AD-A056 984 CANYON RESEARCH GROUP INC WESTLAKE VILLAGE CALIF F/G 5/10
NEW RESEARCH PARADIGM FOR APPLIED EXPERIMENTAL PSYCHOLOGY: A SY--ETC(U)
JUN 78 C W SIMON F44620-76-C-0008
UNCLASSIFIED CWS-04-77A AFOSR-TR-78-0117-REV NL

1 OF 2
AD A056984

AD A056984

AD No.
DDC FILE COPY

LEVEL III

DDC
PROCESSED
JUL 28 1978
D

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

TR 77-25-438

AD A056984

AD No.
DDC FILE COPY

(18) AFOSR-TR-78-0117-REV. (14)

LEVEL III

(1)

(6) NEW RESEARCH PARADIGM FOR APPLIED EXPERIMENTAL PSYCHOLOGY: A SYSTEM APPROACH. Revision

(10) Charles W. Simon

(11) Jun 78 (12) 135 P.

(9) Technical rept. Sep 76 - Sep 77,

Report No. CWS-04-77A

(14)

A50 309

(15)

Research sponsored by the Air Force Office of Scientific Research (AFSC), United States Air Force, under Contract No. F44620-76-C-0008.

The United States Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation hereon.

ACCESSION for
NTIS White Section X
DDC Buff Section
UNANNOUNCED
JUSTIFICATION
Per Basic Rpt.
BY AD-A050209
DISTRIBUTION/AVAILABILITY CODES
Dist. AVAIL. and/or SPECIAL
A

DDC
RECEIVED
JUL 31 1978
D

October 1977
(Revised, June 1978)

Canyon Research Group, Inc.
741 Lakefield Road, Suite B
Westlake Village, California 91361

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

78 07 25 068 391 185 pt.

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: CWS-04-77A
2. GOVT ACCESSION NO.:
3. RECIPIENT'S CATALOG NUMBER:

4. TITLE (and Subtitle): NEW RESEARCH PARADIGM FOR APPLIED EXPERIMENTAL PSYCHOLOGY: A SYSTEM APPROACH

5. TYPE OF REPORT & PERIOD COVERED: Technical Report Sept. 1976 - Sept. 1977

6. PERFORMING ORG. REPORT NUMBER:

7. AUTHOR(s): Charles W. Simon

8. CONTRACT OR GRANT NUMBER(s): F44620-76-C-0008

9. PERFORMING ORGANIZATION NAME AND ADDRESS:
Canyon Research Group, Inc.
741 Lakefield Road, Suite B
Westlake Village, CA 91361

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS:

11. CONTROLLING OFFICE NAME AND ADDRESS:
Air Force Office of Scientific Research
Bolling Air Force Base
Washington, D. C. 20332

12. REPORT DATE: October 1977

13. NUMBER OF PAGES: 123

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office):

15. SECURITY CLASS. (of this report): UNCLASSIFIED

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE:

16. DISTRIBUTION STATEMENT (of this Report)

DISTRIBUTION STATEMENT A
Approved for public release; Distribution Unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

Superseded AD-A050209

18. SUPPLEMENTARY NOTES

CWS-04-77, revised June 1978 from original document, becomes CWS-04-77A.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Experiments
Experimental method
New research paradigm
Human factors research
Experimentalist
Manipulation experiments
Holistic approach
Economical multifactor designs

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

To spotlight the crisis occurring in scientific psychology, warnings, complaints, and criticisms by prominent psychologists and non-psychologists are quoted. The traditional experimental method of the past hundred years does not produce the data needed to solve problems faced by a modern society.

(continued)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

The results from many experiments cannot be combined into a modular, quantitative data base.

The traditional experimental method, as it is applied in engineering psychology, is critically examined. Characteristic procedures and concepts are analyzed to show why they have been counterproductive to the ultimate experimental goals -- prediction of field performance and aggregation of data for future use.

The two distinct approaches to data collection used by empirical psychologists, "Experimental" and "Correlational," are defined and compared. Arguments are presented for merging the more productive features of each. However, contrary to similar suggestions in the past, the combined approach would emphasize the point of view of the Experimentalists. This means that the manipulative method would be retained but in a way that would enable the holistic philosophy of the Correlationists to dominate.

Philosophy, strategy, and techniques are described that can be combined into a new research paradigm for experimental psychology. A sequential process is proposed that enables systematic multifactor experiments to be performed with great economy. Problems are first defined with a real-world orientation. Next, using primarily manipulative procedures, fifty to one-hundred candidate factors from equipment, environment, personnel, and task sources, can be screened systematically to identify the non-trivial ones. These non-trivial factors for the particular task are then subjected to further investigation, the data from which being combined with that from the screening study to produce a response surface as defined by a polynomial of the appropriate degree. This equation is then refined, minimizing both bias and random error, the fiducial limits determined, and the resulting product verified under operational conditions. Further refinement may be required.

The feature that makes this paradigm unique is that the total data-collection process for deriving an equation of all critical variables affecting an operational task will ordinarily be less than that used in four and five factor experiments using traditional methodology. The consequences are that prediction from laboratory data to field performance becomes a reality and a quantitative data base for future reference can be constructed.

The chess-board is the world, the pieces are the phenomena of the universe, the rules of the game are what we call the laws of Nature. The player on the other side is hidden from us. We know that his play is always fair, just, and patient. But also we know, to our cost, that he never overlooks a mistake or makes the smallest allowance for ignorance.

Thomas H. Huxley, Lay Sermons, Addresses and Reviews (1870)

We do not have a simple event A causally connected with a simple event B, but the whole background of the system in which the events occur is included in the concept, and is a vital part of it.

Percy W. Bridgman, The Logic of Modern Physics (1927)

ACKNOWLEDGMENTS

This report was prepared under contract with the Air Force Office of Scientific Research, Air Force Systems Command, United States Air Force, under Prime Contract No. F44620-76-C-0008 with Canyon Research Group, Inc. Dr. Alfred Fregly was contract monitor.

This report summarizes seven years of work and thought on new ways to perform psychological research. A great many people have helped to make the work possible and I have recognized their contributions in the earlier reports of this "advanced methodologies series." Two, however, directly or indirectly during the entire period, were responsible more than any others for the existence and continued support of this program. Dr. Stanley N. Roscoe, University of Illinois at Urbana-Champaign, provided the initial support needed to get the program started, and Dr. Charles E. Hutchinson, formerly of the Air Force Office of Scientific Research, saw that the support remained long enough to enable the program to reach a definitive stage. But more than the gift of time and money, both offered the encouragement needed to sustain the effort and a freedom from bureaucratic pressures that ultimately strangles creativity. Their contributions are gratefully acknowledged.

FOREWORD

Twenty-five years ago in graduate school I refused to do an experiment for a professor because I didn't think the two-factor study would prove anything and because I didn't know how to include nine important factors in the same experiment, except at exorbitant costs. Since then, I have spent a great deal of time trying to find ways of including more factors in a single experiment since, to do less, I believe, seldom provides much useful information.

About a decade ago I came across some novel designs in papers by G. E. P. Box; three years later I obtained a contract to look into improved methods of doing psychology experiments. While Box's work was directed more toward research in the chemical engineering industry, it contained many features that made it particularly appropriate for research in engineering psychology. Even more important than his ingenious experimental designs was his research strategy. From then on, each literature search revealed other techniques never mentioned in school -- and still aren't in psychology departments -- which would give an experimental psychologist exceptional power in sampling an experimental space economically and in analyzing the data more completely. I became aware of the works of Daniel, Hoerl and Kennard, and Gnanadesikan, to name a few. I began to collect classes of techniques -- economical data sampling methods, methods of minimizing irrelevant effects, and methods of analyzing correlated data and handling multiple responses. A whole new way of doing experiments presented itself and for the first time I realized it was practical to do an experiment in which twenty or thirty factors could be manipulated, and critical, uncontrolled variables included. Instead of a mere smorgasbord of techniques, I recognized the nucleus for an approach that represented an oblique departure from traditional experimental psychology.

Less than two years ago, I finally realized that all of these methods actually fit into a connected pattern, a paradigm for research, that combines the most effective features of experiments in which variables are manipulated and controlled and of studies in which data is recorded as it occurs and analyzed for understanding later. Furthermore, although combining these two approaches has been a dream of other psychologists -- most recently Cattel and Royce -- my paradigm is the first to keep the integrity of the "scientific" method -- manipulation and control -- intact, while working in the context of a holistic philosophy. For the first time, insofar as I know, it is possible to include twenty-five, fifty, or even one-hundred factors in a single experiment and derive a mathematical equation defining an operational space from laboratory data. Equally important is the fact that this can be done with an incredible economy in data collection. The paradigm is viable and practical.

This report provides a somewhat prosaic overview of the paradigm. It tells why and what, but not how. "How" must be learned by reading the earlier reports that I have written and some of the original papers from which the techniques were taken, or by attending my "advanced methodologies" seminar. While experience will probably bring changes in specific tactics, the general philosophy and strategy should remain intact. Though some refinement may be required, for all practical purposes, an informed investigator could use the paradigm immediately. If the paradigm is used -- properly -- I am convinced that it will markedly improve the quality and utility of experimentally derived information, and will do so in a highly cost-effective manner.

Charles W. Simon
1977

TABLE OF CONTENTS

Section		Page
I.	INTRODUCTION	1
	CONTENTS OF THIS REPORT	2
	REFERENCING POLICY	5
II.	THE GROWING DISCONTENT	6
	PSYCHOLOGY IN A CRISIS	6
	DISILLUSIONMENT IN SPECIFIC FIELDS	9
	The Secession of Practicing Psychology	10
	Human Resources Research	11
	PSYCHOLOGICAL DATA AS VIEWED FROM OUTSIDE THE PROFESSION	15
	CAVEAT EMPTOR	18
	METHODOLOGY	19
III.	EXPERIMENTAL METHODS OF ENGINEERING PSYCHOLOGY	23
	ENGINEERING PSYCHOLOGY	24
	CONSEQUENCES OF THE "TRADITIONAL" APPROACH?	26
	Avoiding Real-world Complexity	27
	Why Not Look at More Variables?	28
	The "Small" Study Paradox	29
	Testing the Insignificant	30
	Identifying Critical Variables	31
	Ignoring Individual Differences	32
	The Impossible Dream -- Aggregation	33
IV.	THE TWO EMPIRICAL PSYCHOLOGIES	35
	EXPERIMENTAL VERSUS CORRELATING PSYCHOLOGY.	38
	Hypotheses	38
	Manipulation	41
	Universal Laws and Indifidual Differences.	43
	Unilateral Multifactor Studies	44
	Reduction Experiments	45
	Significance Testing	47
	Scientific Orientation	47

Table of Contents (Continued

Section		Page
	CONSOLIDATION -- A NEW EXPERIMENTAL PSYCHOLOGY	48
	Limitations	50
V.	A NEW EXPERIMENTAL PARADIGM	52
	GENERAL STRATEGY	54
	PRINCIPLES	55
	SPECIFIC STRATEGIES	56
	SUMMARY	59
VI.	PHASE ONE: DEFINING THE PROBLEM	62
	A REAL-WORLD ORIENTATION	63
	LIMITING THE EXPERIMENT	63
	OBJECTIVES	66
	INFORMATION SOURCES	68
	TALENTS INVOLVED IN DESIGN OF EXPERIMENT	69
	PRE-EXPERIMENTAL ANALYSIS	71
	Classification	73
	PRELIMINARY EMPIRICAL INVESTIGATIONS	76
	Identifying Primary Factors in Composite Variables	76
	Determining Weights of Multiple-Responses.	77
	Parametric Verification Studies	77
VII.	PHASE TWO: IDENTIFYING CRITICAL VARIABLES	79
	PRINCIPLE OF MALDISTRIBUTION	80
	SCREENING	80
	THE ECONOMY OF SCREENING DESIGNS	81
	CHOICE OF SCREENING DESIGNS	83
	Screening a Very Large Number of Variables	84
	Supersaturated designs	84
	Group screening designs	85
	Screening a Large Number of Individual Variables	87
	ISOLATING INTERACTION EFFECTS	90
	REPLICATION	90

Table of Contents (Continued)

Section		Page
	ANALYZING SCREENING DESIGNS	92
	MULTIPLE RESPONSE (DEPENDENT) VARIABLES	93
VIII.	PHASE THREE: DEVELOPMENT OF RESPONSE SURFACES	96
	FIRST-ORDER RESPONSE SURFACES	96
	Center Point Data	97
	Testing the Adequacy of the First-order Model	98
	Scaling and Transformation	100
	Extending the Screening Plan	101
	SECOND-ORDER RESPONSE SURFACES	103
	Non-critical Variables	104
	Replicating the Second-order Design	105
	Testing the Adequacy of the Second-order Model	105
	ANALYZING CONTROLLED AND UNCONTROLLED VARIABLES TOGETHER	106
IX.	PHASE FOUR: EQUATION REFINEMENT	107
	REDUCING THE UNEXPLAINED VARIANCE	107
	IMPROVING THE FIT OF THE RESPONSE SURFACE	107
	CONFIDENCE LIMITS	108
	EXPANDING THE EXPERIMENTAL SPACE	108
X.	PHASE FIVE: VERIFICATION	109
XI.	CONCLUSIONS	112
XII.	REFERENCES	113
APPENDIX I.	PHILOSOPHICAL DIFFERENCES BETWEEN THE OLD AND NEW EXPERIMENTAL PARADIGMS FOR HUMAN ENGINEERING RESEARCH	121

LIST OF TABLES

		Page
TABLE 1.	CRITICAL MILESTONES LEADING UP TO PSYCHOLOGY'S EMERGENCE AS A DISTINCT SCIENCE OF HUMAN BEHAVIOR	37
TABLE 2.	COMPARISON OF MAJOR FEATURES CHARACTERIZING TRADITIONAL CORRELATIONAL AND EXPERIMENTAL PSYCHOLOGY	39
TABLE 3.	FEATURES OF THE NEW PARADIGM DERIVED BY COMBINING THE BEST METHODOLOGIES OF CORRELATIONAL AND EXPERIMENTAL PSYCHOLOGY	53
TABLE 4.	RELATIONSHIP AMONG PHASES, GOALS, AND METHODOLOGY AS THE EXPERIMENT PROGRESSES . . .	61

I. INTRODUCTION

Thousands of psychologists perform and publish rigorous experiments each year. It is difficult to believe that so many can play the game (Dunnette, 1966, p 344) as strongly as they do without believing that their work has some social significance. Yet, unless they are totally isolated from the "real world," they cannot fail to realize that most of the data being generated is seldom used and, in fact, is often useless. The results from formal psychology experiments have generally failed to provide the data needed to quantitatively predict performance under operational situations. Furthermore, it has not been possible to combine experimental results from related studies into a single cohesive, quantitative data base.

For over a decade, articles have been published in the American Psychologist and other psychology journals, that are critical of our research results and some of our most cherished methodologies. And yet, in these same journals, papers continue to appear that perpetuate the flood of trivial data and improper and inappropriate techniques. The situation has progressed to a point where persons outside the psychological community are reacting and rejecting what was once considered to be time-honored "science."

Analysis of the traditional methods of performing rigorous ("scientific") psychology experiments reveals grossly inadequate rituals, shibboleths, and methods. Experiments in which the primary variables are manipulated have studied far too few factors to ever expect to account for performance

variations under operational conditions, and too often, these few factors have had only trivial effects.

For historical (and to some extent hysterical) reasons psychologists have nurtured a research paradigm for over one-hundred years that, on average, has failed to do the job intended and desired. In the face of mounting criticism, the old paradigm has persisted -- the result of indifference, inertia, ignorance, and most of all, a failure to find a fully satisfactory alternative.

In this report, the need for a new paradigm, its desirable features, and description, will be presented. Its use will markedly improve the accuracy with which performance under operational conditions can be predicted from experimental data and will provide the information needed to build a quantitative data base.

CONTENTS OF THIS REPORT

There are twelve sections to this report. The purpose of the second section is to present to those readers who remain complacent about the informative and social value of formal psychological experimentation, the growing evidence that all is not well. While we produce many experiments, we do not produce much useful information. To quote Koch (1969, p 66), "Throughout psychology's history as 'science,' the hard knowledge it has deposited has been uniformly negative." In this first section, prominent psychologists and non-psychologists who warn, complain, or criticize to some degree the failure of our scientific data are quoted. While the sample is small, its blue-ribbon quality is impressive. In or out of context, these quotations signal the need for change.

The third section takes a sharp look at the traditional approach to engineering psychology. If a change of experimental paradigm is necessary, some understanding of why that is so is needed first. Revered concepts and rituals that psychologists have lived with for more than a century in the belief that these make psychology a "science" and science makes everything right are examined (briefly) critically. Hallowed principles of good research are questioned and found wanting when "good" refers to the quality of the experimental data rather than to the degree to which certain procedures are carried out ceremoniously. Bakan (1965, p 189) wrote regarding the experimental psychologist's love affair with "hypothesis testing": "One is tempted to think that psychologists are often like children playing cowboys. When children play cowboys they emulate them in everything but their main work, which is taking care of cows. The main work of the scientist is thinking and making discoveries of what was not thought beforehand. Psychologists often attempt to 'play scientist' by avoiding the main work."

The fourth section points out the differences between the two principal empirical approaches to the understanding of human behavior. Cronbach (1957) labeled them "experimental," wherein behavior was studied by manipulating it, and "correlational," wherein on-going behavior was analyzed. Since the time psychology became accepted as a science over a century ago, Experimentalists and Correlationists have been "strangers in Paradise," but unwilling to hold hands in spite of occasional efforts over the years to encourage it. Each of these disciplines has some good information-gathering

features which, if combined, could better serve the experimentalists' purpose. Arguments supporting the benefits of merging are put forth, but from the point of view of the Experimentalist.

The fifth section introduces a new paradigm for "scientific research." "An experiment is only a subdesign within the larger design of a total scientific investigation" (Cattell, 1966a, p 11). Even to understand and predict human behavior in a single task, the information-gathering process must take several forms as the investigation progresses. The course of the research program and the methodologies required for each phase are described. The chief feature of the new paradigm is its ability to handle a very large multifactor problem in all its complexity and investigate it systematically using classical manipulative techniques. Philosophy, strategy, and techniques are brought together to create an alternative and more viable paradigm for formal psychological experimentation, particularly as it is employed in human factors engineering research.

The sixth, seventh, eighth, ninth, and tenth sections each cover a different phase of the paradigm. These involve: defining the problem, identifying the critical variables, developing the response surface, refining the equation, and verifying the experimental results, respectively.

The final two sections are the conclusions and the references.

REFERENCING POLICY

In this report, an attempt has been made to provide a description of a complete paradigm which, in fact, has not been completed. Some techniques discussed here have been investigated over the past seven years for the sole purpose of melding them into an overall research methodology; some have not. This distinction is reflected in the referencing procedure. Where methods have been culled, modified, and integrated into the advanced methodology approach by Simon, reference will be made to his reports rather than to the original papers from which the techniques were borrowed. Where methods have not been fit specifically into the new paradigm, but are included here since they appear to be appropriate, reference will be made to the authors of the original papers. This policy is intended to provide the reader with the information in its most relevant form. Once the overall approach is understood, the reader may wish to review all original papers, including those found as references in Simon's reports.

II. THE GROWING DISCONTENT

While some psychologists have always been concerned with applied problems, the majority of those who taught and did research before World War II did so mainly to satisfy their own individual curiosities and to publish. Today, behavioral research has become big business. The federal and state governments support most of the research performed by psychologists. Large laboratories in university and military organizations produce research structured to governmental needs and even "basic" research must be mission-oriented (Bryan, 1972). Today relevance has become the key word; pressure both within and outside the psychological community has increased for research results that can be used to solve the practical problems faced by a complex society. Both practitioners and scientists are being besieged for useful information.

The lack of useful experimental results is bringing about what Deese (1972, p 1) refers to as a "state of crisis" in psychology. The extent of this crisis is reflected in the warnings from prominent psychologists in many fields as well as those outside the psychological community. Only a few of these will be cited here.

PSYCHOLOGY IN A CRISIS

In 1952, the American Psychological Association appointed Sigmund Koch to plan and direct a study of the status of psychology. The study, subsidized by the National Science

Foundation, brought together about 80 scientists to assess the facts, theories, and methods of psychology. Seventeen years later, Koch (1969, p 14) summarized his personal feelings regarding the "science of psychology" in this way:

> Whether as a 'science' or any kind of coherent discipline devoted to the empirical study of many, psychology has been misconceived. This is no light matter for me to confess after a 30-year career given to exploration of the prospects and conditions for psychology becoming a significant enterprise.
>
> But the massive 100-year effort to erect a discipline given to the positive study of man can hardly be counted a triumph. Here and there the effort has turned up a germane fact, or thrown off a spark of insight, but these victories have had an accidental relation to the programs believed to inspire them, and their sum total over time is heavily over-balanced by the pseudo-knowledge that has proliferated.

George Miller (1969), in his presidential address to the American Psychological Association, noted that while scientific psychology has the tremendous potential to influence every aspect of society, the actual contributions of the field of psychology to the solution of the social problems have been disturbingly insignificant.

Morris Viteles (1972, p 601), in a talk before the XVIIth Congress of Applied Psychology, asked: "What does the psychologist know about human behavior to which he can attest with confidence, or at least with a degree of confidence considerably in excess of that characterizing

psychology as a science in the past?" He answered himself by saying: "The search for answers to this question has brought conviction that advances in knowledge during the past 50 to 75 years have been considerably more limited than might be anticipated from reading textbooks or other publications in psychology, and from observing the activities of practitioners of psychology."

Leona Tyler (1973, p 1021), in her presidential address to the American Psychological Association, while reviewing the progress of modern scientific psychology, reiterated the same theme. She said: "As the twentieth century wore on, psychological knowledge increased enormously, and psychologists assumed respected and influential positions. But somehow the hopes for continuous improvement in the conditions of mankind through psychology declined. It became almost naive to assume that what was discovered through research could have much effect on man's nature or institutions. . ."

Cronbach (1975, p 116), in his Distinguished Scientific Contribution Award address, wrote: "Some 30 years ago, research in psychology became dedicated to quest for nomothetic theory.* Model building and hypothesis testing

---

*Cronbach (1975) defines "nomothetic theory" as one that would ideally tell us the necessary and sufficient conditions for a particular result" (p 125).

became the ruling idea, and research problems were increasingly chosen to fit that mode. Taking stock today, I think most of us judge theoretical progress to have been disappointing. Many are uneasy with the intellectual type of psychological research."

DISILLUSIONMENT IN SPECIFIC FIELDS

Disillusionment with the results in specific fields of psychology illustrates just how widespread and close to the grass roots so much of the dissatisfaction really is.

Elms (1975, p 968) writes of the "widespread self-doubts about goals, methods, and accomplishments" of the social psychologists, citing that "similar doubts have been expressed recently within many other areas of psychology, particularly the closely related fields of personality research (Carlson, 1971; Fiske, 1974), developmental psychology (Wohwill, 1973), and clinical psychology (Albee, 1970; Farberow, 1973)."

The title of Robert Lockard's (1971) article: "Reflections on the fall of comparative psychology: is there a message for us all?" speaks for itself. In his opening paragraph he wrote: "What we once knew as comparative psychology has been overrun by a scientific revolution. In the wake of that revolution lies the debris of what was once a traditional branch of psychology, now a confused scatter of views of nature, problems, and methods. The confusion persists for the same reason the revolution occurred; psychologists understood one view of behavior, but

not another, and it was the other that won out." He attributed its demise -- "most psychologists misunderstood what was happening at the time" -- to its irrelevance to the whole of psychology. He attempted to show how a relevant discipline could produce irrelevant results by examining its historical premises and traditions, coincidentally with the rest of psychology.

The Secession of Practicing Psychologists

Disillusionment with "scientific" psychology is expressed in yet another way. In fields where both "science" and practice flourish, the locus of training -- once solidly in the Department of Psychology -- is now being separated, leaving the "scientists" to be trained in the psychology departments and the practitioners to be trained in other departments.

George Albee (1970), in his 1970 presidential address to the American Psychological Association, spoke of the "uncertain future of clinical psychology" Bemoaning the lack of relevance that occurs in the training of clinical psychologists, Albee suggested that perhaps a more effective practitioner might be developed if he were trained separately from the "scientist" aspect as emphasized in current graduate school curricula.

Herbert H. Meyer, (1972, p 608), in his 1971 presidential address to the Division of Industrial and Organizational Psychology, began by saying: "Over the last few years, I have been haunted by uneasy feelings about the future of industrial and organizational psychology. . . . trends in our field indicate that our capability of meeting this

challenge is declining rather than advancing." He tells of the trend to move industrial psychology out of the Departments of Psychology and into the Schools of Business Administration.

Lipsey (1974) surveyed 2340 graduate students and 368 faculty members in psychology and found that although 92% of the students and 83% of the faculty thought academic psychology should be concerned with contemporary social problems, 90% of the students and 79% of the faculty said they did not think that academic psychology was making a significant contribution to needed solutions. Fifty-one percent of the students and 52% of the faculty felt that academic psychology does not yet have much knowledge relevant to social problems.

Human Resources Research

Nor has human resources research, i.e., selection, training, and equipment design, escaped criticism. While testing, learning, and psychophysical experiments are often considered among the most successful types of research, yet there are serious indications that this optimism is exaggerated.

In the area of selection research, Ghiselli (1966) wrote that ". . . though some few specific tests do give reasonably good prediction of job proficiency in the industrial occupations as a whole, the general picture is one of quite limited power." Uhlaner (1967, p 2) expressed his concern ". . .with the limited usefulness of information coming out of many personnel research studies, particularly research studies

dealing with selection, the prediction of human performance, and the measurement of aptitudes and abilities for differential classification."

Nor does training research fare any better. Mackie and Christensen (1967, p 4-5) noted that while research on learning processes represents perhaps the largest single area of investigation presently being pursued by experimental psychologists . . . both academic and practically oriented psychologists agree that a very small percentage of findings from learning research is useful, in any direct sense, for the improvement of training or educational purposes." How to design flight simulators for pilot training has been an important research question for more than two decades, yet Adams (1972, pp 616-617) writes: "I would not consider the money being spent on flight simulators as staggering if we knew much about their training value, which we do not. We build flight simulators as realistically as possible . . . which is a cover-up for our ignorance about transfer because in our doubts we have made costly devices as realistic as we can in the hopes of gaining as much transfer as we can." Psychologists have been working on the problem of transfer and training for more than half a century, yet the results from those experiments provide only superficial guidance in the design of training programs and simulators. Caro (1973, p 508) said it this way: "Perhaps we build simulators as realistically as possible because people who design them do not know much about training. Or, perhaps it is because those who design them know that those who use them do not know much about training, and the safest thing to do is to build simulators like aircraft." In 1977, after surveying

factors affecting training simulator effectiveness, Caro (1977, p 84-85) wrote: "Except to the extent that general learning concepts may be applied to the simulator training situation, few research-based guidelines exist for the simulator training program developer to follow in establishing his training program," and later, ". . . instances were noted in which practices did not make full use of available information about human learning and performance."

But would the judgments be different if the experimental variables were easier to define, as in the problems of equipment design? Not if these comments are at all representative. Adams (1972, p 615) in his presidential address to the Society of Engineering Psychologists, American Psychological Association, stated bluntly: "Our research efforts have been and are insufficient. The future of engineering psychology is in jeopardy unless we examine what we know and how to strengthen it."

Alphonse Chapanis (1963), prolific both as a generator and critic of research in human factors and applied psychology, reviewed the research in Engineering Psychology for a chapter in the 1963 Annual Review of Psychology. He complained that "a distressing amount of literature in engineering psychology is not very good. Moreover, the flaws are not minor methodological faults, but are serious methodological ones which often invalidate the author's conclusions" (p 311). In discussing the gap between research and application, he noted: "In human factors work, however, research appears to take second place to everyday experience in designing, developing, and operating real systems."

Four years later in an article on the relevance of laboratory studies to practical situations, Chapanis (1967) stated: "It appears that if you want to use the results of laboratory experiments to solve practical problems, you should do so with extreme caution. Although the results of laboratory experiments sometimes provide you with ideas and hunches that may be worth trying out in practical situations, you would be rash to generalize naively from laboratory findings to the solution of real world problems." Later in the same article, he observed that ". . . we often do not find in practical situations the results we would have predicted from laboratory experiments."

Meister and Sullivan (1967) studied the extent to which handbooks of human factors information met the needs of aircraft designers and influenced their designs. They concluded that ". . . the human factors discipline is not providing the information required to solve design problems, nor is what it does provide furnished in a manner which is most usable by designers" (p 3).

Few areas so aptly illustrate the inadequacy of our research as do the experiments on visual perception. Originally a classic problem of psychophysics, later one of major concern in experimental psychology, and more recently a fundamental consideration in the research on applied military problems of target acquisition, hundreds of visual perception experiments have been carried out in the laboratory and under operational conditions. Simon (1971, Appendix A) cited comments made over a 14-year period by sixteen persons who tried to collect and synthesize

results of this research so that they could be applied to the design of visual systems. The comments of Greening and Snyder (1968) are typical. After reviewing the studies on visual air-to-ground target acquisition, they concluded:

> "The wide divergence between experimental results from study to study, and the evident importance of many uncontrolled variables, make it unwise to attempt to make quantitative synthesis of existing target acquisition data" (p 73).

Later they stated:

> "No one has yet demonstrated the ability to predict acquisition performance with even modest accuracy over any substantial range of meaningful situations" (p 78).

Little has changed in the intervening years.

Meister (1976) surveyed a representative group of human factors teachers and specialists "of recognized stature" on major issues in human factors. He concluded: "There appears to be almost unanimous agreement that the application of human factors research to system development projects has been less than optimal, and in some cases rather poor" (p 375).

PSYCHOLOGICAL DATA AS VIEWED FROM OUTSIDE THE PROFESSION

In 1971, the U. S. Supreme Court attacked what has always been a virtual monument to the relevance of psychological research -- its personnel tests. Since that time, as a result of the U. S. Supreme Court decision (Curtis, 1971), tests used for hiring and promotion purposes can be challenged

if they evaluate predictors simply by testing the statistical significance of correlation coefficients. Today it is necessary that the pragmatic nature of the test's predictive value be proven. Vitelis (1972, p 604) wrote that ". . . industrial psychologists might well bow their heads in shame in noting that it has been found necessary by the Supreme Court of the United States to remind them of the obligation to validate tests against objective and realistic criteria as a preliminary to their use for selection and classification purposes in industry."

In 1975, Congressmen criticized the National Science Foundation and National Institute of Health for supporting social science programs to study such problems as why children fall off tricycles ($19,200), a dictionary of witchcraft ($46,089), why people fall in love ($132,500), and the use of uterine birth control devices by unmarried college students ($342,000). While admitting that some projects with funny-sounding titles can "have a sound basis for their existence in the budget," in general, projects of this type were referred to as "boondoggles" that waste taxpayers money (Goldwater, 1976).

At about the same time, the U. S. House of Representatives voted to cut millions of dollars of funding from human resources and manpower effectiveness programs requested by the U. S. Department of Defense for 1976 as well as a special Navy exploratory development fund. This cut represented approximately a 50 percent reduction in the funding available for most military human factors programs and

mainly hit programs labeled "basic research." The committee recommending the cuts questioned "both the utility and priority" of such programs (Price, 1975).

In 1977, Congress again threatened to cut over half of the requested funds, nearly $40 million, from the military budget for training, simulation, and related topics (Human Factors Society Bulletin, 1977). This time, as in the first case, a part of this money was eventually reinstated; yet the very acts showed how the value of this research was being questioned. In spite of the fact that 61% of the 1976 Department of Defense Budget went to personnel-related expenses, only one-tenth of a cent per dollar expenditure went to supporting human resources research.

The Controller General of the United States (1977) asked eight Department of Defense research and development organizations to identify human resources R and D reports published during 1973 through 1975 which were intended to support changes to regulations, policies, manuals, training programs, and equipment. Of the 374 that were reported, 164 were not used. In 39 cases, the reason given was because the results were questionable.

Few have expressed their distaste for human factors program as picturesquely as Admirial H. G. Rickover (1970). Asked to comment on a proposal involving a major human factors program in the research, development, engineering, and production of Navy ships, he answered: "It appears that the Human Factors 'program' is another of the fruitless

attempts to get things done by systems, organizations, and big words rather than by people. It contains the greatest quantity of nonsense I have ever seen assembled in one publication. It is replete with obtuse jargon and sham-scientific expressions which, translated into English from its characteristic argot where this is possible, turns out to be either meaningless or insignificant. It is about as useful as teaching your grandmother how to suck an egg."

CAVEAT EMPTOR

Defenders of the faith may argue that these comments represent a biased selection, are taken out of context, and appear more discouraging than those making them intended. In some cases, these criticisms are true to a limited extent. However, too many comments such as these are being made by too many prominent men in too many fields of psychology over too extended a period of time to be ignored. Among the many hundreds of thousands of formal experiments that have been performed, it is too difficult to find a handful that have been directly responsible for definitive solutions to practical problems. If the battle has not been completely lost, at least the odds against us are enormous. The viability of the profession, and our responsibility to our customers, demands that the body of psychologists -- not the few -- make a serious effort to discover why things are not as they should be and do something to correct the cause. Collectively, we are selling tarnished goods. Is it enough to continue to produce as long as we warn: Let the buyer beware? Obviously the answer is "no."

Before ending this section, therefore, it is appropriate that we offer an explanation as to why our experimental results have been unsatisfactory. What has been common to these different branches of psychology, across basic and applied research alike, that could so seriously degrade the effectiveness of their experimental results? The answer is the methodology. This observation has not escaped a number of psychologists.

METHODOLOGY

Today's methods reveal their roots at the beginnings of psychology as a science. Methodology made psychology a "science." Emmanuel Kant had denied psychology that appellation because he believed that quantitative methods could not be applied to behavioral data. Wilhelm Wundt's psychophysical methods made a liar out of Kant, and psychology -- uncertain with its new status -- grabbed at whatever it could find to keep it. The natural science became the model for their experimental methods. The experimentalists who manipulated their variables looked upon themselves as the true scientists. While other psychologists, more concerned with observing natural phenomena, developed innovative techniques with an emphasis on analysis rather than control, the experimentalists maintained the "scientific method" -- with markedly little change until the present time -- whatever the cost. Maslow (1970, p 343) had this to say regarding this rigidity:

> These then are termed the "laws of scientific method." Canonized, crusted about with tradition and history, they tend to become

> binding upon the present day (rather than merely suggestive or helpful). In the hands of the less creative, the timid, the conventional, these "laws" become virtually a demand that we solve our present problems only as our forefathers solved theirs.

The deficiencies in these "scientific methods" have revealed themselves in both applied and basic behavioral research.

Silverman (1971), in an article, "Crisis in Social Psychology," reflected on ". . . why social psychologists have not provided much data that are relevant to social ills." Then, answering his own question, he noted: "If the multitude of social-psychological findings cannot aid the planners of society, it is apparently not because we have been researching the wrong topics. It must be that our data are not generalizable to the objects of our studies in their natural, ongoing states. This is a basic inadequacy of methodology rather than direction, and it will not be resolved by pontifical edicts from any source about what to study and where."

Lipsey (1974, p 553) examined another area of research and concluded:

> The position we associated with the basic researcher--defined by both disinterest in social problems and commitment to experimental methodology--constitutes the dominant tradition which is under attack and susceptible to change. Even its

> methodology, seen by many as the sine qua non of science, receives less support from the upcoming generations of psychologists than among most of its current faculty practitioners.

Bakan (1972, p 86), ever critical of the gossamer nature of our experimental methods, wrote: "I think that now we are in a period of transition -- for the status of the sciences in general and for psychology itself. In the last decade we have begun to question the unquestioned belief that fact-module experimental research is a panacea for man's problems; the payoffs of this research have been smaller than we had hoped for."

Gadlin and Ingle (1975, p 1003) begin a critique of psychological methodology by saying:

> E. G. Boring (1950) once said that the application of the scientific method to the study of human behavior would count as mankind's greatest achievement. Few people today would unhesitatingly agree with such a statement; still fewer could share its opinion. Even those who think that the wish of William James [to help psychology to become a natural science] has been fulfilled are uncertain of the consequences. For a multiplicity of reasons, psychologists are questioning the natural science methodology that has dominated the field since its inception. Much of this inquiry has focused on the laboratory experiment. . . .
>
> Psychologists have come to question the experiment [which they limit to <u>laboratory</u> experiments, which examines dependent

> variables in light of manipulation performed upon independent variables] as a means to describe and comprehend reality.

Perhaps it is not the experimental method that is inadequate, but the psychologists' interpretation of what the experimental method is and how it should be used that is inadequate. "It ain't what we do but the way that we do it" that needs revision. With this direction in mind, let us first begin by examining what we do. Let us take a look at the traditional experimental paradigm to see if we can find what went wrong?

III. EXPERIMENTAL METHODS OF ENGINEERING PSYCHOLOGY

Traditional experimental psychologists employ certain characteristic methods that affect the problems, techniques, attitudes, assumptions, and even myths associated with the design, conduct, analysis, and interpretation of experiments. Typically, the traditional experimental psychologist, in his research:

- Seeks universal laws regarding the behavior of average man.
- States and tests specific hypotheses.
- Manipulates known experimental, independent variables of interest and attempts to hold constant any others.
- Assumes causal relations between independent and dependent variables in unilateral bi- or multi-factor situations.
- Uses reduction experiments in which fewer than five variables are usually investigated.
- Uses factorial designs (or variations thereof) and performs tests of statistical significance.

Of these, the requirement to manipulate and control variables is the characteristic that most differentiates traditional experimental psychology from other approaches.

ENGINEERING PSYCHOLOGY

Engineering psychology is that branch of applied experimental psychology concerned with the appropriate design of devices, equipment, systems, and environments in order to optimize the performance of the man-machine complex. Unlike psychologists involved in selection and training, who try to improve system performance by taking advantage of individual differences among people, engineering psychologists perform experiments to discover equipment characteristics that facilitate the performance of typical people of a particular class. The research methods of this field, on the whole, have remained those of traditional experimental psychology.

One popular textbook on "Research Techniques in Human Engineering" (Chapanis, 1959) illustrates this point. It defines an experiment as "a series of controlled observations undertaken in an artificial situation with the deliberate manipulation of some variables in order to answer one or more specific hypotheses" (p 148). The underlined terms reflect what traditional experimental psychologists have come to accept as important features of experiments in human research. Control helps eliminate extraneous effects and enables the experiment to be repeated by others if desired. The artificial situation enables unusual conditions to be studied at the experimenter's convenience and with more control of extraneous factors than would be the case were the situation studied under operational circumstances. Systematic manipulations of the independent variables help untangle complex

effects and identify causal relationships. Stating a specific hypothesis provides a concrete direction for the experimental effort.

Other statements in the book exemplify commonly accepted concepts and methods that characterize the experimental psychologist's approach to research. For example:

> An experimental design should always be constructed before the investigator actually starts collecting data. (p 151).
>
> The design should yield a measure of the random error in the experiment. (p 151).
>
> Do not confound variables. (p 156).
>
> Factorial experimental designs make up one major class of multi-variable experiments and constitutes one of the most important basic designs you will need in human engineering work. (p 176).
>
> When we say that an experiment is well controlled, we mean that the experimenter has examined all of the possible relevant variables in his experiment and has tried to hold all of them (except the ones he deliberately designs into the experiment) constant. (p 220).
>
> The best you can hope to do [to handle individual differences] is to test <u>enough</u> subjects so that you can get a dependable measure of average performance and some estimate of the amount of variability you can expect to find. (p 236).
>
> . . . it will be a rare human engineering experiment that will give you definitive results with only two or three subjects. (p 238).

Suffice it to say, the approach used in the paradigm presented later differs markedly from the previous statements.*

CONSEQUENCES OF THE "TRADITIONAL" APPROACH?

A survey of 14 years of research published in the journal, Human Factors, (Simon, 1976b) revealed that in 239 experiments, 92% of the experiments studied three or fewer variables; the median number of levels per variable was three. The median number of repeated measures per data point was nine; this means that on average 89% of the effort was spent collecting redundant information. The median numbers of observations used in studies of 1, 2, 3, 4 and 5 variables were 72, 180, 192, 768, and 1200, respectively. Thirty-one percent of the total variance in the experiments was accounted for by the experimental variables and their interactions. In some individual studies, the experimental variables failed to account for even 1% of the total variability in the experiment. Quite often the largest sources of variance were consigned to the "error" term even though they were obviously unidentified subject effects or subject-by-condition interactions associated with such sequential effects as learning and transfer. The analysis also revealed that 24% of main effects, each accounting for less than 1% of the total performance variability in the experiment, were still designated as "statistically significant," which was invariably interpreted to mean "important" by the investigator.

---

*The nature of these differences in research philosophy, while spread throughout the discussion of the New Paradigm, are summarized for those particular statements in Appendix A.

In summary, the traditional experimental method as exemplified by the above data:

1. Looks at too few variables in a single experiment.
2. Collects far too much data for the number of effects that must be estimated.
3. Fails to account for much of the performance variability in the experiment (which means it would account for even less outside the laboratory where many more variables are operating).
4. Studies and identifies many effects which are in fact trivial.
5. Generally considers individual differences to be a nuisance.

Let us examine the consequence of each of these deficiencies.

Avoiding Real-world Complexity

What is wrong with studying only two or three variables in a single experiment? This is the essence of the reduction experiment, so effective in the natural sciences: eliminate all sources of variance to see if the one of immediate interest has an effect. Still, it isn't sufficient if one wishes to describe or predict human behavior; in the real world, phenomena are too complex to be explained by a few variables. In order to obtain results that can be generalized from the laboratory to the operational situation, the experiment must describe that world <u>in all its complexity</u>, rather than deny this complexity by "eliminating" critical variables. In practice, of course, when a three-factor

experiment is planned, the other variables are not actually eliminated from the experiment. Instead, either their presence in the experiment is ignored, or they are held constant, sometimes at a zero value. When existing variables are ignored, the kind of unexplained variance observed in so many human factors experiments will occur and will result in variable error when operational conditions are predicted. On the other hand, whenever variables are held constant in the laboratory at values that are different in the field, a biased error is introduced into the prediction.

Even the laboratory data will be distorted if an interaction between two variables cannot be revealed because one of the variables is held constant. Human behavior is situation-specific; results obtained in the laboratory can only be generalized to comparable conditions in the field. If we limit the number of variables below the number required to adequately describe the complexity of the world, or misrepresent their values when they are not varied, our description of the real world will be incomplete, our predictions erroneous, and our generalizations limited.

Why Not Look at More Variables?

If it is desirable to look at more than two or three variables in a single experiment, why haven't more experimenters done so? What is it about the traditional approach that makes sampling performance in a multifactor space so difficult? Two decades ago, Williams and Adelson (1954) investigated the problem of experimentally determining the design parameters for a variable characteristic, pilot-training simulator. Their analysis indicated that 34

simulator characteristics were critical in the design of the training simulator; they believed these should be studied at five levels each. They noted, however, that the traditional factorial analysis of variance design for that purpose would involve 5^{34}, or 5.8×10^{23} combinations of equipment variables under which performance must be measured. This, they concluded, would be "manifestly impossible." This illustrates quite vividly the absurdities that occur when one tries to extend the traditional approach to problems of this magnitude and complexity. Here cost of doing research is not the problem; such an approach would not be possible at any cost.

The "Small" Study Paradox

Faced with the enormity of conducting a factorial study, Williams and Adelson considered ways of reducing the number of conditions to be investigated. They suggested doing 34 different studies and varying a different variable each time over five steps while holding the remaining 33 variables constant. This would require that performance be measured under 170 experimental conditions. Since more than one measure would be needed to provide some stability to the measures at each of the five levels per variable, they proposed to test 20 subjects at each experimental condition. This plan was discarded when calculations revealed it would require 3400 subjects and 17,000 flying hours. Furthermore, with this approach, there would be no information regarding interaction among variables. Thus this illustrates the contradiction that arises when an experiment is limited to only a few variables in order to make the data collection

task more economical. The economy is a false one, for while less data is collected on the variables of interest, more data, and redundant data at that, must be collected in order to achieve stability in the measures and there is a loss of information about interactions.

Testing the Insignificant

Psychologists have traditionally replicated their designs in order to do tests of statistical significance. Here we have a second paradox noted by Meehl (1967), namely that the very process of replicating to increase the precision of the data, decreases the confidence in generalizing the results from a test of statistical significance. With more replications, the likelihood of finding statistically significant effects increases, while the likelihood that these effects will be critical under operational conditions decreases.

An example of the statistical significance trap can be seen in a study published in Human Factors (Vartabedian, 1971) in which the effects of three variables on seeing letters on a CRT display were examined. The investigator collected more than 3,000 observations. The investigator concluded that one of the three variables was statistically significant. However, this significant variable improved detection performance in the experiment by less than one-half second, which was trivial for the task at hand. In fact, all three experimental factors and their interactions combined accounted for less than 1% of the total performance variability in the experiment. This means that the unexplained variance accounted for 99% of the observed variability. Only because of the enormous number of observations that were made was it possible

to calculate a "statistically" significant effect that neither is of practical importance nor likely to occur in the real world. Numerous authors (e.g., Bakan, 1971; Kleiter, 1969; Lykken, 1968; Nunnally, 1960; Rozeboom, 1960; Signorelli, 1974) have shown how little information significance tests really provide, as well as how frequently they have been misused and misinterpreted, succeeding only in providing an undeserved halo for what would otherwise be trivial effects. Obtaining significance has traditionally overridden every other objective for most experimenters in engineering psychology in spite of the fact it is possible to obtain it for almost any situation by merely increasing the number of replications. Discovering small effects is a worthy endeavor after the large effects are understood.

Identifying Critical Variables

Continuing to search for ways of reducing the magnitude of the effort to study the 34 simulator variables, Williams and Adelson also considered the possibility of limiting their investigation to only those variables that were truly important to the particular training problem. Once again, the limitation of this idea became quickly evident. There is no economical way of choosing the most important variables. The 34 variables that had been proposed a priori already represented without additional empirical evidence, the minimum set that ought to be considered from both a psychological and engineering point of view. This illustrates another weakness of the traditional approach. Because each "experiment" is planned completely ahead of time and is run as an undivided, uncompromising entity unto itself, the functions of identification and description are totally confounded. The set-in-concrete pre-experimental design

stifles the investigative research process. It is unable to cope with the demands to identify and describe through a sequential and iterative experimental process.

Ignoring Individual Differences

Psychologists engaged in equipment design research tend to treat individual differences as a nuisance. When "subject" variance cannot be isolated, individual differences are included as part of the "error" term; when it can be isolated, once calculated, it is usually ignored when the data is interpreted. For example, in a recent transfer of training experiment (Koonce, 1975), almost 80% of the total variance in the experiment was accounted for by differences in pilot performance within conditions and only 2% was accounted for by a statistically significant interaction which occurred when groups trained with different simulator motion conditions performed in the simulator and in the aircraft. The pilot characteristics accounting for these large subject differences were never identified. Yet had they been, the value of the experimental results would have been considerably enhanced were they to be applied to the operational situation. Furthermore, this would have enabled potential interactions between specific subject characteristics and the equipment to be investigated, thereby reducing the possibility of drawing erroneous conclusions regarding equipment design parameters.

Jacobs and Roscoe (1975), in another transfer of training study, took steps to correct this by isolating the effects of pilot aptitude from the data intended to study the effects on

performance of different types of simulator motions. This procedure increased their understanding of the effects of the equipment variable.

The Impossible Dream -- Aggregation

Can the results from small experiments be combined? As an inadequate and inappropriate methodology forced the acceptance of small studies, psychologists began to rely on an implicit assumption that once the results from a great many small studies were obtained, they could be combined in building-block fashion to build a cohesive, quantitative data base. Information could be drawn from this pool of fundamental knowledge to solve new and complex problems. Unfortunately, this hope has never been fully realized in psychology, at least not with any quantification or acceptable precision.

Greening and Snyder (1967) concluded from a survey of visual research data what most seasoned researchers have found to be true in other problem areas. They said: "There is no straightforward way to select data from a number of field and simulator studies and combine the whole into a comprehensive representation of the effects of one or more variables" (p 73), and "It has not been possible to blend the data from either the laboratory studies or the field studies or any combination of the two in order to deduce simple relationships among the important variables" (p 81).

In part, this situation has occurred for obvious reasons. In many experiments, the value of a variable that is held constant is seldom reported, while the values of those ignored are unknown. This prevents the data about the experimental variables from being properly located in a multi-dimensional coordinate space. The results from several studies, therefore, can never be precisely related. Even if this were corrected, the present size of a study is still too small to supply the "clumps" of data required for any stability.

IV. THE TWO EMPIRICAL PSYCHOLOGIES

Experimental psychologists are not the only ones who do research on human behavior. Within its history, scientific psychology has shown a distinctly forked development, "two historic streams of method, thought, and affiliation" which Cronbach (1957) labeled "Experimental psychology" and "Correlational psychology."* These two disciplines differ in their philosophies, methods of inquiry, areas of interests, and loci of application. Psychologists associated with each discipline differ in their training, where they publish, their professional heroes, and even their personalities (Cronbach, 1957, p 671). It is the methodological differences that are of primary concern in this report.

The methods of the Experimental psychologist were copied originally from those used by experimental physiologists and the natural scientists, in particular, nineteenth century physicists. Later psychologists borrowed quantitative methods used in agricultural and engineering research. Correlational psychology, on the other hand, was an outgrowth of the biological sciences, getting its start when Sir Francis Galton, concerned with human heredity, measured individuals on a large scale. To handle his data, he invented the method of correlation. Later methods for studying

---

*Correlationists will argue that their approach is just as "experimental" as that of the Experimentalists. However, in this report, any reference to Experimental psychology or Experimentalists will be in the historical context to refer to neo-Wundtians who manipulate and control their variables.

individual differences were developed, and these in turn sparked the development of more sophisticated statistical tools for analyzing data. Prior to their emergence as distinct disciplines, both Experimental and Correlational psychology had a common heritage in the mathematics of probability and the practical applications of Gauss' normal curve (see Table 1).

Cronbach (1957, p 671) suggests that "the experimental method -- where the scientist changes conditions in order to observe their consequences -- is much the more coherent of our two disciplines. Everyone knows what experimental psychology is and who the experimental psychologists are . . . In contrast to the Tight Little Island of the experimental discipline, correlational psychology is a sort of Holy Roman Empire whose citizens identify mainly with their own principalities. The discipline, the common service in which the principalities are united, is the study of correlations presented by Nature."

However, when he refers to "Correlational psychology" Cronbach does not refer to studies relying on one statistical procedure, but to any effort to relate natural phenomena through post-observational analysis. He says: "The correlator's mission is to observe and organize the data from Nature's experiments. As a minimum outcome, such correlations improve immediate decisions and guide experimentation. At best, a Newton, a Lyell, or a Darwin can align the correlations into a substantial theory" (p 672).

TABLE 1. CRITICAL MILESTONES LEADING UP TO PSYCHOLOGY'S EMERGENCE AS A DISTINCT SCIENCE OF HUMAN BEHAVIOR

Scientific methods: empirical observation and hypothesis testing	Bacon (1561-1626)
Problems of gambling; invented mathematics of chance	Bernoulli (1654-1705)
Requirement for quantitative data in science; denial of psychology as a science	Kant (1724-1804)
Definitive book on probability; method of least squares	LaPlace (1749-1827)
Concept of an absolute threshold or lower limit of sensation	Herbart (1776-1841)
Normal curve applied to scientific observations; means, probable error	Gauss (1777-1855)
Concept of just noticeable difference and just noticeable increment proportional to stimulus	Weber (1795-1878)
Normal curve and elementary statistics applied to methods of biological and social data: astronomy, weather, birth, deaths, marriages, diseases, crime, anthropometric measures	Quetelet (1796-1874)
Psychophysics; $S = C \log R$	Fechner (1801-1877)
Correlation, standard scores, median; invented and applied to studies of heredity individual differences. (Beginning of Correlational Psychology)	Galton (1822-1911)
First psychological laboratory (1879) (Beginning of the Experimental Psychology)	Wundt (1832-1920)

EXPERIMENTAL VERSUS CORRELATIONAL PSYCHOLOGY

Major features that distinguish the research of the Experimental psychologists from that of the Correlational psychologists are shown in Table 2. Let us briefly examine some implications of each in turn.

Hypotheses

Consistent with Sir Francis Bacon's experimental method, Experimentalists have been taught that each experiment must begin with a hypothesis. A hypothesis, whether precisely or casually stated or presented as a statement or a question, does serve to orient the direction an experiment will take and forces the investigator to resolve a particular question. On the other hand, the requirement that a hypothesis is necessary sometimes has created the impression that the purpose of all experiments is to verify hypotheses when in fact some experiments are conducted in order to develop hypotheses.

In practice, hypotheses used by experimental psychologists, when verbalized precisely, are generally very simple, seldom profound enough to justify an expensive formal study and often too limited or too vague to precisely account for any important aspect of human behavior. When psychologists began to use Fisher's analysis of variance for hypothesis-testing purposes, they lost sight of the distinction between scientific and statistical hypotheses. Interest is usually high in the former, but our analytic methods are only capable of testing the latter (Bakan, 1971).

TABLE 2. COMPARISON OF MAJOR FEATURES CHARACTERIZING TRADITIONAL CORRELATIONAL AND EXPERIMENTAL PSYCHOLOGY

CORRELATIONIST	EXPERIMENTALIST
Seeks understanding of how and why individuals differ	Seeks universal laws regarding the behavior of the average man
Asks what happens under observable circumstances	States and tests specific hypotheses
Observes, measures, and classifies situations	Manipulates known independent variables of interest and attempts to hold others constant
Studies relationships between independent and dependent variables in bilateral multivariate situations	Studies relationships -- assumed causal -- between multiple independent and single dependent variables (unilateral multifactor)
Accepts total situation with its realistic complexity	Employs reduction experiments in which fewer than five variables are usually considered an acceptable number
Employs various analytical methods based on a regression model	Employs analysis of variance as a primary model with emphasis on factorial designs
Seeks practical answers	Seeks scientific principles but has had little success in consolidating facts from independent experiments

Even when less formal hypotheses are used, expressed as generalized questions, intuitions, or merely reasons for conducting experiments, they still tend to restrict the problem, the approach, and even the solution. A hypothesis, inferring the question: "Does such-and-such a thing happen?", forces the Experimentalist to be in the position of performing an experiment to determine whether or not Nature has agreed with his perception of the situation. The Correlationist reverses this position and asks: "What does happen?", and performs his studies to discover "what hath God wrought?" While there is undoubtedly a place in psychology for both kinds of questions, in general, psychologists have been premature in their hypothesis testing (Bass, 1974, p 874). Engineering psychologists have continued to use hypothesis testing because they believe it's the "right thing to do," often stopping at the very point -- the test -- where their research should have begun to answer the problems in which they are interested. Their limited repertoire of experimental techniques has made hypothesis testing -- which serves to identify reliable differences -- a means to a final answer rather than a beginning of an investigation to discover functional relationships between operator performance and critical equipment, system, and environmental parameters. Quite often in problems of equipment and system design, having hypothesis-testing as the primary experimental goal, results in an experiment structured to test a limited number of alternative configurations among which the experimentalist hopes a best one will be found. The engineer, forced to balance task-related performance criteria against cost and engineering technology, would be better served if the data were provided as a functional description of all critical parametric relationships.

Manipulation

Manipulating and controlling independent variables are the most important research tools unique to the Experimentalists. By varying the effects of interest (and holding all other sources of variance constant), an investigator can determine how much the response changes when predictor variables are changed by prescribed amounts. The ability to manipulate and control factors so that specific values of each can be studied -- the experimental design -- enables effects of factors and their interactions to be estimated separately although in Nature they might in fact be correlated. This makes the task of interpretation easier and helps identify those variables having the greatest influence on performance. Through manipulation and control, the investigator can be more confident that he has identified causal relationships among variables.

There are some drawbacks however with the manipulative process. For one thing, factors that might critically affect performance cannot always be controlled; they may neither be manipulated nor held constant. Frequently when this is the case, Experimentalists will allow such factors to vary uncontrolled, expecting to compensate for the perturbations by collecting larger quantities of data and averaging, by randomizing their designs, and by performing significance tests in their analyses.

Another difficulty with the manipulation process is that it forces the investigator to consciously decide what to manipulate; in some cases this means he must know in advance which factors have the greatest effect on the particular

performance. This unfortunately will not usually be the case. An investigator may know what he is interested in but this may not be the same as knowing what he should be interested in. As a result he may waste considerable effort investigating trivial factors, ignoring crucial ones.

Of course, the Correlationalists also have difficulty knowing which factors are important. Since they ordinarily do not manipulate or control their variables, the data they collect must not only thoroughly describe performance on the task they are observing but also the situation in which this performance occurs. If they fail to measure the critical aspects of that situation, then they may be no more able to explain the behavior of interest than the Experimentalist. If they should happen to measure critical aspects of the task but not be aware that they are critical, their ability to explain and understand the observed behavior will also be limited. In this case, however, unlike the Experimentalist who must manipulate and control the experimental conditions in advance, two things are in the Correlationist's favor. One, if he is lucky enough to record the right data, the Correlationist may have several chances to identify the critical variables after the fact. He may make iterative analyses of his data trying different variables until he discovers those that seem to explain most of the variations in performance. The Experimentalist, forced to decide before he collects any data, ordinarily has no second chance until he does another experiment. Two, since the Correlationist often measures performance under operational conditions, all critical factors, even if unknown, are likely to be present and to affect behavior realistically.

This enables the Correlationist's estimates of mean performance on particular conditions to be essentially correct. Of course, the unexplained variability about those means will still be larger than desired for precise estimation purposes.

Universal Laws and Individual Differences

Following the examples set by the physical scientists, the Experimentalists seek to derive empirically universal laws of human behavior. They manipulate conditions in the environment to find out how people behave as a function of the conditions being varied. However, since all "people" don't behave the same under the same experimental conditions, Experimentalists claim only to describe the behavior of the average person. With that goal, individual differences are considered to be a source of "error" variance, not suitable for study nor worthy of concern. In practice, the academic rules for obtaining homogeneous subjects are seldom met, and undefined subject performance variability is often greater than treatment variability (Simon, 1976b). "Universal laws" never seem to predict except on a probabilistic basis for large groups of individuals. Behavior is "situation-specific," and the characteristics of the individual may be a major factor contributing to the level of performance being measured. In spite of this, Experimentalists introduce subject characteristics into the experiment only infrequently and seldom consider, as Cronbach (1957; 1975) proposes, the interaction between equipment and subject factors. As a result, the degree to which the experimental data can be used to predict and control behavior is considerably reduced. Bugental (1963) stated it this way:

> The past 50 years have seen a tremendous accumulation of data about people treated as interchangeable units. And yet it is clearly the case that only where we are concerned with masses of people do these data yield useful results. This may seem a harsh judgment but I think it is an accurate one. If psychology is the study of the whole human being, and this I believe is its primary mission, then results which are only true of people in groups are not truly psychological but more sociological. (p 564).

The Correlationists, on the other hand, have concerned themselves with measuring individual differences often under specific treatment conditions. For them, therefore, variations of the test conditions can be as annoying as variations in people are to the Experimentalist. Thus, their measurements have not always been applicable under related but different circumstances.

Unilateral Multifactor Studies

It took Experimental psychologists more than fifteen years to begin to use Fisher's analysis of variance to study multiple independent variables in a single study. In the post World War II period, from 1948 to 1972, Edgington (1974) found in a survey of APA journals that the percentage of inferential studies employing this multifactor approach to psychological problems rose from eleven to seventy-one percent. In 1972, 88% of these were repeated-measure or factorial designs. Research involving the study of multiple independent variables in a single experiment has become common practice for today's psychologists.

When several dependent variables are considered, however, the Experimentalist has traditionally studied them each in separate analyses. Such a procedure can lead to improper interpretations when dependent variables are correlated. In many operational situations, no single dependent measure is sufficient to characterize performance on a complex task. The quality of information obtained from an experiment can be improved when multiple dependent and multiple independent variables are studied in a single bilateral, multivariate analysis. The Correlationists have developed and used these techniques for decades.

Reduction Experiments

Few Experimentalists today deny the importance of a multivariate approach *when predicting* performance on a complex task. In spite of this, relatively few variables are actually studied in a single experiment. This means that fewer factors are taken into consideration than are needed to account for most of the performance variance found in a typical real-world task. Reality is just more complex than that. In spite of this observation, most psychologists have been content to study only a few factors in a single experiment. A major reason for this is the cost of collecting data when many factors are systematically studied using traditional designs. Another reason, however, is that many psychologists do not fully recognize the limitations of the reduction experiment which proved so successful for experimentation in the physical sciences. There still remains the naive belief that data obtained from a study in which only a few of the total number of critical factors are varied (and all others held constant)

is as informative as that from a more complete experiment. In behavioral research, except in the rarest of circumstances, this presumption is incorrect for a number of reasons. One, if the variables included in the experiment are not important under operational conditions, then even significant results in the experiment may be of little predictive value when applied to complex situations in the real world. Two, whenever critical factors are held constant, performance estimates are likely to be biased when the data is applied to real world situations. Three, whenever critical factors are ignored,results in both the experiment and the real world will contain a variable error. Four, the effect of an experimental variable interacting with the variables held constant in the experiment can not be detected. The willingness of the Experimental psychologist to study simplified versions of a complex situation is the main reason why experimental results cannot be applied directly to operational situations without considerable qualification, sometimes to such a degree that the original data cannot be recognized.

Correlationists, on the other hand, by the nature of their problems, have been forced to accept the complexity of the real world. Since they have less opportunity to manipulate the variables, their approach has been to observe, measure, and classify. As a result, the effects of critical factors are often confounded and obscured. But what is lost in clarity is often made up in relevance, and a measurement made under realistic circumstances will often be representative of what can be expected (provided critical factors don't change) under similar circumstances in the future -- even if the underlying causes are unknown.

Significance Testing

Because Experimentalists have used analysis of variance models in much of their research, they have also relied upon tests of statistical significance to help them interpret their data. When they are interested in the functions relating independent and dependent variables, Experimentalists have traditionally been content to plot the mean performance at different levels of one or two effects at a time, usually the ones that were found to be statistically significant.

Correlationists, unable to manipulate their variables, have preferred to use a regression model to analyze the data from their "undersigned" experiments. Correlational techniques are used to unravel and identify entangled variables affecting performance, as well as to provide a multivariate equation, often in polynomial form, that provides a compact and comprehensive summary of the results from all variables. Because this data also can be treated to a variance analysis and even tests of significance, the Correlationists tend to analyze their data more thoroughly than the Experimentalists and obtain considerably more information.

Scientific Orientation

In many respects, the idea that Experimentalists were the scientists created an atmosphere in which attitudes and methods evolved that have only succeeded in degrading the quality of the information produced by the experiment. Some of these have already been described -- hypothesis testing, reduction experiments, and a search for general laws.

Seeking a scientific posture also encouraged the development of theories -- too prematurely, Bass (1974, p 873) has suggested. It also led to a redefining of the meaning of "basic" research. Rather than implying research that would produce data that would be fundamental to many applied problems -- if not today, someday -- instead, the term "basic" for some became associated with research without relevance, now or in the future, to any practical problem. By abstracting reality, the "basic" research of many psychologists became irrelevant research since critical parameters found in the real world were held constant in the experiment at values many standard deviations from any ever to be experienced operationally.

The Correlationists, while believing that their approach is as scientific as that of their Experimentalist colleagues, have tended to emphasize practical problems. Although they too have developed premature theories, used oversimplified experimental conditions, and applied techniques that have led down fruitless paths, on the whole, their research has been somewhat more successful than that of the Experimentalists in meeting the needs of today's society.

CONSOLIDATION -- A NEW EXPERIMENTAL PSYCHOLOGY

Both the approaches used by Experimentalists and Correlationists have contributed to the methodology of scientific psychology. Both have deficiencies when employed traditionally. Ideally, the most effective approach would be to consolidate the best features of the two disciplines. This is not a new idea. Forty years ago, Guilford (1936, p 11) wrote: "In recent years we see more clearly the common ground

existing in those two fields and a number of investigators have been instrumental in bridging the gap that has too long existed between them. It is one of the purposes of this volume to help point out the basic unity of the two fields and to assist in introducing the one to the other." Peters and VanVoorhis (1940, p 357-358), in discussing the place of analysis of variance in research, felt that it "belongs as a first step in a major research where one wishes to make a rough preliminary test of his hypothesis in advance of going to the expense of the elaborate setup needed for a thorough investigation." They felt that "for the positive side of research [meaning that which provides the critical information], the investigator will need the standard procedures of classical statistics, such as correlation, curve fitting, and contrasts of correlated matched groups. Constructive research is just ready to begin where analysis of variance leaves off." Peters and VanVoorhis also comment on how the Correlationists developed tools to help interpret practical and baffling problems, unlike the Experimentalist's -- unidentified but implied -- imitative use of statistics from other disciplines. They suggested that the former is "the ideal toward which we work."

In 1966, Raymond B. Cattell founded the Society of Multivariate Experimental Psychologists and the Journal of Multivariate Behavioral Research to encourage truly multivariate research and to bring out what Cattell (1966b, p 22-23) called "The 'integrated man' -- the new psychologist whose interests will encompass both the structural (individual differences) and the process (perception, learning) laws."

Ten years later, however, no major consolidation has been achieved. Royce (1977, p 135) wrote:

> More than a decade has passed since Cattell's manifesto. How have we fared? As I see it, although progress has been made in the desired direction, particularly in the promotion and publication of a high calibre of multivariate research, the bridging planks of the 1966 challenge have not been significantly implemented.

In this first decade, Royce noted that out of 342 papers, less than 2% could be described as combining multivariate and experimental approaches.

Limitations

When a bridge is built, who will build and who will cross? When one reads Cattell's (1966a) discussion on consolidation, there is a distinct impression that he believes the Correlationists have provided the more sophisticated methodology and it will be the Experimentalists who must change in order to profit from these advancements in technology developed by the Correlationists.

But Experimentalists have never shown a desire to give up their systematic manipulative methods for the uncertainties of mathematical solutions. On the other hand, the Correlationists have shown little sympathy for the reduction experiment. Cattell questioned whether a truly bilateral multivariate study could ever be achieved by Experimentalist even if they employed multivariate analysis of variance models. Speaking of the use of MANOVA techniques by

Experimentalists, Cattell (1966a, p 244-245) noted: "In practice, it is true, the number of independent variables used has seldom exceeded two or three, because the complications of experimental design and computation of higher-order interactions have discouraged investigation. It thus achieves multivariate status in principle, but scarcely performs some objectives of multivariate methods, such as comprehensively sampling large domains of behavioral manifestations." Studying too few variables provides too limited a perception of the situation being investigated. A true consolidation of the meritorious elements from both disciplines -- the manipulative control of the Experimentalist and the holistic coverage of the Correlationist -- is required. A paradigm that achieves this and more is possible and will be presented in the sections that follow.

V. A NEW EXPERIMENTAL PARADIGM

A new approach is needed that will combine the best features of those used by the Experimentalists and by the Correlationists. Such an approach, described in Table 3, would have the following objectives:

1. To approximate from data collected under primarily controlled conditions an equation capable of predicting individual performance on a specific man-machine task under operational conditions.

2. To provide the data in a form that will permit a modular, quantitative data base to be built which can be supplemented with data from other experiments using this paradigm.

3. To achieve the first two objectives at a cost that is justifiable for any important question and which represents a marked saving over that required by traditional methods for information of comparable quality and quantity.

The "paradigm" is a model of the way in which research philosophy, strategy, and techniques can be combined to perform experiments that will meet the above objectives. While a specific plan is described, a part of the philosophy is not to exclude any approach that can materially increase the useful information without adding to the costs. The primary feature of the new paradigm is its ability to

TABLE 3.

FEATURES OF THE NEW PARADIGM DERIVED BY COMBINING THE BEST METHODOLOGIES OF CORRELATIONAL AND EXPERIMENTAL PSYCHOLOGY

- Seeks to describe, understand, and predict the behavior of the individual in his environment
- Asks what happens in specific situations with practical boundaries
- Manipulates independent factors when possible, measures those which vary but cannot be controlled, and records values of critical factors held constant
- Studies relationships *among* multiple independent and multiple dependent variables (bilateral multivariates)
- Seeks to consider all sources of variance that might affect the behavior under consideration in the specific task
- Emphasizes use of regression model without rejecting any design or analysis that could increase the experimental information
- Collects and stores data in a way that builds a store-house of general knowledge which may be drawn upon to answer practical questions

consider a very great number of variables systematically, thus combining a holistic approach with classic manipulation techniques.

GENERAL STRATEGY

Traditional Experimentalists have sought to build a body of information through a series of small experiments. The assumption is made that by conducting enough small experiments -- a few variables at a time -- they can eventually combine the results to form a more complex, multivariate space reflecting the effects of variables. In practice, there are never enough small experiments, results are never quantitatively combined, and no "big picture" ever emerges.

The new paradigm, rather than use a "brick-at-a-time" approach, begins by examining the overall structure of the operational space in order to obtain the big picture first. Additional data is collected to improve the information, to better approximate the operational space. This assumes that by first obtaining an overview, however sparse, considerable economy can be achieved in the data-collection process since it will be easier to determine in what parts of the experimental space further refinement is needed. By including all variables presumed to be of some importance to the task in the initial empirical examination, it is possible to eliminate the trivial ones before a more detailed examination is made to derive a function relating the more critical variables to performance. Where the function fails to reflect reality the most, more data is obtained to correct the model. The

function, in equation form, would serve as the tentative quantitative data base suitable for description and prediction purposes; to this, new data can be added provided the values of all critical variables are known.

PRINCIPLES

The success of this strategy is predicated on certain principles or theories:

1. Equivalence sampling theory. The more closely the experimental world approximates the real world, the more likely experimental data will predict operational behavior. Therefore, the more critical variables that are included in the experiment within operational ranges, the more precise the prediction.

2. Pareto maldistribution theory. Although a large number of variables could conceivably affect results, in fact, only a relatively few will be critical and many will be trivial; the magnitudes of their effects will approximate an exponential distribution.

3. Simple model of human behavior. Human behavior can generally be approximated by a second- or third-order equation; higher-order effects are tentatively assumed to be trivial when proper scaling is employed.

4. Trivial error variance. Most residual variance of any size includes confounded real effects. By accounting for most of the performance variance in a complex task, little error variance will remain in the residual.

5. Minimum replication. In general, collecting data more than once under the same conditions is to be avoided unless the replication can be justified by showing to do so is more informative than to use any other sampling pattern or none at all.

SPECIFIC STRATEGIES

Desired objectives will be accomplished with reasonable precision and accuracy by employing the following strategies and tactics:

1. To achieve relevance, the experimental space will closely approximate the operational space. The limits of the experimental space will be set for all critical dimensions, to match (or exceed) those found affecting performance for a particular operational task. Variables to be considered initially will be based on what expert judgments and empirical analyses suggest might be important operationally.

2. To achieve generality, the study will include all variables believed to have a meaningful effect on the operational task (also defined by multivariate measures), whether related to

the equipment, the environment, the personnel, or the task. Individual differences, for all practical purposes, disappear when those factors producing subject differences on the particular task are included in the experiment as any other factor. With major effects removed, subject homogeneity is now a fact rather than the unwarranted assumption as is often the case in many experiments. Uncontrollable variables considered critical to the task are included as covariates to the controlled variables; this means they must be measured. Since the range of levels has been selected from those found under operational conditions, an equation approximating this space will apply to all subsituations occurring within this space.

3. To achieve modularity, records are kept of the values of relevant but unvaried conditions that might become critical in later studies because of a redefining of the experimental space.

4. To achieve economy, data is accumulated serially, employing different techniques to answer different questions as the accumulation progresses. Major questions are: What factors are critical? What is the simplest model to approximate the response surface? What are the fiducial limits once the approximating equation has been refined? This approach provides a gross overview of the experimental space, which is obtained economically

and subsequently can be refined when and where the need for refinement is noted. It provides for an empirical test to pare early and economically many candidate variables, leaving for further study only those that are in fact critical for the particular task.

This sequential process is achieved by collecting data in blocks, complete within themselves for specific information. New blocks are added only when new information is needed to provide an adequate model of performance. Thus, an experiment can be terminated with a reasonable approximation of the space long before a full factorial design is completed. Blocking may cut across several dimensions:

a. The order of the approximating equation (e.g., the first block collects the data required to approximate a linear response surface; more blocks are added only when tests show that a higher order model is demanded. There is little reason to believe that higher-than-third-order equations will ever be required if proper scaling is employed.

b. Replication is not used automatically. Each replication is treated as a new block of data, employed only when needed. (E.g., it is seldom if ever needed for

precision or for estimating error terms to test hypotheses. It may be used to test conclusions and establish fiducial limits at the end of the study.)

Appropriate scaling of variables and proper use of techniques to diminish irrelevant sources of variance (often introduced by the experiment) also help to keep the required quantity of data low.

SUMMARY

In summary, with this approach the manipulative advantages of the Experimental method are combined with the holistic philosophy of the Correlationists. Mapping a reasonably accurate description of an experimental space that corresponds to a broad operational space increases the probability that the experimental data will relate to field phenomena and generalize across a variety of specific problems. Maintaining a measure of all potentially critical sources of variables (along with the approximating equation) provides a coordinate space within which new data can be fitted. Sequential approaches that collect no more data than necessary to answer the question of the moment -- questions that change as the research program progresses toward the full development of an approximating equation -- enable large numbers of variables to be studied with considerable economy. The techniques and sequencing required to carry out this approach will be described in the following sections.

The new paradigm is divided into five phases intended to:

1. Define the problem
2. Identify the critical variables
3. Develop response surface
4. Refine equation
5. Verify experimental results

The relationship among phases, goals, and methodology are shown in Table 4. Each phase will be described in the following sections.

TABLE 4

RELATIONSHIP AMONG PHASES, GOALS, AND METHODOLOGY AS THE EXPERIMENT PROGRESSES

	1	2	3	4	5
PHASE:	Defining the problem	Identifying critical variables	Approximating response surfaces	Equation refinement	Verification
GOAL:	Exploring and limiting the problem	Building a quantitative data base			Evaluating
LOCATION:	Field*	Laboratory or Field			Field
APPROACH:	Undesigned (No control; measure)	Systematic (manipulation, control; measure)			Systematic or undesigned
METHOD:	•Literature search •Interview •Observe •Experience •Measurement	Fractional factorials: screening -- group, individual measure	Central-composite designs; refinement points	Replication; iteration	Test-residual analysis
MODEL:	-	Res. III** Res. IV	Res. V	Res VI +	-
ANALYSIS:	Correlational: •factor analysis •ridge regression •cluster analysis •etc.	Analysis of Variance: • mean differences • eta squared • ordered graphics • etc.	Correlational: ridge canonical regression		Correlation; significance

*May not be possible if system doesn't exist; simulator may serve instead.

**Res. = Resolution. The Roman numeral indicates which sources of variance are isolated and aliased.

VI. PHASE ONE: DEFINING THE PROBLEM

The first phase of the research program is the least systematic of all and for that reason requires the greatest astuteness, ingenuity, and persistence on the part of the investigator approaching an unknown situation that he wishes to describe quantitatively. Because he usually prefers (or is forced) to work one step removed from the real world in which the situation occurs, he must be selective in what he will study and judicious in how he will study it. Even though he is handed a problem to solve, a question to answer, or a situation to evaluate, the investigator is still faced with a major effort, that of translating a casual expression of the problem to an explicit definition and converting the real-world situation into an experimental plan. In addition, he must see that the subjects, equipment, environment, and task are prepared for the data-collection period. This is the general purpose of the first phase.

To define the problem, the investigator must place limits on a huge multivariate space in which an equally multivariate task is to be performed. The general question that he must answer is: What precisely is the task and under what conditions of the equipment, environment, personnel and certain time considerations it is performed? This he must do in two steps if he expects to optimize his experimental plan. First, he must dimensionalize the problem as it exists under operational conditions. Then, after the first step has been thoroughly worked through, he will render the real-world problem into a viable experimental plan.

A REAL WORLD ORIENTATION

A fundamental principle in the design of any experiment is that the definition of the task and the conditions under which performance will be measured must be based on real-world considerations. Decisions to include or exclude, duplicate or approximate, in the experiment should be made on the basis of their impact were the same thing to occur under operational conditions. Because this relevance is so important, the first step of the problem definition phase is to know reality. Only after the operational analysis has been made should the investigator begin to translate the problem, still conceptualized in real-world terms, into questions and conditions that can be dealt with experimentally.

LIMITING THE EXPERIMENT

In the second step of the problem definition phase, the "reality" of the operational situation will often clash with the "reality" of the experimental situation. This step, therefore, is a time for compromise. This is the time when the requirements for the experiment as defined by the user, the engineer, the investigator, and the operational situation must be balanced against the practical limits imposed by money, time, and availability. As long as the investigator is conscious of the consequences of his decisions, then the trade-offs required can be weighed on the basis of the ultimate criterion: useful information (Simon, 1975b).

Another important principle of the new paradigm is: What's not worth doing is not worth doing well (Hebb, 1974).

The investigator should not only be concerned with translating the real world problem into one that can be systematically studied in the laboratory, but also with whether or not any experiment should be done at all. As an experiment is being formulated, an investigator may recognize the fact that for various reasons he will be unable to get the information desired. To continue with an experiment as planned under those circumstances is unethical. While there are some who would contend that any information is better than none at all, with the high costs of doing research and the dangers of applying erroneous data to real-world problems, a recommendation to terminate the project, revise the question, or increase the resources (when that will make a difference) are all better alternatives than continuing as planned. Some of the circumstances in which a formal experiment would ordinarily not be justified include:

1. Experiments on questions that can only be answered analytically, but never empirically.

2. Experiments in which it can be determined by an informal investigatory study that effects will be trivial.

3. Experiments in which the correct answer will not be obtained because of restrictions placed on the simulation, e.g., critical variables are omitted; variable levels fall outside the range of any practical interest, now or later; there is insufficient time or money or cooperation to do the research

> properly; irrelevant variables can neither be controlled nor measured; critical conditions exist that are not representative of those found in the real world to which the results are to be applied.

Tyler (1973, p 1025) discusses research on problems for which definitive answers can never be obtained. She wrote:

> It is that the applications of what is found out are to a considerable extent out of the investigators' hands. Long after they have moved on to other hunting grounds, in this world or the next, people may be citing their results in support of policies and programs they know nothing about. We need to remember this in making research plans. While it is to be expected that ambiguous results will often turn up, especially in work on new and complex problems, if it becomes clear that the results of a line of research are going to continue to be ambiguous no matter how many successive studies are made because of the impossibility of controlling or correcting for the influence of a crucial independent variable, it would be better not to pursue it further.

She concludes that when it is apparent that there is no way to resolve an issue experimentally, "investigators should give serious thought at the onset to whether the research should be done" (p 1026). Phase One of the new paradigm is to be used in part to make this decision.

OBJECTIVES

The major objectives during the problem definition phase are:

1. Establish the dimensions and limits of the task (or tasks) under investigation, based on real-world consideration.

2. Ascertain that all equipments are operating reliably and accurately as intended and represent the critical dimensions of their real-world counterparts.

3. Check on the availability of the necessary number of subjects (operators) with the correct characteristics for the problem at hand.

4. Prepare for the collection and analysis of data to maximize the ease and accuracy with which this will be done.

In practice, the achievement of these objectives may extend into other phases. Still they illustrate what type of preliminary action must be taken before the experimental design is selected or the data collected.

A list of some of the tasks required to achieve the above objectives is given below. Details are not provided regarding these tasks since to do so would entail a major paper in and of itself. These tasks include:

- Identify the general problem area (Mission).
- Identify the task (or set of tasks) to be investigated. (A task is a particular combination of events occurring consecutively in time, having identical performance criteria which are influenced by essentially the same set of critical parameters.)
- List as specifically as possible the information expected to be obtained from the study when it is finished.
- Identify the measures of task performance.
- Identify the equipment, environment, personnel, and time-related variables that might be expected to critically affect task performance in the real world. (At this point, be liberal but not ridiculous.)
- Determine the range of all predictor variables through which the task is likely to be performed, *now or in the foreseeable future.*
- Determine which of the operationally relevant variables can be created and/or measured in an experimental environment.
- Determine the methods of measuring both predictor and response variables, including the measurement scale that will be used.
- See that the equipment, environment, and tasks are truly representative of the operational situation.
- Determine which measures can be made on-line and revealed during or immediately following an experimental run, and which cannot. Can raw data be used directly or is additional analysis required? How much time delay is there in off-line analysis?
- Check hardware and software to see that performance measures and analyses will be accurate.

- Optimize the Mean Time Between Failures of all equipment. Check system reliability.
- See that the equipment not only meets engineering requirements but also those needed to simplify data collection and enable experimental design changes to be done quickly (flexible). This is also a requirement of any computer software required in the simulation.
- Determine whether there is an adequate supply of truly representative subjects over a long-enough period, and that subject dimensions relevant to the task have been measured.
- Make certain that all research assistants are adequately trained for their job.
- Make certain that planned experimental sequences have been tried to see if there is enough time and distribution of labors to reduce pressures on experimenters during the data-collection stage.
- Make certain that instructions, training, and other techniques for preparing the subjects are adequate.
- Plan for contingencies that might occur during the data collection to minimize disruption (e.g., from equipment breakdown, premature subject termination, environmental interferences).

INFORMATION SOURCES

Dimensionalization of the task, including the conditions under which it will occur, involves an investigation of existing sources of information, and may also involve some

empirical data collection. Sources of information for dimensionalizing the problem include:

Literature review
Interview
Direct observation
Personal experience
Data collection

For a program of any major size, all of the above will probably be required. For some problems, there may be no operational system to be observed, experienced, or measured. Simulation may have to be relied upon in these cases. However, simulation is a retreat from reality, and one must be careful that problem definitions based on observations of performance in the simulator are truly representative of conditions in the field.

TALENTS INVOLVED IN DESIGN OF EXPERIMENT

In any large-scale research program, multiple talents involving knowledge and skills from different backgrounds of training and experience are required for the success of the venture. The concept of an interdisciplinary team is not new, but it has not always been implemented effectively. Ideally, in the beginning of the problem definition phase, each member of the team should present and defend his own parochial point of view. The purpose is to educate the other members and to make certain that no decision is made that will in fact compromise the information to be obtained from the investigation. Eventually, compromises will be made, hopefully ones in which the system point of view prevails and no position is seriously degraded. In practice,

the man with the money most often prevails even though he may not be the most competent to make the decisions which too often are component -- his component -- oriented. Still another principle required to effectively carry on a research program of the magnitude in which the new paradigm is justified is: Multidisciplinary team members who design the experiment must stoutly defend their individual positions but only in terms of system goals.

Members of a research planning team should be experts in the following elements of the experiment situation:

1. The real-world task. At least one participant must be capable of relating all experimental decisions to reality. If something is to be done in the experiment, he must be assured that it will not compromise the value of the experimental results when they are applied to operational situations.

2. The experimental methodology. At least one participant must be able to translate reality into a viable experiment. This implies not merely a knowledge of experimental design and analysis, but also the practical problems that can arise in data collection, and the informational consequences when the experimental paradigm must be compromised.

3. The equipment. When complex equipment is involved, an engineer must ascertain that it will provide the inputs and outputs required by the experiment and will simulate the critical conditions in the real

world. He must be concerned with keeping costs down. He must be prepared to offer compromises in equipment design that reduce costs without sacrificing the requirements of the experiment.

4. The experimental milieu. Someone must be capable of dealing with and representing the users, the administrators, the funders, and others outside the immediate experimental process, but whose opinions can markedly affect the direction the research can and will take. He must be able to explain to them why particular decisions are made and why certain requirements are important. He must be able to represent their point-of-view in the planning phase.

While the necessary talents, knowledge, and skills may all reside in one person, it is not always the case the he will be an expert. It is generally good practice that an investigator seek information outside himself. While there may be circumstances to the contrary, the final experimental plan should be left up to the investigator, who hopefully will combine the inputs from the other sources optimally.

PRE-EXPERIMENTAL ANALYSIS

As a first pre-experimental exercise, the investigator, along with whoever is knowledgeable about the real-world task, should order both the predictor and response variables in terms of their expected importance. With a large number of variables, it may be difficult or even impossible to rank the variables individually; cluster ranking is appropriate.

The important thing is to find out which ones are believed to be the most critical and which are not. The reliability and accuracy of the ranks at the extremes ought to be high for knowledgeable rankers.

As a second pre-experimental exercise, the investigator, along with the real world representative, should mark those variables that might be expected to interact with other variables. He should distinguish between disordinal and ordinal interactions (Simon, 1971, p 21; 1976b, p 62), since it is important that the former be included in the experiment, the latter being dissolvable through proper scale selection.

Third, the investigator and the engineer should rank the variables in order of the following qualities:

1. Ease of simulating (indicating any reduction in difficulty if the range of levels is reduced).

2. Cost of simulating the proposed range of levels, indicating major reduction in costs for reduced ranges.

3. Ease and speed of changing levels.

These analyses, when done independently of the decision to include or exclude a variable, will facilitate making that decision.

Classification

The variables can also be classified in a number of ways that affect the plan of the experiment and the data collection. Important classifications include the following qualities:

- Task specificity
- Manipulability
- Quantitativeness
- Subject attributes
- Subject characteristic selectivity
- Predictor-criteria variations

Whether a variable is general or specific to a particular task is important for prediction purposes, the general ones being found each time the task is performed while the specific ones may or may not be present, but are critical to some extent when they are present (Simon, 1976b, p 57).

Another important distinction among variables is between 1) those that can be controlled and manipulated by the investigator, and 2) those that cannot. The first group (controlled) will be systematically studied in an experiment employing a "screening" design pattern for economy. Equipment variables generally fall into this group. The second group (uncontrolled) will be measured concomitant with the measure of performance and treated as any covariance data might be. Environmental and personnel variables are frequently of this type. Omitting sources simply because they cannot be controlled, or because they are difficult to measure, negates the very purpose of the experiment's primary

objective, i.e., to account for as much of the performance variance as possible. This approach enables us to handle within the single experiment all of the variables judged relevant, whether manipulatable or not.

The controlled variables should be classified in another way, i.e., whether they are quantitative-continuous, quantitative-discrete, or qualitative (categorical). This information will be useful in planning the experiment, as discussed later. Whether or not they are "zero" variables should also be noted. Zero variables are those that can take on a value of zero meaningfully. For example, the "resolution of a visual display" cannot take on a value of zero meaningfully, while "vibration" can.

Subject (personnel) variables can be divided into two types: 1) those pertaining to specific, measurable, simple attributes (e.g., visual acuity, weight, blood pressure), and 2) those pertaining to more generalized, composite characteristics (e.g., pilot/non-pilot, years in service, etc.). The first group can then be handled in the same way equipment variables are treated within the screening design, while the second group must be tested outside the design, with a complete basic screening design being run at each level of the composite subject variables. This is done to minimize the complications that might arise in the presence of disordinal subject-by-equipment interactions. The reasoning here is that interactions are more likely to occur with composite subject variables than with the simple subject variables. However, for any specific situation, the investigator must weigh the alternatives of handling subject variables in these two ways before deciding what to do.

Subject variables also may be controlled or uncontrolled. Controlled variables are attributes that can be obtained by selecting subjects with the correct combinations that must be fit into the screening design matrix. If there are, for example, three such subject variables at two levels each, eight different subjects would be needed to satisfy the eight combinations of high and low conditions of each attribute. Each of these subjects would be tested on the appropriate levels of the equipment parameters, as indicated by the remainder of screening design. Where subject variables are uncontrolled, they must be treated as measured data. In some cases, measures of subject variables are obtained from historical data.

A familiar classification scheme for the Experimentalist is that which separates the independent or predictor variables from the dependent or criteria variables. The dependent variables are the ones that have been most neglected when experiments are designed and the problems defined. The availability of statistics for handling multiple responses in the same multifactor analysis demands that more serious thought be given to this class of variable. Reising et al, (1977, p 221) reviewed over 200 articles in the journal, Human Factors, and concluded that "researchers fail to define the experimental criteria or adequately defend the choice of dependent variables and summary statistics." They propose methods for improving upon the deficiencies reflected in those observations.

PRELIMINARY EMPIRICAL INVESTIGATIONS

During the problem definition phase, certain preliminary empirical studies may be warranted. Among the most common situations are:

1. Identifying primary factors in important composite variables

2. Determining weights of multiple response measures when related to a criterion

3. Performing parametric verification studies

The more effective techniques used in these investigations are those used by the Correlationists. Only a limited number will be suggested here. The reader is referred to books on multivariate analysis, such as that by Cattell (1966), as well as such statistical journals as Psychometrika, Technometrics, Multivariate Behavioral Research, Educational and Psychological Measurement, Biometrics, and so on, for the latest advancements.

Identifying Primary Factors in Composite Variables

Some critical variables are actually composites of a number of more fundamental variables. Usually critical variables may be ordered, but are difficult to measure. An investigator may prefer to introduce such variables into his experiment using several fundamental variable dimensions rather than using the single, complex, composite variable. For example, "background complexity" is a recognized

variable that has considerable influence on the visual detection of targets. But background complexity only can be subjectively ordered on a crude scale, from very complex to plain (solid). Fen Rhodes (1964) attempted to quantify "background complexity" by measuring eleven characteristics found in all terrain pictures and relating them by least squares (multiple regression) analysis to the time required to find the target. Better techniques are available today (e.g., ridge regression analysis -- see Simon, 1975a) but the idea remains a good one for this class of variable.

Determining Weights of Multiple Responses

An investigator may have to use secondary criteria to measure performance under operational conditions because it is too dangerous, too costly, or otherwise impossible for him to measure the ultimate criterion. Yet in the laboratory, he may be able to measure both secondary and primary criteria. He would want to find (in the laboratory) the relationships between secondary and primary criteria, in order that he might apply the empirically determined weights to the field data. Ridge regression analysis (Simon, 1975a) would be a better technique than conventional multiple regression analysis for this purpose.

Parametric Verification Studies

Certain information needed before a large scale experiment begins can only be obtained empirically. To plan an experiment properly, the investigator should discover in situations as close to those that will occur in the experiment itself the following information:

1. How much intra-subject trial-to-trial variability can be expected?

2. How much inter-subject variability among homogeneous groups can be expected?

3. What critical transfer and trend effects can be anticipated over five or ten trials?

4. Can the task be performed under all experimental conditions?

5. Are the data collection schedule and procedures reasonable?

6. Are the performance measures relevant criteria for the particular task?

7. Is the equipment reliable over trials?

8. Are instructions to the subjects clear?

No extensive effort need be made to answer the above questions. The intent is to minimize assumptions and to gain some empirical evidence about these items so that an investigator can correct or be prepared to handle those that show up as being severe. No subtle measures are required. The investigator "plays around" with the equipment and some representative subjects for a day or two, searching for items such as those listed above that might disrupt the data-collection process or distort the results.

VII. PHASE TWO: IDENTIFYING CRITICAL VARIABLES

Experiments in the human factors engineering literature are closer to those found in Phase Two than any other phase. This does not mean that they had been preceded by an elaborate problem definition phase, nor that they would be followed by the function derivation process of Phase Three. Neither are likely. It is because both the old and the new approach involve some version of an analysis of variance model for sampling the experimental space. However, beyond that similarity neither the purpose, the form, the analysis, nor the follow-up effort are necessarily the same.

Whereas the traditional experiment is usually intended to be an entity in and of itself, the data collected and analyzed in Phase Two of the new paradigm is but the beginning of an extended collection/analysis sequence of which Phase Two is a module. Whereas the objective of the traditional experiment has been to identify "statistically significant" effects, the objective in Phase Two is to discover empirically and systematically which of the long list of candidate variables selected rationally in Phase One are really important in the performance of the task. With the techniques described in this section, the data collection required to screen 25 or 75 candidate variables is generally less than that used in some traditional experiments found in the human factors literature (Simon, 1976b, p 26). As few as twice the number of observations as there are variables to be screened may be all that is needed to estimate the effect of each variable independent of any two-factor interaction effect.

PRINCIPLE OF MALDISTRIBUTION

Underlying the approach proposed in the new paradigm is the assumption that the magnitude of the effects of the very large number of variables associated with a particular task will approximate an exponential distribution. This means that for any one task, the effects of a relatively few variables will account for most of the observed variance. This assumption is referred to as the Pareto Maldistribution Theory or assumption or principle (Bunde, 1959; Engineering Statistical Methods Group, 1963). In a limited analysis on some human factors engineering experiments, Simon (1976b, p 55-56) found this assumption to be true.

Thus, by eliminating the smaller, non-critical variables from all subsequent studies, the valuable data-collection time will be saved when the response surface is to be approximated. Since it will be built upon data collected only for the more important variables from a large candidate list for the particular task, the derived equation of the response surface should be expected to predict performance well under operational conditions. The use of screening prior to response surface development therefore not only assures economy in data collection but also increases predictive accuracy.

SCREENING

The first step in planning the screening phase is to divide the candidate list of predictor variables into two groups: 1) those that can be controlled and manipulated by the investigator; and 2) those that cannot. It is the first

group that will be fit into a screening design representing the coordinates at which the multivariate space is to be sampled.

THE ECONOMY OF SCREENING DESIGNS

Economy in data collection is achieved with screening designs because a sequential approach is used. A block of data is collected and analyzed to determine whether or not more is needed to identify the important variables. Since each block is only a small fraction of the total factorial, this iterative process keeps data collection to a minimum. It is almost a certainty that the full factorial space will never have to be sampled, or for that matter, even one-hundredth of it in the screening of 15 or more variables.

The purpose to be satisfied by each data-collection block entering sequentially into the screening study is as follows:

First, collect only enough data to estimate the magnitude of all main effects independently of one another, but confounded with all higher-order effects. (Resolution III design)

Second, collect enough additional data, which when combined with the data from the first block will isolate all main effects from all two-factor interactions. (Resolution IV design)

Third, collect enough additional data, if necessary, to isolate and estimate the effects of specific disordinal two-factor interactions or three-factor interactions that might affect the ordering of the main effects.

Fourth, keep replication to a minimum, if at all.

Two exceptions may occur to the above list: 1) when there is an exceptionally large number of candidate variables (e.g., 75 to 100), a gross preliminary screening may be called for in which critical sources are not fully isolated until later, and 2) when the investigator decides to combine the first and second steps into a single step, thus creating a Resolution IV design immediately. There are pros and cons to this decision (Simon, 1976a, pp 8-11).

In Resolution IV designs, main effects are confounded with three-factor interaction effects, and two-factor interaction effects are confounded with one another in independent strings. In most cases, a Resolution IV design will be sufficient to order the variables correctly since there is evidence to show that with quantitative variables, the effects of three-factor and higher-order interactions are usually trivial, and that these interactions are most likely of the ordinal type (Simon, 1976b, pp 57-65). If the uncommon disordinal interactions should occur, the amount of additional data required to isolate them (as in Step 3) need not be considerable.

This sequential strategy is basic to the new experimental paradigm, i.e., the procedure of collecting as little data as

possible until an examination of the data shows that more is actually needed to reveal additional information. The traditional approach of immediately collecting enough data to estimate higher-order effects is wasteful. Seldom, if ever, are there reliable fourth-order interaction effects and when these do occur, their effects are likely to be trivial. Even non-trivial third-order effects occur infrequently. If under unusual circumstances they turn out to be critical, they can be isolated after the fact has been established, not before.

CHOICE OF SCREENING DESIGNS

An investigator can choose among several forms of screening designs. These are:

I. Designs for screening a very large number (e.g., 100) of variables
   - A. Supersaturated designs
     1. Random balance
     2. Systematic
   - B. Group-screening design
     1. Two-stage
     2. Multi-stage

II. Designs for screening large numbers (e.g., 30) of individual variables
   - A. Box and Hunter designs
   - B. Plackett and Burman designs
   - C. Simon designs

Screening a Very Large Number of Variables

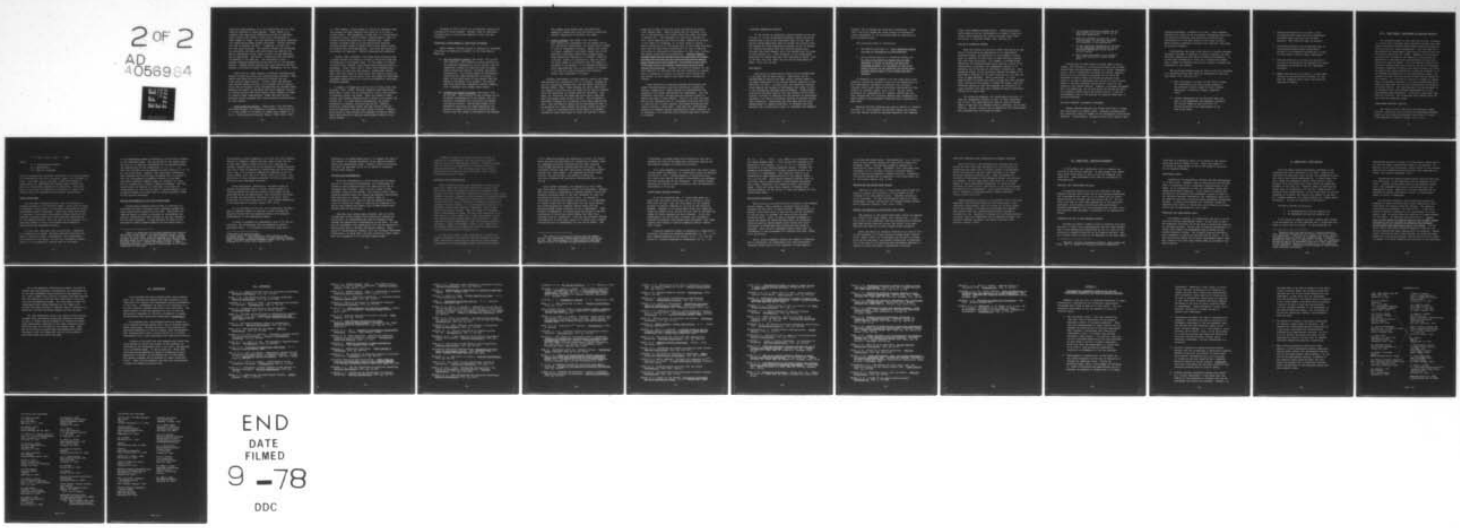
When the investigator has used every rational means during the first phase to pare the candidate list, but finds that he still has seventy-five to one-hundred variables that he can't discard, by tolerating a certain amount of uncertainty he may perform a preliminary screening study to reduce this number to 25 or 30 variables by making approximately 50 observations. There are supersaturated and group-screening designs for this purpose.

Supersaturated designs. These are factorial designs in which there are more variables than there are observations. Mathematically, in this case, it is not possible to isolate every main effect from every other main effect. The matter is resolved for those who propose this method since they assume that the Pareto Maldistribution Theory will be operating. Out of the total number of variables included in the experiment, actually only a few will have a critical effect and many will be non-trivial: thus there will be more observations than there are critical variables, and thus, the critical effects can be isolated. Two approaches in designing these experimental plans have been proposed.

Random Balance designs (Satterthwaite, 1959; Budne, 1959a, 1959b, 1959c) are created by choosing the levels of each variable for each experimental condition at random. As many variables as desired would be included in the description of each condition, a desirable feature when one wishes to locate the condition in the coordinate space. For each condition, the level -- generally one of two alternatives -- at which each variable is set would be selected at random.

AD-A056 984 CANYON RESEARCH GROUP INC WESTLAKE VILLAGE CALIF F/G 5/10
NEW RESEARCH PARADIGM FOR APPLIED EXPERIMENTAL PSYCHOLOGY: A SY--ETC(U)
JUN 78 C W SIMON F44620-76-C-0008
UNCLASSIFIED CWS-04-77A AFOSR-TR-78-0117-REV NL

2 OF 2
AD
A056984



END
DATE
FILMED
9 -78
DDC

Ordinarily approximately 50 or so observations are all that would be required in these designs. Budne (1959b, p 9) suggests that restricted randomization might be desirable, such as having the levels of each factor represented an equal number of times. He also describes a grouping technique with combinations among groups randomized. Scatter diagrams are used to analyze the data. A computer is useful for plotting these diagrams. The largest effects are discovered by eye-balling the data, and after these are removed, the data is replotted so that lesser effects of some magnitude can be identified. Two-factor interactions can and should be examined in the same way. The technique is like a graphic stepwise-regression and may have all of the inherent dangers. Random Balance designs have been used by some and soundly criticized by others (Youden, Kempthorne, et al, 1959).

Booth and Cox (1962) propose a different supersaturated design in which the levels are systematically selected. They assume that there are no interaction effects and only a few critical main effects. In their paper (pp 490-492) they provide designs for up to 36 variables and 18 observations, but as Kleijnen (1975) remarked when he reviewed these designs, the effort to use Booth and Cox's computer routine to develop designs that would handle larger numbers of cases (or require fewer observations) "might very well be prohibitive" and he proposed that group screening designs might be used instead.

Group screening designs. These plans, like the supersaturated designs, are intended to provide a rough first cut at a large number of variables to reduce their number to 30 or so. Group-screening designs (Watson, 1961; Patel, 1962;

Li, 1962) handle a large number of factors by combining them into groups and then treating each group as if it were a single factor. The assumption is made that if a group-factor is found to have a trivial effect (insignificant), then all factors within the group will be insignificant. Those factors in groups found to be non-significant would be studied further; those in groups with trivial effects would be dropped from the investigation. Both size and content of the groups are important. Natural groupings are preferred.

Watson (1961) proposed a two-stage group screening plan in which factors are tested in groups in the first stage and individually in the second stage. If the number of factors is quite large, however, multi-stage group screening might be necessary. Patel (1962) and Li (1962) both proposed plans that allow for more than two stages. Groups that survive after the first stage are partitioned into smaller groups for the second stage, and so forth, until the number of individual variables remaining are of a size to be handled by individual screening designs.

A number of assumptions hold in all of these plans (see Kleijnen, 1975, p 488), the most important and restricting being that a) there are no interactions, and b) the direction of possible effects are known. These are needed to make certain that several effects within a group do not cancel one another out. In human factors engineering, the second assumption is tenable, the first may not be. However, disordinal interactions rather than the ordinal ones are the most important, and the least likely to exist. Unequal group sizes in these designs are possible and may be used to avoid cancellations by putting questionable effects in different groups.

Kleijnen's (1975) article is an excellent overview and discussion of these methods. Whether group or individual screening must or can be used depends on cost and time restraints.

Screening a Large Number of Individual Variables

Three somewhat related choices of designs for screening individual variables are available to the investigator. These are:

a) Box and Hunter's designs (see Simon, 1973, pp 89-101; 105-114). Main effects of each variable can be estimated independently of one another, but are completely aliased with specific sets of higher-order interaction effects, including two-factor interactions. The minimum number of experimental conditions for these Resolution III designs is equal to the first power of two greater than the number of variables to be studied. To isolate main from two-factor interaction effects (Resolution IV designs) this number would double.

b) Plackett and Burman's designs (see Simon, 1973, pp 102-104). Main effects of each variable can be estimated independently of one another, but are partially confounded with two-factor and higher interaction effects. The minimum number of experimental conditions for these Resolution III designs are equal to the first multiple of four greater than the number of variables to be studied.

This number would double when new conditions needed to isolate main from two-factor interaction effects (Resolution IV designs) are added.

c) Simon's designs (see Simon, 1977, pp 8-24). Main effects are independent of one another and of all two-factor interactions, the latter being aliased in sets of independent strings. Designs are robust to linear, quadratic, and cubic trends, and can be adjusted to minimize factor-level change counts. These designs are Resolution IV designs, requiring a minimum number of observations equal to twice the first power of two greater than the number of variables to be studied. It is generally better to leave approximately five or six degrees of freedom to be used for trend adjustments and blocking rather than independent variables.

Plackett and Burman's designs usually require the fewest observations for a Resolution III or IV design. Because of the low correlation between main and two-factor interaction effects (Tukey, 1960), these designs are useful when there are reasons to believe that one might not be able to continue after the Resolution III design data was collected. With them, one could stop the study and still know a good deal about the proper order of main effects. Box and Hunter's and Simon's designs are variations of the same construction plan. Box and Hunter's design can be rearranged so that it is optimized for trend effects and change counts according to a number of available plans, while Simon's is already arranged to make them robust to trend and provide a simple

algorithm to keep it that way while minimizing the factor level change count. Simon's design was not intended to be run in two Resolution III blocks, while the other two forms of designs are. The advantage of a two-block approach is that, after examining the block of data, changes can be made -- in the number of variables, or in their range of values -- before continuing on to the second half if this seems desirable. Data from the first block (Resolution III) may at times be sufficient to terminate the experiment; in that case, running the second block would be uneconomical. ~~However, Simon's designs could also be divided into two Resolution III blocks, provided that the data on the trend overlap and the factor-level change count were modified. For designs larger than 2^{k-p} where (k-p) = 16 or more, relatively little is lost through this blocking.~~ Plackett and Burman's designs are the more difficult to construct and the more difficult to analyze were it necessary to isolate two-factor interactions from one another or main effects from higher-order interaction effects.

Since Resolution IV designs confound three-factor and higher interactions with main and two-factor interaction effects, respectively, these designs are predicated on the assumption that these interactions are negligible. Only the disordinal two-factor interactions are seriously going to modify the order of main effects. Since it is only the order that we are concerned with in a screening study and then only to identify the critical factors, only major disruptions are likely to matter a great deal. Only when the investigation reaches the response surface representation phase, however, is it necessary to be concerned with isolating <u>all</u> major sources of variance.

ISOLATING INTERACTION EFFECTS

If any string of interaction effects appears to be non-trivial, or if the investigator suspects certain two-factor interactions to be of the disordinal type, he will want to collect enough additional data to isolate the critical ones. The purpose for this augmentation process is not to improve the equation derivable from the screening design data, although this would happen, but to be assured that the order of main effects (i.e., the variables) is not disarranged. Techniques for doing this have been supplied by Daniel (1976, pp 246-247) and John (1966) as well as the discussions by Simon, (1973, pp 115-125).

REPLICATION

Replicating an experimental design can be accomplished in two ways and for many reasons. Replication can be accomplished by testing the same subject several times on the same condition(s) or by testing several subjects on each condition, or both. In every case, the additional observations destroy the economy of the designs and result in most of the information obtained being redundant. Traditionally replication has been used to bury the evidence of an investigator's failure to identify critical sources of performance variance or to control irrelevant sources of variance during data collection. Replication has also been used as a misguided effort to improve precision, to estimate the error variance for a significance test, to compensate for sequence

effects, and to average out individual differences. Simon (1973, pp 19-31) showed why these reasons are ordinarily either unnecessary or can be accomplished by effective but more economical means.

Two principles apply to replication:

1. The general principle is: Don't replicate unless the gain in information is cost-effective.

2. The principle specific to screening designs is: As long as the factorial design has not been completed -- which is usually the case with screening studies -- it is better to run a different fraction of the factorial to isolate more aliased effects than it is to repeat the same fraction.

An exception to the general rule that might prove cost-effective occurs when several trials are run sequentially on each experimental condition to minimize trial-to-trial transfer (cross-over) effects from being confounded with the effects of interest (Simon, 1974, p 23). As yet, since techniques for compensating for transfer effects have not been incorporated into screening designs that compensate for trend effects, this procedural technique may be necessary in some cases.

Whenever multiple subjects are used to obtain an "internal validity" or "inter-subject reliability" check, the results from each subject should be examined separately and compared

rather than combined mathematically. Differing patterns among subjects can be interpreted in a number of ways providing clues regarding the results and what future data collection is required (Simon, 1977, Section IV).

ANALYZING SCREENING DESIGNS

When the effects of each main effect and string of two-factor interactions are determined in an unreplicated screening design, the precision of the estimate is often equal or superior to that found in fewer-variable studies that have been replicated. For example, an unreplicated Resolution IV screening design would require a total of 64 observations to estimate the independent effects of 31 variables and 32 strings of two-factor interactions. Each estimated effect is the difference between the means of 32 performance measures on the high level and of 32 on the low level of each variable. That's equivalent in total observations to finding the effect of one two-level variable after replicating the design 32 times; of course, with the screening design we also have measured the effects of 30 other variables and have some information about interaction effects at no extra cost.

For screening purposes, the investigator will want to rank the independent effects in order of their magnitude, but he will also need additional data to help him decide where to draw the line between crucial, marginal, and trivial effects. Several criteria can be applied to the unreplicated data obtained from the screening study:

1. Is the mean difference between low and high conditions of each effect one of practical importance?

2. Does each variable account for a non-trivial amount of the total variance (eta squared) in the experiment?

3. Is the cumulative proportion of variance in the experiment accounted for by all variables designated trivial still a trivial amount?

4. Which variables appear to be "significant" when plotted on a half-normal grid?

Calculating the data needed to apply these criteria (Simon, 1977, Section V) is relatively simple and straightforward when the performance measure is a single, dependent value. Still, classifying variables as crucial, marginal, and trivial is a subjective process; how to weigh the different criteria cannot be decided by precise rules. Since it is easy to recognize very crucial and very trivial effects, the most difficult decisions will occur in the middle of the ordered effects. However, a mistaken assignment won't be disastrous, for at this point the values are small. If a marginal effect should be called trivial, it may be reconsidered later in the program if evidence shows that it was misclassified.

MULTIPLE RESPONSE (DEPENDENT) VARIABLES

Single response measures are seldom sufficient to represent performance on complex tasks. Adequate representation will generally require a number of not necessarily uncorrelated measures. Traditionally, multiple criteria have usually been

analyzed separately, a measure at a time. Since response variables are likely to be correlated, eliminating them from the analysis or holding them constant can distort the interpretation of the data based on the single criteria. Where multiple responses (criteria) are important, they must be analyzed together.

Understanding the joint contribution of several response variables can make it possible to select a smaller but more efficient combination with which to measure performance. It is also more economical to carry out a single test rather than a number of separate tests and it usually increases the generality of the results.

The following methods might be used to rank the variables in a screening design when multiple responses are involved (see Simon, 1977, Section VIII):

1. When the nature or mission of the task is known, the investigator can often assign weights to the multiple criteria based on their relative importance, to derive a single, composite score.

2. Graphic inspection can be used when there are only a few independent and dependent variables. The results based on each criterion are plotted separately and superimposed on the same graph paper.

3. LaGrange multipliers can be used to find the optimum point among multiple independent variables where there are two criteria measures. More criteria might be handled.

4. Step-down procedures can be employed when an investigator cannot assign quantitative weights to his response variables but is able to rank them in order of their importance.

5. Multiple analysis of variance can be used if one wishes to determine the proportion of total dispersion accounted for by the independent variables.

6. Gamma distribution plots permit a visual examination of the multivariate effects in a way that can identify those larger than would be expected by chance.

VIII. PHASE THREE: DEVELOPMENT OF RESPONSE SURFACES

The screening phase is over when the critical variables for a particular task have been identified without fear that hidden interaction effects might change their rank order or, more particularly, their designation. The next step is to approximate the experimental space with an equation with the critical variables as its terms. Whether or not marginal variables are included in the equation at this point is determined by such practical considerations as the availability of time and money. Still an equation based on the critical variables that were selected out of all candidate variables suspected of having real-world effects, is likely to predict well under operational conditions provided it is an unbiased representation of the experimental space. An equation derived from the screening study may be biased for two reasons: 1) a non-linear function is needed to approximate the response surface; 2) critical interactions still remain aliased with non-critical interactions and occasionally main effects. The third phase of the new paradigm is to determine what this unbiased representation of the experimental space should be. To do this, additional data must be collected. Let us see how this combines with the data from the screening study.

FIRST-ORDER RESPONSE SURFACES

The results from a Resolution III screening design provide sufficient data to write the relationship between predictor and response variables in the form of a first order polynomial equation (Simon, 1977, p 71):

$$\bar{Y} = b_0X_0 + b_1X_1 + b_2X_2 + \ldots b_NX_N$$

where:

$\bar{Y}$ = Estimated performance

b_i = Coefficients

X_i = Terms or variables

With an unreplicated, saturated, Resolution III screening design of N observations and (N-1) variables, there is no estimate of error. The term, b_0X_0 is the mean. Each main effect in these designs is confounded with higher-order interactions which are tentatively assumed to be negligible. Until more data (in the form of a Resolution IV design) is taken, main effects cannot be isolated from two-factor interactions.

Center Point Data

For the basic screening design, data is collected at selected corners of a 2^k factorial space. Since each variable is measured at two levels, no non-linear representation of the experimental space is possible. Often when human performance is involved, a non-linear relationship between predictor and response variables might give a more unbiased approximation of the response surface. The next step in the paradigm is to determine whether or not a first-order model adequately approximates the experimental space.

To test this, more data must be collected. Expanding to a Resolution IV design provides some data regarding two-factor, linear-by-linear, interactions, but nothing about the curvature of the space. To get this information economically, it is necessary to collect data at the center

of the experimental space as defined by the critical variables in the screening study. The coordinates of the center point are (0,0,0...0) when the coordinates of the screening design are combinations of the coordinates (±1, ±1, ±1,... ±1). Thus each variable will be measured at three levels (-1, 0, +1) but not factorially. However, this additional information, when combined with the original screening data, is enough to test for the presence of quadratic effects in the data. Individual quadratic effects cannot be isolated, for they are all aliased into a single, composite estimate. But that single source of variance is sufficient to provide the investigator with the clue he needs to decide whether or not he should collect still more data to isolate quadratic terms of the critical variables.

Testing the Adequacy of the First-order Model

Unreplicated screening designs have no provision for estimating the error variance unless untenable assumptions about higher-order interactions are made. Replicating the complete design to obtain an estimate is uneconomical and actually unwarranted as long as data for the full factorial has not been completed. A rough estimate of error can be obtained by taking repeated measures at the center point.*

*There are advantages if multiple measures are taken at the center of the design. Multiple measures permit a Lack of Fit test to be made. They also provide a crude external estimate of error, that could be compared with the internal estimate based on the half-normal plot analyses (Simon, 1977, Sections V and VIII. Another advantage is that it would bring the precision of estimates of performance at the center
(continued on next page)

This measure of error combined in an F-test with the composite measure of quadratic effects has been used to test the fit of the model (Simon, 1970b, pp 32-33; 1977, Section IX. However, usually there are so few degrees of freedom involved that this F-test of statistical significance has little power and to use it as a basis of judging the adequacy of fit is unwise (Simon, 1971, pp 30-33, 44-46; 1976a, pp 14-18). The proportion of total variance accounted for by the lack-of-fit would be a more preferred criterion.

Since qualitative (categorical) variables cannot be scaled, and therefore have no center, center points on a screening design can only be located in the middle of the space defined by the quantitative (and continuous) variables. To include the qualitative variables in the study, there would have to be one center point for each unique combination of the qualitative variables (Simon, 1977, p 51).

If, on the basis of the test, the investigator believes that no quadratic model is required, he may consider the equation derived from the screening design as an adequate approximation of the experimental space.

If there is evidence of considerable lack of fit due to curvature, the investigator must be prepared to collect additional data: 1) to extend the screening study into a

---

(continued from previous page)
of space closer to the estimates at other points in the screening design. Other advantages of using multiple center points have been described elsewhere (Simon, 1973, pp 131-139; 1976a, pp 21-27, 35-41; 1977, Section III).

Resolution V or higher-order plan; 2) to expand the range of each factor to increase the degree of the equation approximating the experimental space. However, before he collects any data, he sees what he can do to reduce or eliminate higher-order effects.

Scaling and Transformation

With the introduction of center points and the possibility of curvilinear relationships, the investigator is forced to think carefully about the scale he will use to represent each variable. With only two levels of each variable, there was no problem. With three or more, then a properly selected scale can change the apparent relationship from non-linear to linear. The simpler model is preferred when we are trying to map the response surface since it reduces the number of observations required to do so and also reduces the chances that aliased higher-order effects might be non-trivial.

When data has already been collected, and it is discovered that certain higher-order interaction effects are non-trivial, an investigator may wish to eliminate them through the appropriate transformation of the data. If this ploy is successful, the investigator will not be required to collect more data to isolate the special effects. Many types of transformation, however, will not eliminate disordinal interactions, in which case, the data collection effort would have to be expanded to fit the more complex model.

A number of papers have dealt with the problems of scaling and transformation, but none have been applied directly to the new experimental paradigm. Special problems arise in applying transformations to multiple factors in multivariate designs. Two techniques that bear further investigation are those by Box and Tidwell (1962) and by Bogartz and Wackwitz (1971).

Extending the Screening Plan

Basic screening designs are Resolution IV plans, which means that enough data has been collected to isolate each main effect from the others and from all two-factor interactions, but leave two-factor interactions aliased in sets of independent strings. Classical central-composite (response surface) designs have used 2^{k-p} fractional factorials of Resolution V, in which all main and all two-factor interaction effects are isolated from one another. If an investigator believes it is necessary to meet this condition completely, then he must add to the original design. When the number of variables under investigation is large, the step from Resolution IV to Resolution V is not a small one. Pajak and Addelman (1975) have determined the minimum number of Resolution III plans required to build a Resolution V plan for various numbers of variables. Draper and Mitchell (1968, p 252) showed that the maximum number of variables that could be studied with a 256 condition Resolution V design is 17.

On the other hand, if the investigator has taken the proper precautions, he will have already isolated the disordinal two-factor interactions as well as any three-factor interactions in strings showing large, non-trivial effects.

If all remaining effects are apparently* trivial, an investigator may have the equivalent of a Resolution V design, since all critical two-factor interactions have been isolated (even if all two-factor interactions have not been). Any polynomial written from that data would include terms representing each main effect, the isolated two-factor interactions, and the strings, the effects of which are composite effects of the two-factor interactions within the string.

Until proven untenable, the assumption is still made that three-factor interaction effects are negligible. However, when one wishes to approximate a response surface, unlike the screening situation when it was only necessary to order the variables and select the critical ones, if strings of three-factor interactions appear non-trivial, it is desirable to isolate those that are responsible for the large effects. This distinguishes the response surface phase from the screening phase, where isolation is not a requirement. Under certain circumstances, an investigator may decide to use the coefficient of the string effect rather than that of an isolated critical interaction. This may not seriously degrade the prediction in this case, since it is likely that a single interaction will be responsible for the entire effect. The

---

*The effect of a string of interactions may appear trivial, yet individual interactions within the string may not be. This would happen if a large positive and large negative effect in the same strings canceled one another.

investigator is always faced with decisions of this sort -- weighing the cost of the added data collection against the anticipated increase in predictive precision.

If it were possible to anticipate which two-factor interactions might be important, an investigator could use Whitwell and Morbey's (1961) "reduced designs of resolution five," with which only certain two-factor interactions are estimable and some effects may not be orthogonal. The reduced design improves the economy in data collection when the Resolution V requirement is to be met.

SECOND-ORDER RESPONSE SURFACES

If we are satisfied that a first-order model or a model with two-factor interaction but no quadratic terms does fits the data, then the third phase of the research process is actually complete at the end of the screening process. However, if an investigator finds that quadratic terms are needed to fit the data he will have to increase the number of levels examined for each variable. To do this economically and yet be able to extract the information required to construct the second-order polynomial, the investigator could employ a "central-composite design" (Simon, 1970; 1976a).

A central-composite design is composed of a Resolution V fractional factorial at selected coordinates (±1, ±1, ... ±1), some repeated measures at the center (0, 0,... 0), and the points of a measure polytope at coordinates (±a, 0,... 0),

(0, ±a,... 0),... (0,0,... ±a), where a is a distance from the center greater than 1. This allows each variable to be measured at five levels: -a, -1, 0, +1, +a, although not factorially in the central-composite design. Instead, the geometric distribution of the data collection points is in the form of a hypersphere. The numerical value of a is determined by other characteristics of the design (Simon, 1973, pp 131-139). With central-composite designs there is sufficient data to approximate a second-degree polynomial with linear, quadratic, and linear-by-linear interaction terms. There are enough extra degrees of freedom (with repeated measures at the center) to test the adequacy of this second-order model.

Non-critical Variables

While critical variables would be included in the response surface design, other candidate variables would be held constant and the value of each recorded. Theoretically, it matters little what value is used for the trivial variables as long as they lie between the limits of the original screening study. Still, in case it becomes necessary to expand the design by collecting more data, the fractional factorial portion should employ fixed values that would correspond to established data points were the study to continue. This use of a "Standard Factors Check List" is fundamental to the development of a modular data base (Simon, 1971, pp 99-102).

While central-composite designs are simple to construct and to understand, an investigator will need alternative response surface plans in his repertoire to meet special

situations that might arise. Non-symmetrical (i.e., two and three or four levels) designs are described by Draper and Stoneman (1968). Less optimum asymmetric designs are described by Lucas (1974). Roquemore (1976) describes more economical (e.g., 46 versus 79 point 7-variable design) hybrid designs for quadratic response surfaces. Sequential third-order designs may be employed if that order model can be anticipated (Simon, 1975, p 146).

Replicating the Second-order Design

There may be some value in replicating second-order designs to improve the precision of the estimates at the extremes of the experimental space. Complete replication is not necessary. To keep the economical quality of these experiments, partial replication of response surface designs may be employed (Dykstra, 1960, Patel, 1963).

Testing the Adequacy of the Second-order Model

The adequacy of the second-order model should be examined by a "lack of fit" test (Simon, 1977, Section IX). If the fit is still not sufficient, and the transformation technique mentioned earlier does not correct the matter, still more data will be required to form a higher-order polynomial.

Since the source of variance referred to as "lack of fit" in this analysis is in fact strings of higher-order interactions, (Myers, 1971), an investigator may be able to isolate which string and which three-factor interaction is contributing to the lack of fit using the same techniques employed to isolate critical two-factor interactions in strings.

ANALYZING CONTROLLED AND UNCONTROLLED VARIABLES TOGETHER

Up to this point, the discussion on analysis has overlooked the fact that some variables can't be controlled or manipulated yet might have a critical effect on performance. These variables will not have been included in the systematically designed screening plan. They can however be treated as covariances to the systematic design and analyzed accordingly. On the other hand, when there is a sizeable number of these uncontrolled variables, it would be more efficient, as well as more informative, if all the variables -- uncontrolled and manipulated -- were treated together as one set of variables and analyzed using "ridge regression" (Simon, 1975a, pp 33-51).

Ridge regression analysis is an improved form of multiple regression that produces equations with more stable, more meaningful coefficients, that are closer to the true coefficients and capable of estimating performance with smaller mean square error than conventional multiple regression analysis will do. Analyses of studies with multiple predictor variables, both the undesigned and the designed variety, and multiple response variables, can employ canonical ridge analysis (Carney, 1975).

IX. PHASE FOUR: EQUATION REFINEMENT

In this phase, the investigator tries to improve the quality of the initial equation. He may attempt this immediately after Phase Three or after he has had feedback from Phase Five. Some refinements that might be considered are discussed here.

REDUCING THE UNEXPLAINED VARIANCE

An equation based on the critical variables and some marginal ones may still leave 20% of the variance unexplained. The investigator will want to try to reduce this by identifying sources of variance that may account for it. He will have to do this by first hand observation of the task being performed, noting those circumstances when performance deviates considerably from the predicted score on each trial.* Any discovery must represent a hypothesis to be subsequently tested.

IMPROVING THE FIT OF THE RESPONSE SURFACE

Although an effort is made to find an equation that fits the data, it still is an approximation over the total surface. Part of the unexplained variance may be due to a lack of fit occurring in specific sections of the response surface* Poor fit is most likely to occur at the extremes of the space where

---

*Residual analysis (Anscombe and Tukey, 1963; Daniel and Wood, 1971) is useful for detecting these circumstances.

less data is ordinarily taken, or at points on the surface where the rate of change is high. Extra data might be collected at these points to see if the shape at the curve can be improved further.

CONFIDENCE LIMITS

Replication of experimental designs has been discouraged up to this point. However, once an equation has been derived that is considered to be a reasonable representation of the space, it is informative to know the confidence limits both within and between operators. Since subject variables are included in the equation and presumably account for most of the variation between individuals, the confidence limits will be based on only minor variations among presumably homogeneous persons. It may be anticipated that different classes of operators will differ in variability and different confidence limits must be determined for each class.

EXPANDING THE EXPERIMENTAL SPACE

For numerous reasons, an investigator may wish to go beyond the original experimental space. He may wish to add a new dimension (variable) or he may wish to expand the range of an existing variable. He may wish to examine performance in a corner just outside the hypersphere space covered by the central-composite design. The original equation provides a basic framework on which any new data can be "hung." In adding data points beyond the original experimental space, some data points within that space should also be included in the add-on design.

X. PHASE FIVE: VERIFICATION

Too many human factors experiments performed in the laboratory are never verified in the field. Results are published and some data finds its way into handbooks without ever having been tested operationally. In most cases, however, these are component results, often trivial, which probably will have relatively little effect on system performance in the long run. Generally they are sufficiently simple that an investigator can use his "common sense" to evaluate their effectiveness. On the other hand, if the new paradigm is followed, the results will appear in the form of a complex equation, not readily subject to "common sense" evaluation. It must be validated in the field.*

Validation serves two purposes:

1. It determines how good the equation is.
2. It determines how bad the equation is.

In an iterative research program, knowing what remains to be accounted for is very important for it signifies that there is still more to be done. If the proportion of

---

*Psychologists have been prone to "evaluate" results from laboratory experiments with results from other laboratory experiments. This is not acceptable for human factors engineering research since the biggest danger -- whatever precautions were taken in Phase One -- is that the laboratory simulation may be an oversimplification of field conditions or that variables nominally the same are in fact quantitatively different. Evaluation must be based on field studies under realistic conditions.

unexplained variance is large, the investigator should search for new variables, higher-order terms in the equation, impurities in his data collection, and/or errors in his analysis. Studying the residuals may provide some clues (Anscombe and Tukey, 1963; Daniel and Woods, 1971).

Evaluating the equation operationally does not mean that every data point in the experiment must be repeated in the field. Instead, verification might be performed by taking only a few representative points distributed within the experimental space. Empirically derived equations are not intended to predict performance outside the experimental space; to do so is dangerous.

Verification studies in which prediction scores are compared to scores obtained under actual operational conditions would not require elaborate designs. The sampling process may be systematic, but need not be. It may be at points of specific interest to the investigator or in the general case, at points literally selected at random throughout the space. An important principle here is: Given a limit on the number of observations that can be made, better information will be obtained by sampling many different points rather than replicating only a few points. For the first time in the entire research sequence, tests of statistical significance might appropriately be employed. Simple linear correlations and t-tests between the values obtained empirically and those estimated from the equation will enable a judgment to be made regarding the accuracy of the prediction.

As in any empirical verification process, one must be assured that the empirical circumstances are representative of the ones presumably being predicted by the equation. If not, then it means that critical variables have been omitted from the equation or that the data collection in the real world was unnecessarily messy. All this means is that the investigator must remain observant at all times to be assured that what he wants to measure, what he thinks he is measuring, and what he should measure are all the same.

For each observation point, the value of each critical variable under operational conditions should be recorded. In fact, it would also be worthwhile to record the values under operational conditions of the other candidate variables that were not critical for the present task. They may be critical in related tasks and keeping the values recorded in both the laboratory and the field enables a solid data base to be built and used.

XI. CONCLUSIONS

A new paradigm has been proposed which, when properly used, will increase the chances that data collected in the laboratory will predict with reasonable accuracy performance in the field. Furthermore the data will be collected in a way that permits a modular data base to be constructed. The chief features of the new approach are that it uses the manipulative approach in a holistic context and is capable of performing large multifactor experiments economically.

As presented here, the paradigm should be viewed as a total research strategy rather than a set of discrete experimental techniques. Because sections of the paradigm were described segmentally in earlier papers, some investigators have inappropriately used a single section as a finished experimental plan, often confounded with traditional tactics.

In spite of the fact that some segments have never been fully worked out within the context of the paradigm (as indicated by the referencing code in the text), the approach is at a workable stage. It can be used now. While modifications may be expected in specific techniques as more experience is gained, the philosophy and to some extent the overall strategy can be expected to remain relatively intact. These are the elements that make the new paradigm a viable and powerful research tool.

XII. REFERENCES

Adams, J. A. Research and the future of engineering psychology. American Psychologist, 1972, 27, 615-622.

Albee, G. W. The uncertain future of clinical psychology. American Psychologist, 1970, 25, 1071-1080.

Anscombe, F. J., and J. W. Tukey. The examination and analysis of residuals. Technometrics, 1963, 5, 141-160.

Bakan, D. Psychology can now kick the science habit. Psychology Today, March 1972, pp 26-28, 86-88.

Bakan, D. The test of significance in psychological research. In Steger, J. A. (Ed.), Readings in statistics for the behavioral scientist. N. Y.: Holt, Rinehard and Winston, Inc., 1971.

Bakan, D. The mystery-mastery complex in contemporary psychology. American Psychologist, 1965, 20, 186-191.

Bass, B. M. The substance and the shadow. American Psychologist, 1974, 29, 871-886.

Bogartz, R. S., and J. H. Wackwitz. Polynomial response scaling and functional measurement. Journal of Mathematical Psychology, 1971, 8, 418-443.

Booth, K. H. V., and D. R. Cox. Some systematic supersaturated designs. Technometrics, 1962, 4, 489-495.

Boring, E. G. A history of experimental psychology. N. Y.: Appleton-Century-Crofts, 1950.

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, 1958, pp 138-190.

Box, G. E. P., and P. W. Tidwell. Transformation of the independent variables. Technometrics, 1962, 4, 531-550.

Bryan, G. L. Evaluation of basic research in the context of mission orientation. American Psychologist, 1972, 27, 947-950.

Budne, T. A. Application of random balance designs. Technometrics, 1959a, 1, 139-155.

Budne, T. A. Random balance: Part I - the missing statistical link in fact finding techniques. Industrial Quality Control, 1959b, 15(10), 5-10.

Budne, T. A. Random balance: Part II - techniques of analysis. Industrial Quality Control, 1959c, 15(11), 11-16.

Bugental, J. F. T. Humanistic psychology: a new break-through. American Psychologist, 1963, 18, 563-567.

Carlson, R. Where is the person in personality research? Psychological Bulletin, 1971, 75, 203-219.

Carney, E. J. Ridge estimates for canonical analysis. Ithaca, N. Y.: Cornell University. Technical Report 75-114, April 1975.

Caro, P. W. Aircraft simulators and pilot training. Human Factors, 1973, 15, 502-509.

Caro, P. W. Some factors influencing Air Force simulator training effectiveness. Alexandria, VA: Human Resources Research Organization, Tech. Rep. No. 77-2, March 1977.

Cattell, R. B., (Ed.). Handbook of multivariate experimental psychology. Chicago: Rand McNally, 1966a.

Cattell, R. B. Guest editorial: Multivariate behavioral research and the integrative challenge. Multivariate Behavioral Research, 1966b, 1, 4-23.

Chapanis, A. Research techniques in human engineering. Baltimore: John Hopkins Press, 1959.

Chapanis, A. Engineering psychology. Annual Review of Psychology, 1963, 14, 285-318.

Chapanis, A. The relevance of laboratory studies in practical situations. Ergonomics, 1967, 10, 557-577.

Controller General of the United States. Human resources research and development results can be better managed. U. S. General Accounting Office, FPCD-77-43, April 22, 1977.

Cronbach, L. J. The two disciplines of scientific psychology. American Psychologist, 1957, 12, 671-684.

Cronbach, L. J. Beyond the two disciplines of scientific psychology. American Psychologist, 1975, 30, 116-127.

Curtis, E. W. Predictive value compared to predictive validity. American Psychologist, 1971, 26, 908-914.

Daniel, C. Applications of statistics to industrial experimentation. N. Y.: Wiley, 1976.

Daniel, C., and F. S. Wood. Fitting equations to data. N. Y.: Wiley-Interscience, 1971.

Deese, J. Psychology as science and art. N. Y.: Harcourt Brace Jovanovich, 1972.

Draper, N. R., and T. J. Mitchell. Construction of the set of 256-run designs of resolution ≥ 5 and the set of even 512-run designs of resolution ≥ 6 with special reference to the unique saturated designs. Annals of Mathematical Statistics, 1968, 39, 246-255.

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.

Dunnette, M. D. Fads, fashions, and folderol in psychology. American Psychologist, 1966, 21, 343-352.

Dykstra, Jr., O. Partial replication of response surface designs. Technometrics, 1960, 2, 185-195.

Edgington, E. S. A new tabulation of statistical procedures used in APA journals. American Psychologist, 1975, 29, 25-26.

Elms, A. C. The crisis of confidence in social psychology. American Psychologist, 1975, 30, 967-976.

Engineering Statistical Methods Group, Increasing the efficiency of development testing. East Hartford, Conn.: Pratt & Whitney Aircraft PWA-2236, July 1963.

Farberow, N. L. The crisis is chronic. American Psychologist, 1973, 28, 388-394.

Fiske, D. W. The limits for the conventional science of personality. Journal of Personality, 1974, 42, 1-11.

Gadlin, H. and G. Ingle. Through the one-way mirror, the limits of experimental self-reflection. American Psychologist, 1975, 30, 1003-1009.

Ghiselli, E. E. Some perspectives for industrial psychology. American Psychologist, 1974, 29, 80-87.

Goldwater, B. M. The coming breakpoint. N. Y.: Macmillan, 1976.

Greening, C. P., and H. L. Snyder. Visual target acquisition study. Supplementary Report A in Air-to-sruface missile technology, 1975-1980 (U), Vol. IV, Supplementary reports, Report R-133, Institute for Defense Analyses, Washington, D. C., December 1967.

Guilford, J. P. Psychometric methods. N. Y.: McGraw-Hill, 1936.

Hebb, D. O. What psychology is about. American Psychologist, 1974, 29, 71-79.

Human Factors Society Bulletin, Cuts threaten combat readiness and defense economy. Santa Monica, CA.: Human Factors Society, May 1977, pp 1-2.

Jacobs, R. S., and S. N. Roscoe. Simulator cockpit motion and the transfer of initial flight training. Santa Monica, CA.: Proceedings, Human Factors Society, 19th Annual Meeting, October 1975.

John, P. W. M. Augmenting 2^{n-1} designs. Technometrics, 1966, 8, 469-480.

Kleijnen, J. P. C. Screening designs for poly-factor experimentation. Technometrics, 1975, 17, 487-493.

Kleiter, G. The crisis of significance tests in psychology. Jahrbuch für Psychologie, Psychotherapie und Medizinische Anthropologie, 1969, 17, 144-163. (Translated by D. P. Barrett, Royal Aircraft Establishment Library Translation No. 1649, The R. A. E. Library, Q.4 BUILDING, R.A.E. Farnborough Hants, England, June 1972).

Koch, S. Psychology cannot be a coherent science. Psychology Today, September 1969, pp 14, 64, 66-68.

Koonce, J. M. Effects of ground-based aircraft simulation motion conditions upon prediction of pilot proficiency. Savoy, Ill.: University of Illinois at Urbana-Champaign, Aviation Research Lab., TR ARL-74-5/AFOSR-74-3, 1974.

Li, C. H. A sequential method for screening experimental variables. Journal of the American Statistical Association, 1962, 57, 455-477.

Lipsey, M. W. Research and relevance, a survey of graduate students and faculty in psychology. American Psychologist, 1974, 29, 541-553.

Lockard, R. B. Reflections on the fall of comparative psychology: is there a message for us all? American Psychologist, 1971, 26, 168-179.

Lucas, J. M. Optimum composite designs. Technometrics, 1974, 16(4), 561-567.

Lykken, D. T. Statistical significance in psychological research. Psychological Bulletin, 1968, 70, 151-159.

Mackie, R. R., and P. R. Christensen. Translation and application of psychological research. Goleta, California: Human Factors Research, Inc., Technical Report 716-1, 1967.

Maslow, A. H. Problem-centering vs. means-centering in science. In D. P. Schultz (Ed.). The science of psychology: critical reflections. N. Y.: Appleton-Century-Crofts, 1970.

Meehl, P. Theory testing in psychology and physics. Philosophy of Science, 1967, 34, 103-115.

Meister, D. Human factors: theory and practice. N. Y.: Wiley-Interscience, 1971.

Meister, D., and D. J. Sullivan. A further study of the use of human factors information by designers. Final Report. Canoga Park, CA.: Bunker-Ramo, March 16, 1967.

Meyer, H. H. The future for industrial and organizational psychology. American Psychologist, 1972, 27, 608-614.

Myers, R. M. Response surface methodology. Boston: Allyn and Bacon, 1971.

Miller, G. A. Psychology as a means of promoting human welfare. American Psychologist, 1969, 24, 1063-1075.

Nunnally, J. The place of statistics in psychology. Educational and Psychological Measurement, 1960, 20, 641-650.

Pajak, T. F., and S. Addelman. Minimum full sequences of 2^{n-m} resolution III plans. J. Royal Statistical Society, Series B, 1975 #37, 88-95.

Patel, M. S. Group-screening with more than two stages. Technometrics, 1962, 4(2), 209-217.

Patel, M. S. Partially duplicated fractional factorial designs. Technometrics, 1963, 5(1), 71-83.

Peters, C. C., and W. R. Van Voorhis, Statistical procedures and their mathematical bases. N. Y.: McGraw Hill, 1940.

Price, H. E. Congressional budget ax chops at human factors. In Human Factors Society Bulletin, Volume 18, No. 11, November 1975.

Reising, J. M., S. L. Ward, and E. P. Roke. Some thoughts on improving experiments. Human Factors, 1977, 19, 221-226.

Rhodes, F. Predicting the difficulty of locating targets from judgments of image characteristics. Dayton: Wright-Patterson AFB, USAF AMRL TDR 64-19, 1964.

Rickover, H. G. Letter written in 1970, as quoted in the Human Factors Society Bulletin, 1977, 20, p 1.

Roquemore, K. G. Hybrid designs for quadratic response surfaces. Technometrics, 1976, 18, 419-423.

Royce, J. R. Guest editorial: Have we lost sight of the original vision for SMEP and MBR? Multivariate behavioral research, 1977, 12, 135-141.

Rozeboom, W. W. The fallacy of the null-hypothesis significance test. Psychological Bulletin, 1960, 57, 416-428.

Satterthwaite, F. E. Random balance experimentation. Technometrics, 1959, 1, 111-137.

Signorelli, A. Statistics, tool or master of the psychologist. American Psychologist, 1974, 29, 774-777.

Silverman, I. Crisis in social psychology: the relevance of relevance. American Psychologist, 1971, 26, 583-584.

Simon, C. W. Reducing irrelevant variance through the use of blocked experimental designs. Culver City, CA.: Hughes Aircraft Co., Tech. Rep. No. AFOSR-70-5, November 1970a, 65 pp.

Simon, C. W. The use of central-composite designs in human factors engineering experiments. Culver City, CA.: Hughes Aircraft Co., Tech. Rep. No. AFOSR-70-6, December 1970b, 52 pp.

Simon, C. W. Considerations for the proper design and interpretation of human factors engineering experiments. Culver City, CA.: Hughes Aircraft Co., Tech. Rep. No. P73-325, December 1971, 135 pp.

Simon, C. W. Experiment simultation. Culver City, CA.: Hughes Aircraft Co., Tech. Rep. No. ARL-72-7/AFOSR-72-3, April 1972, 48 pp.

Simon, C. W. Economical multifactor designs for human factors engineering experiment. Culver City, CA.: Hughes Aircraft Co., Tech. Rep. No. P73-326A, June 1973, 171 pp.

Simon, C. W. Methods for handling sequence effects in human factors engineering experiments. Culver City, CA.: Hughes Aircraft Co., Tech. Rep. No. P74-451A, December 1974, 197 pp.

Simon, C. W. Methods for improving information from "undesigned" human factors experiments. Culver City, CA.: Hughes Aircraft Co., Tech. Rep. No. P75-287, July 1975a, 82 pp.

Simon, C. W. Evaluation of basic and applied research. Pragmatic criteria. Paper presented at 83rd Annual Convention, American Psychological Association, Chicago, IL., 31 August 1975b.

Simon, C. W. Response surface methodology revisited: a commentary on research strategy. Westlake Village, CA.: Canyon Research Group, Inc., Tech. Rep. No. CWS-01-76, July 1976a, 60 pp.

Simon, C. W. Analysis of human factors engineering experiments: characteristics, results and applications. Westlake Village, CA.: Canyon Research Group, Inc., Tech. Rep. No. CWS-02-76, August 1976b, 104 pp.

Simon, C. W. Design, analysis, and interpretation of screening designs for human factors engineering research. Westlake Village, CA.: Canyon Research Group, Inc., Tech. Rep. No. CWS-03-77, September 1977, 220 pp.

Tukey, J. W. Where do we go from here? Journal American Statistical Association, 1960, 55, 80-93.

Tyler, L. E. Design for a hopeful psychology. American Psychologist, 1973, 28, 1021-1029.

Uhlaner, J. E. Human performance, jobs, and systems psychology -- the system measurement bed. Presidential address, Division of Military Psychology, American Psychological Association, Miami, Florida, 6 September 1970.

Vartabedian, A. G. The effects of letter size, case, and generation method on CRT display search time. Human Factors, 1971, 13, 363-368.

Viteles, M. S. Psychology today: fact and foible. American Psychologist, 1972, 27, 601-607.

Watson, G. S. A study of the group-screening method. Technometrics, 1961, 3, 371-388.

Whitwell, J. C., and G. K. Morbey. Reduced designs of resolution five. Technometrics, 1961, 3, 459-477.

Williams, A. C., Jr., and M. Adelson. Some considerations in deciding about the complexity of flight simulation. San Antonio: Lackland AFB, Research Bulletin AFPTRC-TR-54-106, December 1954.

Wohlwill, J. F. The study of behavioral development. New York: Academic Press, 1973.

Youden, W. J., O. Kempthorne, J. W. Tukey, G. E. P. Box, and J. S. Hunter. Discussion of the papers of Messrs. Satterthwaite and Budne (including authors' responses to discussions). Technometrics, 1959, 1, 157-193.

APPENDIX I

PHILOSOPHICAL DIFFERENCES BETWEEN OLD AND NEW EXPERIMENTAL PARADIGMS FOR HUMAN ENGINEERING RESEARCH

Comments (from the text on "Research Techniques in Human Engineering") representing the traditional experimental philosophy were cited on page 25 of this report. The corresponding philosophy of the new approach is given in contrast below.

1. Box and Hunter (1958, p 139) have stated that "the only time an experiment can be properly designed is after it has been completed." They note the indeterminants of most research and the dangers and difficulties of devising experiments that "proceed in accordance with some set of unalterable rules." To handle this paradox, therefore, they suggest: "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." This process of experimental iteration is fundamental to the new paradigm.

2. Some measure of random error is desirable, but not always necessary if the cost of obtaining it exceeds its immediate value. During the screening phase, when a large number of variables is being investigated and observations are at a premium, the measure of random error is of minor

importance. Examining a great number of potentially critical variables will be expected to reduce the bias error (which many psychologists have included in their measure of random error); no test of statistical significance is required at that time since variables are being compared according to the size of their relative effect on performance. Rough estimates of error can be obtained "internally" through the use of graphic plot techniques.

3. Confounding variables can lead to interpretation problems. Confounding sources of variance is a fundamental procedure of the new paradigm, when the sources are from different orders of the total variance package, i.e., main effects, interactions, and so forth. Also, in the early phase of a research program when empirical efforts to identify critical variables require group screening techniques, rational confounding is a necessity.

4. Factorial designs are among the most wasteful designs that can be employed in behavioral research. Seldom if ever will the effects of higher-than-three-factor interactions have to be isolated, so collecting the data required for a complete factorial serves no useful purpose. Fractional factorials are useful and are basic blocks in the iterative process noted in Item #1 above.

5. Holding relevant variables constant will result in a "clean" experiment, in the sense that the effects of the variables of interest will not be confounded with those held constant. However, if

the data were to be used to predict to the operational situation where the variables being held constant are not at the values selected for the experiment, a bias error will be introduced into the prediction. By finding ways of studying a very large number of factors, the new paradigm tries not to have to hold any operationally critical variable constant in the experiment in the function-writing stages. For validation of particular conditions, however, the procedure would be used.

6. One should not "handle" individual differences by testing a large number of subjects. The reasons that individuals perform differently on a particular task should be identified and included as an experimental variable. To fail to do so reduces the prediction power of the experimental results, and may fail to identify important subject-by-condition interaction.

7. When critical personnel factors have been removed, a truly homogeneous subject population should remain, making the uneconomical replication of the basic design less necessary. At the end of the experimental process, a measure of the fiducial limits would be made, but this should be possible with a relatively few subjects, particularly if the preceding steps have been properly taken.

DISTRIBUTION LIST

LCdr. James Ashburn, MSC, USN
NAMRL, Bldg. 1953
Pensacola, FL 32512

Dr. L. E. Banderet
SGDR-UE-CR
Dept. of the Army
U. S. Army Research Institute
of Environmental Medicine
Natick, Mass. 01760

Mr. Vernon E. Carter
Pilot Training Systems
Orgn 3750/62
Northrop Corp./Aircraft Div.
3901 W. Broadway
Hawthorne, CA 90250

Dr. Julien M. Christensen
Chairman, Dept. of Industrial Engr.
Wayne State University
Detroit, Michigan 48202

Dr. Chriss Clark
Honeywell, Inc. (MS R-2340)
2600 Ridgway Parkway
Minneapolis, Minn. 55413

Mr. James Duva (N-215)
Naval Training Equipment Ctr.
Orlando, FL 32813

Dr. Gordon A. Eckstrand
AFHRL/AS
Wright-Patterson AFB OH 45433

Mr. Ronald A. Erickson, Code 3175
Head, HF Branch, Weapons Devel. Dept.
U. S. Naval Weapons Center
China Lake, California 93555

Dr. Marshall J. Farr
ONR, Code 458
800 N. Quincy Street
Arlington, VA 22217

Terrence W. Faulkner
Health & Safety Laboratory
Bldg. 56, Kodak Park Division
Eastman Kodak Co.
Rochester, N. Y. 14650

Mr. Charles A. Gainer
Chief, Army Research Unit
Bldg. 502, P. O. Box 428
Ft. Rucker, Alabama 36360

Dr. Robert A. Goldbeck
Mail Station S-32
Western Development Laboratories Div.
Philco-Ford Corporation
3939 Fabian Way
Palo Alto, California 94303

James E. Goodson, CDR MSC USN
Head, Aerospace Psychology Dept.
Code 15, Naval Aerospace Med.
Research Lab.
Pensacola, Florida 32508

Dr. Tom Gray
AFHRL/FT
Williams AFB, AZ 85224

G. C. Helmstadter, Director
University Testing Services
Payne Hall, B302
Arizona State University
Tempe, Arizona 85281

Dr. Richard Jagacinski
Human Performance Center
330 Packard Road
Ann Arbor, Michigan 48104

Dr. Edgar M. Johnson
U. S. Army Research Institute for the
Behavioral and Social Sciences
1300 Wilson Blvd.
Arlington, VA 22209

AFHRL/ASM (Patricia A. Knoop)
Wright-Patterson AFB OH 45433

Distribution List (Continued)

Dr. Richard L. Krumm
P. O. Box 2706
Main Post Office
Washington, D. C. 20013

Mr. Robert G. Mills
6570th AMRL/HEB
Wright-Patterson AFB, OH 45433

Dr. Frederick A. Muckler (Code 311)
Prog. Dir., Design of Manned Systems
Navy Personnel R. & D. Center
San Diego, CA 92152

Dr. Wallace W. Prophet
Director, HumRRO Cent. Div.
400 Plaza Bldg.
Pensacola, FL 32505

Dr. Clyde R. Replogle
6750 AMRL/EME
Wright-Patterson AFB OH 45433

Charles V. Riche
School of Psychology
Georgia Institute of Technology
Atlanta, GA 30332

Dr. Marty Rockway
Technical Director
AFHRL/TT
Lowry AFB, CO 80230

Dr. Stanley N. Roscoe
Aviation Research Laboratory
Univ. of Ill. -- Willard Airport
Savoy, IL 61874

Dr. Mark Sanders
Department of Psychology
California State University
Northridge, CA 91324

Dr. Dennis E. Smith
Mathematical Statistician
Desmatics, Inc.
P. O. Box 863
State College, PA 16801

Dr. Margaret J. Smith
Naval Education and Training
Program Development Center
Ellyson Field
Pensacola, FL 32509

H. C. Strasel
Chief, ARI Field Unit
U. S. Army Research Institute
P. O. Box 2086
Ft. Benning, GA 31905

Dr. Martin A. Tolcott
Human Engineering Div., ONR
800 N. Quincy Street
Arlington, VA 22217

Dr. Donald A. Topmiller
AMRL/HES
Wright-Patterson AFB OH 45433

Dr. R. Young, Director
Human Resources Office, ARPA
1400 Wilson Blvd.
Arlington, VA 22209

HQ AFSC/DLS
Andrews AFB, MD 20334

HQ AFHRL/CC
Brooks, AFB, TX 78235

Director, Behavioral Sciences Dept.
USAF Academy
Colorado Springs, CO 80840

Flight Dynamics & Control Division
Mail Stop 152
NASA - Langley Research Center
Hampton, VA 23665
Attn: Gary P. Beasley

Department of the Air Force
Air Force Human Relations Lab. (AFSC)
Lackland AFB, TX 78236
Attn: Mark Nataupsky, Capt., USAF
Chief, Evaluation Section
Personnel Research Division

Distribution List (Continued)

Military Asst. for Human Resources
OAD (E&LS)
ODDR&E
Pentagon, Washington, D. C. 20330

Executive Editor
Psychological Abstracts
American Psychological Assn.
1200 17th St. N.W.
Washington, D.C. 20036

HQ USAF/RDPS
Washington, D.C. 20330

AFFDL/CC
Wright-Patterson AFB, OH 45433

Director
USAF Avionics Laboratory
Wright-Patterson AFB, OH 45433

AMD/RDH (Col. George C. Mohr)
Brooks AFB, TX 78235

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 8757
B.W.I. Airport, Maryland 21240

National Technical Information
Services (NTIS)
Operations Division
5285 Port Royal Road
Springfield, VA 22151

Elizabeth Lage Roscoe
2117 Bristol Road
Champaign, Illinois 61820

Dr. J. Robert Newman
Department of Psychology
California State College
Long Beach, CA 90840

Mr. R. S. Easterby
Applied Psychology Department
The University of Aston
College House, Gosta Green
Birmingham B47ET England

Dr. J. Peter Kincaid
Martin Marietta Aerospace
Orlando Division
P. O. Box 5837
Orlando, FL 32805

Dr. W. G. Matheny
Life Sciences Inc.
227 Loop 820 Northeast
Hurst, TX 76053

Dr. Thomas J. Triggs
Department of Psychology
Monash University
Clayton, Victoria 3168
Australia

Dr. James J. Regan
Navy Personnel R&D Ctr.
San Diego, CA 92152