new experimental conditions of the augmentation design are formed by combining the opposite levels of each factor used in the original conditions of the Basic design. The combined Basic and augmentation set — now totalling 16 conditions — is shown in Table [IV-6].

With the sixteen observation points, fifteen effects can be isolated. In this design, these will be the seven main effects and the eight independent strings of three two-factor interactions which remain aliased. Whereas the Basic design was

TABLE [IV-6]. BASIC DESIGN AND A.D.2.

EXPERIMENTAL CONDITIONS	PRIMARY SIGN MATRIX	INDEPENDENT FACTORIAL EFFECTS AND ALIASED INTERACTIONS*														
		I	BG CF DE	AG DF EC	AF DG EB	AE CG FB	AD BC FG	AC BD EG	AB CD EF	1	2	3	4	5	6	7
BASIC DESIGN	A B C D E F G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	1 efg	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	2 a de	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-
	3 bdf	+	-	+	-	-	-	-	-	-	-	-	-	-	-	-
	4 ab g	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-
	5 cd g	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-
	6 ac f	+	+	-	+	-	-	-	-	-	-	-	-	-	-	-
	7 bce	+	-	+	+	-	-	-	-	-	-	-	-	-	-	-
8 abcdefg	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
INDEPENDENT EFFECTS OF BASIC DESIGN		1	2	3	4	5	6	7								
AUGMENTATION DESIGN NO. 1	A B C D E F G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	9 abcd	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	10 bc fg	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	11 a c e g	+	-	+	-	-	-	-	-	-	-	-	-	-	-	-
	12 cdef	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
	13 ab ef	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	14 b de g	+	-	+	-	-	-	-	-	-	-	-	-	-	-	-
	15 a d fg	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
16 (1)	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
INDEPENDENT EFFECTS OF BASIC AND A.D.2. COMBINED		8	1	2	3	4	5	6	7	13	14	15	16	17	18	

*With sixteen observations, identity column can be used for fifteenth effect if signs are reversed in A.D.2.

**Three-factor and higher order interaction aliases not shown. e.g. D=ABC

of Resolution III, this combined design is of Resolution IV. Higher-order interaction effects are still aliased with these fifteen independent effects.

Investigator Logic

A. D. 2. does not identify which of the three two-factor interactions within a string are important. Ordinarily this can be done only by collecting more data. However the amount to be collected can be reduced if the investigator's analytic ability is used to narrow down the possibilities. Youden (52) suggests that the effects of each fraction - the Basic and A. D. 2 - should each be calculated separately as well as combined. A study of the sign patterns of the two sets of data may give a clue as to which interactions are critical.

For example, if the separate estimates of a main effect are both substantial and of the same sign, this supports the conclusion that a main effect is present. If the separate estimates for any given main effect are substantial but of opposite sign, this is equally good evidence that at least one of the two-factor interactions confounded with the main effect is not negligible. By way of illustration, two new sets of fictitious data are shown in Table [IV-7] representing the effects estimated from the data of the Basic and the A. D. 2 designs. To simplify the discussions that follow, a string of effects will be designated by the factor name of the single main effect in the string whenever possible.

Table [IV-7]. Two Sets of Performance Data for Seven Effects

Effects	Aliases	Block I (Basic)	Block II (A.D.2.)
1	A, DE, CF, BG	+12 ←	+9 ←
2	B, AG, CE, DF	+1	+3
3	C, BE, AF, DG	+7 ←	+8 ←
4	D, BF, AE, CG	+2	+1
5	E, BC, AD, FG	+0	+2
6	F, EG, AC, BD	-10 ←	+14 ←
7	G, AB, EF, CD	+3	+2

In Table [IV-7], factors A and C both show large positive effects and factor F shows one large positive and one large negative effect in each set. One might therefore suspect that it is not factor F that is important but one or more of the three interactions, AC, BD, or EG, that are aliased with F. Critical interactions often include at least one of the critical main effects, so one would suspect that Interaction AC is the critical one.

A. D. 3. To help the investigator analytically identify critical main and two-way interaction effects.

Up to this point, augmenting the Basic design enabled the effects of specific main and interaction effects to be isolated. In the case of the A. D. 2. designs, all main effects were isolated from all two-way interaction effects, but strings of two-way interactions remained aliased; additional data must be taken to separate their effects. As an alternative, Youden (52) suggests that instead of separating the main from the two-way interaction effects, an augmentation design could be chosen that would alias the main effects with a uniquely different set of two-factor interactions from those aliased in the Basic design. Properly done, this would permit the investigator to analyze the results and very often detect which main and which two-factor effects in an aliased string are the critical ones, without having to collect more data.

This augmentation design would be created from an entirely different fraction (i. e., a different "family") from the Basic design and the levels of this fraction reversed. A fraction from a different family is obtained by simply repeating the procedure used to develop the original Basic design, Table [IV-4], but aliasing different sets of interactions with each main effect. Thus the Basic design was created from the three-variable factorial design by aliasing $D=ABC$, $E=BC$, $F=AC$, and $G=AB$. An augmentation design from a different family could be developed, for example, by aliasing $D=BC$, $E=AC$, $F=AB$, and $G=ABC$. Beware of merely reversing two aliases. For example, if $E=BC$ and $F=AC$ were used in the Basic, don't use $F=BC$ and $E=AC$.

The defining generators for a Basic design for a new family are:

$$I = ABF = ACE = BCD = ABCG$$

and the defining contrast would be:

$$\begin{aligned} I &= ABF = ACE = BCD = ABCG = BCEF = ACDF = CFG \\ &= ABDE = BEG = ADG = DEF = BDFG = CDEG \\ &= AEF G = ABCDEFG \end{aligned}$$

For A. D. 3. however, the reverse levels of this second fraction should be used along with the Basic design.

To illustrate how an analysis might be made, fictitious performance scores are associated with the aliased main and two-factor interactions in the Basic (Table [IV-4]) and in the A. D. 3. designs separately. These are shown in Table [IV-8].

Each design enables seven effects to be independently estimated, but each effect is the composite of one main effect, three two-factor interactions, and other higher-order interactions which are tentatively considered to be negligible and are not included in [IV-8].

Table [IV-8]. Performance Data for the Basic and A. D. 3 Designs

<u>Basic</u>		<u>Augmentation (A. D. 3)</u>	
A, BG, CF, DE	1.6	A, BF, CE, DG	.5
B, AG, CE, DF	9.2	B, AF, CD, EG	7.5
C, AF, BE, DG	-11.8	C, AE, BG, FG	-3.3
D, AE, BF, CG	3.0	D, AG, BC, EF	1.8
E, AD, BC, FG	7.0	E, AC, BG, DF	5.8
F, AC, BD, EF	3.9	F, AB, CG, DE	1.5
G, AB, CD, EF	2.3	G, AD, BE, CF	7.4

8 Youden (52) explains how an investigator can use the information from these two designs together with his analytical ability to determine which main and two-factor interaction effects are the important ones. To do this, the largest effects found in Table [IV-8] were extracted and shown here along with their signs:

Basic	Augmentation
B, AG, CE, DF +	B, AF, CD, EG +
C, AF, BE, DG -	E, AC, BG, DF +
E, AD, BC, FG +	G, AD, BE, CF +

8 For any main effect to be important, it must be found in these larger effects and the sign for both sets must be the same. This holds for B and E in this example. For any interaction effect to be important, it must be found in these larger effects and the signs for both sets must be different. In the strings beginning with factor C in the first set and factor G in the second, the signs are opposite suggesting an important interaction; BE is common to both strings. On the other hand, although AD is found in a string in each set, the signs for both strings were the same, which does not suggest the presence of an important interaction.

Youden discusses more complex examples and patterns of effects and shows how these can be logically interpreted. In certain cases, where several critical sources of variance exist within an aliased set, some may have positive and some negative effects with the result that they cancel each other. Thus the analysis of patterns may not always be directed toward large effects.

8 Youden warns: "The partial confounding of main effects with fractional replicates does not give something for nothing nor does it solve all the problems of the experimenter. . . . half the information is lost by the partial confounding but the interaction has been identified. The experimenter must choose what he wants. This, indeed, is the real art of experimental design." (52, p. 358) As the number of variables increases, the difficulties in making these logical interpretations increase. However there is nothing to stop the experimenter from collecting more

data. Other fractional designs of both the same of different families would help clarify the situation. Even if this were done several more times, the economy achieved by not doing the complete factorial is still impressive.

Using the Identity (I) Column to Measure Block Effects

If the Identity (I) column is otherwise unused, it can be used to determine whether or not irrelevant sources of variance have crept into the experimental results to create an average shift in performance between the time the Basic and the Augmentation Designs were run. To do this, the signs of the Identity column of the Augmentation Design would be reversed from those in the Basic design. If any change in performance has an equal effect across all of the conditions in the block, the estimates of other 15 effects of interest will be unaffected, since the two designs are orthogonal to one another.

A. D. 4 To add a new factor to the study.

If the experimenter can be assured that no extraneous sources of variance will affect performance differentially between the two blocks, he can use the Identity (I) column of the Augmentation Design to collect information about a new factor. Of course, all of the original aliases of the (I) column will be aliased with the new factor.

The (I) column, rather than being used to estimate the mean, or measure differences between the two blocks, can be used to add an additional factor to the study. In so doing, all of one level of the new factor will be run during the first block (i. e. the Basic design) and all of the other level will be run during the second block (i. e. Augmentation design). This change in factor level automatically reverses the sign of the Identity column between the two parts.

A. D. 5. To obtain unbiased estimates of all main and interaction effects among any three factors if the remaining factors are of no importance.

An augmentation design of the "fold-over" type (e. g. A. D. 2) in combination with the basic, saturated design, results in a fractional factorial of Resolution IV. In the discussion on saturated designs, it was noted that a design of resolution R will provide a complete factorial in any sub-set of the (R-1) factors. This means

that with a Resolution IV design, if only three out of all possible main effects are sizeable and the others negligible, it would be possible to estimate the effects of the three two-factor interactions and the one three-factor interaction for the three factors. It is not necessary to know in advance which three factors will be the important ones.

The main effects and all interactions can also be obtained for any four factors in a Resolution IV design (i. e. the Basic design plus a fold-over design) provided the remaining effects are negligible and provided the four factors are not one of the four-factor interactions in the defining contrast.

STAGE THREE OF THE SCREENING PROCESS: ISOLATION DESIGNS

Let us first review the experimental sequence up to this point. At least N experimental conditions are used in the basic (saturated) designs to estimate $(N - 1)$ effects. However, with these designs, main effects are aliased with two-factor interaction and higher effects. If N additional experimental conditions are collected (augmentation designs), it is possible with certain designs to isolate completely main effects from two-factor interactions. However, strings of two-factor interactions will still be aliased with one another. With other designs, while never completely isolating main and interaction effects, certain critical interactions can be logically identified.

Situations will arise when even supplementing the early efforts with an additional augmentation design will not provide enough information to positively isolate the critical effects. This will more likely be the case when larger numbers of factors are being studied. Consistent with the needs for economy, the final isolation of aliased effects - and these may be two-factor or even a three-factor interaction that is suspected of not being negligible - may be accomplished without collecting an entirely new block of data. Daniel (20) suggests some plans in which only a few additional, properly selected experimental conditions can be examined to identify specific aliased effects which are probably important.

I. D. 1. To separate a single pair of two-factor interactions with one extra condition.

Suppose the experimental conditions of the Basic and A. D. 2 designs have been run and the analysis shows that the effects of factors A and B and the string of the two-factor interactions (AB+CD+EF) are much larger than the other effects. The investigator would guess that the large interaction effect was probably due to the AB interaction. Such a guess may need further confirmation but it is a good beginning. On the other hand, if the analysis had shown the largest effects to be factors A, B, and the string of two-factor interactions (AF+BE+DG) with the remaining effects being small, there is a question as to which interaction (or interactions) of the string is responsible for the large effect. Daniels (20, p. 413) proposes the following approach which combines experimenter analytic skills with experimental design.

The investigator might assume that the effect of interaction DG is negligible since neither factor D nor G show large effects. If DG is negligible then the investigator knows that the effect is a measure for the combined (AF+BE). In order to determine which of these is the responsible one, these effects must be separated. This can be done by adding performance data from an additional experimental condition which will measure the effect of (AF - BE).

How can one determine the experimental condition that will represent among other things the confounded effect of the difference between the AF and BE interactions? The easiest way is to make use of a sign matrix.

In the same way that the size of an experimental effect is estimated by combining the mean performances on each experimental condition, the performance of an experimental condition can be estimated by combining the experimental effects. In other words, a sign matrix can be used in both directions.

To find an experimental condition that will include among several effects one that represents (AF - BE), the following steps are taken:

- 1) Write down only the interactions of interest and all main effects associated with them. Also include the mean, as represented by the Identity (I).
- 2) Put those main effects considered to be negligible in parentheses.

Following these two steps for this example would produce:

I A B (E) (F) AF BE

Next it is necessary to decide on the sign of the arithmetic operations required to combine these effects, i. e., whether to add or subtract them. Signs for this purpose are assigned as follows:

- 3) To include the effect (AF - BE), AF and BE must have opposite signs. In this example, we will use +AF and -BE.
- 4) A minus sign is arbitrarily assigned to all negligible effects.
- 5) The Identity factor, (I), is positive.

These steps result in the following pattern:

+I A B -(E) -(F) +AF -BE

Only the signs for A and B have not been designated. These are determined from the signs already indicated. For example, since the product of the signs of A and F must result in the positive sign assigned to AF, and since F is already negative, then A must also be minus. Similarly, since the product of the signs B and E must result in the negative sign assigned to BE, and since E is negative, then B must be positive. The completed pattern would be:

"Isolation"
condition = +I -A +B -(E) -(F) +AF -BE
number 1

The experimental condition that this pattern of effects represents is b. This was determined by noting those main effects with plus signs (e. g. involving the high level of each factor), and as has been the procedure in the past, naming the experimental condition by the combination of letters representing those effects. There are of course a great many other effects that have been ignored up to this point; however, if this is experimental condition b, then the level of all of the other main

effects must be low, implying a minus sign. In summary, performance for experimental condition b could be estimated by combining the effects of the critical factors and interactions as follows:

$$\underline{b} = I - A + B - (E) - (F) + (AF - BE) + \text{the remaining effects that are negligible}$$

By omitting all negligible effects, that equation can be shortened to:

$$\underline{b} = (I) - A + B + (AF - BE).$$

However, we do not know the effect of (AF - BE), and the performance value for experimental condition b will be obtained empirically. Therefore, the equation will be rewritten as follows:

$$(AF - BE) = \underline{b} - (I) + A - B.$$

The effect of (AF - BE) can therefore be estimated by arithmetically combining the performance value obtained by running one or more subjects on experimental condition b. The effects of A and B, and the mean (I), can be obtained from the results that have already been estimated from the data from the Basic and Augmentation designs (A. D. 2.) combined.

Once the value of (AF - BE) is obtained, it can be combined with the value of (AF + BE) obtained from the earlier data collection to isolate the effects AF and BE as follows:

$$(AF + BE) + (AF - BE) = 2 AF$$

$$(AF + BE) - (AF - BE) = 2 BE$$

Dividing each of these results by two yields the individual effects of each interaction.

Obtaining other conditions. Experimental condition b is not the only one that would enable the two effects to be isolated. Other conditions could be obtained by reversing the signs assigned to AF and BE, or changing the arbitrary signs of the

negligible conditions, E and F, which in turn would change the signs of A and B. Additional experimental conditions, formed by a reassignment of the signs while following the above procedures, are shown in Table [IV-9].

If two experimental conditions are to be added instead of one, they should be selected from those cases where A and B have the same respective signs for each condition but the signs assigned to the interactions are reversed. This would pair: abe and abf, a and aef, and b and bef. Two or four experimental conditions might be used to provide additional precision. Admittedly, the use of so few conditions is fraught with danger, but the purpose is one of identification where making precise estimates is not as critical as determining relative strengths.

I. D. 2 To separate four members of a single string of two-factor interactions with three extra experimental conditions.

Daniels (20, p. 414) uses the same principles for this separation as for I. D. 1., however, in this case it is slightly more elaborate. Let us imagine that as a result

Table [IV-9]. Other Experimental Conditions that Might be Used to Isolate the Effects of (AF+BE)

		EFFECTS							
		FIXED	DETERMINED		ARBITRARY		GIVEN		
		I	A	B	(E)	(F)	AF		BE
EXPERIMENTAL CONDITIONS	2)	+	-	+	+	+	-	+	bef
	3)	+	+	+	+	-	-	+	abe
	4)	+	+	+	-	+	+	-	abf
	5)	+	+	-	-	-	-	+	a
	6)	+	+	-	+	+	+	-	aef

of combining the data from the Basic and A. D. 2. designs it was found that the factors A, C, E, and G are the important ones along with the effect of a string, (AB + CD + EF + GH). In this case, there is no obvious rationale for reducing the number of two-factor interactions as was done in I. D. 1. since one term in each is an important one. With four aliased interactions, at least three additional experimental conditions will be required and their signs should be orthogonal among the strings, thus:

		<u>Effects</u>			
		<u>AB</u>	<u>CD</u>	<u>EF</u>	<u>GH</u>
Experimental Conditions	1)	+	+	-	-
	2)	+	-	+	-
	3)	+	-	-	+

These conditions are orthogonal since the products of the signs for any pair of rows yield an equal number of plus and minus signs. Next, the "minor" factors will be arbitrarily assigned a minus sign. These are the unimportant factors in the interactions, in this case, B, D, F, and H. Next, the signs for the critical factors are also orthogonalized, and along with the minor factors must fit the interaction signs. For example, the product of the sign assigned to A and the minus sign of B (not shown) must yield a plus for AB in the first condition. Thus A has to have a minus sign. With all minor variables having minus signs, the signs of the major variables, based on the interaction sign matrix above, must be as follows:

		<u>Effects</u>			
		<u>A</u>	<u>C</u>	<u>E</u>	<u>G</u>
Experimental Conditions	1)	-	-	+	+
	2)	-	+	-	+
	3)	-	+	+	-

When these two matrices are combined, the three new experimental conditions would be defined by:

<u>Effects</u>									<u>Conditions</u>
<u>I</u>	<u>A</u>	<u>C</u>	<u>E</u>	<u>G</u>	<u>AB</u>	<u>CD</u>	<u>EF</u>	<u>GH</u>	
(+	-	-	+	+	+	+	-	-)	eg
(+	-	+	-	+	+	-	+	-)	cg
(+	-	+	+	-	+	-	-	+	ce

The designation of each condition is obtained by assigning a letter for each main effect in which the high level is used, i. e. the one with the + sign. The effects considered negligible are not shown. Had all effects been shown, the matrix would show an orthogonal pattern.

The performance level for condition eg, for example, could be estimated if all of the effects in the above matrix were known. If the data from the Basic and augmentation design, A. D. 2. were available we would know the effects of I, A, C, E, and G, but not that of the individual effects of the four interactions; up to this point, we only know the effect of their combined sum, (AB + CD + EG + GH). We do not know the effect of the combination, (AB + CD - EF - GH) which would be needed to estimate performance under condition eg.

But we are not estimating the performance for conditions eg, or cg, or ce. The reason for identifying the three new conditions was to be able to run subjects on them and determine the actual performance scores for each. These scores can be combined with the effects already known in a way that will permit the effects of the two-factor interactions to be isolated. Just how this is accomplished is best understood by looking at the sign matrix above combined with the effect of the string of two-factor interactions, as shown here:

<u>Conditions</u>	<u>Source from which value is obtained</u>	<u>Effects</u>								<u>Fictitious Performance</u>	
		<u>I</u>	<u>A</u>	<u>C</u>	<u>E</u>	<u>G</u>	<u>AB</u>	<u>CD</u>	<u>EF</u>		<u>GH</u>
(AB+CD+EF+GH)	Basic + A. D. 2.	(+	+	+	+	+					+3
(eg)	Tested subjects	(+	-	-	+	+	+	+	-	-)	+5
(cg)	Tested subjects	(+	-	+	-	+	+	-	+	-)	-2
(ce)	Tested subjects	(+	-	+	+	-	+	-	-	+	+6

These four sources along with their corresponding performance values should be summed together. This would yield:

$$3I - 3A + 1C + 1E + 1G + 4AB = +12$$

The effects of CD, EF, and GH have been cancelled out. Since the mean (I), A, C, E, and G would be known from the estimates already obtained from the data of the Basic and A. D. 2. designs, by proper arithmetic substitution and simplification, the effect of AB can be determined.

Interaction CD can be obtained in the same way. This time the four sources are combined by subtracting (cg) and (ce) from (AB+CD+EF+GH) and (eg). This causes the signs of all components of (cg) and (ce) to be reversed, of course, and when the four sources are now summed, all of the interactions except CD will cancel out. The remnants of I, A, C, E, and G will be eliminated as before by substituting the appropriate values already obtained from completing the Basic and Augmentation designs.

To isolate the effect of EF, (eg) and (ce) must be subtracted from (AB+CD+EF+GH) and (cg). To isolate the effect of GH, (eg) and (cg) must be subtracted from (AB+CD+EF+GH) and (ce).

I. D. 3. To separate members of a string of three-factor interactions.

No examples will be given, but it is apparent that the same logical approach can be applied to any set of confounded data. In each case, the following steps would be required:

- 1) To reduce the effort the experimenter can first try to logically eliminate certain of the aliased effects.
- 2) At least (N - 1) additional experimental conditions must be used for N aliased effects in a string.
- 3) The rows of signs of the aliased effects to be isolated must be made orthogonal to one another.

- 4) Minor factors in the aliased effects can be arbitrarily given any sign.
- 5) The signs of the major factors, when combined with those of the minor factors, must be made to correspond to the sign pattern of the aliased interactions according to rules for multiplying signs.
- 6) Experimental conditions are identified by those factors with plus signs associated with them.
- 7) Combining effects to isolate the desired ones requires that all lower-order effects already be determined.

When only a few data points are used that are not orthogonally blocked with the preceding portions of the design, some considerations must be given to order of presentation effects. The values obtained for these new I. D. points, being collected after the other data, may be distorted for reasons totally unrelated to the relevant experimental factors. This may be difficult to ascertain, but can be handled in several ways. For example:

- 1) If it is suspected before the experiment is run or after the first block (Basic) that a particular factor will have a large effect and therefore its interactions with other factors might be of interest, isolation data points might be included in with the Basic and/or augmentation designs.
- 2) In addition to the isolation data points, a number of data points that have already been used in the previous designs might be retested when the isolation points are run to determine whether there has been any block shift.
- 3) The value of the individual point could be estimated with a polynomial derived from the basic and augmentation design data and compared with the empirically derived value obtained when the isolation point is tested.

I. D. 4. To isolate the second-order coefficients of a response surface.

Once the most important factors and two-factor interactions have been identified, the final step of a research program is to describe the function relating

these factors and operator performance. The next chapter discusses various economical designs for obtaining these relationships, referred to as "response surfaces," which represent the levels of performance within a multifactor space. One of those designs — a central-composite design — is constructed around a fractional factorial design. Whereas a 2^{k-P} fractional factorial design can only estimate the linear main effects and linear-by-linear portions of the two-factor interactions, second order (quadratic) effects can be obtained by 1) taking measures at the center of the hypercube represented by a factorial or fractional factorial with all factors at two levels; 2) adding $2K$ more data collection points at the apex of a pattern that passes through the center point and through the center of each face of the hypercube. These are called the "star" portion of a design.

Only a part of the data already collected in the screening study may be used in the construction of the hypercube portion of the response surface design. The actual number of useful conditions depends on which augmentation and isolation designs are employed. Therefore in addition to the center points and the "star" portion of the response surface design, more data must be collected to complete a Resolution V fractional factorial. There will be a savings however over the data that would be required were every point of the entire response surface, central-composite design to be collected.

CHAPTER V.
ECONOMICAL DESIGNS FOR QUANTITATIVE FACTORS

Many factors included in human factors engineering experiments are quantitative and can be represented on a continuous scale. Resolution, signal intensity, vibration, field of view, work load, closure rate, and bits of information are all examples of such quantitative variables. Other factors such as background complexity, pilot/non-pilot subjects, and target types, while often treated as qualitative variables, can be dimensionalized, quantified, and described on a continuous scale.

Quantitative factors allow the results of an experiment to be expressed as a function relating operator performance to equipment, system, and environmental parameters. When truly multifactor experiments are conducted, however, the traditional method of plotting the means for one or two factors at a time — particularly if they interact — is not truly informative. Neither the shape nor the values of the curves of the plotted means can be used operationally without knowledge of the effects of the unplotted factors. If there are interactions between the plotted and unplotted factors, simple plots are even more worthless. What is needed, of course, is a multifactor equation that relates all the factors. However inconvenient this might be for immediate interpretation, it increases the effectiveness of the information.

Experimental designs for determining the function relating performance and predictor variables by means of an approximating polynomial, is called a response surface. While these polynomials do not seek to explain underlying functional mechanisms, they do describe the empirical relationship and can be used to estimate by interpolation* the effects of conditions that were not actually studied in the experiment.

*It is inherently dangerous to use a polynomial to extrapolate beyond the limits of the experimental space.

This class of experimental designs employs a regression model and the choice of the coordinates of the experimental conditions is under the control of the experimenter. This latter feature distinguishes them from the "undesigned" experiments in which a regression model is also employed but in which the coordinates of the data collection points cannot be selected by the experimenter.

CHARACTERISTICS OF RESPONSE SURFACE DESIGNS

Box and Hunter (8) describe the characteristics of experimental designs for fitting response surfaces. A good design should:

- 1) Utilize a grid of data points of minimum density over a multifactor space of greatest practical interest.
- 2) Allow for approximating a polynomial of an order tentatively assumed to be representationally adequate to fit the response surface. When no assumption is made of the form of the function initially, one starts with a first-order polynomial model.
- 3) Check on the adequacy of the function by allowing certain combinations of higher order terms to be examined.
- 4) Permit the already completed design of order d to form the nucleus from which a design of order $(d + 1)$ may be built, if the assumed polynomial proves inadequate.
- 5) Lend itself to blocking which
 - a) helps maintain a steadier experimental environment when an experimental program is extended over many data points and time, and
 - b) permits an experiment to be carried out sequentially, so that certain changes can be made in the experimental plan based on information obtained from the previous data collection period.
- 6) Be "rotatable" so that the orthogonal axes of the experimental design can take any orientation without changing the confidence in the prediction made at any given point.

Economy

Response surface designs embody to the fullest the principles of data collection economy. They are planned so as to minimize redundancy and limit data collection to that which is really necessary. They require the experimenter to be continually involved.

Economy of response surface designs is accomplished, in part, by collecting only enough data to estimate the coefficients of the lowest degree polynomial capable of fitting the empirical data.

Theoretically, a minimum of N data collection points are required to write a polynomial of $N-1$ coefficients (plus the mean). Therefore to write a second degree polynomial (Taylor series expansion) for five factors at least 21 observations are required. This is a much smaller number than the 243 observations required to complete a 3^5 factorial design, or even the 81 observations for a one-third fractional replicate. Even when some additional observations are added to the 21, the savings in data collection is considerable and the loss in information is usually negligible. The coefficients that will be estimated are for the mean (β_0), the linear terms ($\beta_i x_i$), the quadratic terms ($\beta_{ii} x_i^2$), and linear-by-linear cross product terms ($\beta_{ij} x_i x_j$).

Further economy is achieved in many of these designs by employing the principle of sequential progressive iteration. In designs that are orthogonally blocked, the data can be collected a block at a time. If each block corresponds to a degree of the polynomial, it is possible to terminate the experiment as soon as the lowest degree polynomial is found that fits the data.

Applications

Response surface designs provide an economical method of conducting experiments that:

- 1) Search a loosely-defined experimental space to discover the coordinates of that combination of parameters which will optimize operator performance

8

for a particular task. This was the purpose for which response surface methodologies were originally developed. It was assumed that the investigator had little or no knowledge about the response surface and therefore did not know if he was investigating an area near the optimum coordinates. Response surface methodologies represent an economical means of exploring the experimental space to find the optimum.

- 2) Describe the function relating operator performance and equipment, system, and environmental parameters. At the sacrifice of some precision, an overview of a complex world can be obtained. This tying together of diverse components into a quantitative function can also serve as a framework within which additional elements can be added or a data base developed that can later be refined.

8

Human factors engineering experiments are seldom required to search for an optimum response through a loosely-defined experimental space. Ordinarily the boundaries of the experimental space are fairly rigidly defined by customer interests, the state-of-the-art in equipment development or anticipated development, the normal conditions of the real world, and/or the results of preliminary testing by the experimenter. Under these circumstances, responses will be mapped over the entire space. A search approach might be employed, however, if there is reason to suspect that the space is so large that a second - or at most, a third - order polynomial will not approximate the response surface. Ordinarily, the narrower the limits of the experimental space, the more nearly linear the relationship will be.

Most human factors data within a multifactor space, as has already been shown, will not be too non-linear, particularly if the factors are properly scaled to begin with. Humans don't show erratic patterns of behavior in these circumstances. When irregular curves are observed, it can generally be traced to either poor data collection techniques or to a curve that is a composite of several underlying factors operating together (44).

8

Another reason that response surface designs will seldom be used in human factors research to search for an optimum is because engineers ordinarily prefer information in the form of trade-off data. An engineer wants to know what will

happen to performance if he uses a little less expensive component or if he improves one factor and degrades another in order, for example, to reduce the weight or size of the equipment. Finding only the optimum parameters assumes that all factors that greatly affect performance are included in the experiment. They seldom are, and for most human factors research any claim of "optimum" results should be suspect until eight or ten factors have been examined.

Response surface designs were devised originally for chemical research and many of the problems associated with carrying out experiments with human subjects were never considered. How to decide the order in which experimental conditions are tested receives cursory treatment in a few designs, yet it represents a major problem in human factors experiments. The presence of qualitative factors and how to include them economically in the response surface designs are not discussed directly. However, qualitative factors can be treated as "dummy" variables when they are to be included in response surface designs. (24)(47)

Types of Designs

Response surface designs can be classified according to their order, the number of levels per factor, and their symmetry. Of the various designs, the central-composite design – which was described in Box and Wilson's (12) introductory paper – has been described most frequently. Simon (43) discusses its application for human factors engineering experiments. The designs that will be considered here will be restricted to those requiring less than 300 observations for a basic design which can be used to study from five to fifteen variables. Specifically, there will be:

Second-order response surface designs

Central-composite designs

Partial replication of central-composite designs

Response surface designs requiring three levels per factor

Non-symmetrical response surface designs

Third-order response surface designs

Response surface designs for "messy" experimental spaces

CENTRAL-COMPOSITE SECOND ORDER DESIGNS

G. E. P. Box and his co-workers (5)(9)(12) introduced response surface designs in the 1950's along with a philosophy and a methodology of research that make this class of design an embodiment of the principles of economical research. Originally a means of discovering the coordinates of independent factors that optimize the response or yield, response surface methodology has also proved useful for mapping an entire multifactor space relating operator response to equipment parameters. A number of excellent papers have been published that describe the rationale and the mechanisms of designing, analyzing, and using these designs (13)(16)(17)(33), including a review of the literature (32). One paper discusses its applicability for human factors engineering research (43). This approach is particularly useful after the critical factors have been selected by the screening process described in Chapter IV.

Construction

The total N experimental conditions in any k -dimensional central composite designs is

$$N = n_c + n_s + n_o,$$

where

$n_c = 2^{k-p}$, the number of points of the "cube" portion of the design representing a two-level factorial (when $p = 0$) or a $(1/2)^p$ fractional factorial of Resolution V when k is five or more factors. Examples of these, suitable for response surface designs, can be found in Appendix II. The coded coordinates of the cube portion are $(\pm 1, \pm 1, \dots, \pm 1)$.

$n_s = 2k$, the number of points of the "star" portion of the design, a k dimensional analogue of an octahedron having $2k$ vertices. The coded coordinates of the star portion are $(\pm\alpha, 0, \dots, 0)$ $(0, \pm\alpha, 0, \dots, 0)$ \dots $(0, \dots, 0, \pm\alpha)$. The value of α determines whether the design will be orthogonally blocked.

n_o = the number of points at the center of the design, with coded coordinates $(0, 0, \dots, 0)$. When an experiment is blocked, those center points associated with the cube portion are referred to as n_{co} and those associated with the star portion are referred to as n_{so} . The number and distribution of the center points will affect orthogonal blocking, rotatability, the uniformity of variance across the response surface, and the power of the goodness of fit test.

The spatial arrangement of the coordinates of a central-composite design for three factors is shown in Figure [V-1]. The coordinates for the eight vertices of the cube, the six vertices of the octahedron, and one center point are shown. In practice, as will be described below, more measures are made at the center.

Features of Central-Composite Designs

Some important features of these designs are:

- 1) The coefficients of a second degree polynomial of the following form can be estimated:

$$Y = \beta_o x_o + \beta_i x_i + \beta_{ii} x_{ii}^2 + \beta_{ij} x_i x_j$$

The coefficients of the linear (x_i) and linear-by-linear interaction ($x_i x_j$) terms are orthogonal, and their effects can be independently estimated. They are also orthogonal to the coefficients of the quadratic (x_{ii}^2) terms. However, coefficients of the quadratic (x_{ii}^2) terms are not orthogonal to one another or the mean (x_o) and their effects are somewhat intercorrelated.

Box and Hunter (8, p. 174) suggest that it is not appropriate to test individual coefficients for statistical significance, with the intention of dropping those that are not significantly different from zero. Instead, the complete equation is the best description of the immediate data, and if a test is to be made, it should be for the adequacy of the combined terms of the same degree. If there is an interest in the contribution made by a particular factor, then a test should be made of the combined

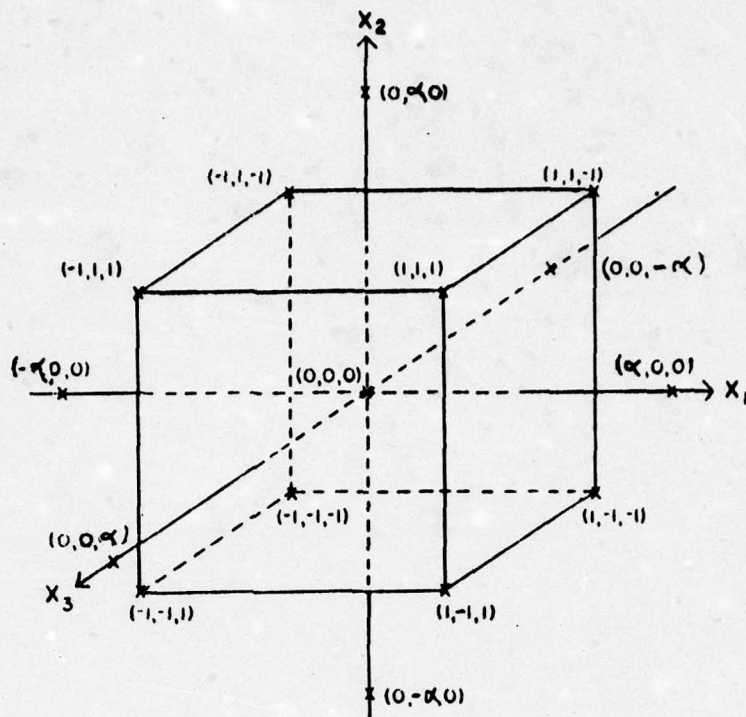


Figure [V-1]. Spatial arrangement of the coordinates of a central-composite design for three factors

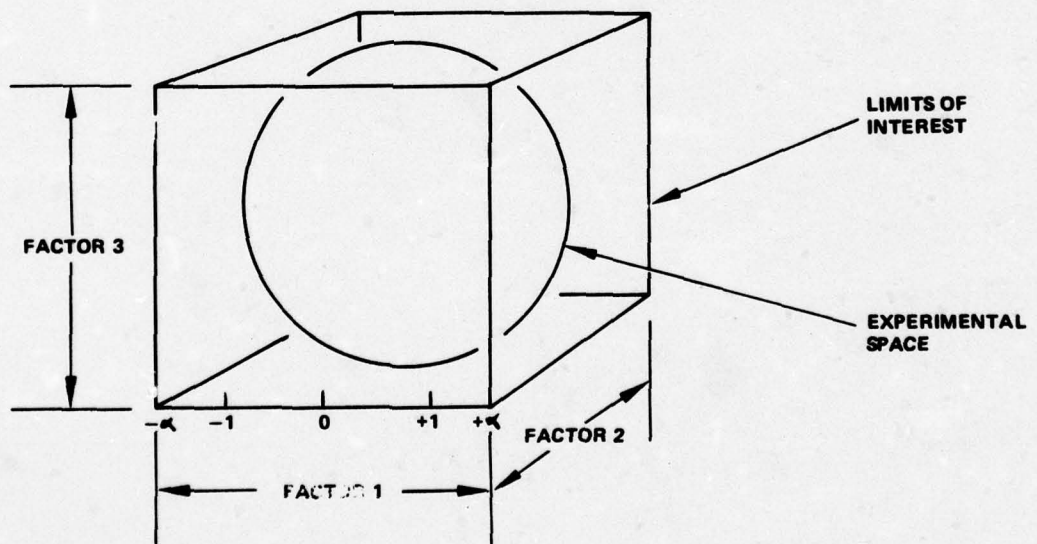


Figure [V-2]. Spherical characteristic of the space covered by a central-composite design

contribution of all terms involving that factor. Of course, the search for important factors should have occurred during a screening period prior to the effort to estimate a response surface.

- 2) The experimental space covered by these designs form a hypersphere. (Figure [V-2].)

Central-composite designs reduce the size of the experiment by not collecting data in the less interesting parts of the experimental space. An experimenter should normally know enough about his problem to be able to localize his experiment around the region of great interest. With central-composite designs, more information is collected at the center of the region with less and less collected the further one moves away from the center. If there is reason to study a corner of the region, auxiliary data points can be added.

- 3) Central-composite designs are constructed so that the "information" is equal for all points equidistant from the center. (Rotatability)

"Information", as the term is used here, is defined as the reciprocal of the variance at any point on the response surface. The feature of rotatability permits the orthogonal axes of the experimental design to be rotated to any orientation without changing the confidence in a prediction made at any given point. For a rotatable design of k factors, the coded coordinate of the length of the arm of the star from the center of the design should equal $\pm 2^{k/4}$; when fractional factorial designs of $(1/2)^p$ are used in place of the hypercube, α should equal $2^{(k-p)/4}$. This is the same as saying that the coded* length of the axial arm from the center ($\pm\alpha$) is equal to the square root of the square root of the number of actual data

* Levels of an experimental factor are coded by standardizing the real world values such that the center point is equal to zero and the point which forms a coordinate of the cube made the standardized unit equal to ± 1 . Any coded value, x_i , equals

$$\frac{X_i - X_{i0}}{U_i}$$

where X stands for a real world value at level i and at the center, i_0 , and U stands for the unit of measurement equal to one standardized unit. The similarity between this and a z-score should be noted.

points in the cube portion of the basic design. This value of α will not always permit the equation for blocking, discussed next, to be satisfied. In that case, α should be adjusted, for an orthogonal design should take precedence over a rotatable design. Usually the difference between the two values, for human factors engineering research, will be of little practical value.

4) Orthogonal blocking enables the techniques of sequential iteration to be employed.

The experimental conditions of the cube and the star portions of the design represent orthogonal blocks within which first-order response surfaces can be estimated. Ordinarily, data is collected on the cube portion of the design and examined to see if a first-order model is adequate. If so, the study ends. If not, data is collected for the star portion of the design against which a second-order model is tested.

Mean differences in performance between these two orthogonal blocks of data will not affect the estimates of the coefficients of the second-degree polynomial. The number of center points assigned to each block and the length of the axial arm ($\pm\alpha$) are important to the orthogonality of the design.

To guarantee orthogonal blocking in the central-composite designs, it is necessary that

$$2^{k-p}/2\alpha^2 = (n_c + n_{co})/(n_s + n_{so})$$

when n_{co} and n_{so} are the number of center points to be added to the cube and the star portions respectively. If the full factorial is used, $p = 0$; if a fractional of $(1/2)^p$ is employed, p takes on that value. Various solutions will satisfy the above equation; however, the total n_o should allow for some replication of center points within at least one block to provide a measure of experimental error (see lack-of-fit test below).

When the number of variables is five or more, the cube portion can be divided into sub-blocks without affecting the coefficients of the main and two-factor interaction effects in the second-degree polynomial. With this additional blocking of the hypercube, the n_{co} center points should be distributed equally among the sub-blocks.

- 5) Central-composite designs provide relatively uniform precision throughout most of the experimental space.

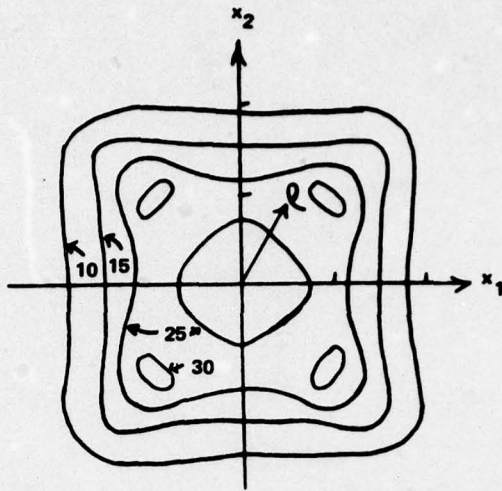
The precision of the response surface (or "information contour") of a conventional 3^2 factorial design is shown in Figure [V-3, A]. With rotatability, this information contour of a two variable central-composite design must look like Figure [V-3, B]. By varying the number of points at the center of the design from one to three, the information profile of Figure [V-3, B] can be changed – as shown in Figure [V-3, C] – to make the precision of information more uniform within the experimental limits, particularly between coded values 0 and ± 1 .

- 6) Central-composite designs provide a test of how well first or second-order models fit the empirical data.

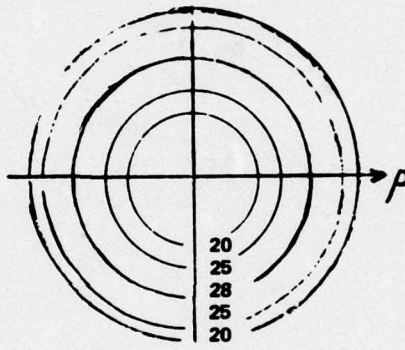
By adding center points to the cube portion of the design, the presence of quadratic effects can be determined, the significance of which can be tested against the estimate of error obtained from the replicated center points. If there is evidence that higher-order effects exist, then data must be taken for the star portion of the central-composite design in order to approximate the response surface with a second degree polynomial. How well this second-order model fits the empirical data can also be tested. If the fit is still inadequate, more data may have to be taken to fit a third-order model. For the central-composite designs, no specific provisions have been made for this last step. Later, some sequential third-order designs will be described, although a large number of data points must be added to approximate a third-order response surface. Since there are only a few degrees of freedom for the error term in a central composite design, the power of the lack-of-fit test is extremely low. While a rejection of the null hypothesis can be considered indicative of the need for a higher-order model, a failure to reject the null hypothesis cannot be accepted without further evidence that the fit is adequate.

Design Parameters

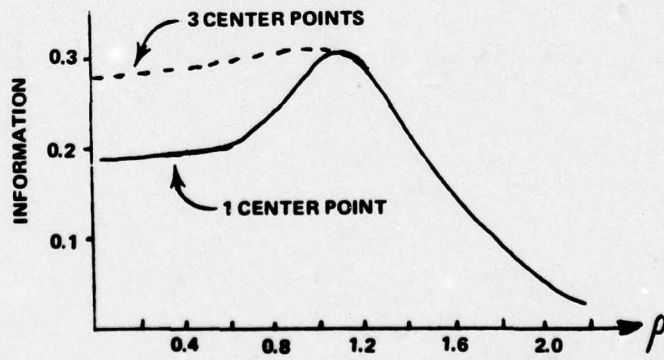
Values needed to construct second order, central composite designs for studying from five to twelve variables are given in Table [V-1]. This includes the



A. Conventional 3^2 factorial



B. Two-factor central-composite design



C. Effect of adding more center points to B

Figure [V-3]. Information contours of experimental designs
 [Adapted from Box and Hunter (8).]

Table [V-1]. Parameters for Designing Orthogonally Blocked, Second-Order Central-Composite Designs

Number of Factors	n_c	n_{co}	n_s	n_{so}	N	α^*	Distribution for Orthogonal Blocking**	
							Cube	Star
5	16	6	10	1	33	2.00	1(16+6)	(10+1)
6	32	8	12	2	54	2.37	2(16+4)	(12+2)
7	64	8	14	4	90	2.83	8(8+1)	(14+4)
8	64	16	16	4	100	2.83	4(16+4)	(16+4)
9	128	8	18	6	160	3.36	8(16+1)	(18+6)
10	128	8	20	4	160	3.36	8(16+1)	(20+4)
11	128	16	22	4	170	3.40	8(16+2)	(22+4)
12	256	8	24	9	297	4.00	8(32+1)	(24+9)

*Length of axial arm of star for orthogonal blocking.

**Given are the number of blocks in the cube portion of the design, the number of cube points in a block, the number of center points for the block containing the cube points; the number of points in the star, the number of center points for the block with the star points. For example, 8(8+1); (14+4) would mean that there are nine blocks altogether, eight of which each contain eight cube points plus one center point and the ninth which contains fourteen star points plus four center points.

number and distribution of the data collection points in the fractional factorial of the cube (n_c), in the star (n_s), and the number of center points distributed to the cube (n_{co}) and the star (n_{so}), and the total number of points (N) in the basic design. In addition, the length of the arm of the star (α) is given for orthogonal blocking. To assure orthogonal blocking, the number of blocks (n_b) into which the cube portion of the design can be divided is shown, followed in parentheses by the number of ($n_c + n_{co}$) points within each block. In the second parentheses are the number of ($n_s + n_{so}$) in the star portion of the design. All of the designs are of Resolution V. To aid in their construction, the defining contrasts for the cube portions are given

in the lists of fractional factorials in Appendix II. Possible modifications of these plans include: 1) Adding more center points in the correct proportions of each block to provide a more powerful test of significance; 2) Reducing by half the number of points in some fractional factorials in the cube portion by allowing a few specific two-way interactions to be confounded with one another.

Partial Replication of Central-Composite Designs

The basic central-composite design provides an estimate of experimental error derived from replicated data points at the center of the design. If it is suspected that variance may not be homogeneous throughout the response surface, it may be necessary to duplicate points in other parts of the experimental space. While such duplication provides more precision in the estimates of the coefficients, more degrees of freedom for an estimate of the coefficients, more degrees of freedom for an estimate of the experimental error, and a more powerful test of the adequacy of the second order model, it also means an increase in the number of runs that must be made.

Dykstra (27) suggests that when non-central replication is desired for the second-order designs, economy can be achieved by duplicating only a portion of the original plan. In his paper, eight types of partially duplicated second-order response surface designs are presented to be used with classic central-composite designs. These designs replicate either the cube (Class 1) or the star (Class 2) portion of the design. If the cube portion of the original central-composite design were a 2^k factorial, the replication may be 2^k , or 2^{k-1} . On the other hand, if the cube portion itself had been only a fractional 2^{k-1} , the replication may be either 2^{k-1} or 2^{k-2} . The $2k$ points of the star are always duplicated. Original plans and partial replicates are selected so as to maintain a Resolution V design in which no main effect nor two-factor interaction will be aliased with any other.

Dykstra's classification scheme for the partially replicated designs is shown in Table [V-2]. The number of data points in the cube and star, the number of center points distributed to the cube or star parts of the design for orthogonal blocking, the values of α for orthogonal blocking, and the total N are given. Only designs for five or more factors are included here.

Table [V-2]. Plans For Partially Replicating Central-Composite Designs

Class	Type	n_c	n_s	Descriptions	Parameters*							Blocks**		
					k	n_c	n_{CO}	n_s	n_{SO}	N	α			
1	I	$2^k + 2^k$	$2k$	Factorial duplicated	1-I	5	64	8	10	8	90	2.828	$8(8 + 1)$ or $4(16 + 2)$; $10 + 8$	
	II	$2^k + 2^{k-1}$	$2k$	Full replication plus half rep.	1-II	5	48	6	10	6	70	2.667	$3(16 + 2)$; $10 + 6$	
	III	$2^{k-1} + 2^{k-1}$	$2k$	Half replication, duplicated	1-III	5	32	8	10	4	54	2.366	$2(16 + 4)$; $10 + 4$	
		"	"	"	"	1-III	5	32	4	10	3	49	2.404	$2(16 + 2)$; $10 + 3$
		"	"	"	"	1-III	6	64	8	12	6	90	2.828	$4(16 + 2)$; $12 + 6$
	IV	$2^{k-1} + 2^{k-1}$	$2k$	Half plus quarter replication	1-IV	7	128	8	14	10	160	3.361	$8(16 + 1)$; $14 + 10$	
		"	"	"	"	1-IV	8	192	12	16	13	233	3.694	$6(32 + 2)$; $16 + 13$
	V	$2^{k-1} + 2^{k-1}$	$2k$	Quarter replication, duplicated	1-V	8	192	12	16	14	234	3.757	$6(32 + 2)$; $16 + 14$	
		"	"	"	"	1-V	8	128	8	16	8	160	3.361	$8(16 + 1)$; $16 + 8$
	2	VI	2^k	$4k$	Full rep., star duplicated	2-VI	5	32	10	20	1	63	2.000	$2(16 + 5)$; $20 + 1$
VII		"	"	"	2-VI	5	32	12	20	2	66	2.000	$4(8 + 3)$; $2(10 + 1)$	
		2^{k-1}	$4k$	Half rep., star duplicated	2-VII	5	16	12	20	0	48	1.600	$16 + 12$; $2(10 + 0)$	
VIII		"	"	"	2-VII	6	32	16	24	0	72	2.000	$2(16 + 8)$; $2(12 + 0)$	
		"	"	"	2-VII	7	64	16	28	0	108	2.366	$8(8 + 2)$; $2(14 + 0)$	
VIII		2^{k-1}	$4k$	Quarter rep., star duplicated	2-VIII	8	64	24	32	0	123	2.412	$4(16 + 6)$; $2(16 + 0)$	
		"	"	"	2-VIII	8	64	24	32	0	123	2.412	$4(16 + 6)$; $2(16 + 0)$	

* k = number of controlled variables

n_c = number of points in cube (factorial portion)

n_{CO} = number of center points for blocks of cube

n_s = number of points in star ($2k$ for class 1, $4k$ for class 2)

n_{SO} = number of center points for blocks of star

N = $n_c + n_{CO} + n_s + n_{SO}$ = total number of points

α = radius arm of star design for orthogonal blocking

** Given are the number of blocks in the cube, the number of center points for the block containing the cube points; the number of points in the star, the number of center points for the block with the star points. Thus $2(4 + 2)$; $6 + 2$ indicates three blocks, two of which contain 4 cube points and 2 center points and the third of which contains 6 star points and 2 center points.

[Adapted from Dykstra (27)]

In his paper, Dykstra supplies the equations needed to determine the distribution of data collection points when larger experimental designs with partial replications are desired. He discusses which designs might be biased were third-order effects not negligible; these are the less-than-Resolution-VI designs. He shows how to calculate the degrees of freedom for the variance associated with the lack-of-fit of the second-order model and the degrees of freedom of the error variance for the partially replicated designs. These partially replicated designs, when compared with the original central-composite designs, will give increased precision in direct proportion to the number of experimental runs. Tests can be made for the heterogeneity of the error variance across the experimental space.

SECOND-ORDER RESPONSE SURFACE DESIGNS WITH THREE-LEVELS PER FACTOR

Box and Behnken (7) recognized that although there exists an infinite choice of levels for any quantitative, continuous independent variable, there may be practical reasons for keeping the number of levels small. Even the five levels required by the central-composite designs may be burdensome in certain applications. They therefore developed a set of response surface designs based on three-level incomplete factorials.

These designs emphasize economy in data collection (e. g. one replicate of a 12 factor design requires 204 data points), allow the coefficients of a second-degree polynomial to be estimated, and provide for a test of the lack-of-fit of the model to the empirical data. Orthogonal blocking is employed whenever possible. The majority of these designs are formed by combining two-level factorial or fractional factorial designs with "incomplete block" designs. Some understanding of how these designs are constructed will aid the reader who may wish to read the original paper to understand how the data should be analyzed. The information supplied here will also be of value in the selection and use of particular designs.

Incomplete Block Designs

An incomplete block design is one in which the experimental conditions are assigned to blocks in a way that eliminates the differences in performance that

arise from the effects of differences between blocks. The number of experimental conditions in a block is less than the total number in the basic design.

One type of blocking was discussed in Chapter III on fractional factorial designs. These designs might be used, for example, if it were not possible for an experimenter to run all of the experimental conditions for the complete design in one day. He would use an incomplete block design to assign the conditions to the two days (blocks) in a way that irrelevant day to day changes in the equipment would not affect the comparisons of interest. In this form of blocking, some higher-order interaction effects are usually confounded with blocks in order to keep the main and two-factor interaction effects from being biased by any differences between blocks.

Another type of incomplete block designs (referred to as Balanced Incomplete Block, or B.I.B. designs) compare the effect of every experimental condition with equal precision. Chapter 11 in Cochran and Cox (16) has an excellent discussion on these designs. In B.I.B. designs, each block will contain the same number of experimental conditions, each condition will appear the same number of times in the complete design, and every experimental condition will occur together within a block with every other experimental condition an equal number of times. This type of blocked design is also used with the factorial to construct the three-level, second-order response surface designs.

In certain cases, the number of replications required to achieve the balance described above may become prohibitively large. Then designs that do not have a complete balance will be used. These are referred to as Partially Balanced Incomplete Block (P. B. I. B.) designs. In P. B. I. B. designs where variations between blocks are large, some pairs of experimental conditions are compared more precisely than others. The simplest of these designs are those with only two levels of precision, referred to as first and second associate classes. Pairs of experimental conditions that are within the same block are called first associated; pairs that are not within the same blocks are called second associates.

Construction

The two-level factorial (or fractional factorial) and the Balanced (or Partially Balanced) Incomplete Block designs are combined to create the three-level, second-order response surface designs proposed by Box and Behnken (7). For example, to develop a four-factor, three level response surface design, the following Balanced Incomplete Block design for four experimental conditions, distributed two at a time in six blocks, is used.

		Experimental Conditions			
		X1	X2	X3	X4
Blocks	1	*	*		
	2			*	*
	3	*			*
	4		*	*	
	5		*		*
	6	*		*	

The * indicates which experimental condition (X_i) is in which block. To make it a completely balanced design, each experimental condition is replicated three times.

A 2^2 factorial design,

X_i	X_j
-1	-1
+1	-1
-1	+1
+1	+1

is substituted for each asterisk in the B. I. B. design. Whenever an asterisk does not appear in the B. I. B., then a zero is inserted instead. In addition, center points (0, 0, 0, 0) are also added.

The complete design would appear as:

		Variables					
		x_1	x_2	x_3	x_4		
Experimental Conditions	↑	1	-1	-1	0	0	Block 1
			-1	-1	0	0	
			-1	1	0	0	
			1	1	0	0	
			0	0	-1	-1	
			0	0	1	-1	
			0	0	-1	1	
			0	0	1	1	
		↓	9	0	0	0	
		10	-1	0	0	-1	Block 2
			1	0	0	-1	
			-1	0	0	1	
			1	0	0	1	
			0	-1	-1	0	
			0	1	-1	0	
			0	-1	1	0	
			0	1	1	0	
		18	0	0	0	0	
		19	0	-1	0	-1	Block 3
			0	1	0	-1	
			0	-1	0	1	
			0	1	0	1	
			-1	0	-1	0	
			-1	0	-1	0	
			-1	0	1	0	
			1	0	1	0	
	↓	27	0	0	0	0	

which can be shortened by combining + and - terms into \pm , such that $(\pm 1 \pm 1)$ implies the four possible combinations of signs in the 2^2 factorial.

± 1	± 1	0	0
0	0	± 1	± 1
0	0	0	0

± 1	0	0	± 1
0	± 1	± 1	0
0	0	0	0

± 1	0	± 1	0
0	± 1	0	± 1
0	0	0	0

Thus the combination of the B.I.B. and factorial designs has created a four-factor, three-levels (-1, 0, +1) per factor, response surface design with a total N of 27 in three orthogonal blocks of 9 each. The characteristics of three-level response surface designs requiring fewer than 300 observations in the basic design for 5, 6, 7, 9, 10, 11, and 12 variables are listed in Table [V-3]. The complete designs, reproduced from Box and Behnken's (7) Table 4, are in Appendix IV. It is necessary to refer to the original paper to learn how to analyze these designs. Designs for 5, 7, and 10 factors, based on a Balanced design with

Table [V-3]. Second Order Response Surface Designs
with Three Levels per Factor

No. of Variables	Total N	No. of Possible Blocks	No. of Center Points*	Type of I. B. Design**
5	46	2	6	BIB
6	54	2	6	PBIB
7	62	2	6	BIB
9	130	5 or 10	10	PBIB
10	170	2	10	PBIB
11	188	1	12	BIB
12	204	2	12	PBIB

*Center points will be distributed equally among the blocks.

**BIB refers to Balanced Incomplete Block designs;
PBIB refers to Partially Balanced Incomplete Block.
When PBIB designs are used, two error terms must
be calculated for the two associate classes.

only one error term, will be easier to analyze than the others which are based on Partially Balanced designs with two error terms. As with the central-composite designs, all of these include the property of rotatability and when this conflicts with the orthogonality of blocking, the latter is favored. Tests of goodness-of-fit are possible. These designs are all of Resolution V, enabling estimates of main and two-factors interaction effects to be estimated independently of one another. In general, they are more economical than comparable 3^{k-p} fractional factorials. The center points are replicated to keep the variance relatively constant within the limits of the experimental space ($\pm\alpha$), and particularly between ± 1 .

THIRD-ORDER RESPONSE SURFACE DESIGNS

Although higher-order effects are usually negligible, there may be times when the existence of third-order effects is suspected. Das and Narasimham (22) published a series of third-order designs for up to 15 factors. For certain of these, called "sequential third-order rotatable designs, it is possible to collect a block of data that provides the coefficients of a second-order surface. If the second-order model failed to fit the experimental data, a second block of data could be added to obtain the coefficients for a third-order surface. The two blocks are orthogonal, and in many of the designs, the data collection can be broken into orthogonal sub-blocks. An earlier paper by Das (21) supplies some support data for understanding, constructing, and using these designs.

Third-order response surface designs will not be described in more detail here because of a number of features that make them inherently poor for most human factors engineering research. These are:

- 1) The number of levels per factor in some of these designs is quite large. For example, fifteen levels per factor are used for a five-factor design.
- 2) The number of observations required for the complete third-order design become quite large for eight or more factors. For example, for eight factors, 480 conditions are needed; for nine factors, 1256 conditions are needed.

There may be circumstances when an experimenter might wish to use these third-order designs to study five or six factors. However, before any third-order design is seriously considered for an investigation of quantitative variables, the experimenter might find it more practical to expend some preliminary effort in finding measurement scales that will enable the relationship between performance and equipment parameters to be expressed by a linear or quadratic polynomial.

NON-SYMMETRICAL SECOND-ORDER RESPONSE SURFACE DESIGNS

There are many situations in which it would not be possible to have the same number of levels for all factors. The symmetrical designs described earlier required that either three or five levels always be employed for the second-order

models. Draper and Stoneman (25), however, propose a set of non-symmetrical designs suitable for handling factors at two and three levels or two and four levels. In addition, they make data collection even more economical by tentatively entertaining the idea that certain coefficients are not required in the polynomial. This allows for the fitting of a polynomial with fewer terms than the number required for the Taylor series expansion (used with the central-composite designs) and results in a further reduction in the number of experimental conditions required.

For example, suppose an experiment was to be planned containing seven factors, four at two levels and three at three levels. A complete factorial design for this $2^4 3^3$ case would require 432 observations. If, however, it were tentatively assumed that two-level factors (factors 1 through 4) exert only first-order effects while three-level factors (factors 5, 6, and 7) exert both first and second-order effects, and the interaction of three-level factors need be represented by only the linear-by-linear component, then the relationship could be represented by the following polynomial:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 \\ + \beta_{55} x_5^2 + \beta_{66} x_6^2 + \beta_{77} x_7^2 + \beta_{56} x_5 x_6 + \beta_{57} x_5 x_7 + \beta_{67} x_6 x_7 .$$

To fit this polynomial, a design consisting of 32 data points in the basic design and a replication of eight additional points is suggested.

The construction of these designs is based on multiple sets of fractional factorials of different magnitudes and degrees of fractionation. The characteristics of a Resolution V design are met — no first or second-order effects of any type are confounded with one another. In certain designs, if this rule is violated in one part of the design, a compensatory effect is introduced in another so that in the final design, no violation exists.

As with the central-composite designs, tests of the lack-of-fit of the polynomial to the collected data can be made. If it is found to be inadequate, additional data must be collected to estimate the other coefficients.

It is beyond the scope of this report to describe these designs in more detail. Draper and Stoneman (25) provide a clear explanation of how the designs are constructed and what can be done in the event more data must be collected. The understanding of the principles of fractional factorials and their applications, described in this report in Chapters III and IV, will enable the reader to understand the Draper and Stoneman article.

Designs for the 2^{p3^q} and 2^{p4^q} cases each have been worked out and presented in detail by Stoneman (46) for all 45 combinations of p and q when $(p + q)$ runs from 2 to 10, ($p, q \neq 0$). Of these, only the $2^1 4^9$ design, requiring 56 coefficients to be estimated in the polynomial, exceeds the 300 observation limit set in this report. This requires 324 observations, still a small amount compared to the 524,288 required for a complete factorial. The next largest design, $2^2 4^8$, requires only 192 observations in the basic design to estimate 47 coefficients of the polynomial. The largest of the two and three level designs, $2^1 3^9$, requires 146 observations in the basic design to estimate 56 coefficients. All other designs of this group require less than 100 observations.

RESPONSE SURFACE DESIGNS FOR "MESSY" EXPERIMENTAL SPACES

Ordinarily in an experiment in which the investigator can select the factor levels, it is assumed that the experimental space will be a regular, symmetrical, multivariate space, with any point within the space a candidate data collection point. There are circumstances, however, when practical problems of feasibility, operability, and availability of the system being studied make it necessary to chop, slice, and gouge sections off and out of the experimental space. Also at times, there may be specific experimental conditions that, by decree, must be included in the experimental design. As a result, it is difficult to superimpose conventional experimental designs on such a space effectively.

Kennard and Stone (35) describe an ingenious plan for constructing a response surface experimental design that considers the constraints mentioned above. No assumption of the correct model of the response surface is required. The design points are selected sequentially in such a way that the mangled space will be

represented as uniformly as possible by the existing points whenever the experimenter decides (for purposes of economy) to terminate the experiment.

Construction

The experimental design is selected from a set of N data collection points in a multifactor space. These N points are the usable points (referred to as "candidates") of a complete factorial plan within the idealized experimental space, i. e. before the inoperable regions were removed. The actual design points are selected from the N candidate points and include any data collection points determined a priori by authority. Kennard and Stone provide the algorithm by which this "messy" design can be constructed using a computer, although they warn that "its spirit is not that of a cookbook, but that of an assistant." (32, p 148).

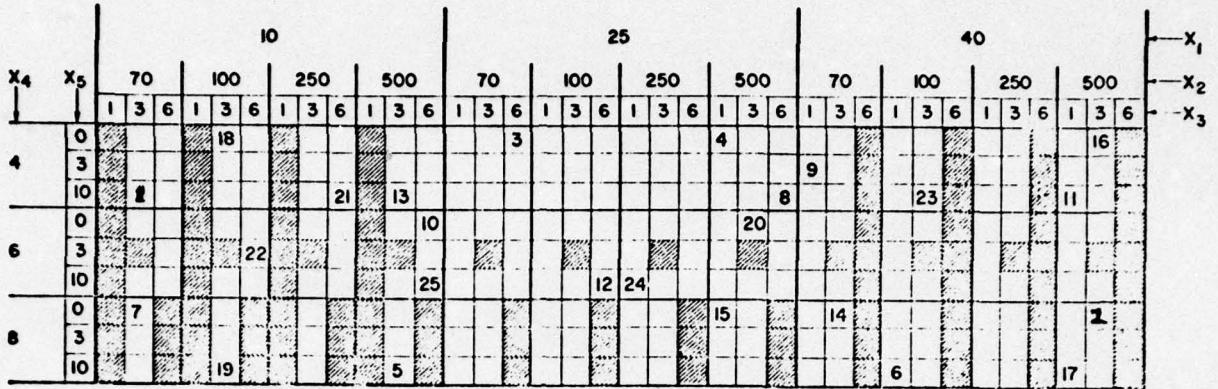
To appreciate these designs, it is helpful to learn how they are constructed. Let us examine some examples provided by Kennard and Stone. A "messy" five factor space originally characterized by a $3^4 \times 4^1$ factorial is shown in Figure [V-4, A]. However, out of the 324 points in the complete factorial, only 216 are actually available and operable. The shaded portions in the figure represent the experimental conditions that cannot be used. The problem is to select a set of experimental design points from the 216 candidate points. The aim is to end up with a set of points in the design "uniformly" spread over the available space. The numbers in the figure indicate the point and the order in which they were selected.

The algorithm employed to build this design is simple and direct. The steps are as follows:

- 1) First find a point at the boundary of the space, e. g. the lowest value of every factor.
- 2) Next find the point that is the farthest distance* from the first point.

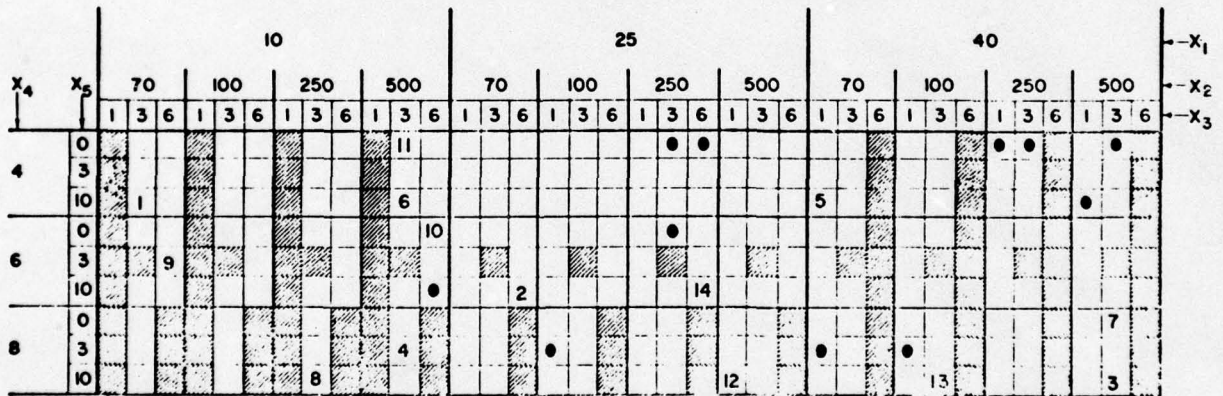
*Distance is defined as the sum of the differences squared between each value of the coordinates of the two points in question. For example, in a three-dimensional space, the coordinates of two points might be (-1, 3, 2) and (2, 2, -2). The differences between them are (-3, 1, 4), which when each value is squared and summed is 24.

A. ORTHONORMALIZED VECTORS



SHADED AREAS ARE NOT REALIZABLE
n = 25

B. II - POINT START - ORTHONORMALIZED VECTORS



SHADED AREAS ARE NOT REALIZABLE
● EXECUTED PRIOR TO DESIGN

Figure [V-4]. "Messy" experimental designs
[Adapted from Kennard and Stone (32)]

- 3) From among the remaining candidates, find the point that is farthest from the points already in the design.

The final step is repeated until the experimenter terminates the design process. When there are ties, the point with the smallest index is arbitrarily selected. When to stop adding points to the design depends in part on the resources of the investigator, that is, how economical he must be.

In the second example, what Kennard and Stone refer to as the "boss option" is exercised. In Figure [V-4, B], the eleven black dots indicate those experimental conditions that were included in the experimental design by decree. The numbers represent the design points selected from the candidate points and the order in which they were selected, according to the previously described algorithm, but after taking into consideration the already selected eleven points. In this case, the first point selected by the computer is actually the twelfth point in the design, the one farthest from the already selected eleven points.

Practical Considerations

The more the usable (candidate) space deviates from the symmetrical factorial, the greater the chance that the "messy" design will be imbalanced with all of the disadvantages of any nonsymmetrical, nonorthogonal design. To correct for this, the authors suggest an orthonormalizing transformation of the distance calculations prior to selecting the design points; this gives the selected points a more spherical orientation and a more uniform coverage. (35, p. 140)

Perhaps the most critical limitation against using these designs for truly multifactor human factors experimentation lies in the computer memory necessary to store the interpoint distances, which can strain even the largest computers. To overcome this the authors state: "For problems having a very large number of candidate points, it has been found that a workable procedure is to first calculate the radius for each point, sort the radii, choose radii bands, and then have only points in these bands as input to the selection procedure." (35, p. 148)

CHAPTER VI.
CONCLUSIONS

If human factors engineering experiments are expected to supply the data needed to accurately predict field performance, then it is imperative to include within a cohesive set of experiments most of the factors that will have a major effect on the performance of a particular task. It is no longer realistic to believe that the results of a large number of small -- two to five factors -- experiments can ever be combined quantitatively to form a unified data base. Nor is the excuse any longer tenable that large experiments are too costly to perform.

The rationale, approach, and designs described in this report provide a practical method of studying a great many factors economically. Basically the designs that have been described are not radically different from those that have been conventionally employed in human factors engineering research, only the manner in which they are applied and the way the results are interpreted are changed. But these changes are the decisive factors that enable "economical" and "multifactor" to be used to describe the same experimental plan. Furthermore the simplicity of its rationale makes the approach so credible: if data-taking is avoided until there is some evidence that it is required and if that which is collected is untainted from as many irrelevant effects as possible, the total amount of effort required to study a great many factors suddenly becomes reasonable. The situation is further enhanced by the fact that ordinarily, for any specific situation, only a relatively few of a great many factors have any important effect on performance. If an experimenter first obtains a less precise overview of performance patterns within his experimental space as quickly and as cheaply as possible, and if he uses this data to identify the important parts of the space, he can afford to expend the effort to probe more deeply where it really counts.

While the design proposed here may not be suitable for all human factors engineering problems, they will be suitable for many, particularly those in which the factors are quantitative. The approaches proposed for employing the designs will certainly increase the successful attainment of many research goals over those that have been employed up to now. The only way to determine the circumstances in which these experimental methods will and won't work is to try them.

REFERENCES

1. Adams, J. Presidential address to the Society of Engineering Psychologists. Amer. Psychol., 1972, 27, 615-622.
2. Anscombe, F. J. and J. W. Tukey, The examination and analysis of residuals. Technometrics, 1963, 5, 141-160.
3. Bakan, D. The test of significance in psychological research. Psychol. Bull., 1966, 66, 423-437.
4. Beale, D. K. What's so significant about .05? Amer. Psychol., 1972, 27, 1079-1080.
5. Box, G. E. P., The exploration and exploitation of response surfaces: Some general considerations and examples. Biometrics, 1954, 10, 16-60.
6. Box, G. E. P., A note on augmented designs. Technometrics, 1966, 8, 184-188.
7. Box, G. E. P., and D. W. Behnken. Some new three level designs for the study of quantitative variables. Technometrics, 1960, 2, 455-475.
8. Box, G. E. P., and J. S. Hunter. Experimental designs for the exploration and exploitation of response surfaces. In Chew, V. (ed.) Experimental design in industry. New York: Wiley, 1956, pp. 138-190.
9. Box, G. E. P., and J. S. Hunter. Multi-factor experimental designs for exploring response surfaces. Ann. Math. Stat., 1957, 28, 195-241.
10. Box, G. E. P., and J. S. Hunter. The 2^{k-p} fractional factorial designs. Part I. Technometrics, 1961, 3, 311-351.
11. Box, G. E. P., and J. S. Hunter. The 2^{k-p} fractional factorial designs. Part II. Technometrics, 1961, 3, 449-458.
12. Box, G. E. P., and K. B. Wilson. On the experimental attainment of optimum conditions. Journal of the Royal Statistical Society, Series B, 1951, 18, 1-45.
13. Bradley, R. A. Determination of optimum operating conditions by experimental methods. I. Mathematics and statistics fundamental to the fitting of response surfaces. Industrial Quality Control, 1958, 15, 1-5.
14. Budne, T. A. The applications of random balanced designs. Technometrics, 1959, 1, 139-155.
15. Campbell, D. T., and J. C. Stanley. Experimental and quasi-experimental designs for research. Chicago: Rand McNally, 1963.

16. Cochran, W. G., and G. M. Cox. Experimental designs. New York: Wiley, 1957 (2nd edition).
17. Chew, V. (ed.) Experimental designs in industry. New York: Wiley, 1956.
18. Conner, W. S., and Shirley Young. Fractional factorial designs for experiments with factors at two and three levels. Washington: National Bureau of Standards, Applied Math. Series 58. U.S. Govt. Printing Office, 1 September 1961.
19. Conner, W. S., and M. Zelen. Fractional factorial experiment designs for factors at three levels. Washington: National Bureau of Standards, Applied Math Series, No. 54, U. S. Govt. Printing Office, 1 May 1959.
20. Daniel, C. Sequences of fractional replicates in the 2^{P-q} series. J. Amer. Statist. Assoc., 1962, 57, 403-429.
21. Das, M. N. Construction of rotatable designs from factorial designs. J. Indian Soc. Agricultural Statist., 1961, 13, 169-194.
22. Das, M. N., and V. L. Narasimham. Construction of rotatable designs through balanced incomplete block designs. Ann. Math. Statist., 1962, 33, 1421-1439.
23. Davies, O. L. (ed.) The design and analysis of industrial experiments. New York: Hafner Pub. Co., 1956.
24. Draper, N. R., and H. Smith. Applied regression analysis. New York: Wiley, 1968.
25. Draper, N. R., and D. M. Stoneman. Response surface designs for factors at two and three levels, and two and four levels. Technometrics, 1968, 10, 177-192.
26. Dunnette, M. D. Fads, fashions, and folderol in psychology. Amer. Psychol., 1966, 21, 343-352.
27. Dykstra, Jr., O. Partial replication of response surface designs. Technometrics, 1960, 2, 185-195.
28. Fenwick, C. A. Development of a peripheral vision command indicator for instrument flight. Human Factors, 1963, 5, 117-127.
29. Finney, D. J., Recent developments in the design of field experiments. III. Fractional replication. J. Agric. Sci., 1946, 36, 184-191.
30. Fisher, R. A., The design of experiments. New York: Hafner, 1949 (5th edition).
31. Hays, W. L., Statistics. New York: Holt, Rinehard, and Winston, 1963.
32. Hill, W. J., and W. G. Hunter. A review of response surface methodology: A literature survey. Technometrics, 1966, 8, 571-590.

33. Hunter, J. S. Determination of optimum operating conditions by experimental methods: Models and methods. Industrial Quality Control, Part II-1, 1958, 15, January, pp. 16-24; Part II-2, 1959, 15, January, pp. 7-15; Part II-3, 1959, February, pp. 6-14.
34. Kempthorne, Q, A simple approach to confounding and fractional replication in factorial experiments. Biometrika, 1947, 34, 255-272.
35. Kennard, R. W., and L. A. Stone. Computer aided design of experiments. Technometrics, 1969, 11, 137-148.
36. Li, J. C. R. Design and statistical analysis of some confounded factorial experiments. Ames, Iowa: U.S. Dept. Agriculture, Bureau of Agriculture and Mechanic Arts, Statistical Section Research Bulletin 333, June 1944.
37. Lykken, D. T. Statistical significance in psychological research. Psychol. Bull. 1968, 70, 151-159.
38. Minturn, E. B. A proposal of significance. Amer. Psychol., 1971, 26, 669-670.
39. Namboodiri, N. K. Experimental designs in which each subject is used repeatedly. Psychol. Bull., 1972, 77, 54-64.
40. Plackett, R. L., and J. P. Burman. The design of optimum multi-factorial experiments. Biometrika, 1946, 33, 305-324.
41. Rozebloom, W. W. The fallacy of the null-hypothesis significance test. Psychol. Bull. 1960, 57, 416-428.
42. Simon, C. W. Reducing irrelevant variance through the use of blocked experimental designs. Culver City, California: Hughes Aircraft Company, Tech. Report No. AFOSR-70-5. November 1970.
43. Simon, C. W. The use of central-composite designs in human factors engineering experiments. Culver City, California: Hughes Aircraft Company, Tech. Report No. AFOSR-70-6, December 1970.
44. Simon, C. W. Considerations for the design and analysis of human factors engineering experiments. Culver City, California: Hughes Aircraft Company, Tech. Report No. P73-325, December 1971.
45. Stat. Eng. Lab., Fractional factorial experiment designs for factors at two levels. Washington: National Bureau of Standards Applied Math Series No. 48, U.S. Govt. Printing Office, Washington, 15 April 1957.
46. Stoneman, D. M. Response surface designs for specified factor levels. Ph. D. Thesis. Madison: University of Wisconsin, 1966.
47. Suits, D. B. Use of dummy variables in regression equations. J. Amer. Statist. Assoc., 1957, 52, 548-551.

48. Tukey, J. W. Where do we go from here? J. Amer. Statis. Assoc., 1960, 55, 80-93.
49. Vartabedian, A. G. The effects of letter size, case, and generation method on CRT display search time. Human Factors, 1971, 13, 363-368.
50. Vaughan, G. M., and M. C. Corballis. Beyond tests of significance: estimating strength of effects in selected ANOVA designs. Psychol. Bull. 1969, 72, 204-213.
51. Yates, F. The design and analysis of factorial experiments. Harpenden, England: Imperial Bureau of Soil Science. Technical Communication No. 35, 1937.
52. Youden, W. J. Partial confounding in fractional replication. Technometrics, 1961, 3, 353-358.

APPENDICES

APPENDIX I

AN ANALYSIS OF THE METHODOLOGY AND EFFECTIVENESS OF SOME REPRESENTATIVE HUMAN FACTORS EXPERIMENTS*

The journal, HUMAN FACTORS, it informs its contributors, "publishes original articles which increase and diffuse the knowledge of man in relation to machine and environmental factors in all their ramifications, pure and applied." An analysis was made of one hundred and forty-one articles published in this journal from Volume 1, No. 1, September 1958 through Volume 14, No. 3, June 1972 in which formal experiments were described and the results presented in some summary statistical tables. Thirty-four analyses of variance in 23 of the articles were excluded from the final analysis because they fell into one of the following categories:

- A partial analysis of a more complete analysis. (n = 8)
- A reprint of an analysis from a study not described in the article. (n = 3)
- A study of a single factor at two levels. (n = 4)
- No data (n = 7), incomplete data (n = 9), or incorrect data. (n = 2)
- Chi square analysis. (n = 1)

As a result of these exclusions, the data in this report is based on the test of the 118 articles and the 239 analyses of variance tables in these articles. Although the data for several analysis of variance tables may have been collected at the same time, either the independent variables or the performance measure changed. Therefore, since this paper is not concerned with content, each analysis of variance is treated as if it is the results of a different experiment, which it is.

* While a great deal of the information supplied by this analysis appears in this present paper as support for the principles of economical multifactor designs, a complete report of these results, along with additional data to be collected for a follow-on contract, will be published next year.

Is this sample representative?

There are a great many human factors experiments produced yearly in industry and government laboratories that are never published in the journal. Many of these have a security classification which limit their accessibility. How representative, therefore, is the group of experiments covered in this report? While there is no way to accurately answer this question, some observations can be made which suggest that in general those papers published in HUMAN FACTORS and those published as company reports and government documents will not differ materially insofar as their experimental methodologies are concerned. For one thing, the human factors community is relatively small, probably composed of fewer than 2000 people of whom less than one-fourth do anything that might be remotely considered to be research. For another, members of the Human Factors Society who publish in the journal as well as those on its editorial staff are among those doing research in industrial and the government laboratories.

Which organizations conducted and funded the research?

Eight types of organizations could be readily identified. These were:

- Army
- Consulting
- Air Force
- Government (non-military)
- Industry
- Navy
- University
- Other (e. g. private research organizations)

Over 34 percent of the work was done in industry, 28 percent in universities, and 20 percent in consulting companies which supported approximately 64, 36, and 31 percent of their own work respectively. The remaining support came from the government agencies -- with military agencies as a group funding approximately twice as much as non-military agencies. Of these publications, the Army published more in-house research in this journal than either of the other two military organizations.

THE DATA BASE

The basic data extracted from the articles includes characteristics of their experimental designs and certain data collection procedures. In addition, each analysis of variance table was reanalyzed to determine the proportion of total variance within the experiment attributable to each identifiable source of variance.

The limited usefulness of tests of statistical significance has received increased recognition in recent years (3)(4)(37)(41). A result may be statistically significant if the sample is large enough. As a result, many effects found to be reliable are also found to be trivial when the proportion of variance it accounts for in the experiment is measured. This measure is referred to as eta squared.

Eta squared is calculated by dividing the sum of squares for the particular source of variance in question by the total sum of squares. The proportion is a descriptive index of the strength of the relationship between a source of variance and performance and is meaningful only within that particular sample. Another measure, omega squared is an inferential measure of how much of an effect a factor would have in the population based on the experimental results. It modifies the estimate on the basis of the error variance and the number of degrees of freedom involved. There are several forms of omega squared depending on the experimental design employed as well as certain statistical assumptions made in developing the equation (50).

For our purposes, however, eta squared is considered to be the more appropriate statistic to use in this analysis because:

1. It provides a direct measure of the quality of the data in the individual experiment and needs to make no assumptions about a hypothetical population. This is not the case with the various inferential measurements (50).
2. Since it uses no error term, a decision need not be made as to what should be used to estimate "error". Nor is it necessary to recalculate the values used in the published data, if the experimenter failed to use the more technically correct error term.
3. It provides the most optimistic estimate of the contribution of each source of variance.
4. The measure is simple, intuitively understandable, and familiar. Its square root is a correlation between a source (variable) and performance. With a 1 d. f. variable, it is a Pearson product moment correlation. With more than 1 d. f., it is a correlation ratio, or eta.

There is an unknown quantity which eta squared cannot supply, and that is how much of the real world is represented by the experiment. While the equipmental factors may only account for 25% of the total performance variance within a poorly-conducted human factors engineering experiment or 90% of the total performance variance within a well-conducted experiment, if the number of factors included in the laboratory experiment are either so few or so unimportant that they represent only 30% of the performance variance found when the task is measured in the real world, then in fact, any observed proportion of variance explained within the experiment will probably be considerably less than that found in the field until better methods of selecting and studying more factors in the laboratory are employed.

THE DATA STRUCTURE

Sources of performance variability in human factors engineering experiments can be segregated into three primary classes:

- 1) Those associated with physical parameters of the equipment, system and environment, referred to as Equipment Variables. (E)
- 2) Those associated with people -- the subjects. (S)
- 3) Those associated with temporal changes from trial to trial. (T)

Since this report is not concerned with the content of the experiments, these sources of variance in the analyses of variance tables were reclassified into the above classes and their interactions.

Two other categories were employed in the analysis for sources of variance not included in the above classes. In a few instances, the experimenter included "Order" as a methodological variable in which the order of presenting experimental conditions was included as an experimental variable and the effect isolated in the analysis. In some experiments, the experimenter did not partition certain interaction effects, but provided some pooled estimate of several such sources of variance. A "pooled" category has been introduced to handle these cases.

The 239 analyses of variance were often examined in subgroups defined by the number of equipment factors in each group. There were seven such groups

involving zero, one, two, three, four, five, and seven equipment factors. Since there were only two experiments with no equipment factors and only one with seven, these were sometimes not included in an analysis; that is, some analyses were based on only 236 experiments. With only four five-factor experiments, these too were occasionally omitted from an analysis and subsequent discussion, when there was insufficient data for the purposes of the particular analysis.

There are a number of ways in which this data can be organized. For example, the studies might have been grouped by the number of any kind of factor -- equipment, subject, or temporal -- in the experiment rather than only by the number of equipment factors. While these different groupings do result in different values for particular groups (particularly the one factor group), as the information is used for this report, the differences between the two analyses are not relevant.

"Unexplained" Variances. Throughout the discussion of this analysis, reference will be made to the proportion of data accounted for by the experiment and the proportion not accounted for, or "unexplained". The term "unexplained" has a particular meaning that should be understood in the context in which it is used. Here, "unexplained" refers to the interactions between subjects and trials and between subjects and/or trials and equipment factors when subjects and trials were treated as replications in the experiment. This is rather conservative definition of the term, since it does not include subject and trial main effects, nor order of presentation effects, when actually their presence in any magnitude reflects a failure on the part of the experimenter to control these unwanted sources of variance. To this extent, "unexplained" as used here is somewhat synonymous with "irrelevant" or "unwanted" sources of variance, not planned for by the experimenter.

APPENDIX II

FRACTIONAL FACTORIAL DESIGNS AT TWO LEVELS*

Plan 2.5.3. 1/2 replication of 5 factors

Factors: *A, B, C, D, E.*

I = ABCDE.

Block confounding: *None.*

Block			
1			
(1)	ab	ac	bc
abcd	cd	bd	ad
de	abde	acde	bcde
abce	ce	be	ae

Plan 2.6.16. 1/2 replication of 6 factors in 2 blocks of 16 units each.

Factors: *A, B, C, D, E, F.*

I = ABCDEF.

Block confounding: *ABF.*

Blocks							
1				2			
(1)	abce	ab	ce	ac	be	bc	ae
abcd	de	cd	abde	bd	acde	ad	bede
bcef	af	acef	bf	abcf	cf	ef	abcf
adef	bcd	bdef	acdf	cdcf	abdf	abdef	df

Plan 2.7.16. 1/2 replication of 7 factors in 4 blocks of 16 units each.

Factors: *A, B, C, D, E, F, G.*

I = ABCDEFG.

Block confounding: *ABFG, ACF, BCG.*

Blocks							
1		2		3		4	
(1)	abce	ab	ce	ac	be	bc	ae
abcd	de	cd	abde	bd	acde	ad	bcde
bcef	af	acef	bf	abef	cf	ef	abcf
adef	bcd	bdef	acdf	cdcf	abdf	abdef	df
cdfg	abdefg	abdefg	defg	adfg	bcdefg	bdfg	acdefg
abfg	cefg	fg	abcefg	bcfg	ae fg	acfg	bcfg
bdeg	acdg	adeg	bcdg	abdeg	dg	cdeg	abdg
aceg	bg	bceg	ag	eg	abeg	abeg	cg

Plan 4.8.16. 1/4 replication of 8 factors in 4 blocks of 16 units each.

Factors: *A, B, C, D, E, F, G, H.*

I = ABCEG = ABDFH = CDEFGH.

Block confounding: *ACD, BEF, ABCDEF.*

Blocks							
1		2		3		4	
(1)	cdgh	bdefh	bcefg	acdefgh	uef	abeg	abdh
abc fgh	abdf	acdeg	ae h	hde	bcegh	fh	cdfg
bcdeg	beh	efgh	df	ahfh	abcdfg	ade	acegh
adefh	acefg	ab	abcdgh	cg	dh	bcdefgh	bef
efgh	cdef	bdg	beh	ard	agh	abcfh	abdefg
abce	abdegh	acd fh	afg	bd fgh	bcf	eg	cdeh
bcdfh	bf g	ce	degh	abeg	abcd e h	ad fgh	arf
adg	ach	abefgh	abdef	cefh	defg	bed	bgh

*Adapted from National Bureau of Standards (45)

Plan 4.9.16. 1/4 replication of 9 factors in 8 blocks of 16 units each.

Factors: A, B, C, D, E, F, G, H, J.

$I = ABCEGJ = ABDFIJ = CDEFGH.$

Block confounding: ACIJ, BEFJ, ABCDEF, BCJ, ABD, CEF, ADEFJ.

Blocks							
1	2	3	4	5	6	7	8
(1)	bdefh	acdefgh	abeg	cdj	berfhj	acfgjh	abdghj
abefgh	acdeg	bile	fh	abdfghj	acgj	bcfj	cdfhj
bedfg	cfgh	abh	ade	bcgj	dfghj	ubcdfhj	acrgj
adefh	ab	cg	bcdefgh	acrfhj	abedj	dgj	bcfghj
bdhj	efj	abcefgj	acdghj	bch	cdcf	abdefg	agh
acdfgj	abceghj	ehj	bilj	afg	abfegh	cdch	bcf
ceghj	bcdfgj	adfj	abehj	dcgh	hfg	acf	abcdh
abefj	adhj	bcdfghj	cefaj	abedef	ach	bgh	dfg
bidfgj	ghj	abehj	acdefj	berfg	cdgh	abdh	afj
acdehj	abefj	fgj	bdefghj	ach	abdf	cdfg	bcfgh
efj	bcdehj	adefghj	abfgj	df	bek	acrfgh	abcdfg
abghj	adefgj	bedefj	chj	ubedgh	acrfj	bfj	dh
efgh	bdg	acd	abcrfh	cdefghj	bcgj	aj	abdshj
abce	acdfh	bdfgh	cg	abdefj	afhj	bcfghj	cdfgj
bedfh	ce	abeg	adfgj	bfhj	dej	ubcdcfj	acfgjh
adg	abefgh	crfh	bed	acgj	abdefghj	drfhj	bj

Plan 4.10.16. 1/4 replication of 10 factors in 16 blocks of 16 units each.

Factors: A, B, C, D, E, F, G, H, J, K.

$I = ABCDEFG = ABCDHJK = EFGHJK.$

Block confounding: ADEFGK, AFGJ, DEJK, BEFH, ABDGHK, ABEGHJ, BDFHJK, HJK, ADEFGHJ, AFGHK, DEH, BEFJK, ABDGJ, ABEGK, BDF.

Blocks															
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
(1)	abdfk	acfgjk	bcdgj	abegjk	defgj	bcf	acdek	fgkh	abdgh	ackj	bcdfhjk	acfgk	bcdf	acdfgh	acdfgh
abefhk	deh	bcghj	acdefghjk	fgjh	abdghjk	ackh	bdflh	bcdfhj	acdehjk	abeghk	defgh	bcfgk	bj	abdfhjk	bdghk
bcdfhj	acdehjk	abeghk	defgh	acfgkh	bcgh	bcdf	abdflhjk	acfgj	bcdfghj	fg	abdghk	bcfg	abefjk	dejk	bcdfghk
acjk	bdffj	fg	abdghk	bcrg	acdefgk	abefjk	dejk	bcfg	bcdefgh	abjk	dfj	acefjk	bcdej	efg	abdefgk
bcfg	acdghk	bcghj	abdghk	bchj	acd/hjk	abfghk	dgh	acdfghk	bcdfghj	efhj	abdehjk	bchj	abfghk	dfg	bcdfghk
acdfghk	bcdfghj	acdfhk	bdcdh	abhk	dfh	bcdfghj	acdfghk	bcdfghj	acdfghj	acdfhk	bdcdh	abhk	bcdfghj	bcdfghj	acdfghk
efhj	abdefghjk	acdfhk	acdjk	ef	abdek	acegjk	bddefgj	efhj	abdefghjk	bc	acdjk	ef	abdek	acdfghk	bddefgj
abfghk	dgj	bc	bcfgk	cdghk	cdghk	bdjk	bcdfghk	cdghk	cdghk	adeg	bcfgk	cdghk	cdghk	cdghk	bddefgj
cdghk	abcejk	adeg	agh	cdghk	abcefg	bcjk	bcdfghk	cdghk	cdghk	cf/hjk	agh	cdghk	abcefg	cdghk	bcdfghk
abcdhj	cf/hjk	bdffghk	oghjk	adefghj	bcfgjhk	cdghk	bcdfghk	cdghk	cdghk	afh	oghjk	adefghj	bcfgjhk	cdghk	bcdfghk
bdhk	afh	abcefgjhk	oghjk	adefghj	bcfgjhk	cdghk	bcdfghk	cdghk	cdghk	bc	oghjk	adefghj	bcfgjhk	cdghk	bcdfghk
adef	bc	cdghk	abcefgj	bdffghk	cdghk	cdghk	bcdfghk	cdghk	cdghk	adef	bc	cdghk	adefghj	bcfgjhk	bcdfghk
bdghk	bcfgj	abcefg	cek	adefghj	bcfgjhk	cdghk	bcdfghk	cdghk	cdghk	bcfgj	abcefg	cek	adefghj	bcfgjhk	bcdfghk
bdghk	bcfgj	abcefg	cek	adefghj	bcfgjhk	cdghk	bcdfghk	cdghk	cdghk	bcfgj	abcefg	cek	adefghj	bcfgjhk	bcdfghk
adfgjhk	bcfgjhk	cdghk	abcefg	bdffghk	cdghk	cdghk	bcdfghk	cdghk	cdghk	bcfgjhk	abcefg	cek	adefghj	bcfgjhk	bcdfghk
adfgjhk	bcfgjhk	cdghk	abcefg	bdffghk	cdghk	cdghk	bcdfghk	cdghk	cdghk	bcfgjhk	abcefg	cek	adefghj	bcfgjhk	bcdfghk
abcefg	bcfgjhk	cdghk	abcefg	bdffghk	cdghk	cdghk	bcdfghk	cdghk	cdghk	bcfgjhk	abcefg	cek	adefghj	bcfgjhk	bcdfghk
abcefg	bcfgjhk	cdghk	abcefg	bdffghk	cdghk	cdghk	bcdfghk	cdghk	cdghk	bcfgjhk	abcefg	cek	adefghj	bcfgjhk	bcdfghk
abcefg	bcfgjhk	cdghk	abcefg	bdffghk	cdghk	cdghk	bcdfghk	cdghk	cdghk	bcfgjhk	abcefg	cek	adefghj	bcfgjhk	bcdfghk
abcefg	bcfgjhk	cdghk	abcefg	bdffghk	cdghk	cdghk	bcdfghk	cdghk	cdghk	bcfgjhk	abcefg	cek	adefghj	bcfgjhk	bcdfghk
abcefg	bcfgjhk	cdghk	abcefg	bdffghk	cdghk	cdghk	bcdfghk	cdghk	cdghk	bcfgjhk	abcefg	cek	adefghj	bcfgjhk	bcdfghk
abcefg	bcfgjhk	cdghk	abcefg	bdffghk	cdghk	cdghk	bcdfghk	cdghk	cdghk	bcfgjhk	abcefg	cek	adefghj	bcfgjhk	bcdfghk
abcefg	bcfgjhk	cdghk	abcefg	bdffghk	cdghk	cdghk	bcdfghk	cdghk	cdghk	bcfgjhk	abcefg	cek	adefghj	bcfgjhk	bcdfghk
abcefg	bcfgjhk	cdghk	abcefg	bdffghk	cdghk	cdghk	bcdfghk	cdghk	cdghk	bcfgjhk	abcefg	cek	adefghj	bcfgjhk	bcdfghk
abcefg	bcfgjhk	cdghk	abcefg	bdffghk	cdghk	cdghk	bcdfghk	cdghk	cdghk	bcfgjhk	abcefg	cek	adefghj	bcfgjhk	bcdfghk
abcefg	bcfgjhk	cdghk	abcefg	bdffghk	cdghk	cdghk	bcdfghk	cdghk	cdghk	bcfgjhk	abcefg	cek	adefghj	bcfgjhk	bcdfghk
abcefg	bcfgjhk	cdghk	abcefg	bdffghk	cdghk	cdghk	bcdfghk	cdghk	cdghk	bcfgjhk	abcefg	cek	adefghj	bcfgjhk	bcdfghk
abcefg	bcfgjhk	cdghk	abcefg	bdffghk	cdghk	cdghk	bcdfghk	cdghk	cdghk	bcfgjhk	abcefg	cek	adefghj	bcfgjhk	bcdfghk
abcefg	bcfgjhk	cdghk	abcefg	bdffghk	cdghk	cdghk	bcdfghk	cdghk	cdghk	bcfgjhk	abcefg	cek	adefghj	bcfgjhk	bcdfghk
abcefg	bcfgjhk	cdghk	abcefg	bdffghk	cdghk	cdghk	bcdfghk	cdghk	cdghk	bcfgjhk	abcefg	cek	adefghj	bcfgjhk	bcdfghk
abcefg	bcfgjhk	cdghk	abcefg	bdffghk	cdghk	cdghk	bcdfghk	cdghk	cdghk	bcfgjhk	abcefg	cek	adefghj	bcfgjhk	bcdfghk
abcefg	bcfgjhk	cdghk	abcefg	bdffghk	cdghk	cdghk	bcdfghk	cdghk	cdghk	bcfgjhk	abcefg	cek	adefghj	bcfgjhk	bcdfghk
abcefg	bcfgjhk	cdghk	abcefg	bdffghk	cdghk	cdghk	bcdfghk	cdghk	cdghk	bcfgjhk	abcefg	cek	adefghj	bcfgjhk	bcdfghk
abcefg	bcfgjhk	cdghk	abcefg	bdffghk	cdghk	cdghk	bcdfghk	cdghk	cdghk	bcfgjhk	abcefg	cek	adefghj	bcfgjhk	bcdfghk
abcefg	bcfgjhk	cdghk	abcefg	bdffghk	cdghk	cdghk	bcdfghk	cdghk	cdghk	bcfgjhk	abcefg	cek	adefghj	bcfgjhk	bcdfghk
abcefg	bcfgjhk	cdghk	abcefg	bdffghk	cdghk	cdghk	bcdfghk	cdghk	cdghk	bcfgjhk	abcefg	cek	adefghj	bcfgjhk	bcdfghk
abcefg	bcfgjhk	cdghk	abcefg	bdffghk	cdghk	cdghk	bcdfghk	cdghk	cdghk	bcfgjhk	abcefg	cek	adefghj	bcfgjhk	bcdfghk
abcefg	bcfgjhk	cdghk	abcefg	bdffghk	cdghk	cdghk	bcdfghk	cdghk	cdghk	bcfgjhk	abcefg	cek	adefghj	bcfgjhk	bcdfghk
abcefg	bcfgjhk	cdghk	abcefg	bdffghk	cdghk	cdghk	bcdfghk	cdghk	cdghk	bcfgjhk	abcefg	cek	adefghj	bcfgjhk	bcdfghk
abcefg	bcfgjhk	cdghk	abcefg	bdffghk	cdghk	cdghk	bcdfghk	cdghk	cdghk	bcfgjhk	abcefg	cek	adefghj	bcfgjhk	bcdfghk
abcefg	bcfgjhk	cdghk	abcefg	bdffghk	cdghk	cdghk	bcdfghk	cdghk	cdghk	bcfgjhk	abcefg	cek	adefghj	bcfgjhk	bcdfghk
abcefg	bcfgjhk	cdghk	abcefg	bdffghk	cdghk	cdghk	bcdfghk	cdghk	cdghk	bcfgjhk	abcefg	cek	adefghj	bcfgjhk	bcdfghk
abcefg	bcfgjhk	cdghk	abcefg	bdffghk	cdghk	cdghk	bcdfghk	cdghk	cdghk	bcfgjhk	abcefg	cek	adefghj	bcfgjhk	bcdfghk
abcefg	bcfgjhk	cdghk	abcefg	bdffghk	cdghk	cdghk	bcdfghk	cdghk	cdghk	bcfgjhk	abcefg	cek	adefghj	bcfgjhk	bcdfghk
abcefg	bcfgjhk	cdghk	abcefg	bdffghk	cdghk	cdghk	bcdfghk	cdghk	cdghk	bcfgjhk	abcefg	cek	adefghj	bcfgjhk	bcdfghk
abcefg	bcfgjhk	cdghk	abcefg	bdffghk	cdghk	cdghk	bcdfghk	cdghk	cdghk	bcfgjhk	abcefg	cek	adefghj	bcfgjhk	bcdfghk
abcefg	bcfgjhk	cdghk	abcefg	bdffghk	cdghk	cdghk	bcdfghk	cdghk	cdghk	bcfgjhk	abcefg	cek	adefghj	bcfgjhk	bcdfghk
abcefg	bcfgjhk	cdghk	abcefg	bdffghk	cdghk	cdghk	bcdfghk	cdghk	cdghk	bcfgjhk	abcefg	cek	adefghj	bcfgjhk	bcdfghk
abcefg	bcfgjhk	cdghk	abcefg	bdffghk	cdghk	cdghk	bcdfghk	cdghk	cdghk	bcfgjhk	abcefg	cek	adefghj	bcfgjhk	bcdfghk
abcefg	bcfgjhk	cdghk	abcefg	bdffghk	cdghk	cdghk	bcdfghk	cdghk	cdghk	bcfgjhk	abcefg	cek	adefghj	bcfgjhk	bcdfghk
abcefg	bcfgjhk	cdghk	abcefg	bdffghk	cdghk	cdghk	bcdfghk	cdghk	cdghk	bcfgjhk	abcefg	cek	adefghj	bcfgjhk	bcdfghk
abcefg	bcfgjhk	cdghk	abcefg	bdffghk	cdghk	cdghk	bcdfghk	cdghk	cdghk	bcfgjhk	abcefg	cek	adefghj	bcfgjhk	bcdfghk
abcefg	bcfgjhk	cdghk	abcefg	bdffghk	cdghk	cdghk	bcdfghk	cdghk	cdghk	bcfgjhk	abcefg	cek	adefghj	bcfgjhk	bcdfghk
abcefg	bcfgjhk	cdghk	abcefg	bdffghk	cdghk	cdghk	bcdfghk	cdghk	cdghk	bcfgjhk	abcefg	cek	adefghj	bcfgjhk	bcdfghk
abcefg	bcfgjhk	cdghk	abcefg	bdffghk	cdghk	cdghk	bcdfghk	cdghk	cdghk	bcfgjhk	abcefg	cek	adefghj	bcfgjhk	bcdfghk
abcefg	bcfgjhk	cdghk	abcefg	bdffghk	cdghk	cdghk	bcdfghk	cdghk	cdghk	bcfgjhk	abcefg	cek	adefghj	bcfgjhk	bcdfghk
abcefg	bcfgjhk	cdghk	abcefg	bdffghk	cdghk	cdghk	bcdfghk	cdghk	cdghk	bcfgjhk	abcefg	cek	adefghj	bcfgjhk	bcdfghk
abcefg	bcfgjhk	cdghk	abcefg	bdffghk	cdghk	cdghk	bcdfghk	cdghk	cdghk	bcfgjhk	abcefg	cek	adefghj	bcfgjhk	bcdfghk
abcefg	bcfgjhk	cdghk	abcefg	bdffghk	cdghk	cdghk	bcdfghk	cdghk	cdghk	bcfgjhk	abcefg	cek	adefghj	bcfgjhk	bcdfghk
abcefg	bcfgjhk	cdghk	abcefg	bdffghk	cdghk	cdghk	bcdfghk	cdghk	cdghk	bcfgjhk	abcefg	cek	adefghj	bcfgjhk	bcdfghk
abcefg	bcfgjhk	cdghk	abcefg	bdffghk	cdghk	cdghk	bcdfghk	cdghk	cdghk	bcfgjhk	abcefg	cek	adefghj	bcfgjhk	bcdfghk
abcefg	bcfgjhk	cdghk	abcefg	bdffghk	cdghk	cdghk	bcdfghk	cdghk	cdghk	bcfgjhk	abcefg	cek	adefghj	bcfgjhk	bcdfghk
abcefg	bcfgjhk	cdghk	abcefg	bdffghk	cdghk	cdghk	bcdfghk	cdghk	cdghk	bcfgjhk	abcefg	cek	adefghj	bcfgjhk	bcdfghk
abcefg	bcfgjhk	cdghk	abcefg	bdffghk	cdghk	cdghk	bcdfghk	cdghk	cdghk	bcfgjhk	abcefg	cek	adefghj	bcfgjhk	bcdfghk
abcefg	bcfgjhk	cdghk	abcefg	bdffghk	cdghk	cdghk	bcdfghk	cdghk	cdghk	bcfgjhk	abcefg	cek	adefghj	bcfgjhk	bcdfghk
abcefg	bcfgjhk	cdghk	abcefg	bdffghk	cdghk	cdghk	bcdfghk	cdghk	cdghk	bcfgjhk	abcefg	cek	adefghj	bcfgjhk	bcdfghk
abcefg	bcfgjhk	cdghk	abcefg	bdffghk	cdghk	cdghk	bcdfghk	cdghk	cdghk	bcfgjhk	abcefg	cek	adefghj	bc	

Plan 16.11.16. 1/16 replication of 11 factors in 8 blocks of 16 units each.

Factors: A, B, C, D, E, F, G, H, J, K, L.

$I = ABCDJK = ABEFJL = CDEFKL = BCEGJKL = ADEGL = ACFGK = BDFGJ = ABCDEFGH$
 $= EFGHIJK = CDGHIJL = ABGHKL = ADFHJKL = BCFHL = BDEIJK = ACEIJJ.$

Block confounding: DEFG, BCFG, BCDE, ACEF, ACDG, ABEG, ABDF.

Blocks

1	2	3	4	5	6	7	8
(I)	abcd	breg	abef	adeg	cdef	acfg	bdfg
abcdefgh	efgh	adfh	cdgh	bcfh	abgh	bdeh	aceh
defgjl	abefjil	bedfjl	abdajl	afjl	cgjl	acdej	brjl
abehjl	dhjl	ceghjl	efhjl	bdceghjl	abdefhjl	bfghjl	acdfghjl
acefjk	bdefjk	ahfajk	bcjk	cdgjk	adjk	egjk	abedgjk
bdghjk	acghjk	cdehjk	adefghjk	abehjk	befghjk	abedfghjk	fhjk
acdghl	bgkl	abdekl	bedefghl	cekl	uefghl	dfkl	abefghl
befghl	acdefghl	efghkl	ahkl	abdfghkl	bedhkl	abceghkl	drghkl
abdfj	cfj	bhj	dej	acdhj	abcej	bedgj	agj
ceghj	abdeghj	acdefaj	abefghj	befaj	dfghj	acfhj	bedefhj
abegj	cdegl	bdefghl	fgl	acefghl	abcedfghl	becfl	adedfl
cdfl	abfl	ael	abdehl	bd	ehl	adghl	brghl
bedek	aek	abcefhk	acdfk	defhk	bfk	abdefgk	cefgk
afghk	bedfghk	dkg	beghk	abegk	acdeghk	chk	adghk
befghkl	adefghkl	abedghjkl	aceghkl	ghjkl	bdeghkl	ahjkl	cdjkl
adehjk	beehjk	efjkl	bedfghkl	abedefghkl	acfhjkl	cedefghkl	abefghjkl

Plan 16.12.16. 1/16 replication of 12 factors in 16 blocks of 16 units each.

Factors: A, B, C, D, E, F, G, H, J, K, L, M.

$I = ABCDJK = ABEFJL = CDEFKL = BCEGJKLM = ADEGLM = ACFGKM = BDFGJM$
 $= ABCDEFGH = EFGHIJK = CDGHIJL = ABGHKL = ADFHJKLM = BCFHLM = BDEIHKM$
 $= ACEHJM.$

Block confounding: DEFG, BCFG, BCDE, ACEF, ACDG, ABEG, ABDF, ABCDEFGHIJKL,
 ABCHJKL, ADEHJKL, AFGHIJKL, BDGHJKL, BEFHJKL, CDFHJKL, CEHIJKL.

Blocks

1	2	3	4	5	6	7	8
(I)	abdfj	abef	dej	berg	bhj	acfg	chk
abcdefgh	ceghj	cdgh	abefghj	adfh	acdefgj	bdeh	abdefgk
acefjk	bedek	bcjk	acdjk	abfajk	abcefhk	egjk	arfghj
bdghjk	afghk	adefghjk	beghk	cdehjk	dkg	abedfghjk	brdaj
bedejlm	aceflm	acdflm	delm	dajlm	cdehlm	abdefgjlm	bedehjklm
afghjlm	bdghlm	beghilm	adefghilm	abefghilm	abfglm	chjlm	arfghjklm
abdfklm	jklm	deklm	abefjklm	acdefghklm	adfjklm	bedgklm	abedfhlm
ceghklm	abcdefghjklm	abefghklm	cdghjklm	bhklm	bceghklm	acfhklm	eglm
defgjl	abegj	ahkl	fgl	bedfjl	acl	dfkl	becfl
abehjl	cdfl	bedefghkl	abdehl	ceghjl	bdefghl	abceghkl	adghl
acdghl	befghkl	cefhjl	aceghkl	abdekl	cfjkl	acdej	ahjkl
befghkl	adehjk	abdghl	bdfhjkl	cfghkl	abedghjkl	bfghjl	cedefghjkl
bedegm	ardghm	abedehjkm	bedefghm	efm	abdejm	befghjm	dfjm
adehm	befghm	fgjkm	ahjkm	abcdghm	cfghjm	adghjkm	abceghjm
abegjkm	defghm	bdfhm	abdghkm	acjkm	bedfkm	abm	acdekm
cdfhjkm	abekkm	acegm	cefghkm	bdefghjkm	ceghkm	cedefghm	bfghkm
9	10	11	12	13	14	15	16
abcd	cfj	cdef	bfk	adeg	ardhj	bdfg	agj
efgh	abdeghj	abgh	acdeghk	bcfh	befaj	aceh	bedefhj
bdefjk	aek	adjk	abcej	cdgjk	defhk	abedgjk	cefgk
acghjk	bedfghk	befghjk	dfghj	abehjk	abegk	fhjk	abdghk
cejlm	bedflm	bfjlm	cdefjklm	abegjlm	abehlm	cefjlm	abedeglm
bedfghjlm	acghlm	acdeghilm	abghjklm	defghilm	cdglm	abdghilm	fhlm
cfklm	abedjklm	abceklm	adilm	bedfklm	befghklm	agklm	bdffghklm
abdeghklm	efghjklm	dfghklm	befghilm	acdhlkm	adegjklm	bedefhlkm	acchklm
bgkl	cdegl	cgjl	ehl	afjl	bd	brjl	adefl
acdefhkl	abfl	abedfghjl	abedfghl	bedefghjl	acdefghl	acdfghjl	bcghl
abcefgjl	adefghkl	cefghkl	acfhjkl	cekl	abedefjkl	abefkl	cedjkl
dhjl	bedhjk	bedhkl	bedghkl	abdfghkl	ghjkl	deghkl	abefghjkl
cdghjkm	bgjm	bedghm	bedhjm	abedefm	cejm	cdm	abefjkm
abfhjkm	acdefghjm	aceghm	aceghjm	ghm	abdfghjm	abefghm	deghjm
adfgm	abcefgkm	abedfgjkm	abedfghkm	bdjkm	afkm	adefjkm	bekm
beekm	dhkm	ehjkm	egkm	acefghjkm	bedeghkm	bcghjkm	acdfghkm

Plan 64.13.16. 1/64 replication of 13 factors in 8 blocks of 16 units each.

Factors: A, B, C, D, E, F, G, H, J, K, L, M, N.

I: Same as for Plan 64.13.8.

Block confounding: AB, AC, BC, ABCEN, CFN, BEN, AFN.

Blocks only: All two-factor interactions are measurable *except* AB, AC, AD, BC, BD, CD, GH, GJ, GK, HJ, HK, JK.

		Blocks*							
		1	2	3	4	5	6	7	8
(1)	ghjk	efjk	cdgjl	edegkl	acghl	acrfghjkl	adfhj	adehk	
	ehklm	egjlm							
	abcdfhjlm	abcdfghklm							
	abcdefjk	abcdefgh							
	efghlm	efghlm							
	efghlm	fghlmn							
	abcdghlmn	abcdeghlmn							
	abcdghlmn	abcdghlmn							

Plan 64.14.16. 1/64 replication of 14 factors in 16 blocks of 16 units each.

Factors: A, B, C, D, E, F, G, H, J, K, L, M, N, O.

I: Same as for Plan 64.14.8.

Block confounding: ABEO, BCK, ACEKO, ABCM, CEMO, AKM, BEKMO, ACLO, BCEL, ABKLO, EKL, BLMO, AELM, CKLMO, ABCEKLM.

Blocks only: All two-factor interactions are measurable *except* CN and JO.

		Blocks*															
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
(1)	abfht	cdjkm	fmn	cdfjkn	ghjk	cdghm	fghjkmn	cdfghn	abcdghmn								
	cdhklmn	abcdfmn															
	acdfghn	bcdeghn															
	acdfghlm	bdghkm															
	bcdfghlmo	aghjmo	abeghjk	abcdcfgh	abefghjkm	abcdeghkmn	abefghjkmn	abefghjkmn	abefghjkmn	abefghjkmn	abefghjkmn	abefghjkmn	abefghjkmn	abefghjkmn	abefghjkmn	abefghjkmn	abefghjkmn
	bcdfghlmo	acdghlmo															
	abcghlmno	cdghlmno															
	abdehjk	dehjk															

Plan 128.15.16. 1/128 replication of 15 factors in 16 blocks of 16 units each.

Factors: A, B, C, D, E, F, G, H, J, K, L, M, N, O, P.

I: Same as for Plan 128.15.8.

Block confounding: ABD, ACF, BCDF, ABCE, CDE, BEF, ADEF, FJ, ABDFJ, ACJ, BCDJ, ABCEFJ, CDEFJ, BEJ, ADEJ.

Blocks only: All two-factor interactions *except* EL and FJ are measurable.

		Blocks*															
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
(1)	acdghk	acdghl	abeklm	bdghkm	ceghklm	adghjkm	abegh	bcdehjl	dgknp								
	acdfghlmn	acdfghlmn															
	acdfghlop	ceghklp															
	acdkmnp	ghkmnp															
	bcdfghlmo	abfghlmo	acjklp	abcdghlmnp	bjmnp	cdghlmnp	acghjkmnp	abdehkn	bcdehklmp								
	bcdeghlmo	abcghlmo															
	abcghlmnp	bdghlmnp															
	abfghjnp	bcdfhnp															

*Other conditions in the blocks can be created by multiplying the single condition given in each block by the conditions in Block 1 according to the rules for multiplication supplied in the text.

Plan 64.13.8. 1/64 replication of 13 factors in 16 blocks of 8 units each.

Factors: A, B, C, D, E, F, G, H, J, K, L, M, N.

I = ABCD = ABFE LN = CDEFLN = ABGHIL = CDGHIL = EFGHIN = ABCDEFGHIN = ABJKL
= CDJKL = EFJKL = ABCDEFJKL = GHJK = ABCDGHJK = ABFGHJKLN = CDEFGHJKLN
= ACEGJL = BDEGJL = BCFGJN = ADFGJN = BCEIJ = ADEIJ = ACFIJLN = BDFIJLN
= BCEGK = ADEGK = ACFGKLN = BDFGKLN = ACEIKL = BDEIKL = BCEIKN = ADFIKN
= ADLMN = BCLMN = BDEFM = ACEFM = BDGHMN = ACGHMN = ADEFGHLM
= BCFGHLM = BDJKMN = ACJKMN = ADEFJKLM = BCEFJKLM = ADGHJKLMN
= BCGHJKLMN = BDEFGHJKM = ACEFGHJKM = CDEGJMN = ABEGJMN = ABCDFGJLM
= FGJLM = ABCDEHJLMN = EHJLMN = CDEHJM = ABFIJM = ABCDEGKLMN
= EGKLMN = CDFGKM = ABFGKM = CDEHKMN = ABFIKMN = ABCDFHIKLM
= FHIKLM.

Plan 64.14.8. 1/64 replication of 14 factors in 32 blocks of 8 units each.

Factors: A, B, C, D, E, F, G, H, J, K, L, M, N, O.

I = ABCDO = ABFE LNO = CDEFLNO = ABGHLO = CDGHLO = EFGHINO = ABCDEFGHINO = ABJKL
= CDJKLO = EFJKNO = ABCDEFJKL = GHJKO = ABCDGHJK = ABFGHJKLN
= CDEFGHJKLNO = ACEGJLO = BDEGJL = BCFGJN = ADFGJNO = BCEIJ = ADEIJO
= ACFIJLNO = BDFIJLN = BCEGKO = ADEGK = ACFGKLN = BDFGKLN = ACEIKL
= BDEIKLO = BCEIKNO = ADFIKN = ADLMN = BCLMNO = BDEFMO = ACEFM
= BDGHMNO = ACGHMN = ADEFGHLM = BCFGHLMO = BDJKMN = ACJKMO
= ADEFJKLMO = BCEFJKLM = ADGHJKLMO = BCGHJKLMN = BDEFGHJKM
= ACEFGHJKMO = CDEGJMNO = ABEGJMN = ABCDFGJLM = FGJLMO = ABCDEHJLMN
= EHJLMNO = CDEHJMO = ABFIJM = ABCDEGKLMNO = EGKLMN = CDFGKM
= ABFGKMO = CDEHKMN = ABFIKMN = ABCDFHIKLMO = FHIKLM.

Plan 128.15.8. 1/128 replication of 15 factors in 32 blocks of 8 units each.

Factors: A, B, C, D, E, F, G, H, J, K, L, M, N, O, P.

I = ABEGN = ACEFN = BCFGP = DEFGO = ABDFNO = ACDGNOP = BCDFOP = ADIHO
= BDEGHKNO = CDEFIKNOP = ABCDFGHKOP = AEFGIK = BFIKN = CGHIKNP
= ABCFIKP = BCHJNOP = ACEGHJOP = ABFIJO = FGHJNO = BCDEFGHJNP = ACDFHJP
= ABDGHJ = DEIJN = ABCDJKNP = CDEGJKP = BDEFJK = ADFGJKN = ABCEFGJKNOP
= CFJKOP = BGJKO = AEJKNO = ABKLOP = EGKLNOP = BCEFKLNO = ACFGKLO
= ABDEFGKLP = DFKLP = BCDGKLN = ACDEKL = BDHLP = ADEGHLP = ABCDEFHLN
= CDFGHL = BEFGHLP = AFHLP = ABCGHILNO = CEILO = ACIJKLN = BCEGHJKL
= EFHJKLP = ABFGHJKLNP = ACDEFGHJKLNO = BCDFIJKLO = DGHJKLOP
= ABDEHJKLNO = CDJLNO = ABCDEGJLO = ADEFJLOP = BDFGJLNO = CEFGLN
= ABCFJL = AGJLP = BEJLNP = CDGHJMO = ABCDEHJMN = ADEFGHJMNOP
= BDFHJMOP = CEFHJM = ABCFGHJMN = AHJMN = BEGHJMP = ACGJKM = BCEJKMN
= EFGJKMNP = ABFJKMP = ACDEFJKMO = BCDFGJKMNO = DJKMNO = ABDEGJKMOP
= BDGMNP = ADEMP = ABCDEFGM = CDFMN = BEFMNOP = AFGMOP = ABCMO
= CEGMNO = ABGHKMNO = EHIKMOP = BCEFGHKMO = ACFIKMNO = ABDEFHKMNP
= DFGHKMP = BCDHKM = ACDEGHKMN = ABCDGHJKLMP = CDEHJKLMP
= BDEFGHJKLMP = ADFHJKLMP = ABCEFHJKLMP = CFGHJKLMP = BHJKLMP
= AEGHJKLMP = BCGJLMP = ACEJLMP = ABFGJLMP = FJLMO = BCDEFJLMP
= ACDFGJLMP = ABDJLMP = DEGJLMP = ADGKLMNO = BDEKLMO = CDEFGKLMOP
= ABCDFKLMNOP = AEFKLMN = BFGKLM = CKLMP = ABCEGKLMNP = GHLMN
= ABEHLM = ACEFGHLM = BCFHLMNP = DEFHLMNO = ABDFGHLMO = ACDHLMOP
= BCDEGHLMNOP.

**COPY AVAILABLE TO DDC DOES NOT
REPRESENT FULLY LEGIBLE PRODUCTION**

APPENDIX III
PLACKETT AND BURMAN DESIGNS

Designs for $L = 2$ (i. e., two levels per factor). The first row of any design is shown opposite N , the number of experimental conditions (divisible by four) that are equal to or greater than the number of factors to be studied. Plus and minus signs represent (as in other factorial designs) the high and low levels respectively of a factor. A complete design is generated from the row of signs by shifting it cyclically one place for $N-2$ more times. This will create $N-1$ rows including the first one if each time a shift is made the new row of signs is listed below the previous one. A final column of all minus signs is added to make a total of $N-1$ columns. For example, if $N = 4$, and the row were $+ - +$, then the complete matrix would be

+ - + -
- + + -
+ + - -

The rows represent the factors and the columns represent the experimental conditions. The signs show which level for each factor is used to make up the particular experimental condition. If there are fewer than $N-1$ factors, only the appropriate number of rows are used and the rest dropped. All columns will continue to be used, allowing for extra experimental conditions for estimating error.

The following designs were selected from Plackett and Burman's (40) paper:

APPENDIX IV

THREE-LEVEL RESPONSE-SURFACE DESIGNS*

Number of Factors (k)	Design Matrix	No. of Points	Blocking and Association Schemes
5	$\begin{bmatrix} \pm 1 & \pm 1 & 0 & 0 & 0 \\ 0 & 0 & \pm 1 & \pm 1 & 0 \\ 0 & \pm 1 & 0 & 0 & \pm 1 \\ \pm 1 & 0 & \pm 1 & 0 & 0 \\ 0 & 0 & 0 & \pm 1 & \pm 1 \\ 0 & 0 & 0 & 0 & 0 \\ \hline 0 & \pm 1 & \pm 1 & 0 & 0 \\ \pm 1 & 0 & 0 & \pm 1 & 0 \\ 0 & 0 & \pm 1 & 0 & \pm 1 \\ \pm 1 & 0 & 0 & 0 & \pm 1 \\ 0 & \pm 1 & 0 & \pm 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$	$\left. \begin{array}{l} 20 \\ 3 \end{array} \right\} \\ \hline \left. \begin{array}{l} 20 \\ 3 \end{array} \right\} \\ \hline N = 46$	<p>2 blocks of 23 BIB₂(one associate class)</p>
6	$\begin{bmatrix} \pm 1 & \pm 1 & 0 & \pm 1 & 0 & 0 \\ 0 & \pm 1 & \pm 1 & 0 & \pm 1 & 0 \\ 0 & 0 & \pm 1 & \pm 1 & 0 & \pm 1 \\ \pm 1 & 0 & 0 & \pm 1 & \pm 1 & 0 \\ 0 & \pm 1 & 0 & 0 & \pm 1 & \pm 1 \\ \pm 1 & 0 & \pm 1 & 0 & 0 & \pm 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$	$\left. \begin{array}{l} 48 \\ 6 \end{array} \right\} \\ \hline N = 54$	<p>2 blocks of 27. First Associates: (1, 4); (2, 5); (3, 6).</p>
7	$\begin{bmatrix} 0 & 0 & 0 & \pm 1 & \pm 1 & \pm 1 & 0 \\ \pm 1 & 0 & 0 & 0 & 0 & \pm 1 & \pm 1 \\ 0 & \pm 1 & 0 & 0 & \pm 1 & 0 & \pm 1 \\ \pm 1 & \pm 1 & 0 & \pm 1 & 0 & 0 & 0 \\ 0 & 0 & \pm 1 & \pm 1 & 0 & 0 & \pm 1 \\ \pm 1 & 0 & \pm 1 & 0 & \pm 1 & 0 & 0 \\ 0 & \pm 1 & \pm 1 & 0 & 0 & \pm 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$	$\left. \begin{array}{l} 56 \\ 6 \end{array} \right\} \\ \hline N = 62$	<p>2 blocks of 31. BIB (one associate class).</p>
9	$\begin{bmatrix} \pm 1 & 0 & 0 & \pm 1 & 0 & 0 & \pm 1 & 0 & 0 \\ 0 & \pm 1 & 0 & 0 & \pm 1 & 0 & 0 & \pm 1 & 0 \\ 0 & 0 & \pm 1 & 0 & 0 & \pm 1 & 0 & 0 & \pm 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline \pm 1 & \pm 1 & \pm 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \pm 1 & \pm 1 & \pm 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \pm 1 & \pm 1 & \pm 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline \pm 1 & 0 & 0 & 0 & \pm 1 & 0 & 0 & 0 & \pm 1 \\ 0 & 0 & \pm 1 & \pm 1 & 0 & 0 & 0 & \pm 1 & 0 \\ 0 & \pm 1 & 0 & 0 & 0 & \pm 1 & \pm 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline \pm 1 & 0 & 0 & 0 & 0 & \pm 1 & 0 & \pm 1 & 0 \\ 0 & \pm 1 & 0 & \pm 1 & 0 & 0 & 0 & 0 & \pm 1 \\ 0 & 0 & \pm 1 & 0 & \pm 1 & 0 & \pm 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \hline \pm 1 & 0 & 0 & \pm 1 & 0 & 0 & \pm 1 & 0 & 0 \\ 0 & \pm 1 & 0 & 0 & \pm 1 & 0 & 0 & \pm 1 & 0 \\ 0 & 0 & \pm 1 & 0 & 0 & \pm 1 & 0 & 0 & \pm 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$	$\left. \begin{array}{l} 24 \\ 2 \end{array} \right\} \\ \hline \left. \begin{array}{l} 24 \\ 2 \end{array} \right\} \\ \hline \left. \begin{array}{l} 24 \\ 2 \end{array} \right\} \\ \hline \left. \begin{array}{l} 24 \\ 2 \end{array} \right\} \\ \hline \left. \begin{array}{l} 24 \\ 2 \end{array} \right\} \\ \hline N = 130$	<p>(a) 5 blocks of 26. (b) 10 blocks of 13. First Associates: (1, 4); (1, 7); (4, 7); (2, 5); (2, 8); (5, 8); (3, 6); (3, 9); (6, 9).</p>

[*Adapted from Box and Behnken (7).]

(continued)

(Appendix IV, continued)

Number of Factors (k)	Design Matrix	No. of Points	Blocking and Association Schemes
10	$\begin{bmatrix} 0 & \pm 1 & 0 & 0 & 0 & \pm 1 & \pm 1 & 0 & 0 & \pm 1 \\ \pm 1 & \pm 1 & 0 & 0 & \pm 1 & 0 & 0 & 0 & 0 & \pm 1 \\ 0 & \pm 1 & \pm 1 & 0 & 0 & 0 & \pm 1 & \pm 1 & 0 & 0 \\ 0 & \pm 1 & 0 & \pm 1 & 0 & \pm 1 & 0 & 0 & \pm 1 & 0 \\ \pm 1 & 0 & 0 & 0 & 0 & 0 & 0 & \pm 1 & \pm 1 & \pm 1 \\ 0 & 0 & \pm 1 & \pm 1 & \pm 1 & 0 & 0 & 0 & 0 & \pm 1 \\ \pm 1 & 0 & 0 & \pm 1 & 0 & 0 & \pm 1 & \pm 1 & 0 & 0 \\ 0 & 0 & \pm 1 & 0 & \pm 1 & 0 & \pm 1 & 0 & \pm 1 & 0 \\ \pm 1 & 0 & \pm 1 & 0 & 0 & \pm 1 & 0 & 0 & \pm 1 & 0 \\ 0 & 0 & 0 & \pm 1 & \pm 1 & 0 & \pm 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$	<p>160</p> <p>10</p> <p>$N = 170$</p>	<p>2 blocks of 85.</p> <p>Second Associates: (1, 8); (1, 9); (1, 10); (2, 6); (2, 7); (2, 10); (3, 5); (3, 7); (3, 9); (4, 5); (4, 6); (4, 8); (5, 10); (6, 9); (7, 8).</p>
11	$\begin{bmatrix} 0 & 0 & \pm 1 & 0 & 0 & 0 & \pm 1 & \pm 1 & \pm 1 & 0 & \pm 1 \\ \pm 1 & 0 & 0 & \pm 1 & 0 & 0 & 0 & 0 & \pm 1 & \pm 1 & 0 \\ 0 & \pm 1 & 0 & 0 & \pm 1 & 0 & 0 & 0 & 0 & \pm 1 & \pm 1 \\ \pm 1 & 0 & \pm 1 & 0 & 0 & \pm 1 & 0 & 0 & 0 & \pm 1 & \pm 1 \\ \pm 1 & \pm 1 & 0 & \pm 1 & 0 & 0 & \pm 1 & 0 & 0 & 0 & \pm 1 \\ \pm 1 & \pm 1 & \pm 1 & 0 & \pm 1 & 0 & 0 & 0 & \pm 1 & 0 & 0 \\ 0 & \pm 1 & \pm 1 & \pm 1 & 0 & \pm 1 & 0 & 0 & \pm 1 & 0 & 0 \\ 0 & 0 & \pm 1 & \pm 1 & \pm 1 & 0 & \pm 1 & 0 & 0 & \pm 1 & 0 \\ 0 & 0 & 0 & \pm 1 & \pm 1 & \pm 1 & 0 & \pm 1 & 0 & 0 & \pm 1 \\ \pm 1 & 0 & 0 & 0 & \pm 1 & \pm 1 & \pm 1 & 0 & \pm 1 & 0 & 0 \\ 0 & \pm 1 & 0 & 0 & 0 & \pm 1 & \pm 1 & \pm 1 & 0 & \pm 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \pm 1 & \pm 1 & \pm 1 & 0 & 0 \end{bmatrix}$	<p>176</p> <p>12</p> <p>$N = 188$</p>	<p>Use 2^{11-4} fractionated on x_2, x_3, x_4, x_5.</p> <p>No orthogonal blocking.</p> <p>BIB (one associate class)</p>
12	$\begin{bmatrix} \pm 1 & \pm 1 & 0 & 0 & \pm 1 & 0 & \pm 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \pm 1 & \pm 1 & 0 & 0 & 0 & \pm 1 & 0 & \pm 1 & 0 & 0 & 0 \\ 0 & 0 & \pm 1 & \pm 1 & 0 & 0 & \pm 1 & 0 & \pm 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & \pm 1 & \pm 1 & 0 & 0 & \pm 1 & 0 & \pm 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & \pm 1 & \pm 1 & 0 & 0 & \pm 1 & 0 & \pm 1 & 0 \\ \pm 1 & 0 & 0 & 0 & 0 & 0 & 0 & \pm 1 & \pm 1 & 0 & 0 & \pm 1 \\ 0 & \pm 1 & 0 & 0 & 0 & 0 & 0 & 0 & \pm 1 & \pm 1 & 0 & 0 \\ \pm 1 & 0 & \pm 1 & 0 & 0 & 0 & 0 & 0 & \pm 1 & \pm 1 & 0 & 0 \\ 0 & \pm 1 & 0 & \pm 1 & 0 & 0 & 0 & 0 & 0 & \pm 1 & \pm 1 & 0 \\ 0 & 0 & \pm 1 & 0 & \pm 1 & 0 & 0 & 0 & 0 & 0 & \pm 1 & \pm 1 \\ \pm 1 & 0 & 0 & \pm 1 & 0 & \pm 1 & 0 & 0 & 0 & 0 & 0 & \pm 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$	<p>192</p> <p>12</p> <p>$N = 204$</p>	<p>2 blocks of 102.</p> <p>First Associates: (1, 7); (2, 8); (3, 9); (4, 10); (5, 11); (6, 12).</p>

Notes

- Unless otherwise indicated the symbol $(\pm 1, \pm 1, \dots, \pm 1)$ means that all combinations of plus and minus levels are to be run.
- Whenever it is possible to use a fractional factorial of Resolution V, it is used in place of the full factorial. This is the case with the design for 11 variables.
- The dotted lines indicate how the designs for 5 and 9 variables can be blocked.
- Designs for 6, 7, 10, and 12 can be blocked by allocating any trials in which the product of the signs of the levels are positive to one block and any in which they are negative to another block. (E.g., $-1, -1, 1, 0$ would be a positive trial; $-1, -1, -1, 0$ would be considered a negative trial.)
- Designs for 9 variables are already blocked according to the dotted lines, but blocking can be doubled by further dividing them into positive and negative trials as per note 4.

DISTRIBUTION LIST FOR REVISED EDITION

Defense Documentation Center
Cameron Station
Alexandria, VA 22314

Education Research Information Center
Processing & Reference Facility
4833 Rugby Ave., Suite 303
Bethesda, MD 20014

NASA - Scientific & Technical Information Facility
P. O. Box 33
College Park, MD 20740

National Technical Information Services (NTIS)
Operations Division
5285 Port Royal Road
Springfield, VA 22151

Executive Editor
Psychological Abstracts
American Psychological Assoc.
1200 Seventeenth St., N. W.
Washington, D. C. 20036

AD-A035 108

HUGHES AIRCRAFT CO CULVER CITY CALIF ENGINEERING EQU--ETC F/G 5/5
ECONOMICAL MULTIFACTOR DESIGNS FOR HUMAN FACTORS ENGINEERING EX--ETC(U)
JUN 73 C W SIMON
HAC-P73-326A
F44620-72-C-0086
NL

UNCLASSIFIED

3 OF 3
AD
A035108



SUPPLEMENTARY
INFORMATION



END
DATE
FILMED
1 -78
DDC

SUPPLEMENTARY

INFORMATION

AD-A035108

To isolate the effect of AB, these four sources along with their corresponding performance values should be summed, as indicated by the signs in the AB column. This would yield:

$$4I - 2A + 2C + 2E + 2G + 4AB = +12$$

The effects of CD, EF, and GH have been cancelled out. Since the mean (I), A, C, E, and G would be known from the estimates already obtained from the data of the Basic and A. D. 2. designs, by proper arithmetic substitution and simplification, the effect of AB can be determined.

Interaction CD can be obtained in the same way. This time the four sources are combined by subtracting (cg) and (ce) from (AB+CD+EF+GH) and (eg). This causes the signs of all components of (cg) and (ce) to be reversed, of course, and when the four sources are now summed, all of the interactions except CD will cancel out. The remnants of I, A, C, E, and G will be eliminated as before by substituting the appropriate values already obtained from completing the Basic and Augmentation designs.

To isolate the effect of EF, (eg) and (ce) must be subtracted from (AB+CD+EF+GH) and (cg). To isolate the effect of GH, (eg) and (cg) must be subtracted from (AB+CD+EF+GH) and (ce).

I. D. 3. To separate members of a string of three-factor interactions.

No examples will be given, but it is apparent that the same logical approach can be applied to any set of confounded data. In each case, the following steps would be required:

- 1) To reduce the effort the experimenter can first try to logically eliminate certain of the aliased effects.
- 2) At least (N - 1) additional experimental conditions must be used for N aliased effects in a string.
- 3) The rows of signs of the aliased effects to be isolated must be made orthogonal to one another.

Corrected page for Technical Report No. P73-326A, "Economical Multifactor Designs for Human Factors Engineering Experiments" by C.W. Simon.