

16.00

RESERVE

HOW PSYCHOLOGISTS DO RESEARCH

6th Edition - 2nd version



Thomy Nilsson

HOW PSYCHOLOGISTS DO RESEARCH

6th edition - 2nd version

**Thomy Nilsson
Professor of Psychology
University of Prince Edward Island**

**2003
U.P.E.I. Central Printing
Charlottetown, Prince Edward Island, Canada
(copyright reserved by the author)**

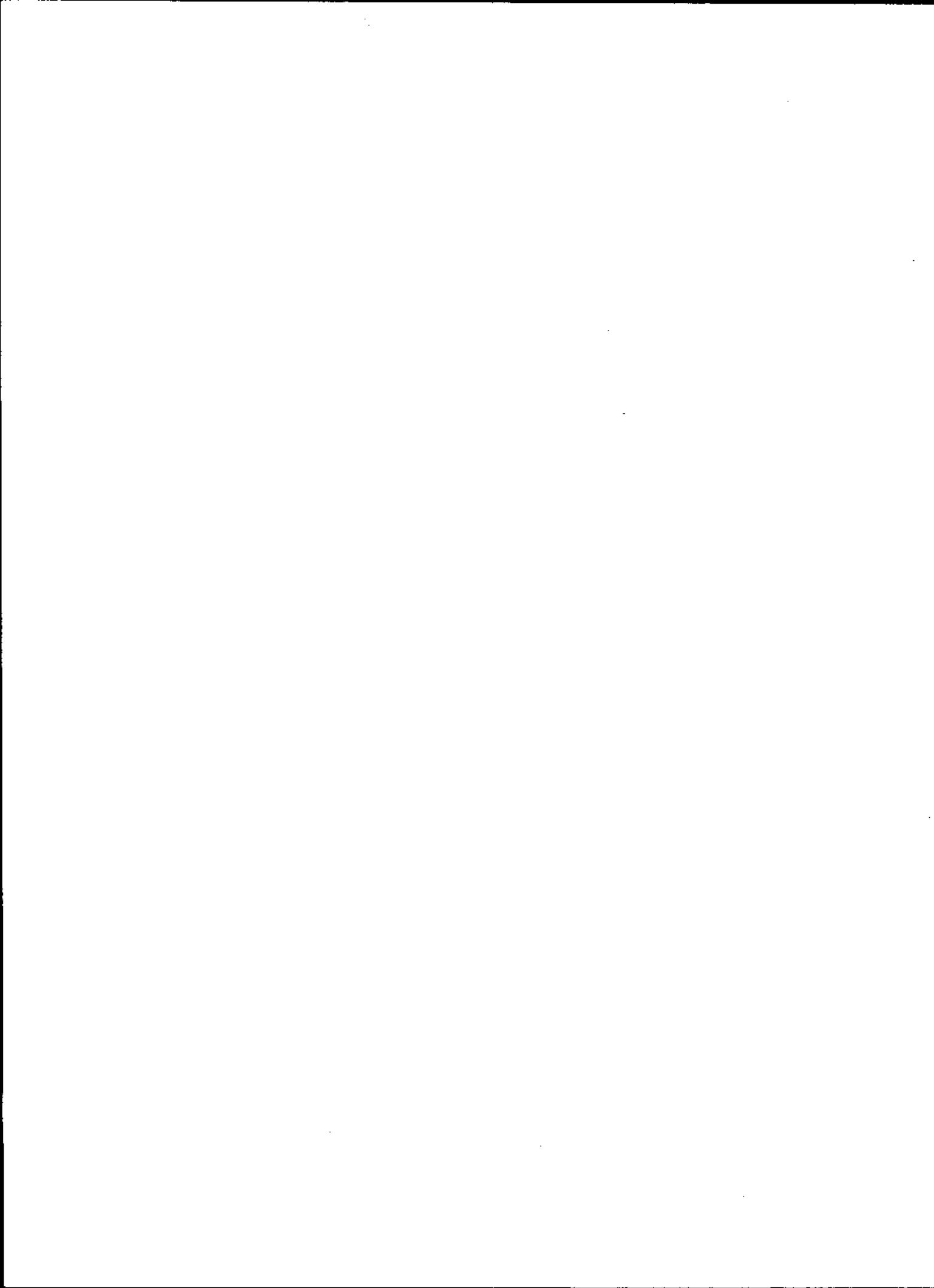


CONTENTS

Chapter 1 INTRODUCTION	1
1.1 Why It's Necessary	
1.2 Sources of Knowledge	
1.3 Writing Research Reports	
1.4 Some Essential Definitions	
1.5 Archives: Research Without Measuring	
Exercises	
Chapter 2 DOING RESEARCH ETHICALLY	29
2.1 Obligations to Humanity	
2.2 Obligations to Your Subjects	
2.3 When Deception Is Necessary	
2.4 Responsibilities for Animal Subjects	
2.5 Professional Responsibilities	
2.6 How Research Ethics Are Maintained	

SECTION I - MAKING MEASUREMENTS

Chapter 3 DIRECT OBSERVATION AND TRACES	47
3.1 To Intervene or Not to Intervene	
3.2 Difficulties You May Encounter	
3.3 Recording Observational Data	
3.4 Who to Observe	
3.5 Traces: "Reading" the World Around You	
Exercises	
Chapter 4 ASKING QUESTIONS	67
4.1 Surveys, Polls, and Subjective Measures	
4.2 What to Ask and How to Ask It	
4.3 Who to Ask	
4.4 How to Conduct Surveys	
4.5 What to Do with Survey Data	
4.6 Structured Interviews - with Philip Smith	
Exercises	
Chapter 5 LABORATORY METHODS	99
5.1 Measuring Performance	
5.2 Measuring Perception	
5.3 How Thinking and Memory Are Measured	
5.4 Measuring What Infants Know	
5.5 Physiological Measurements	
5.6 Using Animals as Subjects	
Exercises	



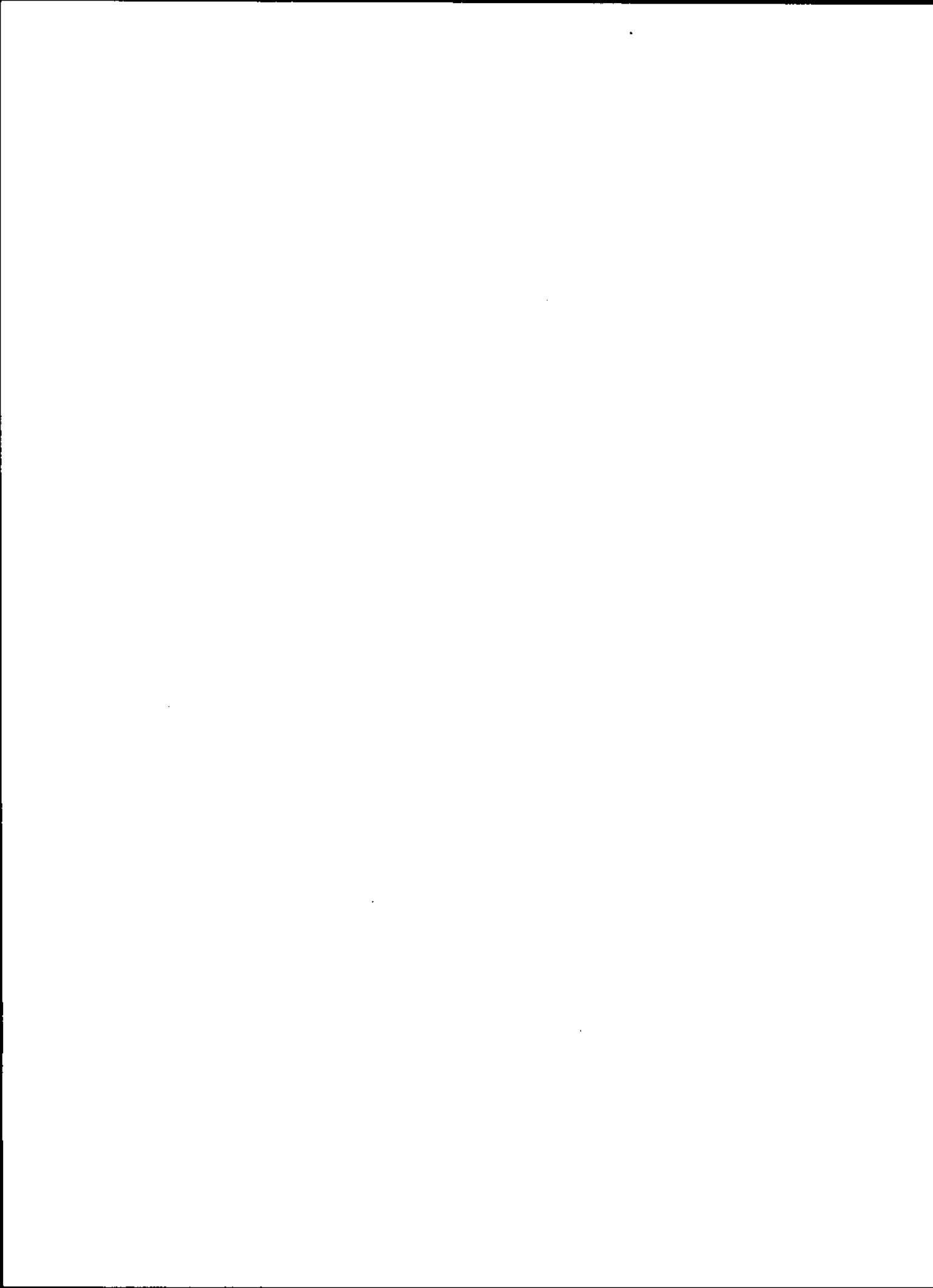
SECTION II - RESEARCH DESIGNS

Chapter 6 DESCRIPTIVE RESEARCH	115
6.1 A School Yard Example	
6.2 Measuring Quality of Life	
6.3 Personality and Friendship	
6.4 In the Laboratory: Reading in Color	
6.4 Case Studies	
Exercises	
Chapter 7 BETWEEN-SUBJECTS EXPERIMENTS	137
7.1 What's So Special About Experiments	
7.2 Obtaining Subjects	
7.3 Assigning Subjects	
7.4 Factorial Designs	
Exercises	
Chapter 8 WITHIN-SUBJECTS EXPERIMENTS	177
8.1 Advantages and Disadvantages	
8.2 Coping with Order and Carry-Over Effects	
8.3 Complete, Incomplete, and Mixed Designs	
Exercises	
Chapter 9 FIELD RESEARCH	217
9.1 Advantages of Working in the Field	
9.2 Variability in Field Research	
9.3 Confounding Effects	
9.4 Out of the Laboratory Into the Frying Pan	
Exercises	
Chapter 10 QUASI EXPERIMENTS	245
10.1 Between-Groups Quasi-Experiments	
10.2 Within-Groups Quasi-Experiments	
10.3 Group Characteristics as Independent Variables	
10.4 Comparing the Three Quasi Designs	
10.5 Tricks of the Trade	
Exercises	
EPILOGUE	268

Appendices

- A - USING THE LIBRARY
- B - HOW TO START USING SPREADSHEETS
- C - GUIDES FOR ETHICAL RESEARCH
- D - A PRIMER ON LABORATORY AUTOMATION
- E - FIVE HANDS-ON RESEARCH PRACTICA - with Paul Grey

RANDOM NUMBER TABLE



A NOTE TO TEACHERS

(Don't read this if you are a student. It contains privileged information about my strategy for teaching with this book)¹

Most of teachers recognize that few students go on to become research psychologists. Thus it has become fashionable to write books that simply try to explain the principles according to which research is done rather than how to actually do it. The intent is to make students knowledgeable users of research information rather than researchers themselves. However, I believe the best way to gain an understanding of something is by learning how to do it. Therefore I have written this book to introduce psychology students by emphasizing the "nuts-and-bolts" aspects of research. If students also acquire some practical research skills along the way, that will likely benefit them when they encounter research requirements in upper level courses or postgraduate studies.

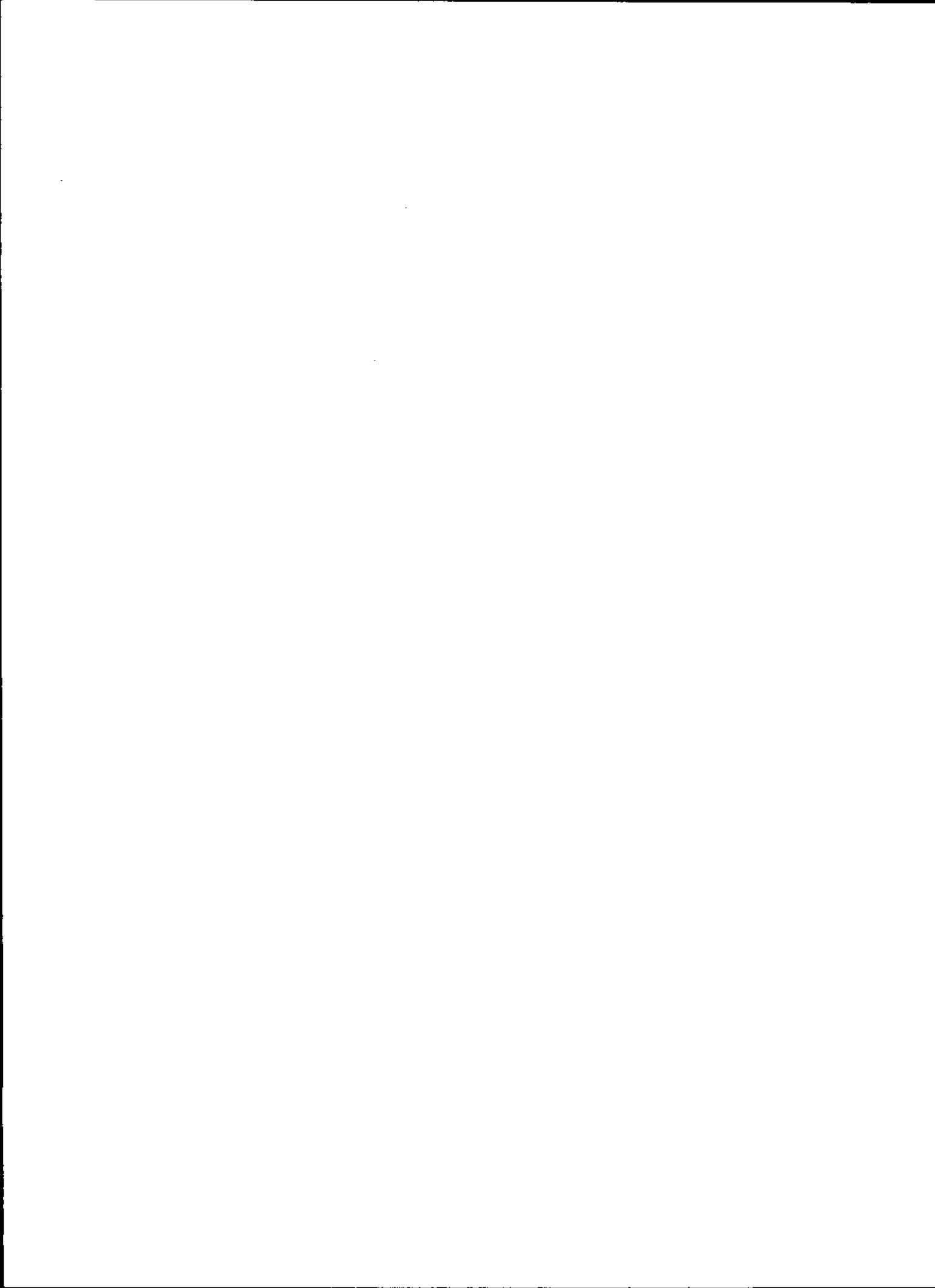
Most chapters have extensive examples of hypothetical (and not so hypothetical) data obtained in different research scenarios. These are used to illustrate in concrete terms various concepts from how numbers can be obtained, subject differences, confounding effects, as well as how to organize and analyze data. In class discussions and on tests, I expect students to create research projects which include details such as where they will obtain their subjects, how many are needed, where the research will be conducted, the sequence in which the subjects will be tested or the conditions presented, and what sort of statistics are appropriate for analysis of the data.

As you looked over the contents, I hope you noticed a major difference in the organization of this text from most other texts in this area. That difference is the distinction between different ways of making measurements and different types of research designs. For many years using other texts, these two aspects were confounded in my teaching. Survey methods tended to be relegated to descriptive designs and correlational analysis; observational methods seemed most suited to quasi-experiments; etc. Small wonder that in my advanced experimental course, students had difficulty realizing that survey measurements were appropriate for a within-subjects experiment, for example. It often took some convincing to get students to recognize the broader possibilities because these were not in their books.

This book is intended for students having little experience doing laboratory and field exercises using social science research methods. In upper level courses, I have encountered too many students who were unable to actually do research even though they had good grades in a prerequisite research course. Each chapter (except the one on ethics) introduces a different approach to obtaining data or a different type of research design. Research exercises are provided at the end of each chapter. A separate laboratory component in my course involves a series of semi-weekly research projects for which the student is expected to collect data and write a brief report. The first week introduces the project and explains how it should be done. The second week's tutorial explains how to analyze the data and write the report. These are described in the Appendix.

A major difference between research in psychology and other sciences is the

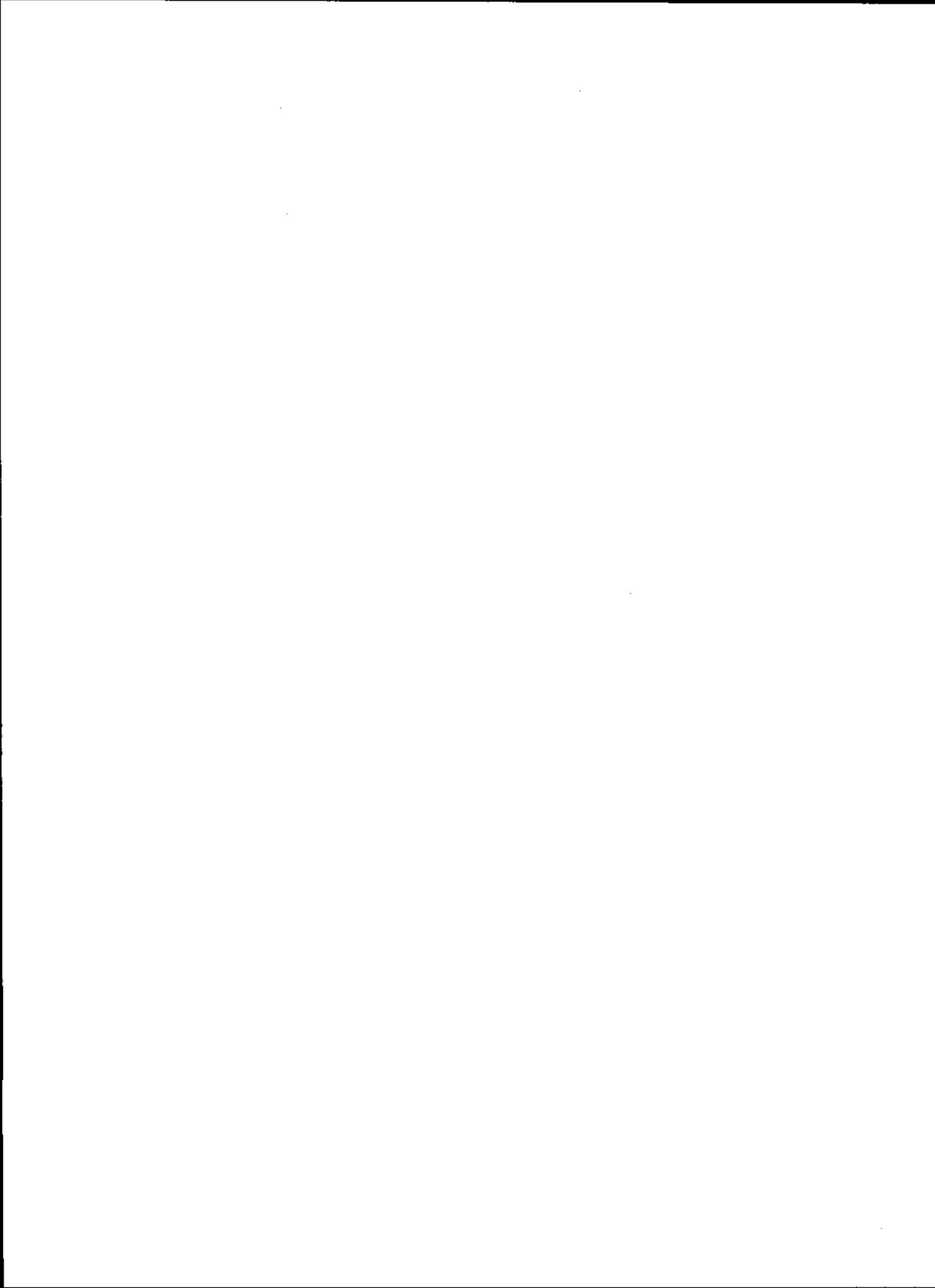
¹ One of the characteristics of a good researcher is curiosity; another is a willingness to break the rules. If you are reading this, consider yourself a good candidate for being a researcher.



need to relate to the subjects whether human or animal. The best way to learn what this involves is to do it. The project for Chapter 3 eases the student into this new realm by having him or her observe the carrying behavior of fellow students. The next chapter puts the student into direct interaction with subjects by having them approach "strangers" and asking them to complete a survey questionnaire. Instructors should be aware that for many students this involves substantial trepidation and calls upon students to muster real effort to carry out the survey - but such is the way of real research in psychology. Hopefully, most students will have developed sufficient confidence from these experiences to be comfortable with the more extended researcher-subject interactions required by experiments in the later chapters.

Many introduction to research books have chapters on the nature of science, theories of measurement, developing hypotheses, concepts of control, power, and limitations of research. Certainly these are valuable topics for discussion with colleagues and advanced students. However, I have found that many of these lessons had little meaning for students just starting to learn about research. Worse, all the caveats left many students confused and wondering whether any psychological research was worthwhile. Most of us had our first research experiences in general science. Consider how you would have felt doing your first chemistry lab if first you had study logical positivism (a philosophical rationale of why we can believe our senses and by extension any measurement), learn about all the personal and social influences that threaten the validity of scientific data, and espouse every ethical strictures on ethical on doing research. I suggest not putting much emphasis on the abstract aspects of the first two chapters. Rather, refer to them throughout the course as issues arise.

Field-research and quasi-experimental designs in the last chapter provide a good opportunity to discuss practical and ethical limits particular to psychological research. Having gained some first hand experience by then, these matters may be more than abstract concepts. I believe the best way to promote ethical research is by showing student how to do effective research. Appropriate wording of survey questions, balanced assignment of subjects to groups, and experience in working with subjects provides more working knowledge of how to avoid theoretical and ethical pitfalls than memorizing principles and rules. "Take care of the small things and the big things tend to take care of themselves."



CHAPTER 1 INTRODUCTION

1.1 WHY IT'S NECESSARY

Most students study psychology because they want to help others and to understand themselves. Research is far from their mind. You are probably taking this course only because it is required for upper level courses in psychology. In the course description and this book's Table of Contents you may see little to dispel resentment at having to study something you are sure you don't need. While you may have felt the same way about taking statistics, by university most of us have learned to accept math requirements - better statistics than calculus. You may be aware that most other disciplines like physics or economics do not require a research course of their sophmores. So why is a course on how to do research required in psychology from the beginning? Here are three reasons:

1. The more complex something is, the greater its potential variability. In studying psychology, you are seeking to understand the most complex thing in the universe - the human brain. This makes psychological measurements more variable than measurements in other sciences. In psychology we need statistics to discern "facts" from fiction. While statistics lets you cope with the variability, interpreting statistical significance (or non-significance) requires understanding how the data were obtained. Statistics can "prove" anything only to those who do not understand how the research was done. So even if you never do any research in psychology, you need to understand enough about research to know what to believe and what to doubt.

2. The best way to learn science is by doing it. Chemistry students do labs using chemicals; biology students use simple organisms. If they make a mistake, the consequences are largely financial. To learn about human behaviour, psychology students do labs using people. If they make a mistake, not only have they wasted the time and respect of the participants, but also the consequences could be physical and emotional damage. That is why the next chapter is about ethics.

3. Other sciences use the latest technologies to push the limits of knowledge. To study itself, the human mind must go beyond its own limits. Psychologists literally "pull" the limits with them. For example, which disciple do you suppose made the first use of a manufactured digital computer? In 1963, the Technical Measurement Corporation introduced the CAT-100. It was used to study human brain waves - exploring the physiological basis of consciousness. That led to the tomography and NMRI technologies which are revolutionizing medicine today. The CAT-100 had been preceded by decades of psychologists pulling that limit manually by thousands of hours of paper and ruler measurements of oscilloscope records to Compute Average Transients. Though it may not always look like technology, psychology requires greater knowledge of "how to do it" than other sciences.

If you find these reasons daunting, perhaps you should consider a career in a simpler science like biology or nuclear physics.

Before going further, I want to point out that you do not have to study psychology to help people. There are many opportunities in the community including such programs as Big Brothers and Big Sisters, various crises centers, and religious organizations that need volunteers right now. Your participation need not be postponed for four to eight years while you pursue bachelor's, masters, and doctoral degrees. These service programs

2 - INTRO

provide a certain amount of guidance so that you would work with more than good intentions. Then why study psychology? There are several good reasons:

- 1) The knowledge gained in your studies will enable you to deal with more difficult human problems.
- 2) It will help qualify you to train and supervise volunteers.
- 3) It can enable you to make a career out of helping people.
- 4) Your studies may provide you with insights that help far more persons than you could ever reach yourself or reach indirectly by supervising others.

1.2 SOURCES OF KNOWLEDGE

Scientific research isn't the only source of knowledge. But have you really thought about how it differs from the other sources? Let's make some comparisons.

1.2.1 OTHER SOURCES

INNATE KNOWLEDGE

We take for granted that other animals are born with various repertoires of information in the form of instincts, yet many have long been believed that humans are above such behaviors. More recently developmental psychologists have found a host of human instincts which indicate we are born with a certain amount of knowledge. Newborns already have the reflexes for walking, comprehending distance, understanding of basic grammar, recognizing facial expression, etc. The latter may be a basis for altruism in everyone. However, well meant but naive attempts to help others can do more harm than good.

COMMON SENSE

We have acquired a great deal of knowledge informally simply by seeing and hearing about what our parents, peers do in everyday life. It is integrated into our thoughts and behavior often without our awareness of its basis. Most of this information is correct and useful. However, because we are not aware of its basis, it is easy to apply it to situations that are not appropriate. An eloquent example is provided in Carl Sandburg's poem Mending Fences:

" - There were it is we do not need a wall:
He is all pine and I am apple orchard.
My apple trees will never get across
And eat the cones under his pines, I tell him.
He only says, 'Good fences make good neighbors.' -"

AUTHORITY

Ever since some early ancestor used threats to get others to store nuts for the winter, certain individuals have been recognized as wiser than others. As their advice was heeded, they became authorities and leaders in the community. Unfortunately, mental ability is only by chance combined with physical prowess. Persons having the latter but lacking the former can impose their will on others and claim authority. Might often is not right. It takes advanced social organization to effectively separate these two types of authority.

An early scientific attempt to test authoritative sources of knowledge was made by King Croesus of Lydia in 550 BC. That Croesus was a genius himself is indicated by his accumulating a prodigious commercial fortune. With this wealth he thought to challenge the Persian Empire. The custom those days was to seek the advice of consultants, called oracles, before commencing any special endeavor. (Lacking diplomas from prestigious universities, oracles alluded to special relationships with certain deities to establish their authority.) According to Herodotus (Burn, 1974), Croesus simultaneously sent messengers bearing rich gifts to the seven greatest oracles of the Mediterranean world. On the hundredth day after departing, they were instructed to ask the oracle what Croesus is doing at noon that day. Only one was correct. The reply of the oracle at Delphi has survived to this day:

"I count the grains of sand on the beach and measure the sea;
I understand the speech of the dumb and hear the voiceless.
The smell has come to my sense of a hard-shelled tortoise
Boiling and bubbling with lamb's flesh in a bronze pot -"

To test the oracles, the king had decided to do something as incongruous as making turtle soup. Certain that he had identified a true authority, Croesus sent another messenger to Delphi bearing fabulous gifts to ask whether Croesus should attack Persia. The oracle replied that should Croesus attack Persia, a great empire would fall. Thus encouraged, Croesus attacked, was totally defeated, but survived to realize that the oracle was correct after all. His own empire had fallen.

Some authorities are clever enough to hedge their answers so they appear to be correct regardless of what happens - just ask a politician. Sometimes authorities are lucky. Scientists know the importance of replicating results before accepting them unequivocally. Some authorities are strong enough to enforce whatever they decide. It is easy to find leaders who have done the opposite of what they promised during elections. Eventually, psychological knowledge may lead to more effective social organizations where authority depends only on wisdom.

REASON

Early human discoveries from knapping flint to building sailing vessels indicate an implicit understanding of connections between cause and effect. However it was not until about twenty-four hundred years ago that discoveries about logic and nature were integrated by Aristotle and other philosophers to show beyond doubt that the world operated according universal principles - principles that mankind could understand. What Aristotle and subsequent scholars failed to recognize for some 1500 years was that even the most brilliant mind also requires knowledge for its deductions. For example, consider the following illustration of a plane that is dropping a parcel in mid flight. Several trajectories, labelled A, B, ..., depict how the parcel might fall. The ability to decide which trajectory is correct does not depend so much on reason as on whether you have certain knowledge about falling objects and how force vectors interact.

[illustration of airplane dropping a package]

The limitations of reason are not limited to applications of physics. Consider "catharsis" for example. In the 1890's, Joseph Breuer found that having patients talk about their problems underhypnosis helped them resolve those problems. Likening it to using a

4 - INTRO

catheter in medical treatment to drain off excess fluids, he called it the "cathartic method". When Sigmund Freud found that simply talking without hypnosis worked as well, he created a model of the mind to explain the treatment. Promoted by the authority of a famous psychologist, who argued that its benefits were a logical result of his theory, catharsis became accepted as the principle procedure of psychoanalysis. A recent study of 3,815 high school students reveals the popularity of the cathartic concept even today. Russell, Arms & Bibby (1995) found that a vast majority believed that participation and observation of aggressive activity reduced subsequent aggression. So much for the impact of decades of research which has repeatedly shown that exposure to aggression increases subsequent aggression! As you might expect, there is little clinical evidence that the cathartic method is effective in treating aggression (Warren & Kurlychek, 1981).¹

[cartoon of psychoanalyst with patient on couch]

Aristotle and Freud show that genius can accomplish a lot using pure reason. However, even genius can be led astray when reason is extended too far from knowledge. Though not common, genius comes for free. It has provided us with tools such as logic and mathematics that are useful for seeking knowledge. Knowledge, on the other hand, requires work.

1.2.2 SCIENCE: LEARNING BY MEASURING YOU DON'T HAVE TO BE A GENIUS

Perhaps you have never considered doing research in psychology because you assumed that one had to be genius. This is a popular myth from the 19th century; a myth understandably condoned by scientists themselves. The modern era has seen a tremendous increase in research accompanied by comparable growth in our collective understanding of the universe. This has not occurred because human intelligence has suddenly increased. Genius is probably no more common today than it was 20,000 years ago. A discovery 400 years ago together with the modern advent of universal education has made it possible for anyone with an inquiring mind and determination to do research.

Early discoveries in agriculture, medicine, and technology depended on combinations of chance events and genius. It took thousands of years to gain new knowledge because discoveries had to occur many times before they became sufficiently established in societies to endure. Gradually improved transportation and more social organization reduced the necessary number of rediscoveries, and knowledge grew faster. Yet as long as the only definitive way to obtain knowledge was by luck or by reason, the seeking of new knowledge was limited to geniuses and very lucky persons. Then, in 1590, Galileo published an account of what he did to find if heavy objects fall faster than light ones. His De Moto is the first formal recognition of systematic observation as a new way of obtaining knowledge.

The power of Galileo's scientific method depends on measurement and deliberately seeking explanations by testing various alternatives. It is important that these tests be done

¹ A landmark review by Bergin & Lambert (1971) found that differences in outcome from various psychotherapy methods were rare. "Clearly some people profit considerably from their therapy experience; but it is disappointing to find that too often persons are not helped or are even hurt by inept applications of the very treatments that are meant to benefit them." (Bergin & Lambert, 1971, p.180)

under controlled conditions so as to eliminate alternative explanations. Much of this course will concern how to control the conditions. It is also important to repeat the observations to make sure a particular observation is not the result of random variability. Because the results contradicted common sense, authority, and reason, Galileo's method had a momentous impact. Within a hundred years Isaac Newton used the method to make a sweeping series of discoveries in physics. Over the next two centuries others applied the scientific method to almost every aspect of human inquiry. The advances in physics, chemistry, biology, and technology far exceeded everything accomplished by our species over the previous 20,000 years.

Application of the scientific method to understanding human behavior took a bit longer. By the mid 1850's Gustave Fechner was developing some of the unique procedures needed to measure human perception, followed shortly thereafter by Ebbinghaus finding ways to measure human memory. Today the scientific method is used in psychology to study every aspect of human behavior. Examples include discovering certain how chemicals control our mood, the patterns of neural activity that evoke the purely subjective consciousness of color, the genetic and social principles that guide our selection of mates and friends.

The scientific method is responsible for the tremendous advances in human knowledge in the past three hundred years. Yet its direct success has overshadowed an indirect effect which has had as much impact. Before the nineteenth century, both genius and wealth were essential to do be able to do research. The scientific method provides a systematic way of seeking knowledge which does not require genius to be successful. Though they tend to be of above average intelligence, the great majority of scientists are not geniuses. The potential for profit and power that can be gained from new knowledge has led to social support systems to foster scientific research. Therefore, one no longer needs to be wealthy to do research either.

1.2.2.2 SCIENCE, SOCIETY AND YOU

The systematic way of seeking knowledge has proven valuable to both researchers and non-researchers. In everyday life, most tasks can be managed by responding intuitively without resorting to formal plans. However, as tasks get more complicated, efficiency can generally be improved by doing more planning, and proceeding in a systematic manner. The suitability of such an approach has been far from obvious. Warfare, for example, has probably had a greater effect on human lives than any other human venture. Throughout history, most battles were entered more on the basis of faith rather than understanding. By objective analyses, there was little doubt about the outcomes. Yet until Clausewitz "wrote the book" on rules of engagement in the early 1800's, successful tactics were seen as the inspired accomplishments of military genius. After Clausewitz, commanders began to systemically analyze information such as the number of soldiers, fire power, and position. The consequences to human history proved momentous. Similarly in the early 1800's, Eli Whitney realized the benefits of standardized parts in manufacturing. It took another century for Henry Ford to see the possibilities of assembly-line production. This systematic approach to making things has revolutionized our physical world as much as the most significant battles.

The early 1800's also brought a major shift in how society was organized. Through most of human history, social organization was synonymous with authoritarian organization. Powerful and wise individuals used their authority to found clans and empires. These persons then used their authority to select their successor and leadership became hereditary. However, strength and wisdom are not inherited characteristics, and the results

6 - INTRO

eventually proved disastrous for the society. The American revolution led the way to a social organization that was founded on critical observation and quantitative measurement of government - the democratic vote. Today we find scores of popular books that extol the advantages systematic approaches to almost every human endeavor from nutrition to starting a business.

A graduate colleague whose dissertation was in brain research found himself working in government due to lack of academic openings. Nevertheless, he proved very successful and rose rapidly in rank to become assistant deputy minister. When I asked him about how he did so well in a field so different from his studies, he replied as follows: "I just used the skills I had as a researcher to dealing with government problems. I systematically identified the problem, compiled information from various sources, and then applied this information to solving the problem. In doing so I made clear exactly how the information applied and what questions it could not answer" This course will offer little opportunity to actually help others or solve problems of society, but learning to use the scientific method of acquiring knowledge will enable you to do so more effectively afterwards.

1.3 WRITING RESEARCH REPORTS

Good scientists are often not good writers. Yet research is of little use to society if it is not reported. The scientific approach to gaining knowledge has also led to a systematic way of reporting what was done. The format that has arisen is so simple "even scientists can do it." You don't need to be either a genius or accomplished author to write good scientific reports. You just need to follow a systematic approach that has four basic steps:

INTRODUCTION - This acquaints readers with the question being asked and brings them up to date on what is already known.

METHOD - A clear description of what was done and how the measurements were made.

RESULTS - A presentation of the results unobscured and unbiased by any inferences about what the results may imply.

DISCUSSION - What the results tell about the question asked, and how the results compare with what was found previously.

Looking over these steps, you may notice that all four are concerned with maintaining clear distinctions:

- a) distinguishing what others have done from what you did
- b) distinguishing what you found from how you interpret what you found

Thus, no matter how muddled the writing may be, at least the reader can keep these basic matters straight.

1.3.1 APA FORMAT

Various organizations of researchers have adopted further particular rules for their reports - rules that deal with details such as formats for tables and graphs, how references should be acknowledged, etc. Most journals that report psychological research require you to use a particular writing style called the APA format - named after the American Psychological Association which developed the format for their journals. Many journals for different disciplines have adopted this or a similar format because it has proven to be particularly effective and efficient.

Maybe that's because psychologists needed an especially good system to make themselves understood. - Editor

It is more likely due to psychologists being more aware of the cognitive problems that arise from poor communication and recognizing effective means of preventing such problems. - Nilsson

This text is intended to be used in a course accompanied by a series of research projects to be done by you. These projects are designed to provide hands-on experience with six major ways of doing research in psychology. You will be using the APA format to report these projects. In the process you will learn not only how to do research but also how to write reports. More about this later.

Swell, but what's in it for those who do not see research as part of their career goals? Not only will you be trained in a systematic way to go about answering any type of question, but also you will have learned professional report writing skills that are rather different from the type of writing done for literary purposes. Many of our graduates go on to successful careers in government and business on the basis of their skill at handling data learned in the statistics course and skill at writing clear reports learned in this course. They rarely use the formal knowledge gained in all their other courses.

1.3.2 CHECKLIST FOR RESEARCH REPORTS

Before you read further, get a piece of tape and fold it over the edge of this page with a bit sticking out. This will make it handy when preparing your lab reports. The format of the checklist illustrates the expected format for reports submitted to journals published by the American Psychological Association, by the Canadian Psychological Association, by other professional organizations such as the Cognitive Science Society, Human Factors and Ergonomics Society, and by many commercial publishers of journals for psychology and other sciences. If you decide to major in psychology, I recommend buying the American Psychological Association's Publication Manual, which is available in most university bookstores.

The following form is used by Mr. Grey to evaluate your research reports. Weekly seminars will explain each practicum and what is expected in the reports. The " - - - " space for each section indicates the grade or points you've earned for that section.

8 - INTRO

title page

REPORT FORMAT 1

Running head: REPORT FORMAT

How Report Format Affects Understanding of Research Results

Your Name

Institutional Affiliation

The running head "REPORT FORMAT", placed in the upper left corner preceding the page number. It helps identify your paper. It should consist of the first 2-4 major words in the "title" and take less than 51 spaces. The title should be 10-12 words that summarize the main purpose or idea of the report. Avoid unneeded words such as "A Study of --" . . .

abstract page

REPORT FORMAT 2

Abstract

Summary of purpose, what was done, what was found. It should be specific, concise, non-evaluative, stand on its own, and be less than 960 characters (about 120 words). (For practicum reports <75 words)

new page

REPORT FORMAT 3

(the **introduction** starts on new page but has no heading)

Explain the purpose in terms of what questions was asked; why it is of interest; what has been done previously of immediate relevance; and how you went about answering the questions.

(For the practicum reports, explicitly identify the independent and dependent variables and their operational definitions as needed.)

(The subsequent sections do not start new pages)

Method

Subjects

Who was studied, how were they recruited, and what instructions were they given?

Apparatus and/or Materials

Brief description of what was used to do the research.

(possible figure page for apparatus)

REPORT FORMAT 5

Figure 1. Figure to illustrate apparatus or materials.
See the Checklist for Figures below for further details. ()

Procedure

Where, when, and how the research was done. Describe steps taken to avoid possible confounding variables.

Results

Describe what were the overall results and any statistical analysis. If you use a table or a graph to present the results, you must still describe in words the main features that the reader should look at. Calculations are shown in an Appendix. (Are the statistics appropriate and correctly done?)

(possible table page)

REPORT FORMAT 7

For student reports and thesis - or when the manuscript is the finished format, place tables and figures on the page following their mention in the report. Papers submitted to publishers will be typeset if accepted, and these should have tables and figures at the end of the manuscript, with figure captions on a separate page from the figure itself.

Table #.

Title of Table Should Summarize the Information Presented.

For further details, see section 1.3.3 Checklist for Tables below. ()

(possible figure page)

REPORT FORMAT 8

Figure 1. Graph to summarize the results.
See 1.3.4 Checklist for Figures below. ()

Discussion

What does your statistical summary and analysis imply? How does this compare with previous findings in similar research? What do you conclude from the results?

(start on a new page)

1st & 2nd TITLE WORDS 11

References

These establish the authoritative source for statements made in your report. They also acknowledge the published contributions of others. Editors are picky about how references are in proper format - therefore instructors must be picky too. See the "How to Cite References" section below.

(Overall grammar, spelling, use of past tense, clarity of expression.)

1.3.3 CHECKLIST FOR TABLES

Whether you are preparing a table for a publication or a test, remember the following publication requirements (American Psychological Association, 1993). With practice you will find that like good grammar, tables just don't look "right" unless done properly.

- Table 1.3.3. - Centered at top.
 - Written on separate line.
- Title - Centered
 - Every Major Word Capitalized.
 - Underlined.
 - Appropriately describes contents
 - Brief (Avoid repeating what's in Headers and Stubs).
- Table Borders - Horizontal line below title and at bottom of table.

HINT: Without getting into fancy stuff, you can make lines for tables with a word processor by using the _____ (underline) or ----- (dashes). Otherwise draw them using a ruler.

- Headers - Appropriate labels for each column.
- Stubs - Appropriate labels for each row.
- Header Lines - Horizontal line separates headers from numbers
 - Horizontal lines group and separate layers of headers
- Data Summarized by Means
- Variability Summarized by Standard Error

- Organization
 - appropriate to purpose
 - Are data for individual subjects appropriate?
 - data organized in ascending or descending order
 - long columns broken up by spacing

- Neatness
 - lines drawn with a ruler
 - Calculations don't belong in presentation tables.

- Note.
 - placed below the bottom line
 - explains, qualifies or provides further information
 - explains abbreviations

Table 1.3.3

How a Few Horizontal Lines Help Organize the Results Presented in a Table:
Effects of Caffeine on Time to Complete the Maze

PLACEBO	CAFFEINE DOSAGE	
	0 mg	150 mg
23.6	23.2	20.5
23.4	22.1	20.3
22.8	20.3	20.2
22.6	20.0	17.4
22.4	18.8	16.0
20.8	18.6	15.5
20.5	17.6	14.4
18.9	16.9	13.4
17.9	16.8	13.3
17.6	13.9	13.2
MEAN: 21.1	18.8	16.4
SE: 0.7	0.9	0.9

Note. Time is in seconds. "SE" is standard error.

1.3.4 CHECKLIST FOR FIGURES

Whether you are preparing a graph for a publication or during your next test, keeping the following publication requirements (American Psychological Association, 1993) in mind will help you do a good job. With a bit of practice you will find that figures just like grammar just don't look "right" unless done properly - then you will no longer need this checklist.

- Title
 - "Figure 1.3.4." - Placed below the Figure.
 - Underlined

12 - INTRO

- Starting on the same line, an appropriate description of figure contents.
 - Only the first word is capitalized.
 - May have additional explanatory sentences.
- Correct Type - Usually a bar graph when X-axis has "categories", an X-Y line graph when X-axis has a numerical or quantitative scale, but clarity is the ultimate criterion.
- Axes - X-axis used for the independent variable
- Y-axis used for the dependent variable
- descriptive label for each axis
- units of measurement indicated in lower case within ()'s
- scale numbers or categories indicated
- appropriately spaced scale markings
- no data points on the axis
- Layout - Are X and Y scales appropriate to show the effects?
- Do X and Y scales produce a graph of squarish proportions?
- Is the space inside the graph well filled without being redundant?
- Error Bars - Vertical lines extending plus and minus one standard error above and below a plotted mean value.
- Organization - Is more than one set of line plots or bars appropriate to compare data from several variables?
- More than one set requires a legend to identify each set in terms of plot marks, lines or bar appearance.
- Is the selection of independent variable assigned to the X-axis and those assigned to separate plots appropriate?
- Can some other type of measurement provide insight? (A second Y-axis has the same basic requirements as the first Y-axis.)
- Neatness - For hand-made graphs, use a ruler to draw axes, line plots, and bars. (*Better yet: Learn to make graphs using a spreadsheet program.*)

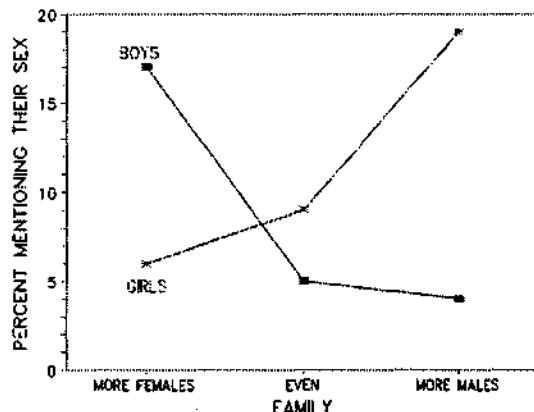


Figure 1.5.4-1. Example of a compound line graph: How often children mention their sex in self descriptions depends on their sex and the sex distribution in their family.

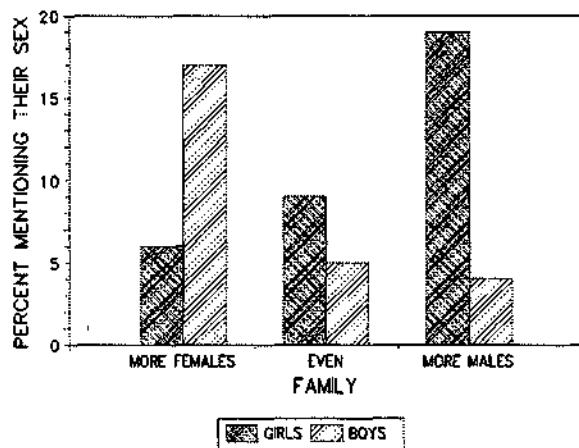


Figure 1.5.4-2. The same results as in the previous figure are shown here using a bar graph.

1.3.5 HOW TO CITE REFERENCES MENTIONING REFERENCES IN TEXT

References should be used to identify specific facts, concepts, or possibilities that are not your own or part of your general knowledge. They are also used to provide authoritative support for your own ideas. There are two basic ways to incorporate references into your writing: 1) "Subjective assessments of fatigue have been found to reflect how long people are willing to continue driving task than various performance, perceptual, and physiological measurements (Nilsson, Nelson & Carlson 1997)." or 2) "Nilsson, Nelson & Carlson (1997) found that subjective assessments of fatigue have been found to better indicators of how long people are willing to continue driving task than were various performance, perceptual, and physiological measurements. The latter format places somewhat more emphasis on the author, which could be appropriate when comparing that authors results to results found by someone else.²

When a reference includes more than two authors, all authors' last names should be included the first time it is cited. Thereafter, only the first author followed by "et al." and the year may be used provided this does not produce ambiguity with respect to other cited papers by that author: "Interestingly, how long Nilsson, et al's (1997) subjects continued to drive did not alter the degree of fatigue they reported prior to stopping." When a

² Certain journals such as *Science*, *Nature*, some medical journals and other science journals do not use the American Psychological Association's format for references. The most frequent alternative is the use of one or more numbers in superscript brackets at the end of a sentence. This is accompanied by a numbered list of references, at the end. Over the years, many journals have been changed their format to resemble APA's. Notwithstanding these alternatives, when preparing a report for a psychology course, always use APA format.

14 - INTRO

reference continues to be discussed, the year may be omitted - provided this does not lead to ambiguity: "Nilsson, et al contend that people are similar in how much fatigue they will endure; what differs is the rate at which they become fatigued."

REFERENCE FORMATS

Like the rules for making tables and graphs, some of details expected in the American Psychological Association's required format may seem picky. However, for the sake of clarity and efficiency, some format had to be agreed upon, so learn the APA format, accept it, and get on with what's really important. Three sorts of references will handle most of your needs:

Journal Papers

1. author(s)' last name(s) followed by comma and initials
2. the year of publication in "(" ")"
3. title of the paper - first letter of first word only capitalized; ended by a period
4. title of the journal - all words fully spelled; underlined; ended by a comma
5. volume of the journal - underlined, followed by a comma
6. first page, "-", last page of paper; ended by a period

eg. "Nilsson, T.H., Nelson, T.M. & Carlson, C. (1997) Development of fatigue symptoms during simulated driving. Accident Analysis & Prevention, 29, 479-488."

Chapters in Books

1. author(s)' last name(s) followed by comma and initials
2. the year of publication in "(" ")"
3. title of the chapter - first letter of first word only capitalized; ended by a period
4. "In" [Initials followed by last name(s) of the editor(s)] followed by "(Ed(s)."
5. title of book - first letter of each major word capitalized; underlined
6. "(pp. first page - last page of chapter)" - followed by a period
7. publisher's location - a major city or else also cite state or province; followed by a colon
8. name of publisher - end with period

eg. "Nilsson, T.H. & Connolly, G.K. (1997) Chromatic Contribution to Shape Perception Revealed in a Non-temporal Detection Task Using Distance. In C. Dickinson, I. Murphy, & D. Carden (Eds.), Colour Vision Research: Selected Proceedings of the International Conference, John Dalton's Colour Vision Legacy, (pp. 197-206). Bristol, PA: Taylor & Francis.

Books

1. author(s)' last name(s) followed by comma and initials
2. the year of publication in "(" ")"
3. title of book - first letter of each major word capitalized; underlined
- (4-optional. "(pp. first page - last page of specific material)" - followed by a period
5. publisher's location - a major city or else also cite state or province; followed by a colon
6. name of publisher - end with period

eg. "Nilsson, T.H. (1999) Statistics: Number Work for Psychology. Charlottetown: University of Prince Edward Island Printing.

If you cite a reference, you are honor bound to have read the original work - not a description of that work by another author, nor just an abstract. When it is impossible for you to read the original, there are several alternatives: You can cite the source by following the reference with the statement "cited by - " and then give the full reference for where your read about the paper. You could also simply cite as a regular reference the source that you did read and explain in the text, "According to (this author), (the other author) found that - ." You can cite an abstract by providing the reference in the usual manner followed by a bracket, the word "abstract", the PsychLit access number, identified by "AN #####", close brackets.

1.4 SOME ESSENTIAL DEFINITIONS

Definitions tend to become literal substitutes for understanding. I try to avoid them, but it just is not practical to tell you about research without introducing some. Otherwise I'd frequently have to use a whole clauses when a word would do - and you'd have more to read. Then also, if I "wrote around" the special vocabulary, you would find yourself at a disadvantage in discussions about research with others. So you will just have to learn some. Here they are:

VARIABLES - anything that changes (i.e. varies) while research is done. Most of the essential definitions involve different variables. Variables can include: 1) the environmental conditions in which the research takes place (eg. temperature, type of room, time of day); 2) stimuli presented to the subjects (color of a flash of light, emotional expression of person offering advice); 3) changes in the subjects (fatigue, prior knowledge of stimuli).

INDEPENDENT VARIABLE - something a researcher changes to see if it affects what is being measured.

The researcher may produce that change by directly **altering** it as, for example, by increasing the brightness of a light, adding caffeine to a subject's beverage, or reducing the apparent social status of a person offering advice. Alternatively, the researcher may obtain the change by **selecting** different subjects such as males and females, selecting different conditions such as the time of the day or where the research is done. Each different state of the independent variable is called a **level**. An independent variable must have at least two levels, otherwise it can not be varied. For example, if room lighting is the independent variable, the levels could be different intensities, or different colors of the light. An independent variable can have many levels. A research project may have more than one independent variable.

DEPENDENT VARIABLE - the particular behavior a researcher measures to learn the subject's present state (descriptive research), or to see how the subject is affected by the independent variable (experimental research).

"Behavior" is taken in the broadest context to include every thing from the subject's degree of curiosity to the subject's production of widgets. (How can one measure curiosity? See the next definition.) The dependent variable is measured

16 - INTRO

at each level of the independent variable to find out if that level has produced a change. Research projects may have more than one dependent variable.

The significance of the following variables requires some explanation about experiments. When an independent variable is changed to find out if it changes the dependent variable, one would like to believe that independent variable caused the measured change. However, it is also possible that something else produced the change. The basic way to ensure that a measured change is due to the independent variable is to keep everything else constant. Since "nothing comes of nothing", the only thing that could cause a change in the dependent variable is the one thing that was changed - namely, the independent variable.

CONTROLLED VARIABLES - *variables that the researcher either holds constant or can balance so that they do not distort the results of the measurements.*

variables held constant - *potential variables that are held constant so that they can not influence your measurements.*

Ideally these would include everything else in the universe except your independent variable. Consider the following types of constants:

- 1) *environmental* : constant temperature, lighting conditions, and noise.
- 2) *apparatus* : equipment is calibrated and functioning properly, test materials remain neat and legible.
- 3) *subjects' state* : ensuring that your subjects do not become less attentive or learn how to behave differently during participation.

BALANCED VARIABLES - *have had each of their levels spread evenly across the levels of the independent variable so that this uncontrolled variable can not bias the measurements at any particular level of the independent variable.*

It is not within our power to hold constant everything that might affect our measurements when doing research. The alternative is to "balance" variables by doing the research in such a way that the each level of the independent variable is tested in each of the levels of those variables that could not be held constant.

To illustrate the use of balanced variables, we'll consider some research that is conducted to see whether increasing an independent variable from LOW to HIGH affects a certain dependent variable. For example, it could refer to whether caffeine affects how long you think you have been talking to someone. The LOW level of the independent variable would then correspond to the subjects having drunk decaffeinated coffee, the HIGH level would correspond to having drunk regular coffee. The dependent variable could be measured in terms of how a written estimate of the number of minutes that had elapsed. If one morning you started measuring all the subjects in the LOW condition, it might well be afternoon by the time you got to measure subjects in the HIGH condition. It is not unreasonable to suppose that a subject's response in a warm sunny room might differ somewhat from their estimate of the same

elapsed time in a darker cool room. Even psychologists can't hold the sun constant during

*And you've told me psychologists can do anything. - Editor
Well, almost everything. - Nilsson*

the research, but you can balance its effects on your research: Make half of the LOW measurements in the morning when the room is sunny and the other half in the afternoon when the room is darker and cooler. Similarly, make half of the HIGH measurements in the morning, the other half in the afternoon.

Balancing a variable does not eliminate its effect on the dependent variable, but it does prevent it from contributing to any change in the dependent variable as the independent variable is altered. In the above example, the sun equally affected measurements in both the LOW and the HIGH conditions. Therefore the sun could not have caused any difference in those measurements. Any difference that was found would have been attributed to the altered independent variable.

Since independent variables are manipulated by the researcher, should they not also be considered "controlled variables" too? - Editor

Yes, you're correct - strictly speaking. However, I think it helps to consider independent variables separately. - Nilsson

UNCONTROLLED VARIABLES - *any variables that have not been held constant nor balanced.*

A researcher would go crazy trying to either hold constant or balance every possible variable, so there will be a lot of uncontrolled variables in any real research. There are three types of uncontrolled variables - one you need not worry about, the second requires precautions, the third is anathema:

NONCONSEQUENTIAL VARIABLES - *uncontrolled variables that have little or no effect on the dependent variable.*

It does not matter that you have not held these constant nor balanced them. You hope that all uncontrolled variables are nonconsequential. One of the things you learn by experience doing research is an appreciation of which variables are likely of no consequence to your measurements. A broad educational background also helps.

RANDOM VARIABLES - definitely have an affect on your dependent variable, but fortunately their effect changes randomly over time.

Since the sum of random effects tend to cancel out and not bias the measurements made at any particular level, these variables are "naturally balanced". However, they will add variability to the results and should be avoided as much as possible. If you are uncertain whether an uncontrolled variable such as temperature, noise level, crowding, etc is random, monitor it

18 - INTRO

and see if it happened to be stronger when you tested some levels and not others.

CONFOUNDING VARIABLES - *uncontrolled variables that systematically affect the dependent variable.*

Consequently you can not tell whether a change in the dependent variable was due to the independent variable or the confounding variable. You do not want confounding variables in your research. (If they can not be held constant or balanced, you can try repeating the study in the same manner while not changing the independent variable. This allows you to measure the effect of the confounding variable alone. The confounding effect can be subtracted from the original measurements to reveal the true effect of independent variable. However, this is getting way ahead of ourselves.)

OPERATIONAL DEFINITIONS - *descriptions of a variable or condition in terms that can be directly observed and/or that tell exactly what was done.*

It helps to think of these as coming in three types:

OPERATIONAL DEFINITION OF AN INDEPENDENT VARIABLE - *whatever the researcher actually altered together with the rationale to justify its connection to the intended change of the independent variable.*

In psychology it is often impossible to directly alter certain independent variables such as "workload" or "friendliness". Say that you wanted to find out whether sensitivity to facial expression depended on the friendliness of the environment. You would have to describe how the environment was changed to make it more and less friendly. An "unfriendly" environment might be *operationally defined* as a corridor in a busy shopping mall; a "friendly" environment defined as a student lounge. As with any operational definition, others may question whether your definition is good. Learning to come up with good operational definitions is a matter of study, experience, and creativity that evolves over ones lifetime. For now, what is important is that you understand the principle of how operational definitions are used in research.

OPERATIONAL DEFINITION OF A DEPENDENT VARIABLE - *describes what the researcher actually measured together with the rationale to justify its connection to an intended change in an unobservable dependent variable.*

It is often impossible to directly measure many effects that concern psychological researchers. For example, how do you hold a meter-stick to love, fear, or curiosity to find out if these have been changed by an independent variable? The researcher must rely on ingenuity to find something that can be measured and which reflects that state of the subject which is the dependent variable of interest. For example, "sensitivity to facial expression" might be measured in terms of the illumination required to permit correct identification of facial expression in a photograph. "Curiosity" might be measured (i.e. operationally defined) in terms of how many unusual magazines a subject peruses while ostensibly waiting to participate in the research project.

OPERATIONAL DEFINITION FOR CONTROLLING A VARIABLE - describes the steps that were taken to render a variable effectively constant.

Some variables need to be held constant or balanced despite being intangible or not directly observable. For example, you might try to hold the subjects' attention constant by providing occasional rest stops and some reward if the subject gets 20 out of 25 questions correct. Alternatively, the level of attention could be balanced with respect to the particular questions being asked across two testing sessions by reversing the sequence of the questions during the second session. This is the most subtle of the three types of operational definitions. Yet a lot of the practical aspects of the various research designs in Section II are actually operational ways of rendering constant or balancing variables that might otherwise be confounding.

Many students find operational definitions one of the trickiest concepts to master in a research methods course - with good cause. Coming up with suitable operational definitions is one of the most challenging and fun aspects of research. It will take much of the course to get a sense of various ways of making the operational definitions involved in doing research.

1.5 ARCHIVES: RESEARCH WITHOUT MEASURING

The rest of this book is about obtaining new knowledge by making measurements. Yet it is possible to seek new psychological knowledge without making measurements yourself. For tens of thousands of years people have passed on epic events by narration and song. Such records are vulnerable to individual interpretations as illustrated by the parlor game of "rumor".³ The invention of drawing and writing provided more permanent records. Since they reflect human behavior and thought, such records can provide insights of psychological significance to events and ideas beyond our direct access.

USING ARCHIVES TO STUDY SOCIAL VALUES PRESENT AND PAST

The Vikings are often regarded as barbarians who cost humanity hundreds of years in rebuilding civilization after fall of the Roman Empire. However, they may also have provided a vital step in the transition from autocratic to democratic civilization. At the beginning of the second millennium, feuds between neighbors were common among European societies. The Vikings settled such disputes and other civil matters in a council, a "Thing", comprising all adult male males in the community. To guide deliberations when one neighbor harmed another, Things referred to a set of rules for compensation called a "blood tax". Accounts of this tax were recorded by Snorre Sturlason (1230), an Icelander whose writings survived the great book burnings during religious conflict on mainland Europe. Nelson (19XX) wondered how these values compared with modern values placed

³ "Rumor" is played as follows: Write down a short narrative such as, "Last summer, William, Robert, Janet, and Elizabeth organized a moon-light hike along the North Shore ---- and as a result they did not get back till sunrise." You ask someone to leave the room and tell them the story. The second person tells it to a third, and the story is passed on successively till the last person, who writes down what he or she has heard. The original and final versions often have amusing differences.

20 - INTRO

on people as indicated by compensation paid by insurance companies today. Since the payments were in completely different units, comparisons were made in terms of relative values based on the maximum payments made in each society. Figures 1.7-1 & -2 shows the compensation for males and females.

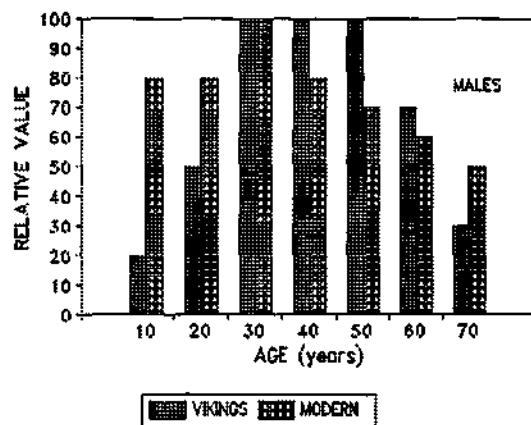


Figure 1.5-1. Compensation for death of males of various ages in Viking and modern societies.

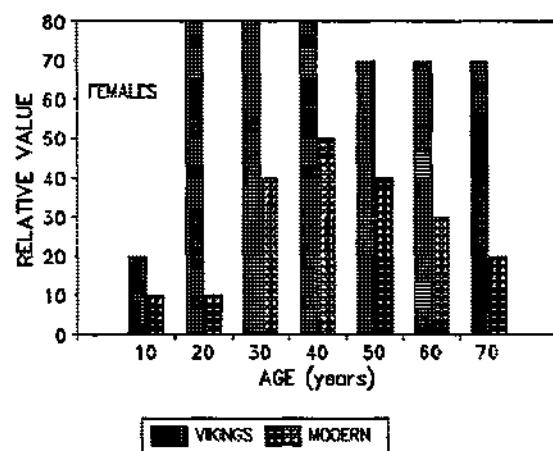


Figure 1.5-2. Compensation for death of females of various ages in Viking and modern societies.

Except for the higher compensation paid for boys in modern times, the values placed on males are roughly similar. A very different picture emerges for females. At all ages and especially for the young and old women, the Vikings placed a higher value on the women in their society than is done today. The difference is attributable to modern values of people being more strongly related to their earning power rather than worthiness of the person.

Some of the earliest writing are Mesopotamian ledgers of agricultural production and commerce. Rulers soon saw the usefulness of records for an aspect of human society as pervasive as life itself - taxes. Ever since, governments have accumulated enormous amounts of quantitative data about their people. In democratic societies, such data is usually available. Combined with some ingenuity and analysis, such data may lead to new insights about human behavior - i.e. psychological research.

Since new knowledge is useless unless communicated, the distribution of research results is a fundamental goal of science. Scientists may spend as much effort preparing reports of their work as they do seeking that new knowledge. As you have learned by now, scientific reporting demands a clear presentation of the research results as well the conclusions that are drawn by the researcher. Accordingly, books and journals contain massive amounts of data that have already been collected and that are available for anyone to use.

Where do you find the books and journals with all this archival data? The library, of course. For a summary of the stuff you can find in libraries check out Appendix A - About Libraries.

1.5.1 GOVERNMENTS AND OTHER ORGANIZATIONS

Government Sources

The largest single source of archival data are the population surveys conducted by many national governments. See the Information Desk at the library for a directory and access to government publications, including those from Statistics Canada. Records of what is said in legislative assemblies of government can provide a lot of information about social organization and the predominant thoughts and concerns of the legislators which reflect social attitudes generally. In Canada the records of many provincial legislatures and the House of Commons are made under contract by a private company called Hansard. In the United States legislatures such as the Congress make their own records.

Foundations

In order to promote research in particular areas of their concern, various private and public organizations gather and organize the results of published research, news, and even conduct surveys and sponsor research on their own. The vast majority of such organizations are concerned with social and medical problems. Therefore, they are particularly relevant to research in psychology. These organizations want researchers to ask them for such information in hopes that this will lead to more research on problems they want solved. However, there is a potential problem with such free information - some organizations may be selective in the data they make available.

The following list of organizations illustrates some sources:

Alzheimer's Society
Canadian Lung Association
Canadian Mental Health Association
Institute for Learning in Retirement
Women's Network

News Media

Newspapers, magazines, and electronic internet bulletins sometimes have articles and editorials about social issues that are supplemented by data. Such data from recent reports by government, foundations, or other agencies may be more up to date than comparable data in research journals which have gone through a review process and then taken their place in line for publication. (The delay in journal publications is often a year or more.)

One of the problems with these fast media sources of data is the lack of review by other professionals whose job is to ascertain the truthfulness of the data and appropriateness of its analysis. You depend solely on the reputation of the author and possible editor that the data are valid. A particular problem with electronic media is the current lack of permanent storage (archiving) of the information. Consequently, information you cite from the internet to support your conclusions may no longer be available when your report is presented.

1.5.2 USING DATA PUBLISHED BY OTHER RESEARCHERS

Sometimes in journal publications, the author(s) present data in tables with sufficient detail for others to analyze. Journal editors have reduced this practice over the years to save space. The understanding in any scientific publication is that authors will provide the data in a publication to anyone. Dissertations more typically include extensive data tables.

22 - INTRO

You can get a good idea of what is in a dissertation from Dissertation Abstracts, which often are more detailed than abstracts of journal papers. If the abstract looks promising you can purchase a copy of the full dissertation from Dissertation Publishers, Inc. (Check your library for access) It is perfectly legitimate to reanalyze such data, perhaps combining and comparing it with data from other publications or your own and write a paper based on your analysis - provided it is not simply a repetition of the original analysis. Of course you must acknowledge the source of data that is not your own. The Canadian Study of Health & Aging is an example of a large scale research project that has a huge anonymous data base on some 10,000 Canadians over the age of 65. The principal investigators encourage others to pursue further analyses of this data - as have some of my students.

1.5.3 ADVANTAGES AND PROBLEMS

The major advantage of using archival data is that you do not have to collect the data yourself. Regardless of the type of measurement or research design, it often costs a lot of money, time and effort to obtain your own data. (Of course it may also require substantial amounts of all three to find exactly the sort of data you need.) Even with good funding, you may not have access to enough persons to study - particularly if they should have special characteristics such as being geniuses, color blind, deep-sea divers, or unconvicted murderers. Certain types of research may require certain circumstances such as weightlessness, extreme poverty, or crises situations. Still other research may require equipment which is costly or usually required for other purposes such as magnetoencephalographs for brain imaging, environmental test chambers, or linear acceleration tracks. Yet, other researchers in special circumstances may have had access to such persons, situations, or equipment; and their data may provide answers to the question you seek to answer.

The major disadvantage of not collecting your own data is that it may not be exactly what you want measured or that you don't know exactly how the data were obtained.⁴ Regardless of occasional news reports, the following problems rarely apply to archival information obtained from scientific publications. The standards and safeguards followed by psychologists and other scientists ensure that you can rely on their data. On the rare occasions when there have been accusations of scientific dishonesty, they have usually centered on a question of whether the researcher had made an appropriate decision about which data to include. There is often no clear answer to such questions. The researcher knows more about what happened when the measurements were taken than can be explained in the report. Yet any decision to exclude certain measurements will always be suspect when it favors the conclusion. That incidents of scientific cheating are sensational attests to their rarity. The following problems apply primarily to other sources of data such as newspapers, magazines, foundations, and some government sources.

SELECTIVE DEPOSIT

"All the news that's fit to print" was the motto of a famous, national news paper. The motto illustrates two problems with what information gets included or "deposited" in archives: 1) It is no longer possible to print "all the news". In the fast paced world of news

⁴ Another MAJOR disadvantage is that it lets others have all the fun.

Coming from you, that figures. - Editor

reporting, decisions on what to include or exclude is often arbitrary by necessity. Yet humans rarely make decisions that are truly arbitrary. As a student of psychology, you are probably more aware than most persons that various human emotions and attitudes influence such decisions. 2) What is considered "fit" may vary greatly depending on social circumstances. Who decides? Regardless of the reasons, you can be sure that some one has selected what data has been preserved in various archival sources and what data was never entered.

The House of Commons in Canada and the United States Congress illustrate two different approaches to archiving information. Congressional members are free to edit their own records before they become permanent. Members of Parliament may not. As you might expect, Congressional members sometimes alter their records to look much more complimentary to themselves. However, as revealed in the "fuddle-duddle" affair involving a reply by Prime Minister Trudeau that contained the "f word", even Hansard and your textbook are subject to a certain amount of selective deposit - to preserve dignity of course. These are poignant examples of selective survival and more alterations that threaten all archival records.

SELECTIVE SURVIVAL

The second oldest profession may well be the "advisor" who knows how to make the boss (leader, chief, king) look good. Depending on their format, archives require a certain amount of maintenance to remain functional. It's all done by accentuating the positive and eliminating the negative during such maintenance. From bureau chiefs to national leaders, the convenient "loss" of damaging records has long been recognized as a means of maintaining control. A researcher depending on archival information should look for record gaps and inconsistencies that may indicate removal of certain information.

BIASED RETRIEVAL

Researchers need also be aware of their own limitations in seeking archival information. Where the potential sources of data are not evident, a search plan should be determined and included in the report. Completion of such a plan helps ensure that you do not stop the search as soon as you have found information that supports your hypothesis. Being systematic and reporting how you obtained your data is a keystone to good research of any kind.

EXERCISES

(A good way to do the exercises in this book is to get together with one or more classmates for a study session.)

1. One way to deal with definitions is to define them briefly in your own words. Do so here or on an index card for the following terms:

INDEPENDENT VARIABLE _____

DEPENDENT VARIABLE _____

CONTROLLED VARIABLES _____

HELD CONSTANT _____

BALANCED _____

UNCONTROLLED VARIABLES _____

INCONSEQUENTIAL _____

RANDOM _____

CONFOUNDING _____

OPERATIONAL DEFINITIONS _____

OF INDEPENDENT VARIABLE _____

OF DEPENDENT VARIABLE _____

FOR CONTROLLING A VARIABLE _____

2. At the library find a current issue of a psychology journal (see Appendix A About Libraries for a guide to the journals) and pick a short research report that looks interesting.

- a) Find an example of each of the following type of variable:

INDEPENDENT VARIABLE _____

DEPENDENT VARIABLE _____

CONTROLLED VARIABLE _____

Was it HELD CONSTANT () or BALANCED ()?

AN UNCONTROLLED VARIABLE (These are usually not explicitly described. You probably have to make some inferences based on what is described.)

2-b) Look through the last paragraph of the Introduction and the Methods section for an OPERATIONAL DEFINITION:

Is it a definition of an INDEPENDENT , DEPENDENT OR CONTROLLED variable?

3. At the library find a current issue of a psychology journal (see Appendix A About Libraries for a guide to the journals) and pick a research report that looks interesting.

a) Use the Checklist for Research Reports in Section 1.5.2 to evaluate the text of the published report.

b) Use either the Checklist for Tables (1.5.3) or for Figures (1.5.4) to evaluate either a table or a graph in the report.

4. Assume that you want to do research that involves the following variables which you can not change, measure, or control or control directly. For various research projects, describe how you could make each one into an independent variable, a dependent variable, and how to hold it constant. Start each operational definition by imagining a particular research project.

"LOVE"

(eg. as an INDEPENDENT VARIABLE: You want to find out if people who are in love recognize the face of their beloved more quickly than the faces of famous people. To do this you need to test couples who are "in love" to various degrees such as "just friends" to recently engaged. You recruit couples at a dance. To assign couples to various levels of being in love, you ask them how many times they have been on a date during the past month. Couples who have dated once or twice are regarded as being "slightly" in love. Couples who have dated 3 to 5 times are "somewhat" in love. Couples who have dated more than 5 times are rated as being "strongly" in love. This definition is not perfect. For example, couples who fell in love at first sight would confuse the results. Assuming that this is a rare phenomenon, the definition is reasonable.)

(eg. as a DEPENDENT VARIABLE: Next consider how to measure love. Do couples feel more in love on sunny days? Observe couples in a park on sunny and overcast days. What could you look for as an indicator of being in love?)

(eg. as a VARIABLE HELD CONSTANT: To compare the effectiveness of three videos on attitudes about family planning, you want to divide couples into three groups of couples that equally in love. What is a criterion could you use to select couples for participation?)

"HAPPY"

"CREATIVE"

"RELIGIOUS"

"SOCIAL STATUS"

"MANUAL DEXTERITY"

"POLITICALLY LEFT (OR RIGHT) WING"

26 - INTRO

5. A frequent concern of students and faculty is whether adequate resources are being provided by society for education. The raising of tuition fees, decreased resources in libraries, non-replacement of obsolete equipment in labs, reduction of staff, and freezing of salaries are universities' main ways of responding to inadequate funding. Governments usually do not admit that their policies are inadequate and refer to funding issues as "restructuring" and "efficiency". Yet governments must file official statistics on their expenditures, and these are reported in documents you can find in most university and larger public libraries. With the help of a librarian you should be able to find such information from Statistics Canada, and provincial publications.

Find some data on government expenditures for advanced education in your province over the past ten years. Compare those figures with one or more of the following: the number of students, the number of teachers, the expenditures by other provinces, the expenditures for health care, the expenditures for industrial development, the salaries for members of the legislature.

CHAPTER 2

DOING RESEARCH ETHICALLY

There are several codes for ethical pursuit of research in psychology. They have been established by professional organizations of psychologists and by various agencies that grant funds for research. The Canadian Psychological Association's Code of Ethics has been recognized by other professional associations as exemplary in its completeness and extensive use of specific examples. Because much of it pertains to the practice of counselling and clinical psychology, I created an abridged version that focuses on the needs of psychology researchers. It is included in Appendix D. Also in Appendix D is a copy of the Human Factors and Ergonomics Society's Code of Ethics. Much of this code pertains to consulting and industrial work, but since it is only two pages, the entire code is provided. Article IV - Subject Precautions most directly pertains to research methods.

Unfortunately, in our increasingly legalistic society there is growing pressure for rules to cover every contingency. Such codes would be lengthy beyond belief - and no more effective. The Human Factors code illustrates that length is not necessary to provide guidelines for professionals. That's because a career in science requires a doctorate and usually a university appointment. Psychologists and other scientists generally have passed a lengthier, more demanding training and selection procedure than that of any other profession, including medicine, law, engineering, business administration. The evidence? When a scientist is suspected of unethical behavior, it is news. By comparison society tacitly accepts that companies take advantage of their employees, that corporations proclaim misleading product information, that political leaders do not fulfil the promises which got them elected.

In this chapter I'll discuss general principles that underlie psychologists' obligations to humanity. Since these are the bases for the more specific concerns, the rest of the chapter is more of an outline. To discuss every issue in detail would take half the book. Doing so would probably not be effective at this stage in your studies, because research experience is needed to make these issues more than abstract concepts. For now, read the codes to get a general idea of their contents. Note how the basic ethical principle of doing unto others as you would like them to do unto you, leads to guidelines for various aspects of research. Having seen how that works, you ought to be able to handle questions pertaining to ethical issues and to avoid ethical errors in your research designs for this course. If in doubt, look it up or ask a professor.

2.1 OBLIGATIONS TO HUMANITY

2.1.1 CONTRIBUTE TO SOCIETY

More than atomic energy or industrialization, the most pervasive and enduring technical development of the last quarter of the second millennium was harnessing electricity. How it was discovered can inspire today's students of psychology. To appreciate the significance of how it came about, picture the world in the mid 1700's. Muscle power from man or beast was the surest form of energy that could be applied to meet human needs. Wind and water power were widely exploited. Chemical power in the form of combustion was just being harnessed thanks to Watt's discovery of negative feedback control. Yet even efficient use of combustion for cooking had to wait till the next century when Thompson's insights into the nature of heat would lead to invention of the stove. Where in that world could one stumble across a totally new form of energy operating on principles beyond imagining? The answer lay in something that was vastly more complex than any human creation up till then - a frog!

A frog's nervous system is more sophisticated than the most modern, automated factory. There it was in the 1700's, readily available for Galvani to observe the twitch of frog's leg, and from that to infer the action of electricity. Now ask yourself the following question: Where in today's world might you find the equivalent of Galvani's frog? Where is that special something which might enable you to discover things we can not yet imagine? Psychology is a study of the most complex entity in the known universe - the human brain. Don't let marvellous accomplishments such as mega computers that can mimic a chess player mislead you to think that we are close to understanding how the brain works. Today's sophisticated electrode implant techniques and brain scanners have barely scratched the surface. What remains to be learned about memory, emotion, creativity, etc. can hardly be guessed. Probably they involve new ways of organizing information using principles which require a leap of insight comparable to going from the twitch of a frog's leg to the discovery of electricity.

While we know more about the world and ourselves than ever before, there is ample evidence that inadequate knowledge can be dangerous. Human history is a legacy of ruins in deserts. Time and again vigorous civilizations have developed, thrived for a few hundred years, and then vanished. Their lands were not deserts when the cities were built. What happened? For insight, we need not look far. Our environment is being destroyed today at a greater rate than ever before in human history. In third-world countries, this may be attributed to ignorance. However, it is just as prevalent in the United States and Canada. For the real answer we must consider human motivation as well as knowledge. Greed, reproduction, intolerance - these are the forces that destroy civilizations. What area of study should we encourage students to pursue to gain the necessary understanding of such forces?

The past hundred years have been accompanied by a revolution in medicine that shows every indication of surpassing itself as physics, chemistry and biology are further integrated into the quest for health. However, by far the largest use of medicine in our society is not treating diseases or injuries. It treats disorders of the mind. Depression, boredom, fatigue, and stress are the price being paid by millions for their participation in modern civilization. Drugs may alleviate the symptoms, but rarely cure. The only real answer, prevention, remains to be learned by psychologists through study of the human mind.

As a university graduate, you can reasonably look forward to a standard of living enjoyed by only the most fortunate 10% of humanity. You'll earn this achievement by years of study and economic sacrifice. For most of us these prove not to be as onerous in retrospect as they may appear to be now. If you continue your studies for a profession in psychology, there is another sort of debt to be paid - an obligation to contribute to the knowledge from which you profit. That is the purpose of theses research as much as their evaluation of your suitability. Many find this experience of discovering new knowledge so rewarding that they make a career of it.

2.1.2 PROTECT HUMANITY

In the age of materialism, the hallmarks of civilization was often identified in terms of physical achievements such as the size of buildings or speed of rockets. As psychologists, we can be heartened by the dawn of what seems to be an "information" age. Information, after all, is solely a product of the human mind. To seek new knowledge about how to use information, even the most materialistic societies have become convinced of the value of studying the most complex information processor in the known universe - the human mind.

In the age of materialism, it mattered not how the buildings and rockets were achieved. The accompanying torment of thousands diminished not the acclaim for such accomplishments. What ancient empires achieved by slavery, modern empires can achieve by slavery too.

Solzhenitsyn's (197) account of science-labor prisons in the former Soviet Union. Yet there are subtler forms of slavery. Your own community has threats to humanity that all students of psychology need to recognize. It is widely recognized that our civilization could be destroyed by unwise use of the knowledge obtained by physicists, chemists, and, most recently, biologists. Surely psychology posses no such risks - or does it?

In the *Commentarii* from Julius Caesar's own hand, we can learn how highly conditioned Roman soldiers supported by superb logistics conquered the less organized "barbarians" of that day. Today those methods are called behavior modification and organizational psychology. We can only speculate about the Celt's perspective on this "civilizing" influence. Gedge's (1978) portrayal of the Roman conquest of Britain suggests the dismay of an individualistic society being defeated by an evil empire using incomprehensible forces. Eventually the Celts learned the power of human organization too. Centuries later their English, French and German descendants would "civilize" the rest of the world.

Skilful use of psychological principles to enslave whole nations is not just ancient history or fiction. The Nazi's successfully used such tactics, fifty years prior to the setting of Orwell's (196) novel *1984*. Solzhenitsyn (197) describes a massive psychological research program underway when he was imprisoned the gulags. It could have proven more disastrous for humanity than any product of the physics or biology laboratories:

In the 1970's, leaders of the Soviet Union became aware of an unanticipated dilemma in their drive towards communism. Their efforts to control individual thought and expression, were proving a barrier to the creativity needed to compete with open, Western societies. While lavish support of geniuses such as Sakarov and Korolev had produced spectacular successes in physics and the conquest of space, these methods were not reliable. It was impossible to predict where new discoveries might be needed. Increasingly, Soviet science was falling behind the West. The need for better communication was evident, but how could this be achieved without also allowing new political ideas to arise? Censoring more open communications would stagger a social surveillance system already over-burdened and inefficient.

Encouraged by early successes in speech synthesis, the Soviet leaders thought they saw an answer - use computers to monitor all communications. That strategy had two requirements the Soviets then lacked: 1) enough big computers to do the job, 2) computerized speech recognition. Computers they could buy from the West, but even the West had not solved the speech recognition problem. Solynitzen (197x) tells of a massive speech recognition project in the gulags. While speech synthesis is mostly an engineering task, speech recognition was a cognitive task that proved to be far more difficult. Twenty years later, prime minister Gorbachev realized the impossibility of maintaining absolute control of a society without ruining its ability to compete in the modern era. The Soviet empire collapsed, and the great experiment of communism was finally recognized as a failure. The outcome might have been different. What if some psychology genius had solved the speech recognition problem before the collapse?

We can hope that solutions to critical problems in the information era will continue to require such freedom of thought and communication that no society can remain competitive if it does not treat its citizens with dignity. Yet we must be aware that this is wishful thinking. The consequences of complacency in these matters could doom humanity. Their training makes psychologists the main defense for recognizing social and physiological means of manipulating society. As a psychologist you should realize that your research has the potential for both greater benefit and greater danger than any other science.

2.2 OBLIGATIONS TO YOUR SUBJECTS

Had not the last quarter of the second millennium also brought the "discovery" of the intrinsic value of the individual, the industrial revolution combined with Roman mass social organization might have proven disastrous for humanity.¹ The breakthrough was led by Immanuel Kant (1724-1804). At that time the empirical revolution begun by Galileo was in danger of formal collapse as its subsequent philosophical gurus, Locke, Berkeley, and Hume found themselves at a dead end to justifying this new form of knowledge. Kant set out to rescue empiricism from the shadow of Aristotle's logic. In the process, he discovered the logical basis of ethics - one that did not depend on religious assumptions. Kant shows that the consequence of consistency in nature leads to the following proposition, "Act only on that maxim whereby thou can will that it should become a universal law." (Greene, 1957, p. 302) From this follows the familiar law Jesus proclaimed in the Sermon on the Mount, "whatsoever you would that men do unto you, do ye even so to them:", (Matthew, chapter VII, verse 12). Either statement adeptly summarizes the rest of this chapter.

Kant's development of that theme is particularly relevant to our present discussion. "So act as to treat humanity, whether in thine own person or in that of any other, in every case as an end withal, never as means only." (Green, p. 309) In the first section of this chapter, we looked at some noble goals that call for more research in psychology. Yet we must never let such goals can over-ride our immediate obligations to those who make the research possible. The demise of communism remains an enduring example of how the noblest of goals become meaningless if one betrays the basic human obligations of respect and kindness to others.

The following outline is a list of don'ts and do's to help you meet your obligations to the subjects who make your research possible. Most of them follow directly from the religious and moral values of our society, so they will come as no surprise - it just may not have occurred to you to think about them in this context before. In Appendix C you will find a more formal statement of many of these guidelines as presented in the Canadian Psychological Association's Code of Ethics.

PROTECT SUBJECTS FROM PHYSICAL HARM

1. Legally Responsible for Health and Safety of Subjects in Your Care
2. Obligation to Know What May be Harmful
3. Test Methods on Yourself and Colleagues Before Using Subjects

PROTECT SUBJECTS FROM PSYCHOLOGICAL HARM

1. Do Not Frighten, Demean, Embarrass, or Confuse
2. Do Not Impose Guilt or Question Personal Values
3. Do Not Attempt to Alter Personality even Temporarily

(An exception may be made in the use of role playing procedure. In these the subject may be asked to act "as if" they had a certain personality characteristic.)

ENSURE THAT PARTICIPATION IS VOLUNTARY

1. Identify Yourself and Your Professional Affiliation.
2. Provide as Much Information as Necessary for Informed Consent

¹ Archeological evidence of affliction and survival of stone age persons indicates that compassion, social care, and ethical behavior are as old as humanity itself.

3. Special Requirements for Children and Cognitively Impaired:
Consent of parents, guardians or principal caregiver.
4. Avoid Coercion
 - a. be aware of the vulnerability of students and subordinates
 - b. be responsible to public respect and credulity for science
5. A Subject Must be Permitted to Leave at Any Time.

PROTECT THE SUBJECT'S PRIVACY

1. Keep Personal Data Records Anonymous Unless Identification Needed
2. Use Codes to Protect Personal Data When Identification is Necessary

MINIMIZE DISCOMFORT

1. Make the Subject as Comfortable as Possible
2. Reduce Boredom as Much as Possible
3. Keep Stress Minimal
4. Provide Rest Periods for Extended Tasks

MAKE PARTICIPATION A POSITIVE EXPERIENCE

1. Explain the Overall Purpose of the Research
2. Be Willing to Explain as Much About the Research as Possible
3. Explain the Research Procedure and the Subject's Role - BRIEFING
4. When Possible Inform the Subject about the Results - DEBRIEFING
5. Thank Subjects for Their Participation.
6. Acknowledge Subjects when Feasible and Privacy is Not an Issue

DON'T WASTE THE SUBJECT'S TIME

1. Research Objectives Should Warrant the Imposition on Others
2. Be Competent with Apparatus and Procedures Before Using Subjects

Much as the above guidelines are important in meeting your ethical responsibilities, additional benefit (The Above are Also Important for Obtaining Valid and Reliable Data)

(As is the case with subject comfort, making research participation a positive experience helps keep your subjects stay alert and motivated to do their best. As steps to ensure good data, caring for your subjects also is part of your goal to contribute new information to society.)

2.3 WHEN DECEPTION IS NECESSARY

The use of deception in psychological research is NOT unethical. Nearly half of social and personality research has used deception (Menges, 1973). Valuable insights into human thought and behavior could not be sought otherwise. Deception is necessary because we can not willingly suspend our cognitive and judgemental processes in order to produce totally objective data. With that understanding, the following guidelines are imperative.

AVOID THE POSSIBILITY OF PSYCHOLOGICAL HARM

1. Temporary harm from deception is not acceptable.
2. Avoidance of harm must NOT depend on debriefing.
 - 1) in case subject leaves before experiment is completed
 - 2) in case debriefing is not successful

DECEPTION MUST BE EXPLAINED AFTER PARTICIPATION

This is an essential part of the "debriefing" (an discussion with a subjects when they have completed their participation in research) when deception was used.

1. Explain what the deception was and why it was necessary.
2. Request subject's cooperation in maintaining confidentiality.

Research on psychological research has found that explaining a deception to subjects protects the effectiveness of a deception better than relying on the subject's gullibility.

2.4 RESPONSIBILITY FOR ANIMAL SUBJECTS

The Canadian Psychological Association, Medical Research Council, National Science & Engineering Research Council and other professional and granting associations have strict codes for surgical, drug, and other invasive methods; for conditioning techniques; and for animal care facilities.

LAW OF SENTIENCE:

Do not use a more sentient species when a less sentient species is adequate. (Sentience refers to the level of consciousness that characterized a species.) Various ways of determining sentience have been proposed including the ratio of brain to body weight. In terms of animals typically used for psychological research, increasing levels of would be approximately as follows: mice; pigeons; rabbits & rats; cats, goats & dogs; new world monkeys; old world monkeys; apes.

RESPECT FOR LIFE

1. Remember, animals used for research are usually killed afterwards.
2. It is your responsibility to ensure that benefits warrant the loss.

AVOID NEEDLESS PAIN

MINIMIZE DISCOMFORT

PROVIDE COMFORTABLE AND HEALTHY ACCOMMODATIONS

These must be appropriate for the size, social instincts, and characteristic behaviors of a species. (Some of our classrooms might not meet these standards.)

2.5 PROFESSIONAL RESPONSIBILITIES

2.5.1 PRESENT HONEST DATA AND CONCLUSIONS

Psychologists and other scientists have rarely presented dishonest data. In most of those cases, it was probably motivated by good intentions rather than for personal profit or fame. Here's why. Perhaps you did not realize that psychologists and most other scientists do not make a profit from the research papers they publish.² Indeed some journals expect authors

² If a paper proves to be tremendously important, it may get republished as part of collection of papers in a book. Since such books are usually relevant only to other researchers with similar interests, the market is rather small. Consequently any royalties after being divided amongst all the authors do not amount to much. There are two ways that psychologists sometimes

to pay a certain amount per page to help cover the costs of producing the journal. It is not unusual for scientists to pay such costs out of their own pocket when grant funds run short. Over the long run, successful publication of many research papers will be a basis for tenure and promotion - both of which do improve ones salary. Yet a single publication is unlikely to make a critical difference. Weighting against temptation to be dishonest for that reason is the greater risk of being discredited and losing their job were the dishonesty discovered. That leaves fame as the other bad motive, but it's a "no win" situation too. If a paper based on fraudulent data has enough impact to make the researcher famous, you can be certain that others will be quick to repeat the study. When those replications fail, the dishonesty will be exposed. About the only way researchers may "succeed" with dishonest data is if their results prove to be of no interest to anyone.

Students face the same consequences if they present dishonest data. The problem is that some students may not realize the consequences. While there may be some sympathy for the poor judgement of a novice, the competition for graduate school or jobs is stiff. Acceptance depends on both high grades and letters of support from professors. A professor can not write a letter supporting a student suspected of dishonesty without jeopardizing the chances of acceptance all future students that they recommend. Consequently a dishonest student is unlikely to have the opportunity to redeem her/him self. That is not harsh; it is necessity. University students are adults and responsible for their decisions.

As for the "good intentions", suppose a researcher believes their data could be valuable to society if only the results were a bit different - just enough to achieve statistical significance. Since most results depend on many measurements and since there is always the possibility of error, the researcher may come to suspect some of those measurements and not others. The problem arises if they then find that eliminating the suspected ones produces the expected results. However, matters are not usually this simple. Psychology is especially vulnerable because its data are more variable than those of other sciences. Subjects do make errors, misunderstand instructions, and occasionally continue participating without paying attention when they have lost interest. This is another reason why those obligations to your subjects are so important. For example, If subjects lose interest, consider whether you made them comfortable, explained the purpose sufficiently, or demonstrated by your behavior and the appearance of the testing situation the professional standards of the research. If some subjects did not understand what they were supposed to do, the instructions should be made clearer.

Here is what to do if you feel that something has gone sufficiently wrong to warrant omitting some of the data you have collected: 1. Explain the problem and ask the subject(s) if they would repeat their participation. This is often not possible due to the subjects' availability or when the original participation may affect how the subject performs when tested again. 2. Discard the data before it is analyzed. 3. Acknowledge that some data was omitted and explain why. For example, "Of the 24 subjects tested, the results from two were sufficiently different (mean = 14.8, sd = 7.4 & mean = 15.3, sd = 8.3) to indicate they had misunderstood

royalties after being divided amongst all the authors do not amount to much. There are two ways that psychologists sometimes get rich from publications based on their research: One is to write a book that captures the interest of the general public. The second is to develop a successful introductory psychology text, where there is a large market potential since the majority of university students take an introductory psychology course.

again using a method that has been improved on the basis of what you learned from the initial attempt. (This illustrates the value of doing practice runs before beginning actual data collection.) These alternatives are better than risking your reputation. If in doubt of how to proceed, seek the advice of a professional colleague. For a student this would be your supervisor or course instructor. Not only can another person view a difficulty more objectively, they also represent a public disclosure of what was done. Whether it's research, teaching, or counselling, honesty is more valuable to a psychologist than any special knowledge or skill. If the thought of being dishonest occurs, remember that motives are not verifiable. Whatever the motive, dishonest data lead to unwarranted conclusions that, in turn, can lead to wasted research by others or worse to unsuitable public actions. You will be held responsible. No psychologist or student of psychology in their right mind would deliberately risk the years of study and economic sacrifice for what can only be some temporary gain.

2.5.2 MAINTAIN OBJECTIVITY

The Problem

Sometimes scientists are portrayed as dispassionate and totally objective seekers of truth. This is rarely the reality. Scientists are driven by the fun of doing science, the pride of doing science well, the challenge of finding what no one else has found before, and the prestige that results from an important discovery. However, it is a reality that the best process for distinguishing truth from uncertainty has a statistical basis which can only prove what is different from chance, but can't prove the lack of an effect. Accordingly, only statistically significant results get acknowledged as contributions. It is also a reality that financial support for research depends on a scientist's demonstrated ability to contribute to knowledge. In such circumstances, the most dedicated scientist can not help but hope for results that are statistically significant.

Problems can arise because the same motives that encourage researchers in their work can also create unintentional errors in data. Whether it involves interpreting a subject's behavior or reading a fluctuating numerical display, one's expectations influence what one perceives. The researcher's tone of voice in asking a question can give clues about the most desirable answer and produce a greater likelihood of that answer without either the researcher or subject being aware of the bias. Even if error in your data is unintentional, the consequences can be same as those resulting from dishonesty. As a researcher, you are responsible for being aware of these possibilities and taking steps to avoid them.

How to Maintain Objectivity

1. Avoid Testing Hypotheses: Many texts emphasize testing hypotheses because it leads to such tidy statistics. A more open-minded approach is to inquire what happens under certain conditions. If the conditions tested are not too restricted and the research is carefully done, whatever happens may prove valuable.
2. Arm's Length Approach: Avoid thinking about the implications of the data as you collect and analyze it. Concentrate on the satisfaction of doing a thorough and accurate job.
3. Professional Conduct: Keep interactions with subjects focused on what is necessary for the task. Supplement these only with what is appropriate to show courtesy and respect.
4. Establish Rules for Decisions: Creativity belongs at the design stage of research, not during data collection. To find out where ambiguities may occur during the observations, use preliminary studies. Make rules to deal with ambiguities in a

consistent manner. While rules sometimes introduce a bias, it is more important to maintain consistency. A constant bias at all levels of the independent variable will not affect a conclusion based on a comparison of results at each level.

5. Objective Recording Methods: Provide specific alternatives for what usually needs to be recorded. Thereby the observer must only make a selection instead of more complex judgements. While computers are dispassionate recorders, their use requires even more specific instructions than would be needed for human observers.
6. Use "Blind" Techniques: If the person making the observations is unaware of either certain hypotheses or of which condition is being tested, they can't inadvertently bias what they record. The same holds for the person doing the data analysis. Using a code known only by a third party for the conditions being tested or analyzed may enable an experimenter to work blindly. When both the recording and analysis of data are done this way, it is called a "double blind" procedure.

2.5.3 DEVELOP COMPETENCY

Obtain the Necessary Knowledge and Skills or Collaborate with Others

2.5.4 ACKNOWLEDGE THE CONTRIBUTIONS OF OTHERS

In scientific writing, you are expected to substantiate specific knowledge pertaining to your work by drawing upon research done previously. The source of that knowledge must be acknowledged for two reasons: 1) to give due credit to the work of others; 2) to enable readers to verify for themselves your interpretation of the "facts". To not do so is called "plagiarism".

Plagiarism is not tolerated in science. It shows lack of respect for colleagues and for the process by which new knowledge is acquired. Moreover, it indicates a dishonesty that questions the truth of the writer's results as well. To understand the significance of this intolerance, you need to recognize the potential fragility of the knowledge that humanity has accumulated over the generations. New knowledge is based on previous knowledge plus some additional data and insight. But what if the previous knowledge is wrong or misinterpreted? A whole realm of misunderstanding ourselves and the universe could develop with potentially harmful consequences.

To see how this works, you may recall from statistics the rational for using at least 95% certainty as the criterion for significance. This is a much higher degree of certainty than we usually require for most decisions in life - including very serious matters such legal guilt, medical diagnosis, or far-reaching political decisions. Yet when 95% certainty forms the basis of future data which is also accepted with no less than 95% certainty, only 13 such steps are required before the cumulative certainty (.95 * .95 * .95, etc.) falls below 50%. On the other hand, the chances of three statistically significant studies all being wrong about the same conclusion are about one in ten thousand. This is why scientists are so concerned about the accuracy of corroborating data and other knowledge before accepting even statistically significant results.

To avoid inadvertently copying what others have said as you discuss their work, I recommend duplicating the first page and the relevant pages of their paper. You can check to see that you have accurately and yet in your own words presented the necessary information. If you must work by making notes, take care to indicate what are the author's exact words with quotations. When you feel that exact sentences in a reference are critical, you may use them provided that you identify them by quotations and the proper reference. As an example, let's say you want to refer to a specific conclusion in some research on a possible way of helping children acquire reading skills. Using a direct quotation, you might write the following:

quotations. When you feel that exact sentences in a reference are critical, you may use them provided that you identify them by quotations and the proper reference. As an example, let's say you want to refer to a specific conclusion in some research on a possible way of helping children acquire reading skills. Using a direct quotation, you might write the following:

---. Comparing the Metropolitan Reading Readiness scores of children who were trained with the special letters with those who trained with ordinary letters, Nelson, Nilsson, Piercy, Johnson, Frascara, Delano, Sone & Bravo (1999) concluded that "Reliable improvement in reading readiness of children aged 5 to 7 years old occurs when the Visually Patterned Alphabet is used in conjunction Alpha-Beta instruction" (p. 528). ---

Alternatively, you could express the same information in your own words and at the same time indicate the results of their statistical analysis as follows:

---. Nelson, Nilsson, Piercy, Johnson, Frascara, Delano, Sone & Bravo (1999) found that the Metropolitan Reading Readiness scores of 5 to 7 year old children who trained with the Visually Patterned Alphabet were significantly higher than those who trained with ordinary letters. ---

Note how this results in more direct and economical description than when the lead-in had to be tailored to fit the quotation.

2.5.5 ASSIST THE RESEARCH ENTERPRISE

1. Help Your Colleagues
2. Review Papers for Submitted for Publication
3. Serve on Ethics and Grant Review Committees
4. Assist the Operation of Professional Associations

2.5.6 TEACH OTHERS HOW TO CONDUCT RESEARCH

2.5.7 DISCUSS SUSPECTED MISUSE OF KNOWLEDGE OR MISREPRESENTATION OF RESULTS WITH COLLEAGUES

2.6 HOW RESEARCH ETHICS ARE MAINTAINED

2.6.1 INTEREST OF COLLEAGUES

Researchers rarely work in isolation, but as members of a university department or institute. They interact on a daily basis and "talk shop". Persons who did not discuss their research with their colleagues would be considered unusual, and concern about their activities would arise.

2.6.2 APPROVAL REQUIRED FROM DEPARTMENTAL OR UNIVERSITY ETHICS COMMITTEES

At all universities, this is required before a grant application can be submitted. Controlling the purse strings effectively controls most research, since most researchers can not afford the costs of doing research without grants. Departmental ethics committees typically review the research methods of student thesis and faculty research that has not been reviewed

elsewhere. Appendix D contains a form used by the Ethics Committee of the Psychology Department at the University of Prince Edward Island. (At UPEI, the ethics of class projects are the responsibility of the instructor. If in doubt about the propriety of a proposed student project, the instructor consults the Ethics Committee.)

2.6.3 GRANTING AGENCIES' REVIEW OF RESEARCH PROCEDURES

This includes a review of ethical factors in the proposed research method.

2.6.4 FEDERALLY SPONSORED INSPECTION OF ANIMAL FACILITIES

2.6.5 PEER REVIEW OF RESEARCH PRIOR TO PUBLICATION

At this stage the focus tends be on the adequacy of the research method and treatment of the data, however, ethical problems in the research method will not be overlooked.

2.6.6 ETHICAL PRACTICE REQUIREMENTS OF PROFESSIONAL ASSOCIATIONS

Professional affiliation is expected when applications for research funds are made. The competency of persons without such affiliation would be seriously questioned. For a psychologist to loss their professional association due to unethical conduct would effectively end their research career.

2.6.7 PROFESSIONALS REALIZE THAT ETHICAL PRACTICE IS THE BEST FOR EFFECTIVE RESEARCH

2.6.8 PERSONAL INTEGRITY

Without this all of the above are of little importance.

EXERCISES

1. For discussion the following is derived from an ethical problem posed by Kant in the Critique of Natural Reason: Consider the case of a student who learns that he/she has certain exceptional abilities which if developed would enable significant contributions to society. The student's family is prosperous. With little effort, the student can look forward to a wealthy life as a partner in the family business. What ethical aspects are involved in the student's career decision?

2. As a new psychology PhD, Sandy is offered a position as Research Director in the Health Department's Alcohol Treatment Division. For the first few months on the job, Sandy gets acquainted with the alcohol treatment programs and studies their records of referrals, treatments, and outcomes. In conjunction with other health records she finds what appears to be overwhelming evidence that the multi-million dollar treatment program is not effective. The main reason individuals do not come back seems to be the occurrence of severe health problems. She also discovers that some past attempts to revise the treatment program met strong opposition and were never been tried. Deciding that the first step to improving the situation was to test the current treatment program, she authorized a temporary modification to the present referral procedure. Designed as a proper within-subjects experiment (see Chapter 7), for the next 3 months new applicants for treatment would be randomly assigned to either the present program or to a minimal counselling condition. After a 18 months, the outcomes in the two groups would be compared. It was not long before the Department had received several complaints from family members of persons seeking treatment and from an influential former patient requesting readmission. They demanded treatment to which they felt they were entitled as taxpayers and which was being provided to others. In short order, Sandy was dismissed.

While the above is not a problem you will encounter in student research, it illustrates a dilemma faced by clinical researchers in both medicine and psychology. To prove that a new treatment is an improvement requires comparing its results with those of other treatments or in some cases no treatment. No one wants to withhold a treatment which might improve a person's life. Yet until properly tested, we do not really know whether a certain treatment actually helps. The more clinically oriented Canadian Psychological Association code is your best guideline for this question. Identify those sections which bear on the above ethical problem. Try to come up with an alternative that Sandy might, at least in principle, have tried.

3. From some recent data showing an interaction between salivary cortisol levels, stress, and physical activity, Bev gets an inspiration for a new means of helping persons cope with stress in their daily lives. Bev develops his ideas into a thesis proposal which is enthusiastically accepted by his supervisor. Funds are allocated which enable Bev to recruit 2 groups of undergraduates to participate for modest payment in a two month counselling and training program, which is followed by a 6 month evaluation period. The data show substantial difference in the stress scores of the two groups, but the difference is not quite significant. In reviewing the follow-up interviews, Bev notes that one subject in the control group and two in the treatment group experienced major changes in their personal lives which could reasonably have counteracted the lack of treatment or treatment, respectively. Sure enough, a reanalysis of the data without those three subjects, produces a statistically significant effect. While Bev can expect to earn his master degree in either case, he believes the results indicate a means of helping people that would not receive attention if left statistically insignificant. Bases on the

chapter and codes in the Appendix, what ethical issues pertain to Bev's decision on whether to include all subjects in his analysis? What would you recommend that Bev do?

4. Go to the library and look over some research papers in one of the following journals: Canadian Journal of Behavioral Science, Developmental Psychology, Ergonomics, Journal of Experimental Psychology: Human Perception and Performance, Journal of Personality and Social Psychology, Memory and Cognition. Select one paper that looks interesting and that you can understand. Assume that you are going to replicate this study. Follow the example of the psychology department's application form for ethical review in Appendix D, and complete it for the paper you selected. Since space in journals is limited some of the needed details about the method may not be specifically spelled out in the paper. Therefore you have to devise details that are plausible.

SECTION I - MAKING MEASUREMENTS

Measurements make psychology a science and separate it from that other study of mind and behavior - philosophy. Until now the only experience that many students have had with making scientific measurement might be weighing something in a test tube or perhaps counting particles while looking through a microscope. What do psychologists measure? Well, sometimes they weigh stuff in test tubes or count things seen with a microscope. They may also measure light rays that produce vision, brain waves that reflect consciousness, the speed of a thought, the strength of a muscle's contraction, the confidence with which a choice is made, the feelings evoked by a certain experience, or the attraction between two persons. No other science uses such an enormous variety of measurements. If you think about it, the introductory psychology course had lots of examples of data obtained by various sorts of measurements. This course explains where those numbers come from. Learning to make measurements yourself will enable you to become a participant, which is more fun than just being a spectator. It is also the best way to gain an understanding of the data on which psychological knowledge is based.

So many methods are used to make measurements in psychology that it is not possible to describe or even mention them all in an introductory book on how to do research. The following three chapters aim to provide enough of an overview to illustrate the principles of how psychologists make measurements. Extra emphasis is placed on the observational, questionnaire, and timing methods because they require no specialized equipment and can be readily understood with a minimum of background knowledge. The growth of insight into psychological measurements and new technologies have expanded the methods available to psychologists so greatly in the last 25 years, that general books on research measurements have become rare. I recommend one of the last books, Kling & Riggs (1971). Though not recent, the principles of the methods they present have been changed little by newer techniques - for the most part. The major difference since "them old days" is the use of computers - not only to automate measurements but also to produce the independent variables. While occasionally more specialized books are written describing measurement techniques in particular areas of psychology, the latest information is readily available in the METHODS section of papers published in journals.

Because there is no better way to learn how to do something than by doing it yourself, most psychologists have learned how to do research by working for and with their professors. This mentoring tradition is older than science itself - or even history when one considers its origin in the apprentice system on which so many trades are founded. To start you in this tradition of learning by doing, this text is designed to accompany a series of "laboratory" exercises such as those described in the back of the book. "Laboratory" is in quotes because several exercises are not done in a laboratory at all, but rather in the real world where people behave naturally. This is also the case for a substantial portion of real psychological research. These semi-weekly assignments are supplemented by some recommended exercises at the end of each chapter. Your teacher may have other ideas to give you hands-on experience. The more you do, the more you will learn than can be learned by simply reading.

Chapter 3

DIRECT OBSERVATION AND TRACES

Not all research is done in laboratories. Measurements can be made in natural (or perhaps somewhat contrived) situations by recording the behavior you see. "Pure" observational research does not involve giving subjects instructions or obtaining data by asking questions. The method's lack of intrusion may reveal behaviors that don't occur when people are given specific instructions or are obviously monitored. (With careful attention to ethical considerations, it is even possible to use this method without the subjects' awareness.) For certain questions about human (and also animal) behavior, these advantages may outweigh the disadvantages of an inability to do things as carefully as is possible in laboratory research. This chapter is about how to reduce these disadvantages and make your measurements more convincing.

What is probably the earliest recorded psychological research project used the method of observation. Herodotus (450 BC) recounts that Psammetichus, king of Egypt around 600 BC, sought to find out which race was the original race of mankind. Since this could not be determined directly, the king reasoned that whatever language was naturally used by infants when they first began to speak would be the language of the original race - provided they were not influenced by hearing their parents speak some other language. According to de Selincourt and Burn's translation (1972, p.129), "He took at random from an ordinary family two newly born infants and gave them to a shepherd to be brought amongst his flocks, under strict orders that no one should utter a word in their presence." Two years later the shepherd heard both children clearly say the word "becos". The children were brought to the king, who heard them say the same word. Not recognizing the word himself, the king made some inquiries and learned that it was the Phrygian word for "bread". Thereby fell a cherished belief of the Egyptians that they were the first people.

Psammetichus' research illustrates several basic aspects of using the observational method. First, we may note that there is no independent variable. It is a descriptive study. While it did provide an answer, in hindsight we can see that the king's research would have been strengthened by more observations. Perhaps all infants from that family or district tend to make a "becos" type sound due to certain inherited oral characteristics. Perhaps any infants brought up in the vicinity of sheep will make some "b" type of sound imitating the sheep. Thus Psammetichus might also have used some additional children from another country and/or had some additional children raised by an isolated fishing family. These flaws were unlikely the result of ethical considerations. The world was seen in much simpler terms. The possibility of confounding variables was no more recognized than cruelties that primitive societies sometimes impose on their members.¹

Psammetichus' dependent variable was the children's first clearly spoken word. He made two assumptions that contributed to the operational definitions of this dependent variable: 1) The "original race" could be identified by its language. 2) In the absence of other linguistic influences, the first words spoken by infants will be that of mankind's "natural" or original language. While these definitions are open to question, the doubt is about what is concluded from the results, not on the results themselves. A critical, confounding variable, the infants' exposure to any particular language, was eliminated by using new born subjects and by raising them in isolation.

¹ In 2,000 years our society will likely be considered primitive too.

3.1 TO INTERVENE OR NOT TO INTERVENE

3.1.1 WITHOUT INTERVENTION

Called "natural observation", this method is preferred when you want to know what happens when research is not being conducted. Certain types of behavior change when people are watched because they want others to think well of them or are shy. For example, research on politeness would probably find more "polite" behavior when the subjects know they are part of an experiment than what normally happens. The method is also used whenever the desired information can be obtained by just looking at the subjects and neither instructions or further information is needed, simply because it is more efficient than making some intervention. Attendance of cultural events, the selection of a consumer product, and the choice of paths through a park may be studied directly in terms of independent variables such as sex and age. Attention to details of behavior or aspects of where and when it occurs may enable you to make inferences about other independent variables such as socio-economic status or the situation that affected the behavior.

When making measurements by observation, a major problem arises when you can not watch and record everything that happens. It is difficult not to let ones attention be drawn to behaviors that strike us as interesting or especially pertinent to a hypothesis. Consequently more of those behaviors can get measured while others that may be more meaningful to the subjects or unfavorable to the hypothesis. Some systematic approaches to natural observation were developed for behavioral research of animals in the wild. A grid is imposed on the area being observed. A schedule is established so that the observer records all behaviors in a certain grid location are recorded during a certain period. Activities in other locations are ignored, no matter how interesting they may appear. Over time, all locations get observed. Scrambling the schedule from session to session minimizes the chance that some behaviors get missed if they tend to occur in certain locations at certain times.

Let's consider how psychologists might study helping behavior in young children. The setting is a day-care center. To learn the effect of role models, some mornings the children are shown a cartoon about Smurfs helping one another; another mornings they are shown a Road Runner cartoon. (These cartoons, by the way, represent two levels of the independent variable "role model". Next I'll describe the dependent variable.) To measure the occurrence of helping behaviors, the psychologists will observe the children during free play periods over the rest of the day. Anticipating difficulty watching and recording the activities of 20 children at once, masking tape was used to mark 6 areas in the play room. A schedule shown in Table 3.1.1-1 was made to tell the observer where to look at any given time. To facilitate recording the activity, an additional card was made with a code for identifying individual children, and a code for various types of interactions that are likely to be seen - see Table 3.1.1-2. The latter is based on a week of making notes while watching the children and on advice from the staff. Like adults, the behavior of children may differ when they know someone is watching. Therefore observations are made through a one-way mirror in a partition placed in a corner. (Since people notice who in a group is being watched, "being watched" is a likely confounding variable. The one-way mirror is part of an operational definition that describes how the researchers held this variable constant.) However, this hampers the watching and greatly reduces auditory cues to the nature of the interactions. These problems might be corrected a video camera surveying the play room from a different angle and with a selectable ceiling-mounted microphone over each area.

Table 3.1.1
Example of a Data Sheet that Might be Used to Record Observations
of Children's Play Activities in a Day-Care Center

DATE:		CARTOON:	
TIME	AREA	CHILDREN	ACTIVITY
10:40	1		
10:45	3		
			ECT.
11:35	6		
LUNCH, GROUP GAME, NAP			
3:10	2		
			ECT.
4:05	5		

Note. The complete counterbalanced sequences for the areas are: AM - 1, 3, 5, 6, 4, 2; 5, 3, 1, 2, 4, 6; PM - 2, 4, 6, 5, 3, 1; 6, 4, 2, 1, 3, 5. These would be rotated from day to day.

Table 3.1.1-2
Codes to Identify Children and Types of Activity

CHILD	CODE	ACTIVITY	CODE	ACTIVITY	CODE	ACTIVITY	CODE
Adrian Wils.	AD	INDEPENDENT		COOPERATIVE		NON-COOP.	
Bjorn Borg	BJ	inactive alone	1	exchanging toy	21	hording toys	41
Cathy O'Rour	CO	playing by self	2	sharing set	22	hording space	42
	ETC.			ETC.			
Lenard Bern.	LE	NEUTRAL		HELPING		AGGRESSIVE	
Leslie MacD.	LS	playing same	11	finding toy	31	pulling at toy	51
Linda Beaton	LI	drawing same	12	showing how	32	pushing other	52
	ETC.			ETC.			
Willy Eliot	WI						

To students acquainted with physics or biology labs, the above experiment may seem simple. Certainly none of the preparatory work and subsequent data collection involves anything like "rocket science." Don't let that fool you into thinking that this sort of research is easy to do. A week of observing the children at play would not be adequate for inexperienced researchers to design the study. Furthermore, when the data have been collected, the results are not evident from simply averaging numbers printed by some ready-made instrument or other. The psychologist can face weeks of reviewing the observation sheets, perhaps comparing their records with video tapes to resolve occasional ambiguities. Item analysis may reveal difficulties introduced by absenteeism. Factorial analysis of variance may be needed to remove a confounding day-of-the-week effect. Trend analysis may uncover the need for more observations to counterbalance a time-of-day effect. Such problems are unlikely to occur in experiments on rolling marbles or crayfish. Are you sure you are up to the challenge?

3.1.2 WITH INTERVENTION FIELD RESEARCH

Some behaviors may occur only in situations that occur infrequently. Waiting for them to occur naturally may tax the researcher's resources. The alternative is to **intervene** and artificially stage the situation. To find out how people react to a maintenance barrier in public passage way, you might consider obtaining permission to put one in place rather than waiting for the necessity occur. However, care must be exercised so that interventions do not interfere with people unduly. Too extensive a barrier might make persons late for appointments or force them onto the street. Intervention enables you to change the situation to produce an independent variable. To better understand the effects of a barrier, you could systematically change factors such as its shape - to find an optimal design, include various notices about its presence - to study various forms of advice, or make the need for the barrier more or less obvious - to study social compliance. Intervention can also enable you to control other variables that might affect your measurements. If barriers were put up for maintenance to remedy an effect of very cold weather, comparing social behavior then with behavior when the barriers were not there might be confounded by the effects of the weather on social behavior.

Another form of intervention involves having the researcher **participate** in the situation being observed. Say you wanted to learn how people react to a stranger slipping on an icy sidewalk. Could you make a section of sidewalk particularly slippery? Certainly not. It would even be irresponsible to seek a particularly dangerous stretch of sidewalk for observation without warning persons of the danger. As solution to such problems, you or an assistant could pretend to slip on a walk and observe how nearby persons react. (To avoid leaving these people feeling concerned or apprehensive, one would immediately afterward tell the bystander subject that it was staged, explain why the research was being done, and thank them for their cooperation.) This sort of intervention also facilitates having an independent variable. If you were interested in how socio-economic status of the victim affect bystander response, you might wait a long time to obtain a similar number of measurements from victims of low, middle and high status. The participating researcher can dress to fit those roles and also stage the incident to balance uncontrolled variables such as time of day and weather that might otherwise confound results obtained by natural observation.

Participation can also provide another advantage. As a participant among the subjects, a researcher may be able to make observations that would be impossible for an outside observer. This can present a dilemma. If your participation is not convincing, subjects may suspect something is going on and not behave naturally. However, if you immerse yourself in the role, you may lose your objectivity in recording data. Participation may also interfere with the subjects' natural behavior. A classic example is illustrated by a researcher who participated so effectively in a doomsday cult that they wanted to make him their leader!

IN THE LABORATORY

Lest you think that observational measurements are only suitable for social behaviors outside the laboratory, here is an example of its use in a tightly controlled experiment involving optical and electronic equipment. Previous research had provided precise measurements of changes in color-hue produced by changing the frequency and duration of monochromatic light flashes. These data led to insight into how color information is carried in the visual system. Several subject reported that in addition to changes in hue, they also experienced changes in color saturation. (A desaturated red appears pink.) However, varying color saturation is a complex matter, and this makes it difficult to measure these effects directly by matching flashed and steady colors. To determine whether saturation changes were reliable and of sufficient strength to be interesting, we decided to first make observational measurements of what occurs.

We devised a subjective rating scale in which "10" was a fully saturated color and "0" was completely white. The measurements were made by having the subject call out their rating of the saturation as wavelength, pulse duration or pulsing frequency were systematically varied. These results showed changes in saturation that were consistent across subjects and that seemed to occur in some pattern as the independent variables (wavelength, duration, and frequency) were changed. Years later, I built an optical system that enabled measuring saturation changes by matching.

It is not uncommon for measurements in science to progress in this manner. A psychologist incidentally becomes aware that something interesting occurs. To learn whether this occurs reliably, systematic observations are made and impressions recorded together with the circumstances under which the effect occurs and does not occur. If these measurements indicate the effect is reliable, then more work and resources are invested to improve the accuracy of the measurements. Many more observational measurements with careful attention to holding constant or balancing a number of variables may be needed to identify which produce the effect and which don't. To measure what happens with greater precision than generally possible by simple observation, some type of test or apparatus may need to be developed assist the judgements or make the measurements more objective.

3.2 DIFFICULTIES YOU MAY ENCOUNTER

3.2.1 SUBJECTS' REACTIVITY

THE PROBLEM

We are all aware of our "company manners" - the tendency to be more polite, neater, etc. when company is present. This effect is not necessarily personal. Simply stand holding a clipboard by the side of a busy thoroughfare and you may create a traffic jam. When drivers notice that they are being monitored, they tend to slow down a bit just to be

sure that they are not exceeding a speed limit. To make you aware of the effect this reaction may have on observational data, we'll consider what tends to happen when people know they are being observed. You can probably recall real life examples for each of the following.

Being "Helpful": Advanced social animals try to help others when possible. Because an understanding of the value of science to create a better world is widespread in our society, perfect strangers are willing to go out of their way to try to help researchers. This may lead to behaviors that subjects think may be expected rather than what they would otherwise do.

Looking Good: It is natural to want others to think well of us, and to a certain extent what others think of us shapes what we become. Therefore we make an effort to behave in more ideal or socially acceptable ways when others are watching.

Enjoying Attention: As social animals, we generally like the attention of others. This can lead to the performance of certain greeting and attention seeking behaviors which interfere with the behavior the research is trying to study.

Shyness: We have all experienced the apprehension of performing before an audience and seen or heard of "stage fright" leaving a people unable to do anything in such circumstances. Whether due to a concern about adequacy or a desire for privacy, the effects inhibit behavior.

Antagonism: Certain types of behavior are 'nobody else's business' and we are angered by an invasion of privacy. A few people have hostile attitudes to research, or indifferently seek amusement by deliberately behaving strangely when they realize they are being observed.

MINIMIZING THESE PROBLEMS

The obvious way to avoid reactivity is to keep the subjects unaware that they are being observed. Being unobtrusive, blending into the environment, letting the subjects get used to your presence when you are not observing, and technology such as hidden microphones or cameras can help. However, these solutions may lead to ethical complications depending on what you wish to observe. As a general guide, behavior that is public does not involve an ethical problem provided that the persons' **anonymity** is respected. When the behavior is not public, **consent** should be obtained beforehand. While obtaining such consent may compromise your data somewhat, there usually is no ethical alternative. Remaining inconspicuous and assuring the subjects that their anonymity will be respected can minimize problems of awareness. In observational research, as in all other methods, **respect** for the rights of others, the **golden rule** of treating others as you wish to be treated, and **courtesy** will minimize most problems that could arise. We'll discuss these matters in more detail when you have become better acquainted the various ways of doing psychological research.

3.2.2 OBSERVER OBJECTIVITY

THE PROBLEM

Sometimes scientists are portrayed as dispassionate and totally objective seekers of truth. This is rarely the reality. Scientists are driven by the fun of doing science, the pride of doing science well, the challenge of finding what no one else has found before, and the prestige that results from an important discovery. However, it is a reality that the best process for distinguishing truth from uncertainty has a statistical basis which can only

prove what is different from chance, but can't prove the lack of an effect. Accordingly, only statistically significant results get acknowledged as contributions. It is also a reality that financial support for research depends on a scientist's demonstrated ability to contribute to knowledge. In such circumstances, the most dedicated scientist can not help but hope for results that are statistically significant.

Deliberate cheating in science is rare as proved by the headlines when it occurs. (Comparable misdemeanors by politicians or business people are hardly newsworthy.) The rarity of the problem attests to the rigor and quality of the training and selection of scientists. Furthermore, the same skepticism that leads scientists to question what is presently believed also leads them to question new findings. Sooner or later errors are discovered - whether honest or dishonest. Scientists know that when they publish their results, others will check those results. Since mistakes reflect badly on them, they strive to avoid them. The real problem with objectivity arises from errors that occur unintentionally. Whether it involves interpreting a subject's behavior or reading a fluctuating numerical display, one's expectations influence what one perceives.

HOW TO MAINTAIN OBJECTIVITY

Simply being aware of the problem puts one on guard, but that is not enough. Here are some active steps you can take.

Avoid Testing Hypotheses: Many texts emphasize testing hypotheses because it leads to such tidy statistics. A more open-minded approach is to inquire what happens under certain conditions. If the conditions tested are not too restricted and the research is carefully done, whatever happens may prove valuable.

Arm's Length Approach: Avoid thinking about the implications of the data as you collect and analyze it. Concentrate on the satisfaction of doing a thorough and accurate job.

Professional Conduct: Keep interactions with subjects focused on what is necessary for the task. Supplement these only with what is appropriate to show courtesy and respect.

Establish Rules for Decisions: Creativity belongs at the design stage of research, not during data collection. To find out where ambiguities may occur during the observations, use preliminary studies. Make rules to deal with ambiguities in a consistent manner. While rules sometimes introduce a bias, it is more important to maintain consistency. A constant bias at all levels of the independent variable will not affect a conclusion based on a comparison of results at each level.

Objective Recording Methods: Provide specific alternatives for what usually needs to be recorded. Thereby the observer must only make a selection instead of more complex judgements. While computers are dispassionate recorders, their use requires even more specific instructions than would be needed for human observers.

Use "Blind" Techniques: If the person making the observations is unaware of either certain hypotheses or of which condition is being tested, they can't inadvertently bias what they record. The same holds for the person doing the data analysis. Using a code known only by a third party for the conditions being tested or analyzed may enable an experimenter to work blindly. When both the recording and analysis of data are done this way, it is called a "double blind" procedure.

3.3. RECORDING OBSERVATIONAL DATA

3.3.1 CONTINUOUS RECORDS

To get the most data from research and not overlook something that may prove to be important, one could record everything that happens. This can be particularly desirable in exploratory research where one is uncertain about what to seek in the data. The growing availability of computer, and video technology makes continuous recording ever more feasible. Its practicality, however, is another matter. Eventually you have to extract numbers from these records. Going over those records even once to obtain numbers usually takes much more time than recording them initially and can be very boring. In considering this approach you should know that, as well as being expensive, it typically produces much more data than is useful. Yet certain measurements such as sequences of behavior, changes in facial expression, or what preceded a certain response that can be difficult to measure selectively. Continuous records, which can be examined backwards and forwards, may be the only way to obtain such measurements.

3.3.2 SELECTIVE RECORDS

When recording everything would be too much, the alternative is to record only what you believe is needed to answer the question. Other behaviors are ignored or noted only incidentally. The catch is that you need to know beforehand exactly what should be recorded. Learning this usually takes extensive experience and one or more preliminary trials to determine what measurements are most promising and also feasible to obtain by observation. During preliminary tests, continuous records may help identify these measurements.

In observational research certain behaviors can sometimes be measured directly in terms such as time, place, and number of items used. Other behaviors may be quantified by the observer in terms such as degree of effort, aggressive or submissive, happy or sad using scales from 1 to 10 or bipolar scales from -5 to +5. Some behavior may first need to be qualitatively categorized. Rather than simply noting that the subject "walked", it may be important to identify whether it was a fast, slow, or delayed walk; whether they avoided, continued as before, or approached a certain location. Later, counting how many times each category of walking occurred provides numbers for analysis.

While one could simply write down these observations each time the behavior occurs, this takes time and may make your presence more obvious. Prepared checklists can greatly simplify and speed up the recording process. They are essential when many subjects need to be observed at once. Most observational studies have been conducted with nothing more than printed forms, pencil, and a clipboard. Today, portable and pocket computers are increasingly used. Computers can be programmed to provide special keys for timing, counting, and indicating categories of behavior. Their scrolling displays make it easy to handle lists longer than what fits on a single sheet of paper. Ratings can be entered on displayed scales with pointing devices. Their memories provide a compact data storage. Computerized or not, it is you, the researcher, who must prepare the checklist for recording the data.

For an example of a study that uses a checklist to record observational measurements, we'll consider how appearance influences people who provide service to others. "Providing service" will be operationally defined as the selling of shoes. Before judging this a mundane choice, think again. Can you name another service of comparable complexity open to public observation performed by and for persons of either gender that

also involves a certain degree of personal care?¹ Several potentially worthwhile studies could be developed from this approach: Insight into gender interactions, assessing the effectiveness of a community information campaign on ageism, or an applied study on improving their stores done under contract to the retail corporation are some examples. Each would involve different independent variables introduced by the researcher(s) participating as the customer with an assistant nearby recording the data.

Table 3.3.2 below presents a draft of a checklist that could be used. It is formatted in stages to reflect the expected sequence of events. Several multiple choice categories are used to record various types of behavior at each stage. Some simple rating scales are used to measure the degree of cordiality at various stages of the interaction.

For ethical reasons, a study like this should only be conducted when a store is not busy, and the sales staff appear to have time on their hands. Even so, the observation period must be kept brief. The importance of protecting anonymity of the sales persons is especially important if this were a study done for the retail corporation, and should be specified by the researcher in the contract from the onset.

¹ Based on superficial reading, media pundits sometimes make scurrilous comments about what they judge to be trivial research without realizing the significance of the operational definitions involved. Perhaps you've seen headlines such as "MP uncovers \$30,000 SSHRC grant to study shoe clerks." They may grab attention, but are misleading when they neglect to explain the real issues being addressed.

While some armchair effort based on personal experience enabled me to create the above example, it is just a beginning. One or more rounds of test trials would be needed to ensure a good checklist. For example, modifications may be needed to accommodate trials when the dialog between salesperson and customer do not proceed as expected. Also, having a colleague present during preliminary testing could help determine whether important aspects are missed or some items could be deleted.

Next, you must consider how to summarize the information on the checklist into a few useful numbers. Simply counting the number of check marks and adding in the three ratings could produce a single "cordiality score". Yet, certain behaviors such as actually getting shoes deserve more credit than simply offering. (Note the weights included in brackets on the checklist to facilitate subsequent analysis.) Furthermore, the last cordiality rating could reflect a more genuine attitude than the initial rating. Trial-and-error work with various weightings may be needed to produce a scoring system that matches general impressions of the acting customer. Alternatively, you may notice more than one aspect of cordiality and devise two or more summary scores to reflect various aspects of behavior, such as a distinction between courtesy and helpfulness.

3.4 WHO TO OBSERVE

Who you observe and how many subjects are needed will depend on the purpose of the research and its independent variable. Without intervention, the above shoe store study could focus on the gender, apparent age, or apparent socioeconomic status of customers to see how these factors influence the salesperson's behavior. You might wait until you had data from 6 of each type of customer then select another salesperson, until you had data from 10 or more sales persons. Subject selection becomes more complex when subject characteristics are an independent variable. For the shoestore example, you might measure courtesy as a function of the salesperson's age, the size of the store, or the socio-economic status of the store as indicated by average shoe price. Then you may want to observe 10 salespersons of each age or in each type of store. Deciding who to observe involves the following considerations.

3.4.1 HOW MANY SUBJECTS ARE NEEDED?

Descriptive and analytical techniques for data usually involve calculating means and standard deviations to summarize the data or test for significant differences. These calculations are generally acknowledged to require at least 6 independent measurements, but, such minimal amounts of data usually don't reveal significant effects. How much data are needed to determine whether an effect is significant? That depends on the size of the effect and on the variability of the data. While you don't usually know these before you do the research, you may be able to estimate them from reports of similar research. The alternative is to make some preliminary measurements yourself.

Let's say that from 6 measurements in each of two conditions X and Y you obtain means of 8 and 10 with standard deviations of 2 and 3 respectively. The standard errors found by $SE = SD/\sqrt{n-1}$ would be 0.89 and 1.34. For a t-test:

$$t = \frac{X_{\text{mean}} - Y_{\text{mean}}}{\sqrt{SE_x^2 + SE_y^2}} = \frac{8 - 10}{\sqrt{0.89^2 + 1.34^2}} = 1.24$$

At $(5+5) = 10$ degrees of freedom, the minimal value of a significant "t" is 2.228. The difference in the these two means is not significant. Hardly surprising given the few measurements and standard deviations of the same magnitude as the difference.

The question is, "How many more subjects would need to be tested to obtain a significant difference if the measurements continue to be similar to the ones obtained?" Assuming that the difference between the means and the standard deviations remain the same, the standard errors decrease as the square root of the increasing number of measurements². Thereby "t" increases even though the difference in means remains the same. A few trial-and-error calculations show that 10 measurements in each condition would yield a significant value of "t". Figure 3.4.1-1 shows how the number of measurements required for a significant difference increases as the difference between two means decreases - assuming the smaller and larger means have standard deviations of 2 and 3 respectively. If you have enough resources to make lots of observations, even differences less than half as large as the standard deviations can be proven as unlikely due to chance.

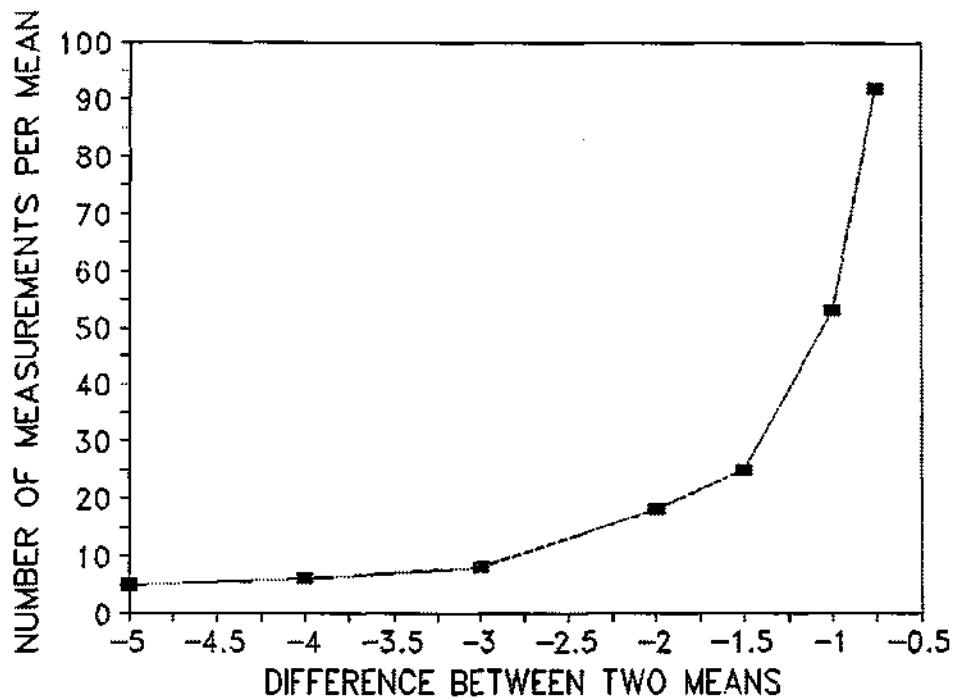


Figure 3.4.1-1. How the number of observations needed to obtain a significant difference between two means increases as the difference between the means decreases. See text.

If your data has more than two means, a similar approach can be taken with ANOVA and means comparison tests like Tukey's HSD. More measurements with the same differences between the condition-means and same variance will increase the condition-mean-square since the conditions-degrees-of-freedom remain the same. Of course the error-sum-of-squares also grows, but the error-mean-square remains about the same

² The statistically savvy may question this assumption since with more measurements, standard deviations are likely to increase as the chances of getting some unusual measurement increase. However, more measurements also increase the degrees of freedom with a corresponding decrease in the critical value of "t". These two effects can be regarded as balancing one and other.

because the error-degrees-of-freedom increase at the same time. Therefore twice as many measurements can be expected to approximately double the value of F. Tukey's "Honestly Significant Difference" between any two means will decrease with more measurements in a manner similar to that shown for t-tests above. All it takes to determine the number of subjects needed are some trial-and-error calculations.³

Psychologists often use observational research to seek associations between circumstances and behavior. These associations are usually described in terms of correlation coefficients. One would not expect more measurements to produce an appreciably larger correlation than one obtained from some preliminary data. However, a table of the critical values of Pearson's Product Moment Correlation shows that the critical value of "r" is roughly halved as the number of measurements is squared. See Table 3.4 below. With a 100 pairs of observations, a correlation of less than 20% is significant. (Though such correlations are significant, they account for less than 4% of the variability, so one should be skeptical about what they mean.)

Table 3.4
Critical value of Pearson's Product-Moment Correlation for 95% Confidence

degrees of freedom	r	degrees of freedom	r
1	.997	16	.468
2	.950	17	.456
3	.878	18	.444
4	.813	19	.433
5	.754	20	.423
6	.707	25	.381
7	.666	30	.349
8	.632	35	.325
9	.602	40	.304
10	.576	45	.288
11	.553	50	.273
12	.532	60	.250
13	.514	70	.232
14	.497	80	.217
15	.482	100	.195

Note: The degrees of freedom for this test equals the number of pairs minus two.

Finally, observational measurements often involve counting which behaviors occur and for such data Chi-Square analysis is required. Higher totals of the same categories of behaviors will not change the critical value of χ^2 , but the χ^2 result will tend to increase. Consider a simple two category situation such as might be encountered dealing a question of consensus. Say that 10 persons were involved, because this is the fewest number of persons that would result in "expected values" of 5 and 5 for and against the resolution if the

³ "Trial and error" calculations are especially easy with spreadsheet programs like Excel, Lotus 123, and QuattroPro. Once the format of the calculations is entered, you can change particular numbers such as the condition- and error-sum-of-squares and the error-degrees-of-freedom to reflect what can be expected if more measurements were made.

results were a matter of chance. How many persons would be needed for a significant majority? The Chi-Square formula is:

$$\chi^2 = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2}$$

where O_1 and O_2 are the number of observed events in categories "1" and "2". where E_1 and E_2 are the number of expected events in categories "1" and "2".

Trying an 80% majority results in the following value of Chi-Square:

$$\chi^2 = \frac{(8 - 5)^2}{5} + \frac{(2 - 5)^2}{5} = 3.60$$

Alas, at 1 degree of freedom, χ^2 must 3.84 so an 80% majority is not significant. With 10 persons selecting from two alternatives, nine of them must agree for a Chi-Square to indicate a significant preference. Figure 3.4.1-2 shows how the number of subjects must be increased for smaller majorities to be statistically significant. If you have some preliminary data to indicate approximately how subject preferences are distributed across two or more alternatives, you can do some trial-and-error calculations to find out how many subjects need to be tested to demonstrate a significant effect. This can be extended to R-values for deciding which categories might differ.

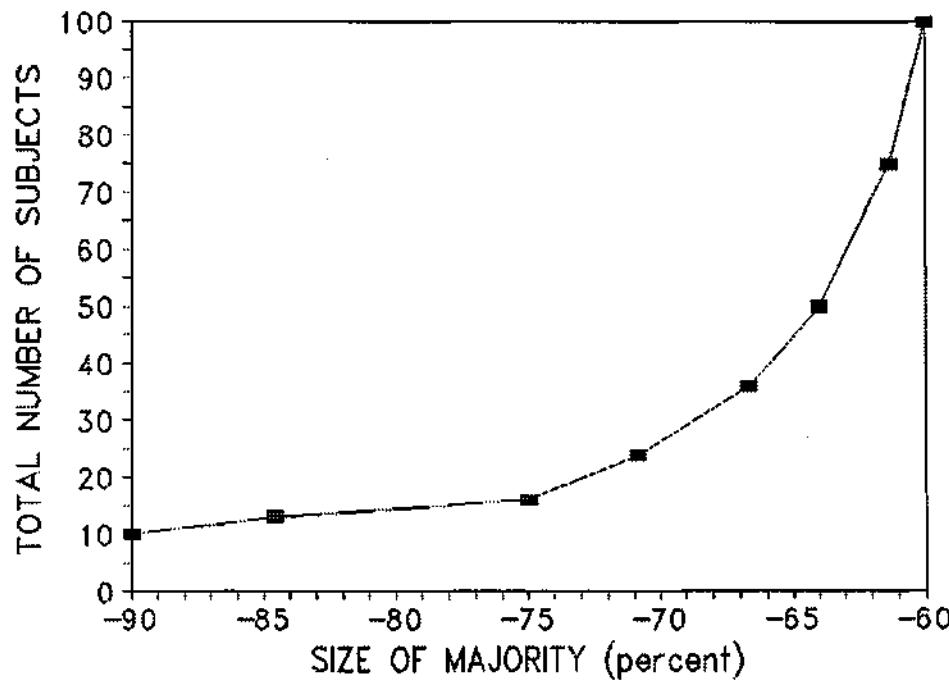


Figure 3.4.1-2. As the obtained majority gets smaller, more subjects are needed for a Chi-Square statistic to be significant..

There is another factor to bear in mind when deciding on how many subjects are needed with categorical data. Chi Square calculations are not reliable when the expected number of observations in any category is less than 5 (Siegel and Castellan, 1988). This may mean testing more subjects or else combining categories.

3.4.2 SELECTING THE SUBJECTS

HOLDING SUBJECT CHARACTERISTICS CONSTANT

One way to reduce variability in your data is to test subjects that are very similar on the basis of age, education, skills, social status, etc. in conditions that do not change from one measurement to the next. Which characteristics to keep constant will depend on what you are measuring. For example, a willingness to help a stranger probably does not depend on the subject's age or handedness, but may depend on the subject's sex or size. Figuring out which factors make a difference involves general knowledge and an understanding of people supplemented by studying reports on related research.

Perhaps it occurs to you that the safest way to proceed is to hold constant as many evident subject characteristics as possible. However, there are drawbacks to being very selective about your subjects. 1) You may have difficulty finding enough subjects. 2) Your conclusions may pertain to so few people that they will not be of interest. Consider a study on how advertising media affect color preferences. Suppose only right-handed, brunette, married, females, who are between 160 and 180 cm tall, and appear to be between over 20 but younger than 30 are selected. If the results were statistically significant, they might well be questioned in terms of their implications for people in general.

The generalizability of data depends on the question being studied. If instead of color preferences, the above characteristics happened to describe three subjects in an experiment involving color difference thresholds, the results from this select sample would be respected. Unlike preferences or attitudes, sensory discriminations depend on primarily on physiological mechanisms that are common to all persons with normal vision, and in this respect those subjects are representative of people in general.

BALANCING SUBJECT VARIABLES

A confounding variable is one that happens to change at the same time as the levels of the independent variable. Such bias is worse than random variability. It is not taken into account by statistics and leads directly to wrong conclusions. The more factors that are not held constant, the greater the possibility that one of these factors will confound the results - particularly if the number of observations is modest. For example, consider the above color preference study being conducted in a mall with no selection of subjects. If it so happened that red and blue colors got tested on cold, rainy days we might end up with those data based primarily on unemployed, older males. Their preferences might differ from those of suburban homemakers who frequented the mall on nice days when yellow and green colors were tested.⁴

Whether subject characteristics are allowed to vary to increase generalizability or can not be held constant, you can avoid letting them confound your results by balancing them across the levels of the independent variable. Balancing can be achieved by **matching** who is tested. For example, test an equal number of old men, old women, young men and young women in each color range or media type. This requires keeping recording these characteristics with the data. However, with observational measurements, that works only for obvious characteristics. What about other characteristics such as education, occupation, etc. that can't be determined by observation?

⁴ A more subtle interaction between weather, mood and color preference could also confound the results if various levels of the independent variable such as colors or media type happened to be tested during different weather conditions. This effect, of course, would not be removed by balancing subject characteristics. You would have to make sure that testing of these levels was balanced for weather conditions.

When measurements are made in a public place, it may seem that the particular persons who happen to pass by are a matter of chance and many observations should ensure matching by randomization. That may not be true. We have already suggested how the weather may influence who is about in a mall. A random selection of those bypassers will not avoid the confounding effect. With observational measurements, it is usually impossible to obtain a random distribution of subjects since that requires a list of names from which a random selection can be made. The only way to balance non-evident subject characteristics is to balance other variables such as weather, time of day, and location that can have a selective effect on who is available to be observed.

In conclusion, by selecting subjects to hold a few characteristics constants, balancing other evident characteristics, and balancing when and where your observations are made, you can reduce variability, avoid confounding effects, and also be able to generalize the implications of the data.

3.5 TRACES: READING THE WORLD AROUND YOU

In east Africa 1.2 million years ago, a female hominoid walked across a mud flat. From the spacing, depth, and other characteristics of her fossilized footprints, some reasonable inductions can be made about not just where "Lucy" was going, but also about her motivation for going there (XXX, 19XX). The spacing of footprints indicates whether the person was running or going for a leisurely stroll. The relative depression of toes and heel reveal whether the person was leaning forward as in search or fleeing with abandon.

During the convention, there appeared to be a neck-and-neck race between the incumbent candidate and a populist new-comer. After the vote was over, pundits were bewildered by the large margin of victory. It was no surprise to the custodians, who noted the huge disproportionate amount of flags, buttons, and party favours and other debris that bore one candidate's name.

It was 5:41 PM, November 13th, 1968???. A combination of natural and engineering circumstances shut down large sections of the northeast power grid. New York was plunged into darkness for 6 hours. Within minutes, police forces were on the alert, but rather than a sudden outbreak of crime, the evening was unusually "slow". The following year a statistician noted a small but statistically significant bump in the number of births registered during the second week of August 1969.

Due to poor planning a major fund raising event by the Environmental Protection League got scheduled at the same time as the national league finals. Which broadcast captured people's attention? A social psychologist had read a report about sudden pressure drops in London's water system during breaks in the English rugby championship game and reasoned a similar effect would occur here. Reasoning that the telethon's breaks would occur regularly on the hour and at 15, 30, and 45 minute times, while breaks in the sporting event would be more haphazard, the researcher saw a means of gauging public attitude without doing a costly survey. A comparison of water system records between the inner city and various suburbs even provided insight into how that attitude depended on socio-economic status.

At 8:43 PM on March 26rd 1997, a person crosses the lawn of Fanningbrook. Thirty minutes later, an RCMP dog picks up the scent, follows the trail briefly in both directions to determine the odor gradient, and then takes off in the direction the person had crossed.

The above are some examples of how we leave traces that bear witness to almost everything we do. To an astute researcher, such traces can provide data about human behavior that can be as clear and objective as the best archived data bases, the costliest surveys, or numbers obtained from precision laboratory equipment. In terms of the insight and

creativity required to obtain data, I consider traces to be the most challenging measurements of all.

*You don't seem to have published any trace research on your own.
Is that why? - The Editor*

Yup. - Nilsson

Trace measurements have two strong advantages over the other methods: 1) They are really unobtrusive. This eliminates any potential "reactivity" problems as discussed above with respect to direct observations. It protects the subjects' anonymity, which solves potential ethical problems. Furthermore, it may enable measurements of behavior that only occurs in private. Not just THAT kind of private behavior, but also of behaviors such as worship, criminal behavior, or caring about the environment at home. 2) Traces can enable one to study behaviors that lies occurs in locations or at times when it is difficult to observe them directly. Examples range from the origin of prehistoric migrations to the observance of fire precautions in remote park areas.

Psychologists also have a duty to be alert for the potential misuses of new technologies such as "nanolables", that could also be used to circumvent civil rights to anonymity.⁵

3.6.1 FINDING TRACES

The big challenge in trace research is seeing how a some type of trace can be observed and measured to provide information about a certain question. Put in the language of science, this amounts to recognizing how your dependent variable can be operationally defined in terms of a trace. Since there is such a diversity of questions studied by psychologists, no general guidelines are possible. The following list of examples should help you get the idea of what is involved in doing trace research.

ON CAMPUS

- bulletin boards messages -> student interests
- orderliness of seats after a class -> regimented thinking
- litter in classrooms -> student lack of courtesy
- cleanliness of blackboards -> teacher courtesy
- the left out in a reading room -> student interests
- coffee cups in lounges -> amount of studying
- lights on in dormitories _> amount of studying
- number of vehicles parked after class times -> library studying
- litter around duplicating machines -> preparation of term papers

IN THE COMMUNITY

- political campaign signs -> voter opinion.
- motor vehicle age year and model -> affluence
- number and state of public washrooms -> concern about hygiene

⁵ Nanolabels are microscopic, color-coded, inert chips that can be placed in wrappings, food, etc (Scientific American). I mention this as much to alert you about their potential for violating privacy as to make you aware of their research potential.

availability and usage of waste baskets -> municipal and public concern about neatness
 bumper stickers -> where people travel, political affiliation
 announcements of fund raisers in newspapers -> community concern for each other
 contents of "Letters to the Editor" -> social concerns of the day
 paths in the snow -> preferred routes and places of interest
 tool shape -> handedness in prehistoric times
 use of recycling bags -> environmental awareness

3.6.2 "STAGED" TRACES

Various methods can be contrived to make it easier for people to leave traces of their behavior. Radio programs use contests to obtain evidence of the size of their listening audience. Similarly, free offers added to messages of public concern presented by various media can provide indicators of public attitudes. Here are some examples of operational definitions of dependent variables that could be used to facilitate measurements.

leave a message pad in toilet stalls -> sexual concerns, profanity,
 the "lost letter" technique -> prejudice
 "Honk if you love Jesus" bumper sticker -> a measure of religious enthusiasm.

Thomas Nelson, head of Psychology at the University of Alberta for many years, had ingenious and novel research ideas. One day he asked for suggestions on measuring what people thought about color. In remote areas of Alberta and the Arctic, we had noticed the greater use of bright colors in clothing and homes. Perhaps this compensated for an environment that was relatively lacking in such colors compared to urban areas replete with color TV (those days) and advertisements. He proposed to ask school children to make crayon pictures of a recent dream and compare the use of color by children in remote communities with those by children in urban areas. "I'll mail a box of crayons and drawing paper for every child in the school with instructions for the teachers." When I pointed out the difficulties using spectrophotometry to measure which colors were used in hundreds of pictures of ambiguous content, Tom said he had simpler way. "They can keep the pictures. In exchange for a new box, I'll ask for the crayons they used to be mailed back. Then I'll weigh the crayons by color." The results showed that children in remote areas used significantly more bright colors like reds, yellows and violet than children in remote areas (Nelson, Allan & Nelson; 1971).

3.6.3 DIFFICULTIES

1. selective deposit
2. selective survival
3. biased retrieval
4. Validity - are you really measuring what you think you are measuring?

EXERCISES

1. OBSERVING INTERPERSONAL DISTANCES

The distance at which two persons communicate with each other may depend on: their: 1) relationship, 2) emotional state, 3) social status, 4) interpersonal attitudes, ..., and many more factors including environmental ones. While many of these factors can not be directly observed, the distance certainly can be measured. For this exercise you are to devise a research study that uses interpersonal distance as its dependent variable.⁶ For your study you could use an operational definition to let the distance represent one of these factors, or you could do a more applied study that looks at how crowding is affected by some factor.

You will need to come up with the following:

- a. A question that can be answered by measuring the distance between two people.
- b. An independent variable - some factor that may alter the distance that two persons maintain from each other. Identify its levels. Is an operational definition required?
- c. An operational definition of what your measure of interpersonal distance represents.
- d. Who are your subjects? Where will you find them? How many do you think you'll need to observe?
- e. Identify at least one variable that you should make an effort to either hold constant or balance to avoid having it confound the results.
- f. Identify at least one variable that is probably inconsequential.
- g. Make up at least a partial data table to illustrate how you think the data would look and how you would organize your measurements.
 - 1) First create some numbers that are plausible results of your measurements.
 - 2) Then organize the numbers in terms of the levels of the independent variable.

2. OBSERVATIONAL CHECKLIST

For a study on altruism, you are going to stage a person looking on the floor or ground for a dropped coin in a busy public place. You plan to be the "victim" and will have an assistant observe the behavior of bystanders.

- a. Think up an independent variable with two levels that could provide a reason for doing this study.
- b. Make a checklist for your assistant to record the observations.
- c. Write a set of instructions to tell your assistant which people to observe and how to use the checklist.

⁶This may seem a backwards way of doing research, but it is not that unusual. Having learned how to do something often leads scientists to think of new uses. See comments on the "rule of the hammer" at the beginning of Chapter 4.

3. TRACES ON DISPLAY

The bulletin boards at various places in a university reveal a lot about the scholarly, social, and other activities of its students and faculty. Go look at some bulletin boards around campus. Taking their location as an independent variable, look for some differences between various locations. Think about what those differences might imply about the different departments or the type of student that frequents those areas. Formalize these thoughts with some operational definitions, and you have the makings of a trace research exercise.

f

Chapter 4

ASKING QUESTIONS

Have you heard about the strange object that landed right in front of the Parliament buildings in Ottawa? An air force general was called in to identify it, but the general had never seen nor knew of any craft like it. Therefore it was declared potentially dangerous and surrounded by tanks and armed soldiers. An engineer was brought to examine it but could find no way to pry it open. A physicist tried to weigh it, but it proved too heavy for any available equipment. A chemist attempted to chip off a sample for analysis but could not scratch it with a diamond, dissolve it with acid, or even melt it with a laser.

They were standing around the object wondering what to do next when from the outside a person called, "Let me try."

"Who are you?", they asked. "What makes you think you can find out what this thing is when a general, an engineer, a physicist, and a chemist have failed?"

"I'm a psychologist." said the newcomer confidently. "I believe I may be able to help."

The general, the engineer, the physicist and the chemist all smirked, but not having any better idea, they agreed to let the psychologist try.

Empty handed, the psychologist walked calmly up to the strange object. The murmuring crowd hushed in anticipation. "What is your name?" said the psychologist in a friendly voice.

"Ralph," it replied.

Much can learned by observing how people or animals behave, but to answer certain kinds of questions such as those dealing with attitudes, plans, knowledge or past experience, might require lengthy observations before any assessment might be possible. Unfortunately, the species which is home to the greatest number of these unobservables, Homo sapiens, is also the species which is most difficult to observe for long periods of time. Fortunately, there is an alternative way to obtain such information from humans - ask them. That's basicly what survey research is about.

To those who don't know much about it, survey research may seem easy to do. Sad to say that is probably why there is so much poorly done research using this method. To begin to give you some idea of what is involved, we'll go through what is involved in doing a survey. I recommend that you get together with one or two fellow students to carry out a survey of your own design. The following sections are intended to help you do just that. But a word of caution: Do not think that this exercise will make you competent at doing survey research. That requires a least a full course in psychological testing, an opportunity to work with someone experienced at doing surveys, and a knowledge of the literature pertaining to the intended subject and how it has been studied previously.

4.1 SURVEY'S, POLLS, AND SUBJECTIVE MEASURES

4.1.1 WHAT THEY'RE GOOD FOR

Questionnaires are suitable for directly learning about what people think and feel. They also make it possible to study questions that are so complex that the variability demands measurements from many people in order to reveal significant trends. Their portability compared to most other research "instruments" makes it possible to study people directly at public events or at their home.

Questionnaires are not poor substitutes for dazzling technology when it comes to laboratory research. In the laboratory, they can be the **method of choice** when technology proves inadequate. (The Australian Government flew me around the world to tell them how a "simple" questionnaire succeeded in identifying fatigue during driving when enough technology to impress a rocket scientist had failed.) Neither are surveys cheap alternative methods of research in the face of cutbacks in research funding. Sure, that piece of paper with 14 questions may look like it could be dashed off in twenty minutes and cost only a few pennies. But consider what goes into making it a valid (it measures what you want to measure) and reliable (it gives consistent results) tool: Months of preliminary testing may be required so see if it works. Locating a large number of potential subjects from which to draw a random sample can be a true challenge in diplomacy involving community and government organizations. (For your study, an opportunistic sample will suffice to demonstrate some of the basics of dealing with subjects since publication is not intended.) People trained to do surveys do not come cheap. Then there are the travel costs to reach your subjects. (Surveys have been amongst the most expensive research projects I have conducted. It cost a \$270,000 to survey 500 Island seniors so that we could find 30 cases of probable Alzheimer's disease.) Finally, keeping track of the results of hundreds of questionnaires and maintaining subject anonymity requires data management techniques that are most efficiently done by computer.

Surveys can address questions that range from international issues, "Should the United Nations maintain a permanent armed force to police the world?", to local issues, "Where should a municipal recycling site be located?"; from social issues, "Should euthanasia be legal?", to interpersonal issues, "What are the characteristics of a good friend?"; from the theoretical, "Is there intelligent life on other planets?", to the practical, "Should there be a 40 minute midday break when no courses are scheduled at this university?" These are typical topics for surveys because the complexity of these questions generally puts them outside the domain of laboratory research. That complexity usually calls for a substantial number of subjects to ensure the **representativeness** of the people you study and to ensure a reasonable chance of **statistical significance** given the variability to be expected from a complex issue. Furthermore, there are so many factors, including uncontrollable factors from a person's life history influencing responses to these questions, that there is little to gain by making these measurements in the immediately controlled conditions of a laboratory.

The survey method need not be restricted to "attitudinal" types of research. With creativity and proper understanding, they may be employed to answer questions about human physiology, memory or cognition. If it were announced that the rare "green flash" phenomenon is likely to occur just after sunset today, a radio phone-in survey might be used to determine the viewing conditions under which the effect was most evident. The Canadian Study of Health & Aging combined many questions about the participants' life style with questions about their cognitive ability and health to obtain insight into the causes of Alzheimer's disease. Professor Cohen added a survey of musical preferences of those participants. She asked them about songs they preferred during various decades back to 1900. Using the subjects' age and data on what music was popular during each decade, she obtained new information on long-term memory and musical cognition. Delmar McLean used a combination of survey and subjective measurements to for an honours thesis on perceptions of causality. He found significant differences in how people of high and low religiosity perceived cause-effect relations in mechanical motion.

Another reason for doing questionnaire research is to study a problem has only been studied by other methods. The value arises from the very nature of scientific inquiry and the meaning of statistical significance. Never forget that statistically significant results can occur by chance 5% of the time. Furthermore, if a significant answer to a problem has been found using only one method of measurement, it is always possible that the method itself involves some yet unrecognized factor that confounds the real answer. There is a special delight when completely different methods of research produce similar results.

4.1.2 WHAT'S THE DIFFERENCE?

Polls ask only one or two questions and usually refer to a preference on one or two specific matters. They usually involve a limited set of possible answers such as YES or NO in response to a referendum, or the choice of a few alternatives such as candidates in an election. However, they can also involve open ended answers, such as a poll on the most outstanding athlete.

Surveys involve a more extensive set of questions. By asking several questions that pertain directly to an issue, surveys can overcome biases that may arise from different interpretations of a single question. They frequently include information about the participant as well as measuring their knowledge, attitudes, and/or preferences. For example, it may be of some interest to learn that 63% of the residents of your town believe that the mayor is doing a "good job". However, if one examines how that belief varies as a function of occupation and income level, one may be able to understand why some people feel that way and others do not. Correlation, multiple regression, and factor analysis techniques can measure the association between two or more answers to different survey questions. In that manner, survey questions are often used like independent variables as well as being dependent variables. Research designed in this way is usually referred to as "descriptive research".

Surveys are also used in experimental research designs. The principle difference being that the independent variable is changed by the researcher and is not part of the survey. For example, a survey could be used to measure fatigue and cognitive performance during simulated driving while the researcher changes an independent variable such noise level or vehicle speed. For an example of a survey used in a quasi-experimental design, consider a survey conducted both before a certain government policy is announced and again afterwards. This design would enable measuring the effect that policy had on public opinion.

Subjective Tests tend to be used primarily in experimental research and usually do not involve enough subjects to be considered representative of some general population. They are more focused than the others in that they measure specific experiences such as perceptual phenomena and emotional states - experiences which may be difficult to measure objectively. The observational measurements of color saturation mentioned in the previous chapter were essentially a single-question subjective test using a rating scale. Table 3.1 shows an example of a multi-item subjective test of human fatigue (Nilsson, Nelson & Carlson; 1997). We believe the success of this test arises from the variety of fatigue related terms that it employs. All the various ways of asking questions to be explained for surveys can be used for subjective measures, but rating scales predominate.

Table 3.1
Form AA of a Checklist to Measure Fatigue - Adapted from Pearson
(1957)

INSTRUCTIONS

We want to know how refreshed or tired you feel right now.

Below are 13 statements about how you may feel. While the questions are similar, they ask about different degrees and types of freshness or tiredness because people interpret words such as "very" or "slightly" in different ways.

Decide whether you feel better or worse than each statement. Indicate how much better or worst by the size of the number you circle.

Answer the questions in order. Work quickly - first impressions are usually the best. Please answer all questions.

WORSE						DO YOU FEEL BETTER OR WORSE THAN				
BETTER										
-5	-4	-3	-2	-1	slightly tired?	+1	+2	+3	+4	+5
-5	-4	-3	-2	-1	bursting with energy?	+1	+2	+3	+4	+5
-5	-4	-3	-2	-1	extremely pooped?	+1	+2	+3	+4	+5
-5	-4	-3	-2	-1	quite fresh?	+1	+2	+3	+4	+5
-5	-4	-3	-2	-1	slightly pooped?	+1	+2	+3	+4	+5
-5	-4	-3	-2	-1	extremely peppy?	+1	+2	+3	+4	+5
-5	-4	-3	-2	-1	somewhat fresh?	+1	+2	+3	+4	+5
-5	-4	-3	-2	-1	petered out?	+1	+2	+3	+4	+5
-5	-4	-3	-2	-1	very refreshed?	+1	+2	+3	+4	+5
-5	-4	-3	-2	-1	ready to drop?	+1	+2	+3	+4	+5
-5	-4	-3	-2	-1	fairly well pooped?	+1	+2	+3	+4	+5
-5	-4	-3	-2	-1	very lively?	+1	+2	+3	+4	+5
-5	-4	-3	-2	-1	very tired?	+1	+2	+3	+4	+5

This chapter will focus on surveys since they are the extensively used method of asking questions and they contain the elements of the other two. Readers may note a certain similarity between examples in the following section on how to ask questions and the checklists used to record observational measurements in the previous chapter. What is important about the different approaches to asking questions (polls,

surveys, subjective tests) as well as the different types of research design (descriptive, experimental, quasi-experimental) is not precise classification, but rather the awareness of the different research possibilities they represent. There will always be methods of measurement and research designs that do not clearly fit in one or another category. Familiarity with the terms simply helps communication.

4.2 WHAT TO ASK AND HOW TO ASK IT

What to ask depends on the survey's specific purpose. The subject matter is up to you. Generally, a question can concern the following:

1. matters of fact: "What is your age?" "Did the candidate vote for the Murphy bill?"
2. matters of opinion: "Which color do you like?" "Which candidate do you favor?
3. matters of attitude: "Do you feel tired?" "How confident are you that candidate XX will do a good job."

The question can refer to:

- a. the person being interviewed - eg. "What is your age?"
- b. other people, places, events - eg. "Did the candidate vote for the Murphy bill?"
- c. general concepts or principles - eg. "Does absence make the heart grow fonder?"

The question can be either:

- 1) direct - referring to the interviewee, certain persons events, or concepts, such as, "Which color do you like?"
- 2) indirect by using hypothetical scenario or story - eg. "The clerk offered Pat a red, yellow, green or blue ball. Which color do you think Pat chose?" This type of question may result in more objective answers because the answer does not refer to the interviewee but rather to certain other persons or events. Then why isn't it used more? Probably because it takes more effort to prepare suitable stories; it takes more time to administer. There is the possibility that the interviewee may misinterpret the story or give an answer that differs from the answer to a direct question, but that also happens with direct questions.

The options provided for a response to the questions on a survey can substantially influence what responses are obtained and how effectively the responses can be analyzed. Here are different ways of obtaining answers to questions.

4.2.1 OPEN-ENDED ANSWERS

The subject simply tells you whatever answer comes to mind. It may appear that these answers would be the most unbiased and provide the most complete information. However, that may not be so. Remember that the interviewee is given and/or willing to take only a certain amount of time to answer each question. Furthermore, the interviewee may not be fully prepared for the questions you present and have other things on his or her mind as well. Some of the answers that are possible might not occur to the interviewee just then. With no guidelines for scaling an answer and no framework for constructing an answer, the interviewee may provide spontaneous estimates that are inaccurate or wander off the topic in replying.

This is the simplest type of question to make, but the hardest to score. To obtain objective quantitative data from open-ended questions requires advanced and very time consuming techniques. See Professor Terry Percival for further information. Such questions are inappropriate for introductory research courses.

4.2.2. FORCED-CHOICE ANSWERS

1. Objective Yes or No:

Did you vote in the last election? YES NO

2. Objective Categories:

What is your occupation?	UNSKILLED LABORER SKILLED LABORER CLERICAL WORKER TRADESPERSON FARMER OR FISHERMAN ADMINISTRATOR PROFESSIONAL
--------------------------	---

3. Subjective Yes or No:

Should euthanasia be legal? YES NO

4. Subjective Yes, Uncertain, or No:

Should euthanasia be legal? YES UNCERTAIN NO

There are trade-offs between #3 and #4. For certain questions, persons may avoid making a decision and produce a lot of "uncertain" answers which may be useless. Sometimes you can get better measurements by forcing the person to make a decision. However, forcing persons to make difficult decisions may also lead to the subjects not taking your survey seriously or refusing to answer. (Remember, you must make it clear to subjects that they can refuse to answer any or all questions at any time.)

5. Multiple Choice:

Should euthanasia be legal?

DEFINITELY NOT PROBABLY NOT PERHAPS PROBABLY DEFINITELY

More alternatives can provide more precise information, but there are trade-offs. More alternatives also take more time. Minimizing your intrusion on the interviewee's time will lead to better answers. Showing respect for your subjects is not only the best policy in the long run - it is an ethical responsibility.

5. Thurstone's Technique: It uses many statements that represent different answers to a question. (Thurstone recommends 20 or more, but less are often used.) These answers are then rated for their degree of strength by a panel of judges, but these ratings are not shown on the questionnaire. The answers are listed in random order after the question. The subject chooses which answer is best.

What is your opinion about euthanasia?

Euthanasia may be appropriate in cases of terminal illness.

Euthanasia is murder - an affront to God and man.

Euthanasia should available with the consent of a lawyer and physician.

Euthanasia should be permitted for persons with incurable pain,

.....

.....

Euthanasia is an entirely private matter that is your right to decide.

6. Semantic Differential: This can be used when the multiple choice represent a continuous range of attitudes. Numbers are provided to help the subject decide on the relative strength of the different choices. The extreme choices are labelled. Intermediate choices may also be labelled. However, avoiding intermediate terms may provide more objective data since there is often more uncertainty about their meaning. the numbers are expected to be a sufficient guide where labels are not present.

A unipolar example:

ABSOLUTELY

MOST
DEFINITELY

Should euthanasia be legal?

NOT

0 1 2 3 4 5 6 7 8 9

A bipolar example:

STRONGLY

STRONGLY
APPROVE

AGAINST

UNCERTAIN

How do you feel about euthanasia? -4 -3 -2 -1 0 +1 +2 +3 +4

7. Rating Scales: These remove all labels (as possible sources of bias or misinterpretation) and rely entirely on the persons ability make r translate their judgements to a purely quantitative scale such as numbers from 0 to 10, or 1 to 9. Bipolar scale are possible but less frequently used.

How tired do you fell right now? 0 1 2 3 4 5 6 7 8 9 10

|---|---|---|---|---|---|---|---|

By rephrasing questions that relate to attitudes about similar things, you can force the subject to think more about his or her selection and also vary the selections to prevent monotony or anticipation of the answer. For example:

Estimate how fast you could run a hundred yards right now:

20 40 60 80 100 120 140 160 180 SECONDS
|---|---|---|---|---|---|---|---|

8. Likert Scale: A number of number of answers to a question are provide. the subject indicates the extend to which they agree (agree or disagree) with each one.

Indicate the extent to which you agree with the following statements about euthanasia:

disagree

agree

It should be available to terminally ill persons.

0 1 2 3 4 5 6 7 8 9

disagree

agree

74 - ASK

It is murder, and must be banned.

0 1 2 3 4 5 6 7 8 9

disagree agree

Make it available through a physician and lawyer.

0 1 2 3 4 5 6 7 8 9

disagree agree

Only available to persons with incurable pain.

0 1 2 3 4 5 6 7 8 9

Note that compared to the Thurstone techniques, the Likert scale provides data on each statement. Its total score for any set of questions can be expected to be robust: yield consistent overall results even though different persons may have some what different interpretations of particular statements. Likert scales can also be bipolar.

9. Ranking: Sometimes when you have several things or answers judged by multiple choice, semantic differential or rating, subjects have difficulty distinguishing them. When you end with most of your data being about "average", or mostly all "good" or "bad", you may be able to force persons to make finer discriminations by asking them to rank order the items in terms of preference or some other characteristic. A good way to do ranking is to provide the subject with the each item printed on a card. Spread the cards on a table and ask the subject to arrange them in order of "preference", "importance", or whatever criterion you want. Note that this only makes sense for things or answers that have something in common. Ranking has three disadvantages: It may take longer to do. The results can be misleading since rank steps do not discriminate a large difference in the interviewee's mind from a trivial difference or possibly no difference at all. Its results are "ordinal numbers", which force you to use less powerful "nonparametric" statistics.

10. Paired Comparisons. Powerful but time consuming for the subject and complex to analyze. Some prefer this method when fine distinctions need to be measured with subjects who are unfamiliar with research. Woodworth & Schlosberg (1954) explain how to summarize these type of measurements. For most circumstances, I believe that rating techniques used with more subjects are more powerful for the same total amount of intrusion and effort.

4.2.3 AVOIDING PROBLEMS

The exact wording of questions also influences how they are answered. Keep them as direct and simple as possible. Beware of phrasing which can bias the answers. Such phrasing sometimes occurs inadvertently because it is difficult to step out side your own thought patterns and habits of expression. The way to avoid that is to request comments from persons not immediately involved with the research.

Unfortunately, unscrupulous persons sometimes deliberately manipulate how a question is asked to increase the likelihood that the results will reflect their wishes. As persons with training in these matters, psychologist have a responsibility to make society aware of attempts to misrepresent of public opinion using improper survey measurements. Beware of the following.

a) **leading** questions: The wording gives the impression that most persons agree or disagree with the proposition - eg. "Do you agree that --- ?" (This wording was used in the recent Quebec referendum. A competent judiciary should have ruled it as unacceptable.) Simply asking, "Should ---?" would have been more objective.

b) **loaded questions:** Unnecessary adjectives and evocative words are used to create a certain impression about the desirability of what is asked - eg. "Should there be a law against the cruel murder of unborn children?" (This is a question about abortion.)

c) **questions with misleading information:** "Since a recent survey found that most female students had been sexually harassed, should violence against women ---?" (Consideration should be given to how "most" and "molested" are defined. "Most" could mean 51% but implies significantly more. "Harassed" could mean any unwanted comment or attention by a member of the opposite sex. Furthermore, the question attempts to equate "harassed" with "violence" by juxtaposition. Finally, the same survey also found that an almost equal percent of male students reported sexual harassment, but this information is withheld.)

d) **double barrelled questions:** Contain two propositions to which a single answer is expected - eg. "Should person who provide narcotics and alcohol to minors be ---?" (Many persons would regard narcotics and alcohol as separate issues. Giving a beer to an 18 year old farm hand is equated with selling cocaine to 12 year olds.)

e) **deliberately vague questions:** The question asks one thing but is sufficiently vague so that another interpretation can be made about the answer - eg. The plebescite on the "fixed link" actually sought permission to study its feasibility. It did not ask Islanders whether they wanted it built.

4.2.4 QUESTIONNAIRE FORMAT

Having some idea of how to create questions for a survey, the next task is to organize them into a form that is suitable for the subject to read themselves or for an interviewer to administer in a consistent manner. There is a bit more to a questionnaire than just a list of questions. After we have consider some other basics like who to survey and how to go about doing so, we'll look at a survey example in Section 4.5 as we consider what to do with the data. Meanwhile here is an outline of what a survey should include:

TITLE

It should be informative, accurate, and worded to capture the subject's interest. The challenge is to accomplish this in 16 words or less.

IDENTIFY WHO IS CONDUCTING THE SURVEY

Your name and institutional affiliation.

BRIEF STATEMENT OF SURVEY'S PURPOSE

This should explain what you are seeking learn from the survey data. That may include briefly stating some social problem or other phenomenon you are studying. For students, it also expected that the survey's connection with their studies be identified.

INSTRUCTIONS TO THE SUBJECT

Well worded questions usually require little explanation. Even elderly persons these days are sufficiently acquainted with rating scales, that these usually require little additional explanation. Always explain that they may refuse to answer any and there is no obligation to continue. Conclude with an acknowledgement of appreciation for their participation,

THE QUESTIONS

Only ask what you expect to analyze. Minimize "fishing expeditions" that waste the subjects time. Nilsson's Rule for Surveys: "A typical questionnaire produces about 7 + or - 2 really good responses, depending on the subject's interest. The more you ask beyond that, the lower the overall quality of the responses."

This rule does not include basic questions about the subject such as age, etc. They can go either at the beginning or end. At the beginning, they can help lead into more difficult questions. At the end they may interfere less with the attention devoted to other questions. As with some much of psychological research, rarely is there a simple solution - that's what makes it fun.

DEBRIEFING

Depending on how the survey is conducted, further information about the survey, an explanation of any deception and why deception was needed, answers to "frequently asked questions", how to learn about the results when the study is completed, and thanking the subject for their participation may be included in writing at the end of the survey or provided verbally. Even when provided verbally, the information in print can help the subject understand it and/or help the researcher complete the interview. Since the questionnaire is usually retained by the researcher, debriefing information could be printed separately for the subject to keep.

4.3 WHO TO ASK

4.3.1 POPULATIONS AND SAMPLES

In your study of book-carrying styles, you could not observe every one on campus. Therefore you observed just some students. These were a **sample**. All the students, who would have been suitable to observe were the intended **population** for that study. In surveys it usually is not possible to ask everyone to participate either. The different ways that you could have obtained your book-carrying sample are similar to the different ways you can obtain a sample for a survey. The main concern about your sample of just a few persons is that it be **representative** of all the people whom it would be suitable to include - the population. Only a representative sample allows you to claim that the results you obtain are generally true - not just specific to the particular people you studied.

Being representative does not mean that your sample has to be proportional to the population with respect to every possible characteristic. For example, in a survey about attitudes on abortion, hair color or handedness are probably not important to representativeness, since there is no reason to expect that these characteristics are related to the attitude. On the other hand, gender, age, and religious orientation are likely to be related to this attitude. Therefore it would be important for the sample's proportions of persons with those characteristics to be proportional to the population's. Knowing which characteristics are important to include in a sample and which are not requires general knowledge as well as specific knowledge about the area being studied. The alternative relies on **random sampling** to achieve representativeness, but this is more difficult than you might think,

4.3.2 Non-Random Sampling

Self-Selected Samples

You could put up signs and ask people either to come and participate in your or ask them to invite you to come and give them the survey. These kind of samples should be avoided if at all possible. People who want to participate in a survey may differ significantly from, those who don't. There fore, you would not know if your sample was representative of the population (all students).

Opportunity Samples

While not desirable, a lot of survey research is based on opportunity or accidental samples. It is probably the method used in your observational research on book-carrying. You study whoever happens to pass by. The problem you will face is convincing others that your sample was nevertheless representative enough to warrant generalizing the results to some larger group outside your sample. That representativeness depends critically on exactly how you selected your non-random sample. A satisfactory selection requires knowledge about the group that you seek to describe. You will need to know where most persons belonging to the group may be found and when they may be found there. You will need to know what persons in the group will not be found in those places at those times. Your report should than acknowledge what sort of persons you were unlikely to have sampled and what this implies for your conclusions.

Purposive Samples

Applied research often requires learning about the characteristics, knowledge and attitudes of the people in a certain factory or community, members of a certain organization or occupation, or persons with a certain social responsibility or status. One example is a survey that compared assembly line workers at one factory with those at another factory. In those situations it may be possible to test everyone concerned - the whole population of persons with the specific characteristics needed for a particular research purpose.

Snowball Sampling

Recently the PEI Centre for Health & Aging sought to learn about the concerns of elderly parents who were caring for developmentally retarded children - children now themselves becoming elderly. The Association for Community Living who provided the grant estimated that about 1% of Island families were in this situation. To use random sampling technique would have required successfully contacting some 4000 families to locate 40 families in this situation. The cost would have been prohibitive. Professor Weeks suggested we try snowball sampling. After some discussion, we decided it might work particularly well on PEI where community contacts are strong and considering that we would be addressing a population of families could be expected to be strongly motivated to participate.

With this technique you start with a small group of persons whom you want to survey (purposive sampling) and ask them to suggest other persons to participate in the survey. Obviously there is danger of such a sample being biased in terms of kinship and other social affiliations. However, we felt that if we contacted at least 85% of all such families, there would be little opportunity for bias. Though we suspect that it only works in a well organized society, snowball sampling may prove to be the method of choice for studying rare phenomena. Time will tell.

4.3.3 Random Sampling

Random Selection

This is harder than you might think. You can NOT simply pick people by what may seem to you to be random according to who walks in the door, raises their hands, or happens to be where you were looking. This is called "opportunistic" or "accidental" sampling. Though such non-systematic approaches may prevent you from deliberately biasing the sample, it by no means assures a random sample. Any number of factors beyond your immediate control can bias opportunistic samples. Who just "happens" to walk in the door may vary considerably in his or her characteristics depending on such factors as the time of day and where the door is located. Whether the door be to a library or shopping mall, for example, some persons may never enter due to their study habits or income level. Those that do enter at a certain time such as 10:00 AM on any weekday, may not be taking a full course load, or do not have a full time job. It is difficult to even estimate how opportunistic samples are biased, because you only know who is there - not who isn't.

The only way to achieve a true random sample begins with a list of all possible subjects. For example, such lists could be of all the students enrolled in an introductory psychology course who have volunteered to participate in experiments either out of interest or for some grade credit to acknowledge the experience, or it could be of everyone over 65 years of age in a certain county, or of all persons whose names begin with "M" listed in the telephone directory. The source will depend on the purpose. You then use a random number table or some equivalent computational method to select your subjects from that list.

Bear in mind that random selection from a list only ensures that your sample from that list is unbiased. It can not, of course ensure that the list is an unbiased sample of some larger group. Volunteers from a particular introductory psychology course may not be representative all introductory psychology students depending on such factors as the time the course is offered. Due to all the Mac's on Prince Edward Island, surnames that begin with "M" are more likely to over represent persons of Scottish background than would be case with some other letter. Also, one might find an under representation of single females who live alone, due to some tendency for such persons to have unlisted phone numbers. Nevertheless, selection from a list is greatly preferred over opportunistic selection because the source is identified and you can consider how that list might not be representative of some larger group.

More Complex Random Techniques:

Say that you want to ensure that the results of your survey can be generalized to both males and females. A small sample size of some 30 persons selected randomly from a list might by chance happen to contain substantially more females than males. Indeed, this would be likely to occur regardless of sample size if the list had more females than males. What should you do if you had any reason to suspect that gender might affect the results? Having become aware of the problem, you can probably guess - "randomly select an equal number of males and females." While this is the correct idea, exactly how to accomplish it may not be obvious. You could keep selecting random samples until you happened to get one that came out even, but there is a better way. A better way is to divide your list into male and female sublists and randomly select an equal number from each. This is called **stratified random sampling**. When several different characteristics of the persons on your list may produce different effects, you can divide or stratify your list into several layers of sublists. For example, if the influence of both gender and age was

of concern, you could have the following sublists: females under 19, females 19-23, females 24 and older; males under 19, males 19-23, males 24 and older.

The following methods gets us into some advanced research techniques. They are included here to make you aware of further possibilities.

Say you want your sample to include persons with relatively rare characteristics such as Alzheimer's disease, or certain physical disabilities. A randomly selected sample is unlikely to include any such persons. What should you do if there are not enough such persons to make an equally stratified sample possible? Deliberately include them in your otherwise random sample. Calculating the proportion of these persons to other persons in your sample enables you later to "correct" the overall results to take into account the bias you introduced. This is called **weighted stratified sampling**. Weighted sampling also makes it possible to include more persons with certain uncommon characteristics to help you deal with greater variability in their responses. Once again, because you controlled the proportions during selection, you can "correct" for these proportions in the final analysis.

Some groups you might want to survey are so large that a complete list does not exist, or it would be impractical to test a random sample from such a list. The Canadian Study of Health & Aging faced this problem in seeking a national survey of persons over 65. The solution was to establish a number of regional centers each of which selected a random sample from a list of seniors in their immediate vicinity. The overall approach is called **cluster sampling**. Even though there is a random selection within each "cluster", this in itself does not make the overall sample representative.¹ The overall representativeness of cluster sampling depends on how the clusters are established. Locating the clusters at random might seem an ideal way to do so, but in psychological research this is usually not practical - many places are not suitable sites for conducting research. Deliberate selection of cluster sites requires attention to the characteristics of the persons available at each site to achieve a representativeness of the whole study. The Canadian Study of Health & Aging in a sense stratified Canada into different regions based on both population and provinces. Additional attention was paid to selecting persons from both urban and rural environments, since most research centers were located in urban areas.

4.4 HOW TO CONDUCT SURVEYS

4.4.1 REACTIONS TO SURVEYS

The previous chapter pointed out some natural reactions that people may have when they realize that they are being observed - eg. wanting to be helpful, concern about looking good, etc. Such reactions can be magnified in the answers given during a survey for several reasons. Compared to being observed at a distance, interviews are: 1) MORE PERSONAL - therefore a person may feel more obliged to give answers they think are expected or believe could be remembered; 2) LESS SPONTANEOUS - the interviewee can consider his/her answer while hearing or reading the question. 3) VERBATIM - exact records are more intrusive. Researchers can do little to avoid these effects except to be courteous, professional, and protect the general reputation of survey research by scrupulously protecting the anonymity and confidentiality of the data.

¹ In this respect, cluster sampling is similar to a random selection from an opportunistic sample. Random selection from a biased group can not correct for the bias in that group.

Researchers should be aware that certain questions may be considered inappropriate for surveys. These include topics of a very personal nature such sex and grief. Questions concerning potentially criminal behavior not only should meet refusal, but could put you and the person at legal risk. Avoid questions that may press for answers which are more specific than the subject's opinions. Also avoid questions whose choice of answers does not include an answer that the person finds appropriate. Besides relying on your own careful scrutiny and experience, there are three additional steps you can take: 1) Examine the research literature for examples of questionnaires that have been used for purposes similar to your research. 2) Ask colleagues for their opinions on your questions. 3) Test the survey on a few members of the public or the group you intend to survey.

Always bear in mind that the person being surveyed has the right to refuse to answer any question and to end participation whenever they chose. Furthermore, inappropriate or poorly worded questions can result in your survey not being taken seriously and lead to spurious responses which could be worse than no responses.

4.4.2 DIRECT INTERVIEWS

For most purposes this is the preferred way to do a survey. Go to the selected persons home, place of work, or, particularly for opportunity sampling, accost the person in the street or other public place. Introduce yourself, including your professional affiliation, and request to speak. Quickly explain that you are conducting a survey and ask if they would participate - please. If they ask for further information about the survey, either explain verbally (have this rehearsed) or offer to show them your questionnaire. If they agree proceed with the survey. If not, apologize for the interruption and move away.

When possible, it is desirable to make an initial contact by mail to explain that they will be asked to participate in survey, the survey's purpose, and provide some information about the type of questions and time required. Announcing the survey in the newspaper helps prospective subjects to be aware of the significance of the research. Telephone calls can be used to schedule interviews, but mail with a letterhead, return address and signature is preferable for the initial contact because it clearly indicates who you are. These extra steps help ensure high participation and reduce reactivity.

Surveys may be conducted orally with the interviewer asking the questions and writing the answers. Surveys may also be conducted by handing the interviewee a questionnaire and asking them to read and write their answers. (A clipboard and a black pencil with good eraser are recommended.) A written survey often takes less time, but may not be suitable for elderly persons and is less personal than an oral one. Whether oral or written, the survey should begin with a brief explanation of its purpose. (Be careful that the explanation does not bias how the interviewee responds.) You must tell the person that they may refuse to answer any or all of the questions.

For most surveys, it is expected that you keep the data anonymous to protect privacy. Then you can explain that his or her responses will be kept anonymous. This can be done by identifying the subjects on the data sheet only by code number. Also let the subject see evidence of this anonymity by either putting the completed question in a large envelop yourself or hand the envelop to the persons. (Keeping at least a few questionnaires in the envelop helps indicate your intent to honor the anonymity of their answers.)

Don't forget to thank the person for her or his cooperation and then move away so as not to intrude further. Politeness and a professional manner the key ingredients beyond

the questionnaire itself. Avoid chatting with the person to minimize your intrusion and especially before the survey is completed to prevent biasing their responses. Remember the task at hand is to conduct a survey - not to socialize.

The direct interview is the method of choice for conducting surveys because of its high success rate and the variety of questions it can use. However, it is also the most expensive way. It takes time to talk to each person; it can be costly to travel to reach your subjects. The average cost was about \$85 per interview for the Canadian Study on Health and Aging, which involved an initial telephone contact to schedule the interview, travel to the subject's residence within a 30 kilometer distance, and an oral interview that lasted about an hour. However, direct interviews can result in 85% success rates. This is much higher than any of the other methods. High success rates are important because they reduce the chances of biased results. Remember, you have no control over who refuses to participate in a survey. If those who refuse have certain characteristics in common, that reduces your ability to generalize from your sample to the intended population.

4.4.3 TELEPHONE INTERVIEWS

For two reasons, even greater attention must be paid to the questions asked over the telephone than in direct interviews. 1) Your access time to subject is generally shorter since there is less social facilitation to maintain the interaction. Figure on 10 minutes as the maximum acceptable duration. 2) They must be clearly readable since there are no visual cues from print or lip movements to supplement the auditory information.

Other than the questionnaire, all you need is a telephone and time. Numbers can be selected from phone book using a random number table to count down to the next subject. Random dialling within an exchange is more problematic as it won't discriminate homes from businesses, offices, etc. Some geographical selection is often possible based on street address or the initial three-digits of the phone number. After briefly explaining your purpose, it is also possible to select your subjects by asking if you could speak to someone with particular characteristics such as age, gender, occupation, etc.

With telephone interviews you can expect success rate of up to 70% are possible. Good "telephone manners help. These you can acquire by experience. Success can often be improved by sending a letter of introduction to arrive a few days before phoning the subjects. A letter head helps establish your identity as a proper researcher. The content of the letter is your opportunity to explain the purpose of the interview, and why it is important. When the spontaneity of the answers is not critical, the mailing could include the questions to be asked over the phone. This enables the subjects to consider their answers instead of giving the first reply that comes to mind.

Private companies charge about \$10 for each completed, local interview, lasting no more than 10 minutes. They are also willing do some selection of the subjects for a modest additional cost per call.

ADVANTAGES - These are the fastest way to conduct surveys. Compared to direct interviews, they are relatively inexpensive because you save transportation costs and travel time. Some subjects appreciate the minimal time and effort of a phone interview compared to being directly interviewed or reading, filling out and posting a mailed questionnaire.

DISADVANTAGES - The anonymity of the interviewer, lack of the social reward provided by face-to-face contact, the dislike that some persons have for talking to strangers on the

telephone, the difficulties some persons have in using telephones - all result in higher refusal rates than direct interviews. Even when successful, answers may lack objectivity because the person being interviewed has no direct evidence that his or her answers will be kept anonymous.

4.4.4 MAIL SURVEYS

This is the cheapest way to do a survey. Unfortunately it also has the lowest success rate - 20% may be good for a public survey of this type. With most of the sample refusing for reasons beyond your control, the obtained answers can hardly be considered representative of the intended population. Only when the subjects have personal or professional interest in the topic are you likely to get better than 50% success. Prepaid postage and preaddressed response envelopes are an absolute necessity. Research on conducting mail surveys suggest a few techniques that have been found to improve the response rate (Dillman, 1991). These include:

- 1) A pre-questionnaire letter of introduction explaining the study and asking for cooperation.
- 2) Arranging the questions to capture the subjects interest at the beginning.
- 3) Providing a financial reward. A \$5.00 reward accompanying the questionnaire was more effective than the promise of \$10 on receipt of a reply.
- 4) Personalizing the envelope and using real stamps had a modest effect.
- 5) Following up the initial request with a post card reminder one week later.
- 6) Sending a second request and questionnaire if there has been no reply after four weeks.
- 7) Sending a third request and questionnaire if there has been no reply after seven weeks.

Any one of these generally improved the response rate by a few percent. Taken together, success rates of up to 70% have been obtained.

Personal experience with a mail survey that announced a forthcoming telephone call to answer questions and gave the recipient the option of completing the survey over the phone resulted in nearly an 80% success rate. The majority chose to mail back their survey.

4.4.5 USING THE INTERNET

The potential for conducting surveys by e-mail is fantastic. Nothing could be faster or cheaper! Some research using internet surveys has already been published. However this format is not without its problems.

About 30% of Canadians now have access through home computers or their work place. They will continue to represent a special population for many years to come. Addresses of persons belonging to certain groups may be obtainable from bulletin boards, interest groups, professional networks. As yet there appears to be no general way to identify the addresses of such organizations. Known members of an interest group or professional group should be able to obtain a reasonable response rate to surveys of interest to that group, but I have seen only a few figures on their success rates.

Rules and etiquette for doing surveys on the internet for surveys remain to be established. While personal web "pages" could be used to solicit responses to surveys, this would produce a sample that was highly self-selected. In general, the transient nature of

e-mail can be expected to render this format more vulnerable to self selection than surveys using the postal system.

If developing a questionnaire for the internet, bear in mind the variety of e-mail systems and word processors in use. Use an older word processor that does require the latest software to be recognized. Better yet, prepare the question directly in the e-mail message to avoid the delay of downloading. These steps may limit the accompanying graphics and other "bells and whistles", but in most cases would be more convenient for the recipient. Limiting the number of lines in the survey to what can be seen on a monitor without scrolling would seem likely to encourage a reply.

4.5 A SURVEY EXAMPLE

Here is a short survey about study habits of university students. Some explanation of the questions is provided below. We'll then consider how to get numbers from the questionnaire to enable a systematic summary and analysis. Hypothetical results obtained from 20 males and 20 females are summarized in Table 4.5-1. These will be the basis of some representative statistical analysis.

A Survey of Study Habits and Factors that Influence Their Effectiveness

Adrian McGinty and Kim Ralston
 Department of Psychology
 University of Prince Edward Island

This survey seeks to learn how study habits are associated with academic performance. It is conducted as an exercise for a Psychology 271 - an introductory course on how to do research. We respect your right to privacy will keep the answers anonymous. Don't feel obligated to answer questions you prefer to skip, and you may of course discontinue at any time. However, we would appreciate your taking a few minutes to consider the following questions.

1. Do you usually study a) alone,
 b) with another person,
 c) or with a group?
2. On the average, about how many hours do you study per week? _____
3. Do you have job during the semester? YES NO
4. Are you a member of a sports team or other extra-curricular activity that practices or meets regularly on a weekly basis? YES NO
5. How many hours do you work and/or are you involved in organized extra-curricular activities during a typical week? _____
6. If you listen to music while studying, please indicate which type:
 a) none - I usually don't listen to music while studying.
 b) rock or modern
 c) folk or country
 d) classical

84 - ASK

7. How important would you consider having the library open 18 hours a day, seven days a week?	NOT IMPORTANT	VERY IMPORTANT
	0 ----- 1 ----- 2 ----- 3 ----- 4 ----- 5	
8. What was your average grade last semester?	<input type="checkbox"/> above 85 <input type="checkbox"/> 75-85 <input type="checkbox"/> 65 to 75 <input type="checkbox"/> below 65	

This concludes the survey. If you have questions about it, we'll try to answer them. The results will be posted in the Psychology Department entrance hall after the winter break. Thank you for participating.

4.5.1 THE QUESTIONS EXPLAINED

Many factors could influence how effectively people study. This survey addresses a social aspect, the effects of working or extra-curricular activity that would compete for time available to study, listening to music, and whether the library is used. These could all be considered independent variables that affect studying. The principle dependent variable is the effectiveness of the studying and is operationally defined in terms of the average grade in the previous semester. The number of hours studied could also be considered a dependent variable in terms of whether the person also works or is active in extracurricular activities. Yet studying time could also be an independent variable in its own right affecting dependent variable of average grade. This independence can be assessed by comparing its relation to average grade, its relation to time devoted to other activities, and to the relation between grades and time on other activities.

The survey demonstrates the use of several question formats: Yes-no is used in questions 3 and 4 about whether the subject has a job or is involved in extracurricular activities. Categorical multiple choice for social context question 1 and music question 6. Numerical reports are requested for hours studying or doing other activities. Question 8 could have used this format too for the average. Instead four numerical ranges were used. Why?

It was recognized that asking about grades could be construed as somewhat personal. The use of ranges make this question less intrusive. It is often important in surveys to obtain information about income because it bears on some many personal, health, and social factors. Since most persons regard this as very private matter, a category approach is highly recommended over a direct numerical answer. Keeping the number of categories as small as possible minimizes the intrusion and decreases the likelihood of refusal. For most purposes information about whether the person has a below average, average or above average income will suffice.

Question 7 on the library uses a rating scale. It also illustrates an indirect way of reducing a social desirability effect that might affect the answer to directly asking if the subject used library materials to support studying. Presumably students who rate access as more important do so because they use the library more than those who rate access less important.

4.5.2 GETTING NUMBERS FROM QUESTIONS

With rating scales like question 7 on the importance of library hours, with Likert scales, the Thurstone technique, and of course with direct numerical estimates of time, distance, etc. you get numbers you can use directly. When semantic differential answers do not include a numerical scale you can substitute numbers for the various degrees of strength implied by the answer indicated - either unipolar: 0 to 4, 5, ..., 9 or however many are needed; or bipolar: -5, -4, ..., 0, ..., +4, +5 or whatever range is appropriate. When answers such as those for question 8 about the subject's grades involve a choice of numerical ranges (above 85, 76-85, 65-75, below 65), substituting a mid-range number (90, 80, 70, 60) will usually prove satisfactory for a concluding verbal description.

Subjective "yes", "uncertain", or "no" answers can be converted into +1, 0, and -1 respectively. Subjective "yes" or "no" questions may be converted to +1 for "yes" while a "no" could be either a "0", or "-1" depending on the question. For a question like, "Do you have a job during the semester?" a "YES" here clearly merits +1. A "NO" simply means the person does not have a job, so a zero is appropriate. However for a question like "Should euthanasia be legal?" where there are strong opinions may predominate, "NO" probably indicates opposition and would be appropriately represented as "-1".

When the answer refers to categories, it usually makes no sense to convert the category answers into numbers. Instead make a list for each category and enter a "1" under that category each time a person selects that category as the answer. Use a different row for each person and you will be able to tell which answers to one question were accompanied by which answer on other questions. Counting the number of times each category was selected will give you numbers.

4.5.3 SUMMARIZING THE DATA

A spreadsheet program is the most convenient way to work survey data. Table 4.5-1 is set up like a spreadsheet. Each possible answer has its own column. Each subject is assigned a different row. In that row, her/his responses are entered in the appropriate column for each question. Descriptive summaries and statistical tests can then be done right on the spread sheet without ever having to type those numbers again.

Once you have the all responses organized, you can examine them and look for trends or relationships. Where the answers are categorical as in question 1 whether they study alone or with others and question 6 about music, add up the number of responses in each category. With a QuattroPro spreadsheet, this would done using the @SUM() function. Questions that involve ratings or numerical information can be summarized in terms of their means, @AVG(), for each level of the independent variable and their standard errors, $\text{@STD}(\text{ })/\sqrt{n - 1}$. Graphs of these summaries can help you interpret the results. (Yes, spreadsheets also do graphs.)

Table 4.5-1.
A Spreadsheet Summary of Answers to Questions on the Study-Habits Survey

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	SUB #	SEX	QUESTION												
2			1*	2	3**	4**	5	6*	7	8***					
3	ALONE	2	3 OR MORE	STUDY HOURS	JOB	EXTRA-CURRIC	WORK & E-C HRS	NO MUSIC	ROCK MODERN	FOLK WEST	CLASS	LIBRARY			Avg Grade
4	1	M	1		35	0	0	0		1	1	5	90		
5	4	M	1		22	1	0	24	1			1	70		
6	5	M	1		21	0	1	15		1		2	60		
...															
23	38	M	1		24	1	1	12		1		4	90		
24	SUM		13	5	2		9	8		3	6	11	5		
25	MEAN				26.3			22					2.9	74	
26	SE				1.3			3.2					0.4	1.6	
27															
28	2	F		1	29	1	1	32		1	2	60			
29	3	F	1		27	0	0	0		1		4	60		
30	6	F	1		34	1	0	16	1			1	80		
47	40	F		1	22	0	1	12		1		3	80		
48	SUM		8	6	6		7	11		7	6	5	4		
49	MEAN				27.2			19					2.2	75	
50	SE				1.5			2.0					0.3	1.5	

Notes: * selection = 1; ** indicates that 1 = YES, 0 = NO; *** indicates the mid-range value.

4.5.4 ANALYZE THE DATA

For categorical questions, the category sums can be analyzed using Chi Square tests to determine if the distribution of choices differs from chance. A simple Chi-Square test of the data on whether students study alone or with others indicates that significantly more students studying alone ($R = 2.1$). Chi Square calculations can also tell you if number of responses in each category differs significantly for various levels of the independent variable. The organization indicates that more males study alone while more females study with two or more persons. However, a Chi Square contingency test does not indicate that this is significant. If the above trend continues, trial-and-error calculations reveal that some 140 students of each gender would need to be tested to demonstrate this effect as significant.

For questions having a numerical response for each person, differences between the averages can be tested for significance using t-tests or analysis of variance with a multiple comparison test such as Tukey's or Duncan's tests. It makes no sense to determine whether the mean hours studying differs significantly from the mean hours at work and extra-curricular activities. The only independent variable evident in the present organization is gender. Visual comparison of the mean hours of study, hours of work, or average grades for males and females indicates no significant differences there. Males do rate the importance of more library hours higher than females by a difference that about equals their standard errors, so a t-test was tried here. Alas, 't' only equalled 1.4. However, if twice as many students had produced this result, the difference would be significant.

A more interesting comparison with the present organization examines the association between hours of study, hours of work, and average grades. This done by calculating correlation coefficients. In QuattroPro that involves highlighting two sets of numbers on the spreadsheet and using the "Regression" TOOL. With 38 degrees of freedom, 'r' must exceed .31. For these hypothetical data, the correlation between hours studying and hours at work or in sports and other extra-curricular activity was -.28. Time spent doing other activities does reduce study time significantly. Had 50 students been tested, this correlation would be significant. However, it is correlation squared which represents the amount of variability accounted for by an association. Even a significant 'r' of -.28 accounts for less than 8%. This indicates there are other, probably more important, factors which determine study time. The correlation between hours studying and grades was +.64 - clearly a significant association that should be born in mind by all aspiring psychology students!

*Hold on there Nilsson. These are 'hypothetical' data.
You just made them up that way. - the Editor*

*And just how else would you expect me to make them up?
Tell you what. Let's have the students do a real survey. I'll bet you
my royalties on the next edition that the data confirm these 'hypothetical'
results, if you'll bet a full-color printing of the next edition. - Nilsson*

Got nothing to say? Aha! - Nilsson

To see what these data can tell us about the effects of certain study habits such as studying alone or listening music, they need to be reorganized. To determine whether listening to music effects grades, the data would be sorted into four groups based on the answers to question six. With a spreadsheet program this readily done with a 'Sort' TOOL. Once sorted, you can calculate separate mean grades for each type of listener, then do an ANOVA followed by Tukey's or Duncan's test on the means. Since ANOVA's with unequal amounts of data for each level are an advanced topic, I'll demonstrate a n effect where a t-test is appropriate. Let's look further at the relationship between extra-curricular activities and grades. Table 4.5-2 shows what the data look like when sorted into two groups based on the answer to question 4.

88 - ASK

Table 4.5-2.
Data from the Study-Habits Survey Sorted in Terms of Extra-Curricular Participation

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	SUB #	SEX	QUESTION												
2			1*			2	3**	4**	5	6*			7	8***	
3			ALONE	2	3 OR MORE	STUDY HOURS	JOB	EXTRA- CURRIC	WORK & E-C HRS	NO MUSIC	ROCK MODERN	FOLK WEST	CLASS	LIBRARY	AVG GRADE
4	2	F			1	29	1	1	32				1	2	60
5	5	M	1			21	0	1	15		1			2	60
6	7	M		1		26	0	1	24	1				4	80

22	40	F			1	22	0	1	12		1			3	80
23	SUM							19					1		
24	MEAN														77
25	SE														1.7
26	---														
27	1	M	1		1	35	0	0	0			1	1	5	90
28	3	F	1			27	0	0	0			1		4	60
29	4	M			1	22	1	0	24	1				1	70

47	39	F		1		16	1	0	18		1			3	60
48	SUM							21							
49	MEAN														72
50	SE														1.4

Notes: * selection = 1; ** indicates that 1 = YES, 0 = NO; *** indicates the mid-range value.

A t-test of the two means indicates a significantly higher average grade for those who participate in extra-curricular activities compared to those who do not participate ($t = 2.27$, $p <.05$ at 18 and 20 d.f.). (Note the proper format for reporting a statistical test in a report.) Before reaching a conclusion based on this analysis, look over the questionnaire and the data again. There is a confounding effect.

Some students who participated in extracurricular activities also participated in sports. Consequently, the higher mean grade could be due to sports and not other extracurricular activity. To find out the extra-curricular effect alone, we should compare students who only participated in the extra-curricular with those who didn't. For the above data, this could be done by simply deleting those students in both. However, that could get tedious and is prone

to error with hundreds of subjects. Here is a simple spreadsheet calculation that can do it efficiently: Add a new column between G and H. In the new column's row for the first subject subtract the response to question 3 from question 4. A YES will result in a zero. Copy this calculation to the other subjects in the YES portion of the spreadsheet. Now sort this portion based on the new column. Now it is easy to find a mean and standard error for the separated subjects who were only in sports or other extra-curricular activities.

Numerous additional analyses could be done on these data. Some examples include: 1) What is the relationship to grades of work and extra-curricular time? 2) Is there an optimal amount of extra-curricular or work activity? Insight into this complex relationship might be gained by graphing grades as a function of time spent not studying. 3) Do different music listening habits result in different grades?

Would you have thought so much potential and work could arise from this apparently simple, little 8-question survey? As one proceeds, ideas develop for alternate or additional questions, for more subjects - perhaps with respect to particular characteristics such year of studies or their major. No, that piece of paper with some questions doesn't look like rocket science, but making and using it well may be even more challenging and may contribute more to society.

4.6 STRUCTURED INTERVIEWS

These are like a survey done by interview, except the subject is not limited to a certain selection of answers. This enables the questions to be more open since their semantics do not have to be directed towards the specific choices provided. The respondent has opportunity to answer in their own words, which can convey different information and shades of meaning than can provided by any set of pre-defined answers. The answers can also be longer and thereby provide greater detail. Perhaps best of all is the opportunity to request further explanations or ask additional questions depending on the subject's responses.

Compared to an open discussion, the structure provided by an agenda is important to keeping the interview directed toward the obtaining the information that is needed for the research. Structure enables the researcher to describe the method by which the results were obtained. It enables the research method to be tested by others. Agreement between the researcher and subject on the content and sequence of questions reduces reactivity. The process provides opportunity for the researcher to explain the agenda of the interview, modify that agenda depending on the subject's reactions, and ask the respondent to suggest related questions that had not occurred to the researcher. This in turn reduce reactivity and it lets the subject know the amount of time and effort that will be required. Though structured interviews require more time and effort than surveys, that may be more acceptable to the respondent since they also share control of the process. As with surveys or any other research participation, it must be made clear to the interview that she or he may stop the interview at any time and for whatever reason. If the length of the interview exceeds 1 1/2 hours, be sensitive to the subject's possible tiring and be prepared to suggest a break or continuing another day. Not only does this ensure the quality of the data you obtain, it is this your ethical duty.

4.6.1 HOW TO PREPARE

An interview is structured by the researcher preparing a set of questions before hand. The questions are organized into a sequence or "script" that the researcher believes will keep reactivity minimal and provide the most objective data. Based on prior experience the

research may anticipate various alternative answers and prepare certain follow-up questions to ensure that specific information is not missed. The script can also provide guidelines on subsequent questions that can be skipped as no longer relevant depending on certain answers. Saving time and avoiding unnecessary questions also reduces reactivity.

Exactly what questions to use and the order in which to ask them depends so greatly on the purpose of the interview that only some rough guidelines are appropriate at this level of your studies:

1. Prepare yourself by becoming familiar with the topic and how it has been studied by other researchers.
2. Study published reports that have used structured interviews for similar research.
3. Outline the objectives of your research and the particular kinds of information needed from the interview to meet those objectives.
4. Keep the questions as specific as possible. Consider how to phrase questions so that the interviewer's guide can list a number of alternative answers that could simply be checked off. Consider the recommendations provided in Chapter 3 - Surveys.
5. Avoid open-ended questions on broad or complex issues by approaching the issue in steps.
6. Attend to the ethical guidelines outlined in Appendix C.
7. Discuss the interview method with an experienced colleague.
8. Test the structured interview with a colleague or friend serving as the subject. This will provide insight on how others interpret your questions, the sorts of answers you can expect and where follow-up questions may needed. It will also give you an idea of how long the interview will take and whether editing to reduce the demands on the subject is desirable.
9. Learn as much as you can about your subject before the interview. this can help to establish rapport and avoid unnecessary questions.

4.6.2 AN EXAMPLE

The structured interview technique is used in clinical practice as well as in many fields of research. In a clinical setting, data gathered from clients can be compared with normative data to aid in assessment and treatment planning. In a research setting, the structure interview is particularly valuable in exploratory research when the researcher hasn't yet identified all the important variables. In both settings, the structured interview promotes interaction with the respondent, allowing clarification of questions the respondent didn't understand as well as follow-up on ambiguous or interesting responses. Administration of a structured interview requires correspondingly more skill than does the survey technique.

The following agenda for a structured interview was prepared by Philip Smith, who does professional counselling, as an exercise for students to learn about this research method.

STRUCTURED INTERVIEW EXERCISE ON STUDENT CONCERNS

This example provides the opportunity to experience what it is like to conduct a structured interview. You will derive the most benefit by finding a volunteer respondent who is not well known to you. The interview is about potential matters of concern or satisfaction to university students. The interview begins with collection of demographic data and then moves into exploration of matters that might be sources of worry or satisfaction to the respondent. The respondent is then probed to volunteer an additional source of worry to students. Respondents who indicate that some of the matters noted are sources of concern

are then probed about who (if anyone) they have talked to about their concern(s). Before doing the interview, you should be thoroughly acquainted with the sequence of questions in the outline presented below and how to check off the answers.

The interview begins with an explanation of its purpose: to identify what things are sources of concern and of satisfaction to university students and also to give you experience doing a structured interview. Also explain to the respondent that their names or other identifying information will not be recorded and that their answers will be kept in strict confidence. To break this confidence would be a serious breech of ethics!

AGENDA

(gender: male; female)

1. "I need to gather some basic information about the people in participate in the interviews, although I will not be recording your name or any other identifying information. If you prefer not to answer any question, you can pass on it, and you are of course free to end the interview at any time. I will keep in confidence the source of all information given to me.

"First of all, what year of study are you in? (1st, 2nd, 3rd, 4th, other)

2. "Are you a ___full-time or ___part-time student?

3. "In what faculty are you registered? (___arts, ___science, ___education, ___business,
___engineering, ___veterinary medicine)

4. "What is your age? _____

"We're interested in learning students' opinions about a number of areas. Please tell me how you feel about each of these matters by choosing one of the answers from this card."

Hand respondent the answer card (shown below).

5. "First of all, your academic performance, is it a source of great worry, some worry, neither a source of worry nor satisfaction, some satisfaction, or a source of great satisfaction?

6. "About your personal finances, are they are a source of worry or satisfaction? (great worry, some worry, neither a source of worry nor satisfaction, some satisfaction, great satisfaction)

7. "When you think about your employment prospects, is that source of worry or satisfaction? (great worry, some worry, neither a source of worry nor satisfaction, some satisfaction, great satisfaction)

92 - ASK

8. "In general would you say that your relationship with your peers is a source of worry or satisfaction? (great worry, some worry, neither a source of worry nor satisfaction, some satisfaction, great satisfaction)
9. "In general, would you say that your relationship with your family is a source of worry or satisfaction? (great worry, some worry, neither a source of worry nor satisfaction, some satisfaction, great satisfaction)
10. "About your relationship with professors, In general is that a source of worry or satisfaction? (great worry, some worry, neither a source of worry nor satisfaction, some satisfaction, great satisfaction)
11. "I've asked about a number of possible sources of worry or satisfaction for students, Do any other likely sources of worry occur to you?"
(Record only the first identified source.) _____

-
12. "Would you say that for you personally, (the first other source of worry reported by the respondent) is a source of great worry, some worry, neither a source of worry nor satisfaction, some satisfaction, or a source of great satisfaction?

(possible questions 13 -19: PART A)

Ask these questions if the respondent has indicated that the specified issue is a source of great worry, or of worry for him or her. For each 'yes' response, move directly to part B (below) before returning to the next question in part A.

"You mentioned earlier that:

- (13) your academic performance -
(14) your personal finances -
(15) your employment prospects -
(16) your relationship with your peers -
(17) your relationship with your family -
(18) your relationship with your professors -
(19) (first other source volunteered) -

- is a source of worry for you.. Have you ever discussed this concern with someone?"

- (13) yes, no (academic)
(14) yes, no (finances)
(15) yes, no (employment)
(16) yes, no (peers)
(17) yes, no (family)
(18) yes, no (professors)
(19) yes, no (Other)

(possible questions 13 -19: PART B)
 "Have you discussed this concern with :

WHO	SOURCE OF CONCERN						
	ACADEMIC	FINANCES	EMPLOY-MENT	PEERS	FAMILY	PROFES-SOR	OTHER
FRIEND							
FAMILY MEMBER:							
PROFES-SOR							
COUNSE-LOR							
SOMEONE ELSE:							

20. "We're just about done! Some times people feel that they would like to talk with someone about a matter that worries them, but they don't feel comfortable doing so. Do you sometimes feel that way?" (yes, no)

If 'yes' go to 21. Otherwise go to END.

21. "Do you sometimes feel that way about talking to a

- friend?
 - family member?
 - professor?
 - counsellor?
 - someone else? Who might
-

that be in terms other than the person's name?"

END. Thank the respondent for his/her participation.

Answer Card for Respondent

ANSWER CARD				
A SOURCE OF GREAT WORRY	A SOURCE OF SOME WORRY	NEITHER WORRY NOR SATISFACTION	A SOURCE OF SOME SATISFACTION	A SOURCE OF GREAT SATISFACTION

4.5.3 ANALYZING THE RESULTS

Structured interviews require analysis which can differ radically from what is suitable for the numerical data obtained with most other measurement techniques. However, it is usually possible to summarize some results in terms of numbers. There are several such possibilities in the above example. 1. By counting the number of each type of answer chosen from the "Answer Card", one could do a Chi-Square analysis to find out if the respondent gave a significantly greater number answers from one category. If several males and females were interviewed, a Chi-Square contingency test could determine whether gender affected the type of answer. Converting the worry-to-satisfaction answers to a -2 to +2 numerical scale opens several possibilities: 2. The similarity of two respondents could be calculated in terms of a correlation of their answers to questions 5 to 10. The similarity of the average answer to each of these questions could be compared for males and females, or younger and older students. 3. Analysis of variance could be determined whether there was an interaction between gender and field of study with respect to the amount of worry-satisfaction about the six areas referred by questions 5 to 10. Counting the number of different types of people indicated in the answers to questions 13 to 18 is another source of numerical data. 4. On the basis of their answer to question 5 about academic concerns, respondents could be classified as "worried" or "satisfied" - with those answer "neither" omitted. t-test could tell you whether they differed significantly with respect to how much they talked to other about their concerns.

Non-numerical techniques loosely called "cluster analysis" can be used to seek relations between the answers to various questions. Basically these can be as simple as tables that are organized to facilitate simultaneously looking at many answers to certain questions. Visual study may then reveal that certain answers tend to occur together (i.e. in "clusters"). Considerable trial-and-error work may be needed to come up with the right organization to reveal clusters. Computer programs can do this too, but their use and interpretation go well beyond introductory courses.

4.5.4 SPECIAL PROBLEMS

Generalizability

You want the results of your case study to be seen as representative of results that could be obtained from persons having the same special characteristics as your subjects or who have experienced the same levels of an extrinsic independent variable. You convince readers of generalizability by explaining how you selected your subjects, describing their particular characteristics as completely as is relevant to the study. You may also refer to the characteristics of persons in related research and compare them to your subjects. The difficulty of generalizability also depends on the degree to which the independent variable concerns physiological, cognitive or personality characteristics.

Subject Reactivity

Overcoming reactivity is the biggest problem in doing case studies. The personality of the subject and also of the interviewer are largely unknown - except as brought out in the interview record. Researchers select subjects for case study based on personality judgements as much as subjects agree to participate based on their own personality and judgements about the researcher. Keeping the data anonymous and explaining this to the subject goes a long way in reducing reactivity.

If you have never done interview research, you may be concerned about whether you have the skills needed to reduce subject reactivity. I believe that, if sufficiently motivated, most persons can indeed conduct satisfactory case studies - including structured interviews. The key is to follow the everyday guidelines that are also the basis of professional ethics: observe the normal social courtesies, showing respect for the subject, and treat subjects as you would want to be treated. If you consider yourself as too shy or too brash, play the role that is expected as a professional obligation - you may fool yourself as well. Finally have faith in the "script" prepared for a structured interview. With a bit of trial and error testing before the interview, the script can help carry you through, where memory, spontaneity, or charm might fail you.

Researcher Effects

This is the second major problem with case studies - or at least those conducted by "other" researchers. Lengthy personal interaction with subjects presents many opportunities for the researcher to unconsciously influence the data depending on the questions asked and just how they are put. The danger can be reduced by interviewer deliberately maintaining a cognitive distance from any hypotheses he or she may have about the research - just don't think about the research aspect and concentrate on the interview itself. The structure of a structured interview is particularly helpful in keeping the questions and their order consistent.

EXERCISES

1. In a journal, find a paper has used a survey to obtain its data and includes a copy of that survey. Answer the following:
 - a) Who were the subjects? How were these people selected? How many people were surveyed?
 - b) Who was the survey delivered? eg. by interviewer coming to their home, by telephone, etc. Were the questions read by the subject?
 - c) What were the independent and dependent variables in the study? Identify a factor that was held constant either in terms of subject selection or how the survey was done.
 - d) What was method used to summarize or analyze the data? What did this indicate?
 - e) Take one question in the survey and 1) try to improve it by rewriting - explain your improvement; or write an additional question and explain why it should be added.
2. Assume you wish to do a survey of kindergarten teachers. The survey has 10 multiple-choice and rating questions. Describe how you would select your sample. Estimate how much it would cost to conduct the survey by various methods, and explain your estimates.
3. For one of the following, propose an agenda for a structured interview:
 - a) with high school principals about improving Science Fair participation
 - b) with garage mechanics about the safety features in modern automobiles
 - c) with travel agents to learn about public preferences in accommodations
 - d) with a Dean about reforming the grading system

Chapter 5

LABORATORY METHODS

So far we have considered methods that did not require any special apparatus to make the measurements. The following sections introduce some of the measurement techniques typically used by psychologists in a laboratory. For further information start with the general books and journals recommended on page 12. Also available in most libraries is Sidowski's (1972) book on apparatus and techniques in many areas of research. Use PsychLit to track down journal papers about a particular type of research and read the Methods sections. Specialized books too numerous to mention are continually being published for particular areas such as auditory, imagery, or electroencephalography research, but these often are not acquired by the average library. Your best guide to such books and other information is a professor doing research in that area. Finally, while the focus is on laboratory techniques, bear in mind that just as subjective questionnaires and the other previous methods can be used in the laboratory, many of the following techniques that use little apparatus can be used outside a laboratory setting.

5.1 MEASURING PERFORMANCE

There are more ways to measure performance than there are things that you could ask persons to do. Many such as the number of problems solved or number of times a target is missed can be measured directly by counting and require no measurement instrument. Besides facilitating the control of independent variables and holding constant or balancing other variables, laboratories can enable you make measurements that require conditions or apparatus that are difficult to obtain in the 'real world.'

Do not let yourself be intimidated by the apparatus aspects of laboratory research even if you are "technically challenged." It is all a matter of learning: reading the instructions, asking questions, and trying, trying, trying. Saying that you "can't" just because you don't do something correctly the first time is mostly an excuse (all too often socially condoned) for being lazy. Whether it's writing poetry or operating an oscilloscope, keep trying till you get it correct; that's what learning is all about. Laboratory research is not about using apparatus. Simple or complex, the apparatus is just a means. Its what you use it for that matters. This section concentrates on making measurements of performance using just a stop watch. I think you'll see that it is imagination not technology that first limits research.

5.1.1 STOPWATCH RESEARCH

With a bit of ingenuity, a lot of things can be learned about human behavior, cognition and physiology using nothing more than a stop watch. One of the "Holy Grails" of psychology is to find the exact moment at which recognition occurs after we have been presented a stimulus. To illustrate what is possible, I shall explain how using nothing more than a stop watch you can obtain an estimate of this moment that will rival the latest brain scanning technology.

5.1.2 REACTION TIME

(footnote about computers and display scans)
as measures of cognitive difficulty
as measures of attention

100 - LAB

implications of speed-accuracy trade-off

5.1.3 PERFORMANCE INDICATOR

duration and achievement
task suitability - ergonomics
skill - sports psychology

5.1.4 ATTENTION AND MENTAL EFFORT

tapping rate and attention

5.2 MEASUREMENTS INVOLVING PERCEPTION

In the 1850's Gustave Fechner, a physicist (there were no psychologists those days), became intrigued by the challenge of finding the relation between the physical energy and our subjective experience of basic phenomena such as light and sound. For reasons that are still debated today, he never completed his quest, but the measurement methods he invented in the process were the first uniquely psychological measurements. Within 50 years others adopted, adapted and added to his methods making psychology a science unto itself. Despite Fechner's failure, his "psychophysical methods" continue to provide insight into the mind by measuring how it uses information in terms of the physical characteristics needed for perception and recognition. These measurement methods have proven to be consistent for normal subjects when perception or recognition does not depend on knowledge or attitude. This makes possible have standards for normal acuity, color, auditory, odor, and touch sensitivity - even for different species, since the basic physiological processes are very similar.

5.2.1 TAPE MEASURE RESEARCH

How far away you can read the warning on a package or recognize the emotion portrayed in a photograph can reveal how effectively the warning was worded or your own emotional state. Therefore it is possible to obtain such complex psychological measurements with little more than some space and a tape measure. However for consistent results, consideration must be given to how the measurements are made. The versatile Fechner's **method of limits** provides a good example of how a tape measure can be used, even as the tape measure illustrates the method:

For this example we'll place an illuminated easel at the end of a hallway and put our subject in front of it on a chair with rollers. A tape measure attached to the easel runs out to the chair. The experimenter places a photograph on the easel, begins to move the subject back, and asks the subject to report as soon as the facial expression becomes unrecognizable.

Wouldn't it be better to have the subject sitting still, and have the experimenter move the easel? - The Editor

Yes, that might be a good way to do it too. There often many ways to do research, and figuring out what works best is one of the challenges that makes it fun. A number of factors such as uniformity of lighting,

and how well the chair rolls, may influence that decision. - Nilsson

The distance is recorded and the chair is moved still further back. The subject is now moved forward towards the easel and asked to report as soon as the facial expression becomes recognizable. Repeat this back-and-forth procedure several times and calculate the mean distance. This average generally proves to be consistent when measurements are repeated the next day, month or year.

HOW LIMITS DATA IS ANALYZED

Simply averaging an equal number of back and forth distance provides a reliable description of the subject's sensitivity (or picture's effectiveness.). However, having taken introductory statistics, most readers will realize that mean values are seldom adequate for a conclusion in psychology and that the variability of the data must also be taken into account. You could simply calculate the standard deviation of all the distances, but this would be misleading since the back distances refer to a different perceptual effect than the forth distances. Averaging the distance at which the picture becomes unrecognizable with the distance at which it becomes recognizable does give a true estimate of the **threshold** distance at which the transition occurs. This threshold tends to be more consistent than either of the direct measurements because those distance vary depending on how sensitive or certain the subject chooses to be. One can obtain a set of threshold measurements by averaging successive back and forth distances - called "running averages". Though it requires a bit more work, the standard deviation of these running averages is a more accurate summary of the variability of such data. The running averages themselves would be used for analysis of variance.

OTHER USES

Combine this procedure with any of an almost unlimited number of independent variables and you have an appropriate operational definition to answer questions in every area of psychology. Here are some examples:

1. Environmental Psychology - Vary the types and level of lighting to find out which is most effective for ease of reading. Readability at greater distance defines greater reading ease. Then reduce the level of the lighting to see how much energy could saved without appreciably reducing readability.
2. Ergonomics - Vary the font and colors used in a traffic sign to find the most effective design. Readability a greater distance defines a more effective design.
3. Personality - Are extroverts more sensitive to the emotional expression of others than introverts? Compare two groups in terms of the mean distance at which they can recognize the facial expression in photographs. Greater distance corresponds to greater sensitivity.
4. Social Psychology - Does similarity of sex, age, or facial features affect the ability of a person to recognize the facial expression of others. Greater ability is operationally defined in terms of greater recognition distance.
5. Physiology & Social - By asking the subject to fixate (look directly at) on a point either to the left or the right of the easel, it is possible to selectively stimulate the right or left

cerebral hemisphere. With this additional variation, one could determine if the well known right hemisphere advantage for recognition of emotion is influenced by how familiar the subject is with the persons in the picture.

5.2.2 INSIGHTS FROM PERCEPTUAL LIMITS

What is the softest sound you can Hear? Does its pitch make a difference? Does how rested or stressed you are make difference? Questions like these can be answered with just a cassette player by systematically adjusting its volume control.¹ While the methods of limits could be used to make these measurements, with many types of equipment, it can be easier to have the subject, rather than the experimenter, operate the control. Then the procedure is called the **method of adjustment**. The subject successively adjusts the stimulus strength until it can no longer be detected. When this setting is recorded, either the subject or (preferably) the research further reduces the stimulus strength. Then the subject increases the strength till they can first detect the stimulus. These up-down adjustments are repeated several and running means calculated just as in the method of limits. The method of adjustment is the fastest of the perceptual measurement methods and more sensitive than the method of limits. There is some danger that the subject may simply learn certain control positions, but this is usually not a problem with cooperative subjects. The main reason for not using this method usually has do with the technical difficulty of providing a suitable way for the subject to adjust the particular aspect of the stimulus or a suitable way for the researcher to record the settings.

Perceptual limits usually differ depending on various stimulus characteristics such as frequency, shape, location, background, adaptation, etc. Patterns in these differences are the first stage cues to learning how the senses function and how the brain processes information. Figure 5.2.2 shows the relative strength of air pressure fluctuations needed to hear a sound as a function of frequency. The two discontinuities in the function indicate that three distinct mechanisms are involved, even though they overlap considerably in their range of operation.

[Figure 5.2.2. Auditory thresholds.]

5.2.3 DIFFERENCE THRESHOLDS

"Please adjust the color of the lower rectangle to match the color of the top." This simple instruction is the basis of a second stage insight into how the brain processes information. No matter how carefully this is done, subjects hardly ever obtain an exact match. Yet the errors are not random. Their variability changes systematically with most stimulus characteristics such as strength and quality. The standard deviations of multiple matches multiplied by 0.65 yields a measure of the change in strength or quality needed

¹ The volume control on some players may not be sensitive enough to permit smooth adjustment. This can be solved by adding a second, external potentiometer in series with a pair of wires. Add markings to the dials of both potentiometers, and then the loudness for various settings can be calibrated with a sound level meter, which is usually available in psychology departments.

to perceive a difference, a **difference threshold**, which is comparable to the results obtained when subjects are asked to report when they can first see a difference or stop seeing a difference using either the method of limits or method of adjustment (Woodworth & Schlosberg, 1954). The matching method of measuring difference thresholds is usually the easiest of the three methods.

Among the reasons why difference thresholds can provide a deeper insight into information processing is that they enable us to study how perceptual systems work under normal conditions. For example, while we occasionally strive to detect a faint sound or identify the color of a dim light, most of our perceptions involve stimuli that present no difficulty in detection. Difference threshold functions are often more complex than absolute thresholds because they involve additional levels of processing. Absolute visual sensitivity to different wavelengths of light, for example, shows only modest discontinuities similar to auditory thresholds. Color difference thresholds, as can be seen in Figure 5.2.3, change markedly as a function of wavelength. They leave no doubt about the presence of multiple mechanism in color perception. The price for this additional insight is more work. Difference thresholds require a careful calibration of apparatus to ensure that varying the particular stimulus quality such as color does not also vary brightness and saturation.

[Figure 5.2.3. Color difference thresholds.]

5.3 PROBLEM SOLVING AND MEMORY

5.3.1 MIRRORS, LEARNING & FORGETTING

It is popularly believed that, "You never forget how to ride a bicycle." That same is not said about a list of twenty words that have been learned suggests there are major differences between memory for skills (implicit memory) and memory for facts (explicit memory), but why is not understood. These differences can be studied with a minimum of apparatus. Implicit learning is often demonstrated by having a person trace an outline of a five pointed star while looking at the star and her or his hand only through a mirror. The accuracy with which this task is performed can be measured in terms of the number of times the subject erroneously crosses the lines of the outline. Evidence that learning takes place is demonstrated by the decrease in number of error crosses on successive trials.

Explicit learning can be demonstrated by having a person briefly read a list of words and then repeat the words from memory. The number of incorrect or missing words decreases rapidly with every repetition. To make the two tasks more similar, the words could be read through a mirror as well. Repeat both tasks until the person makes no mistakes on either and you have a learning curve for each task. (Since there is no way to equate the drawing and verbal for tasks for difficulty, one or the other task may require more repetitions to achieve no errors.)

To measure forgetting, have the subjects repeat both tasks after an hour, 3 hours, 6 hours, 10 hours, 15 hours (Note: This arithmetic progression increases the retention interval between successive measurements by one hour and takes into account that each test reminds the subject of what is to be remembered. Other progressions such as a geometric doubling of the time of subsequent tests could also be used on the assumption that information decays at a rate proportional to the amount that remains.)

5.3.2 MEMORY PSYCHOPHYSICS

The above method of counting the amount of information retained, is limited in that it only measures whether the information is correct or wrong. The research can not measure how close the memory is to the original information. Some success in adapting Fechner's psychophysical methods to the study of memory shows it is possible to go beyond this limit. "Adjust the angle of this line to match the angle of the line you just saw." Repeating these measurements for lines at different angles and after various waiting intervals produces a "time map" of what happens to information stored in memory. To precisely define the intervals over short time periods, the subject only looks at the original and the matching stimuli for one second. Repetition is used to give the subject multiple opportunities to produce a match. Unfortunately, this requires many hours to obtain a complete set of measurements. There fore I have only made such measurements for memory intervals up to 30 seconds so far, though this is long enough to bridge the realms of short-term and long term memory.

[Fig 5.3.2-1. Mean and sd of line orientation matches]

Figure 5.3.2-1 shows that the average match hardly shifts at all as memory time increases. This indicates the memory storage itself was free from bias. Note, however that the variability of the matches increases steadily with time. Evidently accuracy deteriorates the longer the information is stored. Fitting an exponential decay equation to this function indicates that the half life of this information is about 24 seconds. I originally thought that this decay was a random process. Plotting the memory difference thresholds were plotted as a function of both angle and time revealed that their accuracy changed with line angle in the same way as the perceptual difference thresholds that do not involve memory. See figure 5.3.2-2. The remarkable thing is that these functions do not become random with longer storage, rather they become more pronounced. This suggests that pattern of neural activity retrieved from memory must be very similar to the pattern of activity produced by these stimuli during perception!

[Fig 5.3.2-2. Variance of angle matches as function of time in memory]

5.4 MEASURING WHAT INFANTS KNOW

Not so long ago, it was believed that until children began to talk, they had little thought or knowledge - at least there was little scientific data to the contrary. When your subject can not talk nor follow instructions, it might seem impossible to find out what they can perceive and whether they think. Yet what can not be expressed by words or deliberate actions may be expressed by reflexes. Within a few weeks after birth, it is evident that infants can control their eye movements and move their head to direct their gaze. They also have an innate curiosity which leads them to attend to things that are novel or complex. This orienting reflex is observable indicator of a cognitive response. In the 1960's a few psychologists realized that by organizing what is placed in front of an infant and carefully watching its eyes, it was possible to discern what captured its attention and what didn't. A few other attentional reflexes including reaching and sucking were also found to be usable.

From such slender threads hangs a wealth of insight into the functions of the infant mind and understanding of its perspective of the world. Scientists have learned that subtle

effects are not necessarily unimportant effects. A revolution in physics and consequently our understanding of the universe was launched when Roemer noted a slight difference in the period of Jupiter's moons when the planet was close to Earth or on the other side of the solar system. For an even more remarkable example closer to home, just think about how this information is reaching your brain. Each 40 millivolt, one-thousandth of a second nerve impulse represents a change in ion concentration of less than one-part in a million. Who knows how many slender threads are still waiting to be grasped in other areas of research?

5.4.1 HOW IT'S DONE

How that one reflex can be used illustrates the essential characteristics of all good scientific research: careful observation, organized presentation of the independent variable, attention to details that result in control of other variables, and, especially for infant research, patience. The basic procedure seems so simple. Show two or three pictures or objects and note which receives most of the infant's gaze. To do this requires a means to presenting the stimuli simultaneously to avoid bias. This is accomplished in various ways such as opening a curtain, or illuminating the presentation area, while the infant is oriented in the general direction. To be able to tell where the infant is looking the researcher must view the eyes from directly in front, without getting in the way or being a distraction. This is accomplished by viewing through a peep hole in the center of the presentation area either directly, or with a video camera, or both to obtain a permanent record.

[Figure 5.4.1 Illustration of an infant in a preferred viewing apparatus.]

Now let's consider what happens in a basic "preferential viewing test". A picture of a human face is presented on the left and a picture of similar face-like parts in some other arrangement is presented on the right. When displayed, the infant's eyes briefly wander back and forth, then settle to gaze at the face. Rules have to be established for what is considered initial wandering and what is an interested gaze. The dependent variable may be the selection itself or the total amount of time the gaze is directed at each picture. Since a single trial is not sufficient to establish any preference, the procedure must be repeated several times to establish the probabilities for each alternative. The location of the human figure in the two stimuli must be alternated randomly to rule out the possibility of a positional bias. By the time a preference has been established for this one set of stimuli, the infant may need to rest, be played with, or otherwise distracted to prevent boredom. Therefore, it may take several visits to obtain a complete enough data for even a few levels of an independent variable.

In infant research, more than half of one's time is required to arrange for parents to bring their infants to the laboratory, to discuss the research with the parents, waiting while other needs such as food or diapers are met, and just plain waiting for the infant to appear ready for the next presentation. This is research that requires the patience of a saint and the salesmanship of a carnival huckster. When more than descriptive research is pursued, ingenuity is an asset too.

5.4.2 MAKING INFERENCES

Perception

For an example, we'll consider how to determine if an infant sees colors? Present three colors, two the same and one different. The infant shows some preference for looking at whatever color is different. Make the two colors more and more similar, and at some point the

infant no longer shows a preference for the one that is different. At that point you can conclude that you have reached the limit of its ability to distinguish this color. Next start with both colors being the same and increase the difference until a preference is shown. Assuming you have also done all the other right things such as presenting each set of differences several times and randomizing the position of the different color, you have now succeeded in using the "method of limits" to measure a color difference threshold. A basic study of the color spectrum detailed enough to reveal the characteristics of trichromatic vision would require such measurements at some twelve regions of the spectrum. Many visits and babies are needed to obtain one set of data (Ross, 19XX). Yet it is possible, and the same procedure can be used to measure sensitivity to line orientation, acuity, depth perception, and even whether infants experience visual illusions.

Memory

To measure memory requires more inference and patience. An object's novelty depends on memory - in this case the lack of memory. Conversely, when an object no longer elicits a novelty reflex, we can infer that it is remembered. As an object is presented repeatedly together with various other objects, infants start to ignore it. How long that ignoring behavior lasts indicates how long the infant remembers the object. Sounds easy until you consider that presenting the object once to test for memory, immediately refreshes memory. Therefore, measuring how long a memory lasts requires a succession of tests. This technique makes it possible to determine how various characteristics of objects and various situations affect memory.

Cognition

Used another way, the novelty reflex can reveal what an infant expects to happen. What is expected in turn depends on understanding. For example, when a toy is rolled behind an obstruction, the infants looks for it to emerge on the other side. This suggests that the infant's brain has calculated a trajectory for that movement. If the toy is exchanged for a different one while briefly hidden, the novelty reflex is exceptionally strong and resistant to fading. This indicates recognition that objects are supposed to permanent.

5.5 PHYSIOLOGICAL MEASUREMENTS

5.5.1 HEART RATE, SKIN CONDUCTION, AND LIES

Though heart rate measures have limited application in psychological research, instruments for measuring heart rate have become so inexpensive with modern technology, that they are worth considering as secondary measurements in certain situations. A sound sensing device intended to for exercise monitoring can now be bought in local drug stores for \$100. Since muscular work is directed related to metabolic rate, which in turn involves oxygen and glucose consumption, which is delivered by the autonomic nervous system making the heart beat faster, hear rate does reflect how much work a person is doing. However, as demonstrated by the well know example of astronauts on a certain moon walk, high heart rates do not always correspond to the subjective effort that a person is making. Some researchers report that the size of moment to moment heart rate variations is related to stress. (XXX, 19XX). The autonomic nervous system reacts to stress in a variety of ways. One of these is to increase sweating, which reduces the skin's electrical resistance and gives rise to the galvanic skin response, which is what "lie detectors" measure. While this may seem a promising way to

measure stress or what people feel, many persons including persons with psychopathic disorder do not show typical autonomic responses to stress or threat.

5.5.2 NEUROPSYCHOLOGICAL TESTING

These include a variety of questions of basic facts, such as, Who is the prime minister?; ability, such as, "What is 3 times 9?"; questions of what the subject feels when lightly touched on the different parts of the body such as left sole or right wrist; requests to recall a few words such as "hand", "dog", "shoe" at several occasions during the examination; observations of or for certain reflexes like the knee-jerk; observations of how the eyes track a moving finger; observation of how accurately a person can point to a certain place on his body or table with eyes closed; etc. While test is quite simple, in total the results of a 15 minute exam can surpass the sensitivity of a computerized axial tomography (CAT) scan at detecting neurophysiological abnormalities.

5.5.3 NEUROELECTRIC RECORDING DIRECT MONITORING OF BRAIN ACTIVITY

A nerve impulse is a 40 millivolt pulse that lasts about a thousandth of a second. Though the current in a single impulse is infinitesimal, solid state amplifiers have little trouble picking up such small changes. Making small enough "microelectrodes" that won't damage a neuron and getting such an electrode into one of the proper neurons present greater difficulties. Though special equipment is required to produce either very fine steel needles or glass tubes, these can be bought ready made. Also useful are somewhat larger needles that push their way between neurons and record from a cluster of neurons in the immediate vicinity of their tip. Both types of electrodes are put directly into the brain of an animal using a "stereotaxic" instrument. (The same type of device is also used for human brain surgery. There is no pain, since the brain has no pain receptors.) These instruments consist of a means of precisely moving an electrode along three axes and have fine graduated measurement scales. Their use involves an initial calibration relative to a reference point on the skull, and then the electrode is moved to a certain location in the brain using "stereotaxic atlases" which have been prepared for various species - included humans. The brains of individual animals are sufficiently similar that this is practical.

[Figure 5.5.3-1. illustration of a stereotaxic]

NON-INVASIVE MONITORING

Even the simplest experience such as a light flash involves repeated impulses by thousand of neurons. This would be a sizable electric effect if 1) their impulses all occurred at the same time - neurons in the -70 millivolt resting state tend to cancel out the + 40 millivolt impulse effects, 2) there were not many other cells doing other things at the same time, 3) the strength of the electric field did not diminish with the square of the distance from the active site through the skull and outer skin. Consequently the only reliable brain activity that can be monitored as it occurs are very general effects that reflect alertness and various stages of stages of sleep and comma. Fortunately, statistics provides another way for more specific monitoring.

Briefly present a red flash or a picture of Daffy the duck, and thousands of neurons are responding in some characteristic manner that differs from their response to a green flash or Donald Duck. Simultaneously millions of other neurons are involved in various other tasks

such as remembering a Christmas tree, or a movie theater from your childhood, or planning what to have for lunch. An electrode pasted on the scalp picks up all these activities indiscriminately. Consequently, the response displayed on an oscillograph is completely different each time one of these stimuli are presented. Yet present one of these stimuli repeatedly at intervals of a few seconds some sixty times. Save the response that is produced following each stimulus onset and add them up. The patterns of activity not produced by the repeated stimuli will differ each time and tend to cancel out when added. The responses to the repeated stimuli will be nearly the same each time and get larger with each addition until by sixty repetitions they will stand out clearly. Computers can readily do this kind of signal summation provided you feed them responses through an "analog-to-digital converter" which changes the momentary voltage fluctuations into numbers. Eureka! ²

[Figure 5.5.3-2. Evoked responses to different colors (Shipley, et al 1964).]

5.5.4 PSYCHOPHARMACOLOGICAL TECHNIQUES

For human research, non-invasive measurement of neurotransmitters, drugs, or nutrients that influence and are products of behavior is still in its infancy. Yet great strides can be expected in the next twenty years. Though you have hopefully not had to take one, you have doubtlessly heard of the breath analysis test that can determine the percent of alcohol in a person's blood. Along similar lines, a respiration meter can measure the amount of exhaled carbon dioxide. Since the amount of CO₂ a person exhales directly reflects the level of metabolism, a simple formula enables one to accurately estimate energy consumption in calories/sec. In turn the rate at which the body uses energy reflects the amount of effort being expended. While such measurements have generally been aimed at the physical effort exerted by muscles, refined techniques should make it possible to detect changes in overall neural activity and thus perhaps enable measuring cognitive effort when muscular effort is held steady. A saliva test that can determine from a cubic centimeters whether a person has smoked tobacco during the past 48 hours is now available commercially. Tests for other drugs are under development. An ear clip is now available to measure a person's blood oxygen level in terms of infra-red transmission through the skin. You can expect further refinements in that technology to enable measuring relative amounts of other substances in the blood.

[Figure 5.5.3-3. Heart rate and energy expenditure KKK p.109]

5.6 TECHNIQUES FOR RESEARCH WITH ANIMALS

It may take a bit of ingenuity, but all of the above measurement methods that have been described using human subjects can be done with animals as well. Unless you have had some

² Warning. When I proposed an honors thesis that involved recording brain activity of subjects looking at a flashing light, professor Gilray Kandle arranged for me to visit the Albany State Medical Center to learn the basics. The doctor who helped me first issued the following warning: "Are you certain you really want to learn how to do this? Most people are never quite the same afterwards. They keep pursuing these techniques in the belief that they will discover something truly remarkable about the brain." He was right.

experience along these lines, the methods required for animal measurements can be more difficult to picture, so we'll look at some specific examples of how these measurements are made. However, first I'll outline some reasons why you might want to make such measurements using animals.

1. Similar Yet Simpler - Genetically speaking, chimpanzees are 99.8% similar to humans. In that respect even a rat is 92% similar. This underscores the fact that at the cellular level their digestive, muscular, endocrine, circulatory, neural systems would be very difficult to distinguish from ours. Consequently most basic functions including homeostatic motivation, movement control, learning, and perception are operate according to much same principles as our own. The generally smaller size and overall lesser complexity of these systems in animals can facilitate understanding those principles and lay the foundation for ourselves.
2. Availability - For some research problems it is desirable to have subjects available for many hours, days or even months of repeated measurements to obtain the required data. Even with ample financial resources available for salaries, human subjects demand evenings, weekends, and holidays away from your research. From infancy to old age, animals subjects are available around the clock, 365 days a year, provided that you provide adequate time for the meals and rest periods needed to maintain their health. (However, against this convenience must be weighted the time required to train animals to perform the tasks required by the research. Then there is also the researcher's responsibility to provide the animals' "room and board" and any other needs important to their comfort and well-being.)
3. Invasive Techniques - When warranted by the research objectives, only animals may receive drugs, have electrodes planted to stimulate, record or lesion their central nervous system, or undergo environmental conditions such cold or isolation. Subsequently it may also be necessary to sacrifice the animal to permit studying the effects of the experiment on its body and nervous system. While the consequences for the animal are often regrettable, much of our understanding of how the human brain works (not to mention life saving medical techniques and medications) would have been impossible with the use of animals for research. For guidelines on the ethical use of animals in research see Appendix A - Doing Research Ethically, Section A.4 - Responsibility for Animal Subjects.
4. Comparative Study - One way to work out how the human central nervous system works is comparing its structure and function to that of other animals. When anatomical or neurochemical differences between species are systematically accompanied by differences in their behavior, one may be able to relate those learn how that behavior is produced.

5.6.1 OBTAINING ANIMAL SUBJECTS

For certain experiments such as those involving surgical techniques, drugs or restricted environments, animals are the only suitable subjects. Yet, when delays arise because human subjects are unavailable or forget to come, the use of animal subjects may seem an attractive alternative. With animals you can decide precisely when and how often they participate. The great drawback of using animals is that it usually is not them but rather us that we really want to understand. Yet, ingenious methods continue to be developed that enable experiments to be done with animals on questions such as problem solving, volition, and consciousness, that a few years ago would have been considered only possible with humans.

110 - LAB

Just because animals don't have to be recruited, compensated for their time, or debriefed, do not think that animals are a more economical and convenient alternative to human subjects. As was pointed out in the ethics chapter, animals require housing facilities that are costly. While you don't have to pay them for participating, you have to pay to obtain these subjects and then feed, clean, and otherwise maintain them for the rest of their lives. Present standards for facilities to house and test animal subjects are far more stringent and costly than those for humans. (This may seem ironic, but animals can't complain or refuse to participate.) The following table shows some costs based on purchase from a laboratory breeder and maintenance in a university operated facility:

Table 5.6.1-1.

Some Approximate Costs for Purchasing and Maintaining Animals for Research

ANIMAL	PURCHASE	DAILY FOOD AND CARE
RAT	\$10	\$0.20
CAT	\$200	\$2.20
DOG	\$450	\$6.00*
MACAQUE	\$2000	\$6.60

* Includes minimal walk twice daily.

Animals obtained from a laboratory supplier have the advantage of being disease free and of known genetic background. The ability to request litter mates and genetic uniformity are advantages in most research designs. However, do not make the mistake of thinking that all rats of a given strain are alike. Differences in intelligence and "personality" even in rats make it important to use one or more of the procedures discussed in Chapter 7 to ensure that differences between individual animals do not confound the results.

5.6.2 HOW ANIMALS "TELL"

Observing the behavior of animals can tell you directly whether they are hungry, thirsty, afraid, or interested in mating. To measure what animals perceive, remember or understand requires the researcher to train them to perform different responses depending certain prearranged alternatives. Pressing an easily moved lever by paw, beak, or tentacle is a widely used response. In the simplest format, the animal presses the lever to indicate that it perceives or understands the significance of a certain stimulus event and otherwise ignores the lever. Since you can not explain to the animal that it is important to press the lever under those circumstances, it is necessary to provide a direct reward for doing so correctly. That reward is usually food, drink or a treat, but it can also be to remove an irritating sound or mild shocks. (To get any animal, including humans, to do something by punishing them if they refuse does not generally work.) To keep the animal from wandering away, the testing area is enclosed. A common type of such testing apparatus is called a "Skinner box" after the psychologist who

extensively developed this type of research.³ The ability to learn certain sequences of moving in response to various cues can be tested using mazes of various types.⁴

How do you inform an animal that pressing a lever or selecting a certain object will provide a reward. It is possible to simply leave the animal with the apparatus and eventually by random movements or deliberate exploration will usually discover the behaviors that produce rewards. However, this can prove frustrating for subject and experimenter. To speed up and ensure that animal learns that a certain response can be rewarding, a training procedure called "shaping" is generally used: The experimenter waits for the animal to wander close to lever and then directly operates the mechanism to deliver a treat. The animal hearing the sound and smelling the reward will investigate the dispenser and find the reward. Thereafter, the animal tends to hang around the dispenser in hopes of finding another treat. Subsequently, the experimenter requires the animal to be ever closer to the lever before a reward is dispensed. Before long the animal accidentally moves the lever enough to activate the dispenser and soon makes the connection. It typically takes a rat a few sessions to learn the lever pressing response; much of that time is probably spent becoming familiar with the "box". After that, the experimenter can introduce rules such as the presence of a certain stimulus when a lever press produces a reward. By changing the stimulus, one can measure the animals sensory discrimination; by waiting, measure memory; or by changing the rules, measure learning.

5.6.3 ANIMAL PERCEPTION

I'll illustrate what is involved by describing how the color vision of trout, *Salmo guardirei*,⁵ can be measured. A two-way maze was constructed by building a starting chamber with two gates in the middle of a long, glass fish tank. A small rear-projection screen at each end of the tank could be illuminated by either white light or narrow-band spectral light from a monochromator. From the central compartment a trout could swim either left or right to approach either a white or colored screen. The left or right position of the color was random. Initially a mild shock helped get the trout moving out of the starting box. When it arrived at the colored screen, a few crumbs of food were dropped to the trout. Trout learned to swim

³ Apes are sufficiently intelligent to learn rules for various complex responses such as selecting a certain object and to be motivated by sheer curiosity or tokens that can later be exchanged for treats. With them a less automated Ames General Testing apparatus can be used. It is simply a platform which slides from behind a screen to present objects and permit selections in a controlled situation that facilitates the observation and timing of responses.

⁴ The Morris water maze, which has the animal swim to a hidden platform is proving successful for research with rats because it prevents the animals from leaving odor trails and because it is self motivating - they will seek the platform simply to get out of the water. By changing visual cues to the location of the platform, perception, learning and memory can be tested.

⁵ Since there several varieties of trout, rats, etc, it helps to identify the genus, species and when possible the strain or breed used.

to screens of various colors, in a couple of weeks. Then the room was dimmed and the luminance of the both the white and colored screens was systematically decreased. When the screens were sufficiently dim, the trout's choices became indicating it could no longer discriminate the colored screen from the white screen. The luminance at which confusion occurred varied as a function of spectral wavelength. Three 'bumps' in this sensitivity curve resemble bumps in the human spectral sensitivity function suggesting that trout discriminate colors like we do.⁶

5.6.4 ANIMAL COGNITION

In 1988 several persons in Montreal all suddenly developed sever memory loss and some died with a few days. A recent factor common to all was that had eaten mussels shortly before the symptoms began. Analysis of shipments of mussels revealed in a few the presence of a rare neurotoxin, domoic acid. While the search was on for how this got into mussels, physiologists fed mice this toxin in comparable doses and found that many shortly developed convulsions and died. Autopsy revealed extensive damage to the hippocampus, a portion of the brain that was suspected to be essential to memory storage. To determine the actual effects of domoic acid on memory, some students and I injected rats with various doses of contaminated mussels (Pinset, Alexander & Nilsson, 1989). (The lethal level for rats was still unknown.) Most rats receiving the higher doses developed signs of convulsion within two hours; most receiving the highest dosage died. Prior to the injection all the rats had been trained in a Skinner box to press a lever when a round light on one side was turned on. Tested two days after injection, those rats that had developed some convulsions responded significantly less to the light onset than rats receiving low doses or injections of neutral saline. This demonstrated clearly that domoic acid impaired memory. We also tried to teach all rats to press a lever to the lever to a new stimulus. Interestingly, the memory impaired and non-impaired rats did not differ significantly in the number of trials required to make this new connection. This suggested that the toxin selectively impaired memory but not the performance of movement or learning.

5.6.5 INVASIVE PHYSIOLOGICAL MEASUREMENTS

The electrical activity arising from nerve impulses can be measured using electrodes (either fine needles of iron or of hollow glass filled with a conducting gel) implanted in clusters of neurons or even individual cells in the brain or nerve pathways. When the measurement site lies within the brain and is not directly visible, a three dimensional atlas of the typical brain for various species may be used in conjunction with a three dimensional placement apparatus called a stereotaxic instrument to position the electrode. (We know from

⁶Lest you think that psychologists do animal experiments with certain species just to be different, the goal of this experiment was to find out if non-lethal exposures to methylmercury nevertheless damaged the neural system of survivors (Hawryshyn, MacKay and Nilsson, 1981). Rainbow trout are the criterion species for clean water. By definition of the Department of Environment, if exposure to waters from a certain lake or stream kills less than 50% of the trout, the water is not polluted. We found that exposure to mercury levels well below this level produced obvious loss of sensitivity to red and green light, but not blue light. A similar pattern of color loss can be expected to occur amongst people who have been exposed to mercury, but this has yet to be determined.

human medical trials that electrodes implanted in the brain do not cause pain. Unfortunately, there are individual differences in brain structure even of simpler animals such as rats, so it is usually necessary to kill the animal afterwards and examine brain slices to be certain that the measurements were actually obtained from the intended location. Therefore such research is not done for trivial reasons.) What is actually measured is either the magnitude or pattern of electrical activity of a cluster or the frequency and timing of impulses in a single neuron. By combining these measurements with observations of the animal's behavior, with electrical stimulation or lesions of other parts of the nervous system, or with certain drugs, researchers are gradually figuring out the function of various areas of the brain.

For an example consider the following experiment done to locate the switch for "attention". A cat has been trained to obtain food by pressing a lever when a flash of light occurs in the absence of a tone. When the tone is present, the cat ignores the flashes. Electrodes are placed in the optic nerve, thalamus - the relay center for sensory information to reach the cortex, primary visual cortex, secondary visual cortex, and a region of the frontal lobe that is active just before the cat makes a lever pressing movement. At which of these locations are responses to the flashes diminished when the flashes have no meaning? Recordings reveal that the switching occurs in the thalamus before the information even reaches the cortex. Since there are no direct connections between the visual and auditory relay centers of the thalamus, by itself it can not know whether the flashes are important. This suggests that signals to ignore flashes must be sent to the thalamus from some other area.

With hollow glass electrodes, it is possible to obtain samples of the neurotransmitters which are present in a particular location. In the above example such electrodes could reveal whether an additional inhibitory transmitter is present in the visual thalamus when the tone is present. Since such hollow electrodes could also be used to inject drugs that mimic or control neurotransmitters, one could confirm the blocking function of that particular neurotransmitter by injecting it when the tone is not present and determining if responses to flashes are still diminished at primary visual cortex but not in the optic nerve. These electrodes could also be used to inject a dye into the thalamus. That dye would be taken up and passed along only those nerve fibers that enter and leave that particular region. By examining sections of the brain that include the thalamus and the adjacent sections, one can expect to trace the dye in nerve fibers back along the optic nerve, forwards to the visual cortex, and along any other fibers including those carrying the ignore signal that are connected to this region of the thalamus. Such dye tracing could reveal where the ignore signals originate. In principle one could now continue this line of research by recording from those originating areas in other cats, and so on.

If you have followed this mini-odyssey in brain exploration so far, consider the following. It is also observed that cats trained in the above manner also ignore the flashes when they are not hungry. Whether responses to flashes when not hungry are blocked at the same place as when the cat is hungry could be determined by the same recording technique described initially. How might one then find out where the "not hungry" signals originate?

EXERCISES

1. Go to the library and browse through the current periodicals section looking in journals such as: Behavioral Neuroscience, Canadian Journal of Experimental Psychology, Canadian Journal of Behavioral Science, Developmental Psychology, Journal of General Psychology: Human Perception & Performance, Journal of General Psychology: Human Learning and Memory, Journal of Experimental Psychology: Animal Behavior Processes, Journal of Personality & Social Psychology, Memory & Cognition, Perception & Psychophysics, Perceptual & Motor Skills, Psychobiology.

Find a paper that involved research done in a laboratory and used one or more of the following types of measurement: a) reaction time, b) a perceptual limit (threshold) or difference threshold, c) memory, d) infants, e) animal learning, f) electrophysiological recording.

For each paper, prepare a short outline containing the following information:

- a) a proper reference for the paper in APA format (i.e. author(s), year, title, journal, ..).
- b) the purpose of the research
- c) identify the independent variable(s)
- d) describe the subjects and how were they recruited or obtained
- e) the conceptual aspect of what the researcher(s) wanted to measure - i.e. the dependent variable
- f) the operational definition of that independent variable - i.e. what was actually measured.
- g) describe briefly the type of apparatus used to make the measurements.

SECTION II - RESEARCH DESIGNS

Recall how the first chapter presented the scientific method as the means by which non-geniuses could discover new knowledge? So far we have considered the measurement aspects of research. Measurement techniques made it possible for anyone to see for themselves what otherwise might only have been discerned by the most astute observer. On the basis of a few casual measurements of the period of a swinging lamp, Gallileo was able to make the intuitive leap about the nature of gravity that led to the Tower of Pisa experiments. Lacking such genius, most of us would need a more extensive set of observations systematically organized to get us much further than simply acknowledging, "That's interesting." Now that you have been introduced to various ways that psychologists make measurements, you need to learn how to organize the measurements.

In the "old days" (that's when I was a student) it was not unusual for medical researchers with costly equipment to collect large amounts of data that had little value because they had no training in research design. These days many psychologists get involved in medical research because they understand how to organize the measurements to minimize doubts about what they mean and to maximize how efficiently resources are used. Psychologists had to become good at research design because things like human motivation, perception and cognition are so much more complicated than anything encountered by biologists or physicists. Without the use of the research designs introduced here, even the most careful measurements would have produced little "firm" data for psychologists to build an understanding of human behavior. Again, if you are daunted by the prospect of learning how to organize measurements, switch into chemistry, where you don't have to worry about counterbalancing a within-subjects design!

It is important to distinguish the various methods of making measurements described in Section I from this section's various research designs. Any measurement method may be used in any research design and vice versa. While there is a tendency in the literature for descriptive studies to use observational measurements and for measurements requiring laboratory apparatus to use experimental designs, do not let your thinking be limited to the conventional. First aim for whatever measurement is best to obtain the information you need. Use whatever design is most suitable for subjects you want to test. Then consider what is feasible and within your available resources. Knowing about other measurement-design combinations may help you find alternatives that are more feasible for the resources available.

Chapter 6 DESCRIPTIVE RESEARCH

So you want to do research on a certain topic but find little information or previous research.¹ What to do? Consider it an opportunity for real exploration. Start by making your own measurements to obtain an accurate description of the behavior involved. To avoid missing something which may turn out to be important, be both systematic and broad in your measurements. Refrain from focusing on the first interesting results or jumping to conclusions till all measurements are completed. Remember, the significance of a remarkable finding often depends on comparing it with many plain findings.

There is no independent variable explicitly involved in descriptive designs. That comes later when you have learned enough to ask more specific questions. However, an extensive set of measurements may lead you to wonder about certain relationships within the data. For example, did these results occur only when these other results also occurred? To answer such questions, you can treat some results as if they represented levels of an independent variable. When two sets of measurements are paired, they can be compared for similarity using correlation. If they are not paired, you can compare means using t-tests and Tukey's HSD test to tell if the means are different.

This chapter presents four quite different types of descriptive research. Since there is no independent variable, the design of descriptive research focuses on the measurement techniques and the subjects who participate. Until one knows more about what is happening, the representativeness of the subjects tends to be more of an issue than in experimental designs. Yet the number of subjects that suffice for "representativeness" vary greatly depending on whether physiological or social measurements are involved. (Subsequent chapters focus on specific methods which help insure that the effects of levels of an independent variable are measured without bias.)

6.1 A SCHOOL YARD EXAMPLE

To illustrate how to do descriptive research, we'll start with a study on what children do during lunch recess at school. Arrangements, of course, have been made with the school principal and teachers before hand.² Observations could be made in a non-scientific manner by simply wandering around the school yard watching what happens, perhaps talking to a child now and then. Doing this long enough would give you a substantial impression of what

¹ By now research has been done on almost every conceivable topic in psychology. Unfortunately, there are two barriers to finding older reports. 1) Computerized abstracts only go back to the late 70's, which limits "easy" search to the last quarter of the 20th century. 2) The typical university library has holdings only for the last 30-40 years. This can make it expensive to read earlier materials. Yet diligent digging can be rewarding. Earlier publications often contain a wealth of ideas and details on how the research was done.

² In addition to obtaining permission, you visit each class and briefly explain to the children that you are studying how children play and will be in the school yard at lunch time over the next several days and asking some children at random a question about how they feel.

happens. After thinking about what you have seen and heard, you may summarize your experiences with in a descriptive essay.

So what is wrong with that? (Clued in to that word "non-scientific" and inferred something is wrong, eh?) Well, even months of observing and questioning would probably not enable most of us to accurately describe the booming buzzing confusion of children at play. First, since you can't be everywhere at once, you may be unaware of events elsewhere that influenced what you saw. Second, most of us would not be able to remember everything we did see. Third, it would take uncommon discipline not to let your attention get attracted to the more unusual behaviours or perhaps dwell on aspects consistent with our expectations.

A descriptive research design helps overcome these problems. The key components of the scientific method are "systematic" and "measurement". The measurement possibilities you have already studied. Just keep in mind that descriptive research can use more than observational data. Questionnaire data obtained by brief interviews would certainly be appropriate for a school yard. Use a checklist and/or audio recorder. Some unobtrusive measuring instruments could also extend the range and type of data available. A counter attached to the swings and a motion detector aimed at a certain place are some examples of how to measure activity.

The "systematic" part of the scientific method is what design is all about. This playground study was chosen as an example because it illustrates descriptive designing in terms of location and time factors that can be easier to organize than complex factors such as social status or generosity. If you have ever seen an archeological "dig" you may have noticed a gridwork of strings running from peg to peg over the site. That may not be practical for a school yard, but you could map the various areas of potential activity to help you cover each area evenly. You may be tempted to ask, "Wouldn't it make sense to spend more time describing what occurs where more children are?" No, unless you study all areas equally, you won't be able to say with any certainty that there was more activity in some areas than others.

The mobility of our subjects presents problems not faced by archaeologists. Therefore we'll also adopt a time sampling technique used by ethologists to study animals in the wild - an appropriate analog, some would say. To cover the entire school yard, a schedule is made that defines how long observations are carried out in each place. It may seem dumb to leave a particular place when a lot is happening and go another that may even be empty, but this is necessary for an accurate description overall. If the activity you miss in one place is not unusual, you are likely to encounter it again another day. If it is unusual, you do not want it unduly emphasized in the description.

What is being discussed here are, of course, techniques for unbiased **sampling**, since you can not be in all places all of the time. Some practical aspects of sampling were introduced in chapters 3 & 4 to get you going with the observational and survey projects. Descriptive research usually involves sampling in a broader context with respect to both subjects and measurements than is usually needed in other designs.

6.1.1 METHOD

Preliminary observations indicated four school yard areas, the swings, the field, the steps by the entrance, and a hill which together accounted for over 95% of the children. Figure 6.1 is a map of these areas. Over the 48 minutes of outdoor recess at lunch, the researcher visited each place three times as shown in Table 6.1-1. The sequence was changed each day, but we'll consider the results for just a single day.

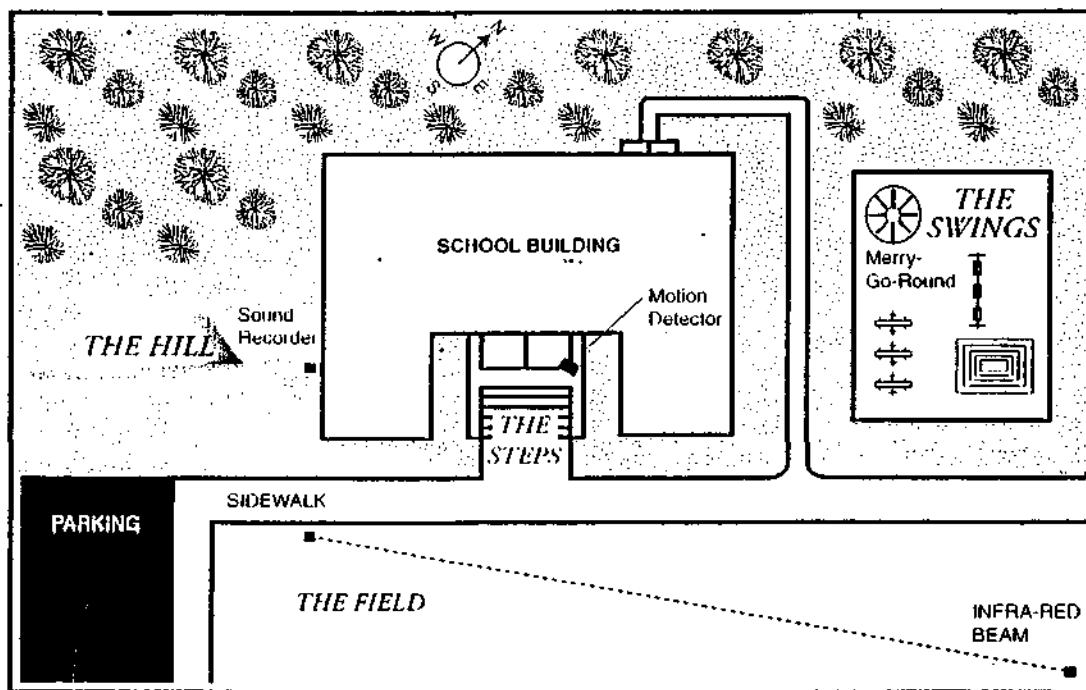


Figure 6.1-1. A map of the school yard showing the various places where measurements were made. The "X" in each place indicates the location used to select the child who was asked to rate how boring or fun they considered that place.

These were the measurements: At each place the researcher counted 1) boys, 2) girls, 3) children alone, 4) pairs talking or engaged in inactive play, 5) pairs in active play, 6) groups of three or more talking or in inactive play, 7) groups in active play. At each area a particular spot was randomly chosen for that day. The child closest to that spot would be asked to indicate on a -5 to +5 scale how bored or happy they felt just then. A data logger automatically recorded some aspect of activity at each play area. The one at the **swings** recorded the work in kilo-joules done to rotate the merry-go-round. At the **hill**, a tape recorder proved a continuous sound record. This analyzed to provide an average loudness for every 4 minutes. An infrared beam spanned the **field** as shown on the map. The number of beam interruptions was recorded every 4 minutes. A motion sensor mounted above the **steps** provided a measure of the amount of time that there was movement. (This may represent more gadgets than a typical descriptive study. I took advantage of the opportunity to illustrate some possibilities that do not require expensive technology.)

The rumor that you are an apparatus nut is substantiated by a vitae which shows that nearly one-third of your publications are about apparatus and methods. - The Editor

And proud of it! My mentor, Dr. Gihay Kandel, acknowledged that the "electric trains" aspect was undeniable but valuable attraction of research. - Nilsson

Table 6.1-1
Measurements of Children and Activities Obtained at Various Places and Times During
Lunch Recess at the Glen Cove Elementary School

TIME	AREA	COUNTED BY RESEARCHER				CHILD		DATA LOGGERS					
		BOYS GIRLS		ALONE	TWO	TWO	GRP	GRP	RATED	SWING	HILL	FIELD	STEPS
		TALK	PLAY	TALK	PLAY	B - F	WORK	LOUD	X'S	MOTION			
12:00	SWING	16	14	11	2	6	3	8	2	120	65	14	3.8
12:04	HILL	13	4	10	2	2	0	3	3	150	74	32	3.6
12:08	FIELD	14	10	1	4	0	7	12	0	170	72	54	3.3
12:12	STEPS	3	20	2	8	2	11	0	-1	180	76	61	2.7
12:16	SWING	9	17	9	4	4	3	6	-1	190	73	79	2.6
12:20	HILL	20	5	10	2	4	5	4	4	200	81	92	2.1
12:24	FIELD	16	14	0	2	2	4	21	3	160	79	87	2.4
12:28	STEPS	2	13	1	10	0	4	0	-3	110	77	83	1.5
12:32	SWING	7	6	5	6	2	0	0	0	80	73	96	1.8
12:36	HILL	14	1	6	0	4	0	5	2	60	68	94	2.7
12:40	FIELD	15	16	1	0	2	3	25	5	40	74	78	3.1
12:44	STEPS	12	24	3	10	6	17	0	-1	60	67	81	3.9
MEA		11.8	12	4.9	4.2	2.8	4.8	7	+1.1	126.7	73.3	70.9	2.8
SD		5.2	6.7	4	3.4	1.9	4.8	8	2.3	53.7	4.6	24.8	0.7

6.1.2 RESULTS

Table 6.1-1 presents some hypothetical data from a descriptive study of school yard behavior. Look for patterns in where boys and girls go during the recess, the types of activity at various places, and the degree of activity. However, it is not easy to compare how many children are in each location, activity counts, fun ratings, and indicators of activity level because of the different ranges and variability of the measurements. A simple way to put all the numbers on a common basis is to transform them into Z-scores by subtracting the mean value of each type and dividing by the standard deviation. These Z-scores are shown in Table 6.1-2 below..

The Z-scores make it easier to see the changes in popularity and activity over the recess period. Positive numbers greater than +0.8 indicate above average popularity or activity. Negative numbers below -0.4 suggest less popularity or activity.

Table 6.1-2
The Measurements in Table 6.1-2 Transformed into Z-Scores

TIME	AREA	COUNTED BY RESEARCHER				CHILD		DATA LOGGERS					
		ALONE BOYS	ALONE GIRLS	TWO TALK	TWO PLAY	GRP TALK	GRP PLAY	RATED B - F	SWING WORK	HILL LOUD	FIELD X'S	STEPS MOTION	
12:00	SWING	0.81	0.30	1.53	-0.64	1.66	-0.37	0.12	0.39	-0.12	-1.80	-2.30	1.36
12:04	HILL	0.24	-1.20	1.28	-0.64	-0.44	-1.00	-0.50	0.82	0.43	0.16	-1.57	1.09
12:08	FIELD	0.43	-0.30	-0.99	-0.05	-1.49	0.47	0.62	-0.47	0.81	-0.27	-0.68	0.69
12:12	STEPS	-1.67	1.20	-0.74	1.12	-0.44	1.31	-0.87	-0.90	0.99	0.60	-0.40	-0.12
12:16	SWING	-0.53	0.75	1.03	-0.05	0.61	-0.37	-0.12	-0.90	1.18	-0.05	0.33	-0.26
12:20	HILL	1.58	-1.05	1.28	-0.64	0.61	0.05	-0.37	1.25	1.36	1.69	0.85	-0.93
12:24	FIELD	0.81	0.30	-1.24	-0.64	-0.44	-0.16	1.75	0.82	0.62	1.25	0.65	-0.53
12:28	STEPS	-1.86	0.15	-0.99	1.71	-1.49	-0.16	-0.87	-1.76	-0.31	0.82	0.49	-1.74
12:32	SWING	-0.91	-0.90	0.02	0.54	-0.44	-1.00	-0.87	-0.47	-0.87	-0.05	1.01	-1.34
12:36	HILL	0.43	-1.65	0.27	-1.22	0.61	-1.00	-0.25	0.39	-1.24	-1.15	0.93	-0.12
12:40	FIELD	0.62	0.60	-0.99	-1.22	-0.44	-0.37	2.24	1.68	-1.61	0.16	0.29	0.42
12:44	STEPS	0.05	1.80	-0.48	1.71	1.66	2.57	-0.87	-0.90	-1.24	-1.36	0.41	1.50

Another trick you may have picked up in "stats" is using tables to rearrange data in various ways to find a perspective that makes the results easier to understand. The sequence in which the measurements were made tends to interfere with seeing what happens at various places over time as far as popularity and the fun ratings. On these aspects, comparisons between places are only meaningful in durations of 16 minute periods, since was as often as measurements were repeated. Though the automated activity measurements were continuous, relating them to popularity or fun calls for averaging the activity data over the 16 minute periods.³ Based on these considerations the data in Table 6.1-2 were reorganized in terms of 16 minute periods and the activity measurements averaged for that period. However, comparing the two tables with respect to the number of boys and girls and the fun ratings suggested a further modification from Table 6.1-2. The number of boys and girls is nearly the same and remains almost constant over the recess.⁴ Therefore, the Z values of these counts offer no particular advantage. The "bored - fun" ratings also seem as easy to interpret as actual ratings as Z-scores, since they are already +/- values relative to a neither boring nor fun. There in the next table, we'll revert to the original gender and fun measurements.

³ In comparing the averaged activity levels with the single measure of popularity or fun, we are implicitly assuming that the single measurements at each location are representative of that measure for the entire 16 minute period. That assumption is certainly not entirely correct. It's simply the best approximation possible with these data.

⁴ The totals are not quite consistent presumably because a few children might be late leaving the building, some leave the school yard to use the washrooms, and some might have moved from one place to another at the same time as the researcher. These are aspects of real world research you must accept. In this set of data the resulting inconsistencies in total number of children from one-third period to the next is much less than 5% so we need not be concerned about these comings and goings confounding the results.

Table 6.1-3
Data from Figures 6.1-1 & 2 Reorganized According to Place by 16 Minute Periods

16 MIN PERIOD	PLACE	BOYS	GIRLS	ALONE	TWO TALK	TWO PLAY	GRP TALK	GRP PLAY	RATE B - F	Avg ACTIVITY
FIRST	SWING	16	14	+1.53	-0.64	+1.66	-0.37	+0.12	2	+0.53
MID	SWING	9	17	+1.03	-0.05	+0.61	-0.37	-0.12	-1	+0.71
LAST	SWING	7	6	+0.02	+0.54	-0.44	-1	-0.87	0	-1.24
FIRST	HILL	13	4	+1.28	-0.64	-0.44	-1	-0.5	3	-0.33
MID	HILL	20	5	+1.28	-0.64	+0.61	+0.05	-0.37	4	+0.93
LAST	HILL	14	1	+0.27	-1.22	+0.61	-1	-0.25	2	-0.6
FIRST	FIELD	14	10	-0.99	-0.05	-1.49	0.47	0.62	0	-1.24
MID	FIELD	16	14	-1.24	-0.64	-0.44	-0.16	+1.75	3	+0.58
LAST	FIELD	15	16	-0.99	-1.22	-0.44	-0.37	+2.24	5	+0.66
FIRST	STEPS	3	20	-0.74	+1.12	-0.44	+1.31	-0.87	-1	0.75
MID	STEPS	2	13	-0.99	+1.71	-1.49	-0.16	-0.87	-3	-0.87
LAST	STEPS	12	24	-0.48	+1.71	+1.66	+2.57	-0.87	-1	+0.11

Note: Data not in bold are Z-scores. AVG ACTIVITY is the mean of four measures of activity obtained at 4 minute interval over the 16 minute period. Different measures were used in different places. See text.

Let's look for some patterns in these data. However, do not expect to see patterns immediately. This sort of looking is work. Since the typical short-term memory capacity is 7 ± 2 , small wonder that we just put all the information together at once. You have tools to work with: inductive reasoning - putting selected data together to see what comes out; and deductive reasoning - jumping to conclusions on the basis of a few data then examining more data to see if the hypothesis is warranted. With that in mind let's begin with factors on the left and proceed systematically across the table:

1. Looking at places, The Swings and rest of the playground were the most popular in the First Period. By Mid Period, popularity shifts to The Hill and The Field presumably due to the games developing at both places. In the Last Period, the children gathered at the Steps.
2. Looking at differences related to gender, The Hill and The Field are most popular with boys, while boys tend to avoid The Steps. Few girls are attracted to the tumble play on The Hill while The Steps are their most popular place.
3. Looking at social interaction, the Z-scores show that children tend to be alone mostly at The Swings and The Hill. Large negative Z-scores in the Alone column indicate The Field has the most group play. The Steps were the most popular place to talk either in pairs or groups.
4. The most fun was reported in The Field during the last period and at The Hill during the mid period. An "eyeball" summation reveals that The Steps were considered most boring. Both findings must be viewed with caution since they are based on data from only one child at each location at each period.

5. Because the level of activity was measured by different techniques at each place, it is not possible to accurately compare the absolute levels at different places despite the Z-transformations. (Note that the mean Z-scores at each place sum to zero - as they must one considers that they are calculated on the basis of the measurements at each place.)

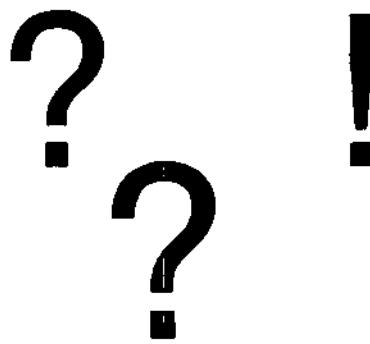
6.1.3 DISCUSSION

The shift in popularity from The Swings to The Field and The Hill over time could indicate that the attraction of mechanical things is short lived. Still The Swings increase in popularity and activity during the mid-period, and that is attributable to more girls playing there while many boys have left. The popularity of The Steps in the last period could indicate an eagerness to return to class. Note, however, that The Steps were sampled last. Being within a minute of the end of recess might inflate their popularity compared to places further from the entrance. Other explanations are also possible: The children could simply be tired and seeking a place to sit. The children could be cold and are attracted to The Steps because they are sheltered from the wind. The latter possibility indicates the importance of including data on the weather whenever a descriptive study is done outdoors,

The present data do not provide much insight into gender differences with respect to either social activity or fun. Rather than simply counting the number of boys and girls in each place, counting how many of each are alone, talking or playing would be helpful in further research. Similarly it would be helpful to have a fun rating from both genders at each place. Unfortunately these additional measurements would require two researchers unless more time is spent at each location. That however would reduce the visits to each place from three to two.

Another limitation in the present procedure is that it only permits linking fun to the places. To link fun to the type of activity would require Interviewing one alone child, one talking child, and one playing child at each location. Double that to be able to describe any interaction with gender. I hope this gives you some idea of the challenges involved in carrying out supposedly "easy" descriptive designs.

6.1.4 META-CONCLUSION (DON'T SKIP THIS PART)



One day, two researchers met a conclusion.

It's a "meta-conclusion", because unlike a research report, the conclusion to be drawn here is not about the research example but rather about what the example is meant to illustrate besides some specific techniques about doing descriptive research. By now it should be coming clear that this research could be improved in several respects if only more resources were available to obtain more measurements, to repeat them over several days, and/or for more researchers. This is particularly true for descriptive research since one does

not have sufficient information to know where to focus one's efforts. Combine that demand on resources with an inability to actually prove anything for a publication and you may understand why comprehensive descriptive research is not done more often.

6.2 MEASURING QUALITY OF LIFE

The previously mentioned Canadian Study of Health & Aging (CSHA) used a combination of various surveys followed up by psychological and medical tests to study the cognitive abilities of some 10,000 Canadian seniors. While a major objective was to describe the prevalence and incidence of Alzheimer's disease, the extensive amount of data on so many seniors presented many opportunities to describe other aspects of medical, psychological, and social factors pertaining to aging. The Prince Edward Island Centre for Health & Aging decided to see how data on quality of life were related to health and access to medical and social services (Clyburn, Clyburn & Nilsson, 1995). It was a descriptive study in that we had no particular independent variable which would be varied. We were aware of possible differences between urban and rural dwelling seniors in access to services. We wondered about the influence of economic and social factors, but we did not select participants on these bases. Our plans were simply to look and see what types of answers were related to quality of life and what factors were not related.

6.2.1 METHOD

We recontacted a random sample of 150 of the 500 persons who had participated in the 1990 study on the Island and expressed willingness to participate again in such research. In this interview we repeated a number of the questions about quality of life used in the CSHA and added more questions that had been used by the Manitoba Centre for Aging for a similar purpose. The latter included a number of questions about seniors' use of medical and social service. Examples of these questions are shown in Table 6.2.1-1.

Table 6.2-1
Examples of Questions Used to Survey Seniors About Quality of Life

As you know, I'm going to spend some time talking with you about your health and general circumstances. I will begin by asking you some questions about you and your family. We would like to get a sense of the family networks and support available to Canadian seniors. I will then ask a few questions about routine activities that you do. As we said before - all of this information is confidential. Do you have any questions before we start?

1. How is your eyesight (with glasses or contacts if you wear them)? Is it excellent, good, fair, poor, or are you completely unable to see? (Cue)
 1 Excellent 2 Good 3 Fair 4 Poor 5 Unable to see 6 R
 5. People often have one or more individuals they can count on for help and support. Can you think of someone like this in your life?
 1 Yes 2 No 3 R 4 DK
 IF YES: How many such people? Number: _____
 9. How many people do you spend time with on a regular basis?
 (At least once a month; in your house or theirs; do not include phone conversations)
 Number of people: _____ 0 No-one 77 R 88 DK
-

11. Activities of Daily Living QUESTIONS

Now, I would like to ask you a few questions about things that we all need to do as a part of our daily lives. I would like to know if you can do these activities without any help, or if you need some help to do them, or if you can't do them at all. Please tell me about your situation today when answering these questions.

- A. **Can you eat . . .** (Cue)
 2 without any help? 1 with some help? 0 or are you completely unable to feed yourself?
- L. **Can you go to the bathroom . . .**
 2 without help? 1 with some help? 0 or are you completely unable unless some-one helps you?

THE 3MS (a screening test for cognitive impairment)

Now I am going to ask some questions of a different kind. Some of the questions that I ask you will be easy; others may be more difficult. They are all routine questions that we ask of everyone. I may also ask you the same question twice. Just answer all of them as best you can.

14. COUNTING and SPELLING BACKWARDS

- /7 **COUNTING FORWARDS** Can Can't
 5 to 1 (write their answer _____)
 5 4 3 2 1
- SPELL "WORLD"** Can Can't
 "World" backwards (Print letter _____)
 D L R O W

25. THREE-STAGE COMMAND

- /3 **Take this paper with your...**
 left/right hand, 1 0
 fold it in half, and 1 0
 hand it back to me 1 0
 Not completed: Physically unable 66

36. Now, I would like to ask you a few questions about how you have been feeling over the past week.

- D. **Do you prefer to stay to stay at home, rather than going out and doing new things?**
 1 Yes 2 No
- K. **Do you think it is wonderful to be alive now?**
 1 Yes 2 No
- O. **Do you think that most people are better off than you are?**
 1 Yes 2 No

Now I am going to ask you a few questions about your activities

37. **Last summer, how often did you...** (Cue)
- a) **go to visit friends or relatives?**
 1 Often 2 Sometimes 3 Never 7 R 8 DK
- d) **work in the garden, yard work?**
 1 Often 2 Sometimes 3 Never 7 R 8 DK

We are interested in how things are going these days. Please answer the following questions "yes" or "no". During the past month have you ever felt:

38. Depressed or very unhappy?

Yes No Other R DK

43. Generally satisfied with the way your life has turned out?

Yes No Other R DK

The next questions have to do with more general life experiences

44. I am just as happy as when I was younger

Yes No Other R DK

45. As I look back on my life I am fairly well satisfied

Yes No Other R DK

48. Life is hard for me most of the time

Yes No Other R DK

Service Utilization - Have you used any of the following services in the past 6 months?

If yes, how often? (code actual number of times)

SERVICE	FREQUENCY	SERVICE	FREQUENCY
General Physician	_____	Pharmacist	_____
Specialist MD	_____	Optometrist/Optician	_____
Hospital Emergency	_____	Nutritionist/Dietician	_____
Day Program	_____	Audiologist	_____
Walk-In Clinic	_____	Community Health Nurse	_____
Social Worker	_____	Clergy	_____
Veterans Affairs	_____	Home Care	_____
Dentist	_____	Fitness Programs	_____
Chiropractor	_____	Mental Health Clinic	_____
Physiotherapist	_____	Podiatrist	_____

75. How well do your income and assets currently satisfy your needs?

1 - completely adequate

2 - somewhat adequate

3 - somewhat inadequate

4 - totally inadequate

9 - not answered

76. a) If you had some additional income, would you spend it on any of the following?

01 - better housing 05 - more or better furniture

02 - housing repairs 06 - medical needs _____

03 - more or better food 07 - recreation

04 - more or better clothing 08 - transportation or new car

05 - travel 10 - other _____

b) Which of the above would you do first? _____

I have enjoyed talking with you. In this study I have talked with many seniors and learned something from each one of them. In particular, many have commented on the sort of things that have contributed to their long life. What do you think makes people live long and keep well?

6.2.2 RESULTS

Simply counting the responses to various questions provided information about many aspects of seniors lives. Here are some: 45% lived alone; 58% had some to help living with them or within walking distance; 79% reported themselves as being in good or excellent health; yet only 13% reported two or less health problems, while 63% reported five or more; 84% were at least satisfied with their finances, but 59% reported incomes of less than \$20,000 per year. Correlations of many answers with life satisfaction are listed in Table 6.2-2.

Table 6.2-2
Correlates of Life Satisfaction for Prince Edward Island Seniors

VARIABLE	r	cases	VARIABLE	r	cases
HEALTH SATISFACTION	+.44*	147	INDEPENDENT HOUSING	+.15	143
PERCEIVED HEALTH STATUS	.39*	146	NUMBER OF CHILDREN	+.15	143
BEING MARRIED	.30*	73	RELIGIOUS SATISFACTION	+.13	138
SATISFACTORY RECREATION	.30	142	INCOME LEVEL	+.07	135
GOOD FAMILY RELATIONS	.28*	142	SENSE OF CONTROL	+.07	142
GOOD TRANSPORTATION	.27*	141	FREQUENCY OF ACTIVITIES	+.06	143
SENIORS LIVING NEARBY	.26*	140	GENDER (male)	-.01	143
SATISFACTORY FRIENDSHIPS	.25*	141	YEARS OF EDUCATION	-.05	142
SATISFACTORY HOUSING	.24*	143	LIVING ALONE	-.15	143
NUMBER OF HELPERS	.23*	138	MOVEMENT PROBLEMS	-.19	111
FINANCIAL SATISFACTION	.19*	143	NUMBER HEALTH PROBLEMS	-.30*	98
SATISFACTION WITH INCOME	.16	142	ACTIVITY LIMITATIONS	-.31*	98

Note: * identifies significant correlation

6.2.3 DISCUSSION

Several aspects of the data revealed strong ability of the seniors to generally overcome factors that were associated with specific limitations. Life satisfaction correlated only +.07 with income level, which is fortunate considering that 59% had incomes below \$20,000/year. This must contribute to the ability of a full 84% of Island seniors to report that their life was satisfactory or better. Health was the most important factor to their life satisfaction. Nevertheless 87% reported having 3 or more health problems. Clearly an adequate health care system is critical to seniors being able to maintain a satisfactory life despite the limitations in their lives and incomes. "Reforms" to the health care system that ignore these findings may create substantially greater demands to maintain seniors quality of life by other means.

6.2.4 META-CONCLUSION - 2

Persons in government may be more influenced by descriptive designs which don't involve the more theoretical interpretations possible with statistical tests of experimental data. Plain facts speak loud and clear to every one. Psychologists should be aware of the potentials of their research to influence public policy and thereby improve the quality of life in our society.

6.3 PERSONALITY AND FRIENDSHIP

Like astronomy and geology, the study of personality is a descriptive science. You can not (ethically) manipulate personality to see its effects on human thought and behavior. You can only carefully observe (that is measure) people's behavior and describe relationships between behaviors.

The Myers-Briggs Type Inventory (MBTI) is probably the most popular descriptor of personality in the world. It is based on how people answer 126 questions about their general behaviour, preferred expressions, and social interactions. Factor analysis has shown that everyone tends to answer these questions in patterns which can be organized into four dimensions that represent opponent personality types. These dimensions include the well known introversion-extroversion distinction, plus feeling-thinking, sensing-intuitive, and judging-perceiving. Various combinations of preferences among these bipolar dimensions result in 16 basic personality types. After that interaction between dimensions and strength of preference within each dimension make matters more complex - appropriately so given the tremendous complexity of human personality. These interactions distinguish the MBTI from another well known personality test, the Minnesota Multiphasic Personality Inventory, popularly known as the "MMPI" and widely used by high school guidance counsellors.

Previous research had shown that persons having different MBTI characteristics tending to differ systematically in how they interpersonal conflicts (Johnson, 1977). This suggested to Percival and MacBeth (2001) that different personality types might also differ in their perspectives on friendship. To measure how people think about friendship, Percival developed a "Friendship Maintenance Questionnaire" (FMQ). The research consisted of giving both questionnaires to many people and then seeking relationships between respective scores.

6.3.1 METHOD

The MBTI and FMQ were administered to 125 volunteers who were residents of Prince Edward Island, 20 to 70 years old, and had various occupations though 15 were unemployed and 4 were retired. Tables 8.3-1&2 illustrate the kinds of questions used on these tests.

Table 6.3-1
The Sorts of Questions Asked on the Myers-Briggs Type Inventory

PART I. Which Answer Comes Closer to Telling How You Usually Feel or Act?

- | | |
|---|---|
| <ol style="list-style-type: none"> 1. Do you prefer to be with <ol style="list-style-type: none"> a) imaginative people, or b) realistic people?
 2. At a party, do you tend to <ol style="list-style-type: none"> a) talk with one person at a time, or b) join in group discussions? | <ol style="list-style-type: none"> 25. Do you generally <ol style="list-style-type: none"> a) show how you feel, or b) keep your feelings to yourself?
 26. When travelling, do you like <ol style="list-style-type: none"> a) to map out the route, or b) head to what looks interesting? |
|---|---|

PART II. Which Word in Each Pair Appeals to You More?

27. a) absolute approximate (b) 70. a) what why (b)
 28. a) party sport (b) 71. a) friendly considerate (b)

PART III. Which Answer Comes Closer to Telling How
You Usually Feel or Act?

72. As a supervisor Is your first concern
a) to meet production standards, or
b) satisfy customers?
 73. Given a choice to work on two projects,
would you prefer the one which,
a) has exciting possibilities, or
b) will definitely advance knowledge?
 125. A judgement should be
a) fair
b) be compassionate
 126. Is it better to
a) get it right the first time, or
b) do it quickly then make
adjustments till it's correct?
-

Table 6.3-2
Examples of Questions on the Friendship Maintenance Questionnaire

Each of the following statements presents a way of maintaining a friendship. The statements answer the question, "What makes a friendship last?" Read each statement, and mark a check in the box that indicates your degree of preference for that way of maintaining a friendship.

1. I prefer to accept and respect my friend just as he or she is, without trying to change my friend in any way.

[]	[]	[]	[]	[]
no preference	slight preference	moderate preference	strong preference	very strong preference

2 I prefer to share intimate thoughts and feelings with my friend.

[]	[]	[]	[]	[]
no preference	slight preference	moderate preference	strong preference	very strong preference

20. I prefer to choose friends who make definite plans and keep them.

[]	[]	[]	[]	[]
no preference	slight preference	moderate preference	strong preference	very strong preference

The MBTI scores were used to select 80 persons whose scores on the various dimensions most clearly identified them in terms of one of the 16 possible personality types, with 5 persons of each type. For the each FMQ question, a point-bi-serial correlation was calculated between the scores on that question and scores on each of the four opponent-personality dimensions. There were no significant correlations with the sensing-intuitive scores

so the data were collapsed across this dimension, which resulted in 10 persons for each of the 8 remaining combinations of personality type. Analysis of variance followed by Duncan's means comparison test then showed where there were significant differences FMQ mean scores amongst the 8 personality types.

6.3.2 RESULTS AND DISCUSSION

The most evident and significant differences in relationship between personality preferences and friendship preferences occurred for the extroverted-introverted and the thinking-feeling dimensions. Persons strong on the extroversion and feeling directions emphasized the importance of harmony and were willing to maintain that harmony by accepting their friends' different values. Persons expressing extroversion and thinking preferences felt that fairness and reciprocity were most important for friendship. Introverts with feeling considered having common values with friends to most important. Introverts who expressed thinking preferences felt that trust, respect and tolerance were the critical factors in friendship.

6.3.4 META-CONCLUSION - 3

The personality and the quality of life examples show the tremendous of amount of data analysis that can be required by descriptive research designs. Such extensive analysis are needed to find which relationships turned out to be interesting when the independent variable has not been predetermined in the design. They represent one side of a trade-off faced by all researchers attempting to good research. One either puts effort into producing different levels of an independent variable, or one collects a great deal of data with less preparation and does more analysis afterwards.

6.4 IN THE LABORATORY: READING IN COLOR

So far we have considered descriptive designs that emphasized observational and questionnaire measurements and that could be conducted in community settings. Descriptive research is also done in laboratories with all the attendant apparatus needed for measurement and control. Indeed it was descriptive research on rats during the 1960's that started a revolution in psychology. Observations of rats when they were not pressing bars or running mazes to obtain food led to the "discovery" that animals were not just driven by hunger and sex, but were also acted out of curiosity and developed what could be called concepts. In a decade these ideas combined with research on human memory for pictures and on perception to produce cognitive psychology.

The following describes an applied, descriptive study. There was no expectation of any theoretical impact. The only thing unusual was making the observations more representative of consumer behavior by using a non-conventional measurement. As sometimes happens with non-conventional methods, the results lead to a theoretical breakthrough in understanding color vision.

Some years ago the Health Canada discerned the need for more forthright warnings about the dangers of smoking. Based on some descriptive studies by consultants, phrases like. "Smoking causes cancer" were to be put on cigarette packages in big, bold letters. To placate the tobacco companies, it was agreed that the warnings could be printed in colors that harmonized with the package design. Months later Health officials were chagrined to find on the market new cigarette packages with the warnings printed in gold on cream colored backgrounds, in silver on bright red backgrounds, etc. When they complained to the companies that these warnings were hardly legible, they received a reply that legibility was a

matter of opinion. The companies claimed the warnings looked perfectly legible to them. They knew there were no standards for the legibility of colored letters on colored backgrounds. So one day Health Canada called and asked if I could think of a way to measure the legibility of colored letters on colored backgrounds. The method would have to withstand a court challenge of its validity. As a "closet" descriptive scientist I was intrigued.

6.4.1 METHOD

With 16 million colors, spectrophotometric criteria were clearly not the answer. The criteria would have to be based on perception. The method of choice for measuring legibility in vision laboratories used the minimum time needed to recognize words. Not only would such apparatus be difficult to define in legal terms, but it could rightly be argued that such criteria were not representative of reading in the marketplace. Therefore I suggested defining legibility in terms of the maximum distance at which a warning could be read. A legibility track was built (see Figure 6.4), and some preliminary research demonstrated that the method produced results that were consistent for different observers and consistent with subjective impressions of legibility.

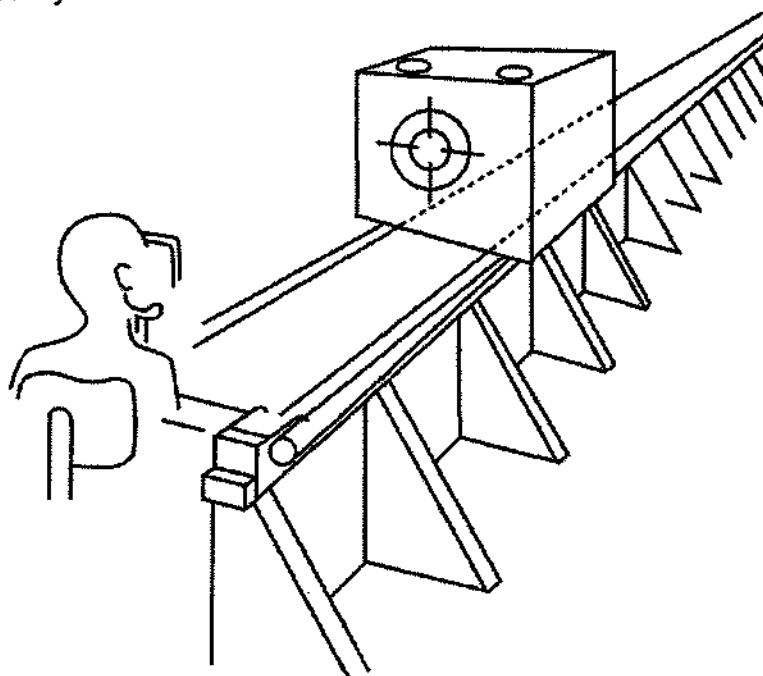


Figure 6.4 Schematic drawing of the legibility testing track.

The initial testing used three subjects, who were screened for normal color vision and normal or corrected to normal acuity. Several practice sessions were held to acquaint them with the measurement procedure. (Subsequent research with two groups of 40, unpractised subjects using different instructions produced similar results.)

A subject was seated at one end of a 7.3 meter track that carried a large box containing two, shielded, 100 watt, light bulbs which uniformly illuminated a cigarette package that could be seen through a view port. A computer controlled motor moved the box at uniform speed and sensed its position within a millimeter. We tested eleven commercial packages with the same warning, "Smoking during pregnancy can harm the baby" in various color combinations. Since this was a **within-subjects** design, the packages were tested in **quadratic counterbalanced orders** to avoid effects that arise when one item is tested before another item. (The terms are explained in Chapter 8.)

Maximum legibility distance was measured using the Method of Limits. With the box close enough for the warning to be easily read, the subject pushed button send the box down the track. When the subject could no clearly see the message, a second button told the computer to measure the box's distance and to let it travel a random distance further. Now the subject signalled the box to approach and indicated when the message could first be clearly read. This procedure was repeated six times, then the next package was tested.

6.4.2 RESULTS

Table 6.4 shows how color and lightness contrast of the letters and background affected the distance at which the same message could be read.

Table 6.4
The Various Brands, their Messages, Nominal Colors, Mean Legibility Distance,
and Relative Legibility at 90° Viewing

BRAND	COLORS OF WARNING	LEGIBILITY DISTANCE	SD	DUNCAN'S RANGE	RELATIVE LEGIBILITY	LIGHTNESS CONTRAST
Sweet Caporal	BLACK / WHITE	127	1.1	1	1.00	82
B & H Lights	GREEN / SILVER	117	7.6	2	.86	62
MacDonald Select	SILVER / CREAM	117	4.1	2	.86	23
Peter Jackson	BLACK/WHITE	115	6.9	2-3	.83	62
Matinee Slims	GOLD / WHITE	111	5.8	3	.77	32
Medallion	GOLD / YELLOW	109	8.6	3	.75	34
Peter Jackson	WHITE / SILVER	109	5.2	3	.73	22
Player's Special	GOLD / DK. BLUE	104	2.7	4	.68	52
Player's Medium	GOLD / WHITE	103	7.8	4	.66	35
MacDonald Export	GOLD / GREEN	103	4.4	4	.66	10
B & H Special	BLACK / GOLD	98	4.9	5	.60	25
Player's Plain	GOLD / BLUE	91	0	6	.52	02

Notes: All brands had the same warning, "Smoking during pregnancy can harm the baby", printed in the same font. Legibility distance is in cm. The Sweet Caporal warning had more space between its letters.

When legibility was operationally defined in terms of maximum reading distance, well-spaced black letters on white background were significantly the most legible. Lightness contrast of the letters and their background correlated +.76 with legibility distance. However, when the Sweet Caporal warning with its extra-spaced font was omitted this correlation dropped to +.62. Duncan's range test indicated that six groups of packages differed significantly in legibility distance. Yet, the best warnings seemed easier to read than the worst by a greater amount than indicated by the distance difference. The reason for this discrepancy becomes evident when one considers how distance affects the size of the retinal image. Image area is proportional to the number of neural pathways needed for perception, and image area decreases with distance squared. Therefore, I calculated how the image areas of each brand compared to the image area of the most legible warning. These results are listed as RELATIVE LEGIBILITY in Table 6.4. In

terms of relative legibility the least legible warning was about half as effective as the best - a difference more consistent with subjective impression.

6.4.3 DISCUSSION

The results showed that legibility of the various warnings could be statistically differentiated into six ranges based on legibility distance. The evidence that legibility depended on lightness contrast was consistent with the most recent legibility research based on reading speed (Knoblauch, Ardit, & Szlyk, 1991). It was strong enough to convince Health Canada that the warnings should be printed in black and white as they are today. Evidently it was sufficient for the tobacco companies too. They never challenged it.

This study also demonstrated the feasibility of defining legibility directly in terms of human perceptual performance. Given a standard sample for comparison, the simple technology of measuring legibility distance means that the legibility of any warning could be tested directly rather than trying to provide guidelines for an astronomical number of color and font combinations. Legibility distance has another advantage over minimal reading time measurements. Legibility distance can be directly related to visual acuity as commonly measured using the 20/20 system, which is also based on distance.⁵ This enables legibility criteria to be established relative to an accepted range of visual acuity of the general public such as 20/50 acuity which is required for drivers' licenses. It is hoped that this research will lead to legal standards for legibility of all warnings, including instructions on medications and safety advice at work sites.

There were, however, some disquieting effects in the data. That correlation of +.62 between lightness contrast and legibility was much lower than what Knoblauch, et al (1991) found. It meant that less than half of the difference between the legibility means was attributable to lightness differences. With the effects of font differences essentially removed, the only other difference in the warnings was their colors. Data obtained by Knoblauch et al (1991) and all the previous studies they cited showed that color had no effect on legibility - once colors were corrected for lightness contrast. Yet my data showed that the standard black/white Peter Jackson warning was no more legible than 5 other warnings which used green, yellow, gold, and silver. Except for the B&H Lights, all of these had substantially less lightness contrast. These four were also significantly more legible than Players Special, which had more lightness contrast. Evidently legibility distance measurements produced results that differed from reading speed measurements. Was there some flaw in the distance measurement technique? Could all previous studies be incorrect about the effect of color on legibility?

An opportunity to do some descriptive research for the Canadian Space Agency answered one of these questions (Nilsson & Connolly, 1997). It involved systematically measuring how the color of words symbols and line drawings affected recognition. There was no doubt about these results. When corrected for lightness contrast, color had a major effect on legibility. Why had this not been found previously? Once such clear data were available, the explanation became obvious. Almost all previous legibility research had used some measure of legibility that depended on reading speed. Yet, physiological studies of the visual system has revealed that color information is conveyed by slow parvo-cellular neurons, while lightness information is conveyed by fast magno-cellular neurons. Therefore, when legibility is defined in terms of speed, only the information about lightness gets used.

⁵ "20/20 vision means being able to see at 20 feet what the average person can see at 20 feet. A person with poor acuity such as 20/200 vision can see at 20 feet what the average person can see at 200.

Some years later an invitation from NATO presented an opportunity to check the validity of the distance measurement method by comparing its results with the best methods anywhere. After a single training session, five UPEI students were able to distinguish the effectiveness of more examples of camouflage in terms of recognition distance than were 60 NATO trained observers using recognition speed (Nilsson, 2001).

6.4.4 META-CONCLUSION - 4

Descriptive research is often regarded as a more menial type of research than experiments testing hypotheses. Practical but boring. Fine for those lacking in theoretical ability. No wonder it is a well kept secret that a lot of laboratory research resembling experiments is at heart descriptive. Yeah sure, these studies may have one or more independent variables, but that's because lab researchers usually can't obtain various levels naturally the way you can find different play areas in a school yard or persons of different personality. What makes such research descriptive is that it is not geared to testing a specific hypothesis, but rather to describe what occurs under a broad range of conditions. From such data the researchers may inductively arrive at some theories or discover some as-yet-to-be-explained effects. As happened in a few rat labs in the 60's, Psychology needs more descriptive research to go beyond the confines of present theory. Unfortunately, descriptive research also tends to be more expensive than little hypothesis testing studies. It does not thrive in times of cutbacks in research and threats to tenure.

There is a second lesson here. "You can lead government to water, but you can't make drink more than the minimal amount needed to solve an immediate problem." There has been no follow-up to suggestions for the need to establish legibility standards.

That's telling 'em, Nilsson. But is it really appropriate for students know? - The Editor

You betcha! - Nilsson

6.5 THE CASE-STUDY APPROACH

A case study is one in which the researcher chooses to study a single person or group extensively as compared to studying many persons in a more limited manner. Data may be collected by direct interviews, by observation, from writings and other records made by or about the person or group. The approach has been more typical of psychological research with a clinical focus and general research done in Eastern Europe as compared to the laboratory and social research traditions of North America.

6.5.1 AS A SOURCE OF IDEAS

Each of us is a "case" in that we are unique and special in various ways. Since all ideas arise as consequence of a long history of our own experiences, case studies are always the source of research ideas. Keep that in mind even as we consider a more specific use. The typical research project is usually well focused; the conditions are controlled as much as possible; the dependent variable is defined to answer a specific question. In these circumstances, any differences from the typical result is regarded as error variance to overcome by averaging. In focusing on a single individual and collecting as much information as possible rather than just what is needed to answer a specific question, case studies can reveal details about human behavior that tend to be overlooked by the other research designs. When such details are followed up, they sometimes lead to insights and new leads that might otherwise have

taken years to emerge from a series of studies testing various hypotheses. This brings up an important distinction in doing research: deductive versus inductive approaches.

Deductive research is when one starts with a hypothesis about what will happen and then proceeds to do research to determine whether the hypothesis is true. This is the approach emphasized in most introductory statistics books. A simple view of statistics is possible when research is designed so that a difference between means supports some hypothesis. Then all that is required of the statistics is to disprove the "null hypothesis" by showing that the difference is unlikely due to chance. However, once one gets beyond the t-test stage, the multiple comparisons possible with analysis of variance techniques enable broader perspectives to be rigorously pursued. One can take a "let's see what happens" approach. **Inductive** research is when one makes a number of observations. Various descriptive and analytic statistics are then used to help develop a hypothesis to explain the results. In reality, science proceeds following both approaches - sometimes simultaneously, often alternating with no clear distinction of "which came first."

Case studies exemplify the inductive approach. They are often initial research efforts in new areas, where there are no hypotheses. Because the clinical areas of psychology are the most complex and least understood, examples of case studies continue to present us with new examples of behaviors that will eventually lead to a better understanding of the mind.

6.5.2 THE REPRESENTATIVE CASE

The principle reason for collecting data from many persons is to ensure that the results are not unique to a particular person whom we happened to study. Yet it is not necessary to study thousands of people to gain a reasonable understanding of many human behaviors. The influence of individual differences depends greatly on what aspect of human behavior is being studied. All thoughts and actions are ultimately traceable to chemical changes and nerve impulses. Our bodies are 99.9% genetically identical to that of chimpanzees. The chemical reactions that enables our neurons are in most aspects identical to that of the simplest organisms with nervous systems. Therefore studying neural functions even in a mollusk can yield momentous insights into how all human nervous systems work. Basic functions such as color sensation and short-term memory are mediated by neural processes which are common to all persons regardless of ethnic origin. Provided that the subjects do not have some specific abnormality such as color-blindness or agnosia, it is reasonable to study basic aspects of human perception and memory based on the reports by one or two cases.

For certain types of knowledge such as how cognitive functions emerge in infancy or how social movements develop, the researcher may require enormous amounts of information over a long time to gain a new insight. It simply would not be possible to repeat such observations many times with different individuals or groups. Nevertheless, on the basis of other background information and broad experience, the researcher has good reason to believe that the data from one or a few cases represent an effect that occurs widely. Other sources of information and/or reason may be used to support this contention.

Based on years of clinical practice, Abraham Maslow felt that people occasionally were motivated at a level that did not involve needs like hunger, safety, love, or fame. When motivated in this manner, people seemed to function at their creative and productive best. To learn more about this motive, Maslow reasoned that the most effective way would be to study persons who functioned at this level more often than most people. He felt that such people could be recognized by their outstanding accomplishments. "If you want to know how fast a human being can run, then it is of no use to average out the speed of a 'good sample' of the population: it is better to collect Olympic gold medal winners and see how well they can do." (Maslow, 1970)

Accordingly Maslow set about interviewing and studying the biographies of especially creative and productive persons like Albert Einstein and Eleanor Roosevelt. He supplemented this with studies of personally known people (including students) who also seemed motivated in this manner. He made it clear that was not seeking something unique to famous people but a motivation potential in all people. "The highest possibilities of human nature have practically always been sold short. Even when 'good specimens' have been available for study, the temptation too often has been to consider them not human but supernaturally endowed." (Maslow, 1970) Upon summarizing the information from his case studies, Maslow was apparently to find many common characteristics in people functioning at this level despite their diverse backgrounds and achievements. Maslow (1970) lists seven characteristics which are summarized in Table 6.5. This led him to propose a single, new type of motive he called "self actualization."

Table 6.5
Characteristics Maslow (1970) Found Common to Self Actualized People

-
1. Creativity and inventiveness
 2. Problem centered rather than ego centered. Capacity for concern about larger problems of society and humankind.
 3. They feel that their life has a purpose.
 4. Objectivity and detachment; acceptance of self and others.
 5. High tolerance of the unknown and ambiguity.
 6. Mystical or peak experiences: recognizing events that are of a special quality in that they serve to organize and give direction to one's life.
 7. Freedom from prejudice and cultural conventions; an unconventional morality about what is right and wrong.

6.5.3 AN EXCEPTION TO "PROVE THE RULE"

The old saying about exceptions, is frequently misunderstood because in older usage "prove" means "to test." For example, until about 1970 laboratory research on human thought had been largely confined to methods that had grown out of Ebbinghaus' 1888 experiments on memory using word lists. It was long argued that language was a unique type of symbolic processing that made it possible for humans to employ reason or "think" (Whorf, 1958). Since only humans possess the vocal apparatus and special brain areas for the production and understanding of speech, they were the only animals considered capable of thought involving more than simple declarations such as, "I hunger;" or "Stay away." To challenge this theory, a few researchers raised chimpanzees from birth with their own family to give a chimp every opportunity to acquire linguistic concepts the way human infants do. Since a chimp's vocalization repertoire is too limited to convey complex thoughts, they taught the chimps to communicate symbolically with gestures and tokens as do humans born mute. These exceptions to Whorf's rule demonstrated that verbal processing was not necessary to be able to reason, recognize abstract concepts, and ask questions. Thus case studies of two chimps, Washoe and Sarah, provided the exceptions that left the Whorfian hypothesis untenable. This break-through directed attention to research on human thought that did not involve words. The result spurred the revolution which produced cognitive psychology.

Case studies are a strong tradition in clinical research. No doubt our prehistoric ancestors relied on case studies of unusual behaviors like illness to arrive at remedies within their means. "Six months ago, my sister had terrible pain in her belly here and here. It went away then came back worst a day later. Third uncle had similar pains before he died last summer. I was

desperate. I told her to chew on the leaves of this plant because their peculiar taste seemed right to me. The next day she was feeling much better and the pains did not come back." That studies of exceptional clinical cases have contributed greatly to understanding human disorders is attested by the names of those who reported the first cases - eg. Alzheimer, Jacob-Creutzfeldt, Parkinson, etc. Following this tradition, Freud's writings show how strongly his theories were influenced by relatively few case histories of his patients. Unfortunately Freud did not realize that his cases were indeed exceptions, and he used what he learned from them to create rules rather than test them.

Case studies of persons with Williams disease illustrate a more positive use of rare clinical cases. In 1961 a heart surgeon, J.C.P. Williams reported that several of his child cases had not only some specific cardiovascular abnormalities in common but also an unusual facial appearance: a turned up nose and a small chin. They were also mentally retarded. Subsequent case studies of similar children revealed some unusual mental characteristics indeed (Lenhoff, Wang, Greenberg & Bellugi (1997). They generally have below average IQ's with particularly poor reading, writing, and mathematical abilities. They have very poor drawing skills (see Figure 6.5) but are unusually adept at recognizing faces. Their speech skills are superb. Many can extemporaneously speak in rhyme. They evidence attention deficits on most tasks except tasks involving music. Though unable to read music, they display extraordinary musical talents. An unusual number have perfect pitch and can play and sing complex compositions from memory having heard them only once. Unlike most savants, their interpersonal skills are outstanding. "As a group, they tend to be empathetic, loquacious, and sociable." (Lenhoff, et al, 1997, p.68)

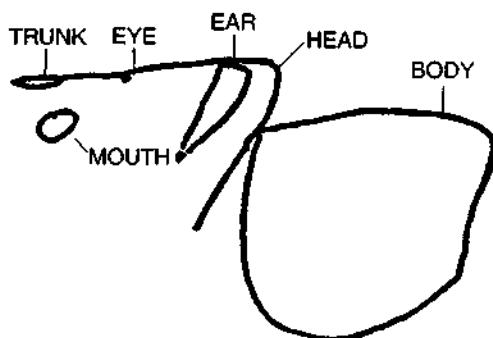


Figure 6.5. Drawing of an elephant by a teenager with Williams syndrome, who could verbally describe an elephant in great detail at considerable length.

There is evidence of historical reference to people with these characteristics in folk tales handed down for generations. Their behavior bears strong resemblance to persons identified as elves - kindly, enchanting story tellers and musicians. Such enduring and specific characteristics would seem likely to have a genetic basis. Genetic analysis has indeed uncovered a deleted segment in chromosome 7 of 95% of persons with Williams syndrome. One of the some 15 genes within that segment helps produce elastin - protein that helps give various tissues structure. That lack is consistent with their facial characteristics, cardiovascular abnormalities and other organic disorders that persons with this syndrome are prone to develop. However, the lack of elastin does not explain the remarkable cognitive abilities and deficits. (Individuals with genetically linked elastic deficits alone do not have such cognitive abnormalities.) The complete human genome has not been mapped, but it will take decades, probably centuries before we understand what the various genes do. Detailed study of special cases like persons

with this very rare syndrome will play a major role in deciphering the links between mind and genetics. There are momentous opportunities for psychologists in such research.

6.5.4 META-CONCLUSION - 5

In summary, the case study may seem attractive to novice researchers because it has the potential of breaking stereotyped ideas about research, can be personally very rewarding in its one-on-one relationships with the subject or group being studied, and dispenses with the rigors and constraints of other research methods. Yet case research is doubtlessly the most difficult type of research to do well.

Doing research without predefined measurements leaves one vulnerable to a host of potential problems:

1. Getting sidetracked from the original goal and ending up with a superficial, hodgepodge of information.
2. When many questions are pursued and large amounts of data collected without bounds, it becomes increasingly likely that some unusual and apparently profound effect will occur due entirely to chance. The problem is distinguishing chance effects from substantial ones. You likely won't be able to tell without repeating the case study.
3. Because the synthesis of large amounts of unstructured data involves considerable subjective judgement, it is easy to fall into the trap of arriving at some hypothesis prematurely, and then become prejudiced to seeing only data that support the hypothesis and overlooking contrary data.

Case study designs require a tremendous depth and breadth of knowledge, an extraordinary discipline, and an ability to see relationships amongst diverse bits of information that borders on the quality of genius. It is definitely not the design of choice for novices, and even geniuses can be led astray. Had Freud had a greater breadth of knowledge, he probably would not have created such warped, ideosyncratic theories.

EXERCISES

Based on the School Yard Study:

1. Rearrange a subset of the data to answer the following question: Did children play alone less as the lunch recess progressed?
2. The fact that The Swings were always tested first and the steps last during each period led to questions about the meaning of the measurements. Outline a testing schedule for an extended study which would have the measurements repeated over several days to solve this problem. Create hypothetical data just for the number of boys and girls in each place to illustrate how your procedure works.
3. Make a graph to show how average activity varies as a function of time over the lunch recess. Develop a hypothesis to explain this function.
4. How closely is fun related to the level of activity?
5. The measurements of activity level are confounded by another factor that was also measured.
Identify the confounding factor and create a new column which corrects this problem for Table 6.1-3.

Chapter 7 BETWEEN-SUBJECTS EXPERIMENTS

7.1 WHAT'S SO SPECIAL ABOUT "EXPERIMENTS"

For many persons scientific research is synonymous with experiments. That is not true. The essence of scientific research is using measurements to obtain knowledge - preferably taking those measurements in a systematic manner. Thereby astronomy, geology, and descriptive psychological research are scientific but not experimental. In science the word "experiment" has a more specific meaning. Here is how experiments differ from descriptive research:

7.1.1 THE INDEPENDENT VARIABLE

In experiments researchers obtain different levels by producing them or by assigning subjects to different existing levels. Descriptive research generally does not have independent variables. Indeed, in certain disciplines such as archaeology and ethology considerable effort is taken to avoid adding an independent variable into the measurements.

ADVANTAGE: When you impose the levels on groups of subjects who you know (from using matching or randomization - see below) to be equal in all other respects, you can be certain that any changes in the results are caused by the change in levels of the independent variable.

(The significance of that word "cause" can not be emphasized too strongly. Experiments are the only way to actually prove that something has caused something else. Proof of cause is the surest form of knowledge we have. By this means in a few hundred years scientists have been able to provide the knowledge that has made it possible to go from speculation about atoms to technology, from herbal remedies to medicine, from brain anatomy to grasping for the meaning of consciousness.)

DISADVANTAGE: More work is required to produce the levels of an independent variable than to simply accept existing conditions. Alternatively, it may not be possible to assign subjects to certain conditions such as the number of siblings, where they went to school, their annual income, favorite sport, etc.

7.1.2 DEPENDENT VARIABLE

Experiments can use the same types of measurements as used in observational, survey, and trace studies. (For example, one could vary the noise level or the cleanliness of study carols and measure how this affected the amount studying in terms of traces such as the number of books left in the carol. One could conduct a survey on how motivation and fatigue were affected by working at different temperatures in an environmental test chamber.) Additionally, experiments can use a vast variety of other measurements which usually require a laboratory.

7.1.3. CONTROLLED VARIABLES

The greater involvement of the subjects usually permits the experimenter to control (by holding factors constant or balancing factors) the situation in which the measurements are obtained. This ensures that these factors will not confound the results.

What to control? It is not feasible to hold constant or match all conditions in an experiment. It is less important to control factors which do not change systematically as the levels are changed. They simply add variability, but not bias. Discretion based on general and specific knowledge is needed to identify which variables need to be controlled.

7.1.4. SUBJECTS

While it's desirable to have subjects representative of humanity in general or even representative of certain groups such as the freshmen at a certain university who are taking an introductory psychology course, this is usually not attempted in experiments. There are 2 reasons:

1. Experiments usually require more time and effort per subject than other designs. This generally precludes (due to lack of resources) testing enough subjects to ensure that they are representative.
2. Enough is often understood about the behaviors measured in experiments to be able to relate them to physiological and cognitive processes and permit the screening subjects for "normality".

7.2 OBTAINING SUBJECTS

Before you can ask people to be subjects in your experiment, you need to obtain approval from your teacher, supervisor or departmental ethics committee. They will check that your research method is ethical and that you have made the necessary arrangements to deal fairly with subjects.

7.2.0 RANDOMLY?

In experiments, subjects are rarely selected at random. Not that it is undesirable. Random selection of subjects would help to ensure that they are representative of the general population (or some portion of the general population such as third year psychology majors at your university.) Rather, random selection usually is not practical. There are several reasons:

1) It is difficult to obtain lists of potential subjects in a group sufficiently large to merit the claim that the random selection was representative of some general "population". A random selection from a group of twenty classmates in an experimental psychology course could hardly be considered representative of some general population such as 3rd year psychology majors.

2) Of course the above difficulty is also encountered when psychologists do descriptive research. However, it becomes especially difficult with experiments because these designs require that the subjects be given some form of "treatment" - whether it be the level of noise in a reading room, the amount of caffeine in a beverage, or some preliminary information about a video. Experimental treatments generally require more involvement on the part of subjects than research lacking treatments. For example,

administering a treatment may require the subjects to come to a certain place at a certain time or perform some type of task before the measurements are made. The greater involvement often results in a substantial number of randomly selected persons who decline participation. When that occurs, you have a "self-selected" sample instead of a random sample. Therefore, experimenters usually do not try to obtain random samples.

3) Treatments introduced in experiments may also require that the subjects have certain characteristics besides a willingness to participate. Such characteristics might include having a drivers license, being bilingual, or having normal color vision. While you probably could obtain a list of all students in the introductory psychology course at your university (the single largest source of subjects for psychology research), you are unlikely to obtain a list of introductory students with driver's licenses, for example. Determining whether potential subjects have the necessary characteristics increases their involvement which in turn increases the refusal rate. Such selection, of course, also diminishes the randomness of your sample when you select only certain subjects. For example, it is a mute point to claim a "random sample" of all bilingual, right-handed, male, 1st year students in certain sections of an introductory psychology course who have a driver's license, normal color vision, and don't play an instrument. (That's overkill, but gets the point across.)

7.2.1 FINDING VOLUNTEERS

Subjects for psychology research are recruited from Introductory Psychology classes, classmates, friends, campus posters, newspaper ads, community organizations. Do such means work? Not always. That is why psychologists are concerned about the experiences that subjects have when they participate in research. Chapter 2 discusses ethical guidelines, which set basic standards for dealing with subjects. In addition, you should endeavor to make participation a positive experience for subjects, because psychology research can not continue without persons willing to volunteer to be subjects. People in general are quite willing to participate in research reflects a strong public belief in the value of science, trust scientists, and understand the need to know more about ourselves in order to deal with the problems of the world.

Many new students of psychology (including this professor, long ago) do not find it easy to overcome a certain shyness about asking strangers or even acquaintances to help them conduct research. Since you are considering a serious study of psychology and likely some related career goals, learning to interact with strangers is an important aspect of the discipline. So if you are daunted, *act* brave. With experience you will gain confidence. The best preparation you can make is to review for yourself why the research is worth doing. (Even if the immediate purpose is simply to teach you about research, that will help you contribute to society's needs in the future.)

If you have difficulty recruiting subjects, don't take it personally. Ours is a busy society. People often have difficulty finding the time to do everything they would like. When participation requires more than a nominal amount of time, other incentives may be needed to encourage enough people to volunteer. Here are some incentives typically used by psychologists:

- 1) Students in introductory psychology courses are often offered some modest course credit for participation in research projects. Most instructors recognize that

research participation is a good way to learn about the basis of psychological knowledge. Arrangements for such credit and guidelines for their use exist in most psychology departments. Typically it is a matter of discussing your need for subjects with the instructor or course coordinator. You may then be invited to make a brief presentation explaining the research to the class and pass out a "sign-up" sheet. Such sign-up sheets usually ask for the subjects name, and either a phone number or e-mail address to arrange an appointment or a schedule of times and places. Contacting subjects the day before their participation not only reminds them of their appointment, but also indicates your interest in their participation.

2) You may consider offering a reward such as a big fresh apple, special pen, T-shirt, or a nominal sum of money. Local industries may be willing to donate pens with their logo. A special pin ("I SURVIVED KIM'S MAZE EXP") may tickle your creative talent and become a collector's item. Two to five dollars may be appropriate when the participation involves substantial effort and/or patience.

3) It is typically the case that some of your fellow students will also need subjects for their experiments. Therefore informal understandings between students to help one another as subjects are not uncommon in psychology departments. This sharing also contributes to a sense of comraderie among psychology students. Undergraduate psychology clubs can facilitate this social spirit by maintaining lists of volunteers and arranging informal (i.e. not professor supervised) colloquia where students exchange research ideas and present results from term projects and thesis.

7.2.2 HIRING SUBJECTS

When research participation requires more than an hour or that subjects return several times, payment on an hourly rate may be appropriate and necessary. In addition to hiring research assistants - so the professor has time for teaching and writing - and buying equipment, funds to pay subjects are a major reason why professors need research grants to do research. Where would do students get money to pay for their research? Student research is usually done as part of a course or thesis. Your psychology department likely has some funds available for such purposes - speak with your instructor.¹ Psychology departments usually have

¹ What if you want to do research completely on your own? To be honest, while this would seem ideal scholarship, it is more difficult than you might think. Apart from the ethical considerations, undergraduates are generally not considered competent to independently pursue research with either human or animal subjects. (That's what a PhD is for.) But there are exceptions. If you seriously want to pursue independent research within the university, go talk to some of the professors whose interests match your own. It will be up to you to convince them that your idea is worth pursuing with some of their grant funds or helping you apply for the necessary funds. Since such research may also involve other expenses such as laboratory space, test materials, and the professor's time; it would not be inappropriate for him or her to expect to share credit for the outcome. Do not be surprised if a professor requests some agreement on such matters from the onset. This is professional courtesy, and you are being treated as their colleague.

arrangements and salary guidelines for hiring subjects. For amounts less than \$100, the university may be able to arrange for such payments to be classified as "honoraria" or as compensation for travel expenses to avoid income tax hassles. While you might make payment contingent on completion of the experiment, this may raise local issues too complex to discuss here.

7.2.3 KEEPING SUBJECTS - INTERESTED

You can not "keep" subjects, but you can try to keep them interested. Regardless of how you recruit subjects, never forget that a significant reason for participation is curiosity and their willingness to contribute to science. It is your responsibility to earn their respect and meet their expectations. In many types of experiments, the subject's task can be down right boring.

I'll say! When I was at university, I was in an experiment that required sitting in total darkness and reporting the color of a light flash every few seconds. The experiment took five sessions - each lasting for nearly an hour. There is no way that such an experiment could be made interesting. - Editor

So why did you continue? - TN

Well, it was an opportunity to earn a few bucks. And also, I guess because the student explained that the results might lead to some insight on color blindness. - Editor

I explain to my assistants and subjects that an awful lot of good science is boring - whether it be counting particles through a microscope or crawling on your hands and knees in a desert with a toothbrush for days on end. Real science is often not glamorous. Indeed when things are really boring, that's a good sign that everything is working well and data is getting collected. - TN

Making participation in your experiment as interesting as possible will help ensure that your subjects will take their participation seriously and make the effort required to follow instructions. Word also gets around. You will find it easier to obtain subjects when it gets known that your research is interesting and/or worthwhile.

This would be a good time for you to refresh your memory of the main points on ethics in Chapter 2. Good research involves more than not violating the appropriate code of ethics, and respecting the rights of your subjects. It is also up to you to make the subjects as comfortable as possible. Pay attention to what may seem like small details such as the appearance of the laboratory, a nice chair, and convenience of any equipment the subject uses. Details like this convey to the subject that you value their participation. If the participation can not be made intrinsically interesting, it is vital that subjects understand the potential value

of what they are being asked to do. Here are some guidelines that go beyond the basic ethical considerations:

1. Behave professionally in conducting the procedure.
2. Explanations and debriefings provide the opportunity for you to explain the why the research is worthwhile.
3. If deception is necessary, explain why afterwards
4. To make an experiment more interesting, it is often possible to provide some feedback to the subject regarding their performance. Even a boring task can be motivating when you try to see how efficiently or consistently you can do it.
5. Put on a "good show". Think of yourself as an actor whose job is also to entertain the subjects. (Good teachers often take this approach to putting on a good lecture.)

7.2.4 SCREENING AND SELECTING SUBJECTS

As mentioned initially, for experiments it is unusual for psychologists to attempt to find subjects who could be considered a random sample of some population. Faced with a limited number of volunteers, experimenters often resort to other methods that ensure that subjects are not unrepresentative of a general population. Therefore you may want test the volunteers for normal characteristics and abilities that are relevant to the experiment. This can include standard tests such as those for visual acuity (a Snellen chart - E, Q W, CDTR, PHGKL) hearing sensitivity ("Now write numbers that you hear in the third column."), physical fitness (heart rate change after moderate exercise such as rapid stepping), spatial reasoning (Mental Rotations Test (Vandenburg & Krus, 1979)), level of anxiety (Taylor Manifest Anxiety Scale (Taylor, 1953)).

Do not get overly enthusiastic about screening subjects for "normality". Normality is based on average performance. If you began to test a whole classroom of students using every standard test available, you would soon find that there are very few, if any, who are "normal" in all respects. To be practical, screen subjects only for characteristics that are relevant to the experiment.

THE MYTHICAL AVERAGE PERSON

Next time you are with a large group of friends, try this: Select only those who do not wear glasses or contact lenses. Then select only those males between 5 foot 4 inches (163 cm) and 6' 1" (185 cm) the mean for males plus and minus one standard deviation) and those females between 5' 1" (155 cm) and 5' 10" (178). Then select only those remaining who are right handed. Then select only those who have not had individual music lessons. How many persons do you have left?

If you are testing fellow students, never forget that through their studies they may have learned about specific theories or discussed in depth certain issues in a manner that would be representative of most students or the general population. Therefore it may be appropriate

to ask volunteers whether they have specific knowledge or strong opinions about certain theories or issues that you may be testing in your experiment.

For certain experiments, it may be important to test only persons having certain skills or characteristics. For example, if your experiment sought to measure moral behavior by testing how carefully persons obeyed traffic signs in a simulated driving task, you may want to restrict your subjects to persons having a drivers license. Other specific aspects of the subject's task may require a height limit, minimum reach length, or required strength.

For some experiments, the subject's ability to do the experiment may not be evident they actually try it. Preliminary practice may identify such problems. If the subjects report that they are uncomfortable doing the task or if their results indicate difficulties, these are suitable reasons for not having them participate. A WORD OF CAUTION: You want to avoid any possibility or even the perception of selecting subjects because their performance does meet certain expectations of a hypothesis you may testing. Therefore it is imperative that subjects be selected based only on preliminary testing and not after a full set of measurements have been obtained. The procedure for selection must be clearly described in the METHODS section of your report.

Another reason for selecting your subjects may be that the question you are trying to answer may only pertain to certain types of persons such those with a second language, a part-time job, married, over 65 years of age, etc.

When you have to average your measurements across subjects, you may try to reduce the variability by testing as homogenous group of subjects as possible. Therefore, you may want to test only males, only right handed persons, only persons between the ages of 18 and 24, only psychology majors, only persons who have done a similar experiment previously, etc.

Some characteristics of persons who have volunteered may not be evident or even known to the subjects themselves. These include personality variables such as introversion or extroversion, sensory abilities such as tone discrimination, cognitive abilities such as memory span for a series of numbers, skills such as reaction time. Yet these characteristics may affect the results. When such characteristics can not be held constant, by selective screening, they must be prevented from confounding the results of a between-subjects experiment by balancing them evenly across the groups of subjects assigned to each level of the independent variable. (We'll shortly discuss how to do balanced assigning.) To obtain such information will require preliminary testing of the subjects before they are brought into the experiment. When this information is of a personal nature, blind techniques using code numbers assigned by a third party to the subjects can maintain anonymity.

In all situations where you have screened or selected subjects, it is important that the type of screening or the selection basis be reported. This would usually be done in the SUBJECTS section of your report and possibly further explained in the INTRODUCTION.

7.3 ASSIGNING SUBJECTS TO LEVELS

This is the most important and section. Make sure you understand it thoroughly.

Don't say that! Readers will think the rest of the chapter is unimportant. - The Editor

Perhaps, but I figure they might need some encouragement to really work through the examples step by step, rather than just skimming over it and silently nodding their heads. - T.N.

7.3.1 RANDOM ASSIGNMENT PROCEDURES

Random numbers may be obtained from a random number table - see Appendix F at the back of this book.² They can also be obtained by rolling die, flipping one or more coins, drawing numbers from a hat and replacing them after each draw. Many general-purpose computer programs such as BASIC, and spreadsheet programs such as QUATTRO-PRO and EXCEL have a function which produces a random number each time the function is used. In QUATTRO-PRO it is "@ RAND"; see Appendix C - How to Start Using Spreadsheet Programs. Some more operations added to this function can generate random numbers of any specified range.³

A. PURELY RANDOM

HOW TO DO IT

Random assignment of subjects to levels can be done as follows:

1. Prepare a table with column headings for each level of the independent variable.
2. Assign a number to each level.
3. Arbitrarily select a starting place on a random number table and decide which way you will move along the numbers. (For all following examples I'll use a zig-zag pattern: starting by going right along the first row, down to next row at the right end, then left along the next row; then down; etc.)
4. Move along the numbers until you reach a number that is one of the numbers assigned to the levels. Ignore other numbers.
5. Make a list of these numbers as you read them off.

² You may think that you can produce random numbers off the "top of your head." Research has shown that people do not generate numbers that are anywhere near being random.

³ To generate random numbers from 1 to 4, multiply @rand by 4, add 0.5, and round off the result to nearest whole number:

$$@ROUND((@RAND*4) + .5, 0)$$

6. Pair up the first subject to the first number on that list; the second subject to the second number, and so forth.⁴
7. Make a table putting each subject in the column of the level represented by the subject's random number.

AN EXAMPLE

Consider an experiment on how noise affects problem solving. Let's say you want to test the effects of two levels of noise, low (LOW) and moderate (MOD), plus a control condition with no noise (NON). These 3 levels of the independent variable require 3 groups of subjects. Here is a list of some relevant information about 18 people who have volunteered to be subjects: (All volunteers just happened to have paired initials and signed up in alphabetical order.)

Table 7.3.1-1
Initials, Gender, and Reaction Time (RT) Information About Persons Who
Have Volunteered to be Subjects in Our Experiment

SUBJ	SEX	RT	SUBJ	SEX	RT	SUBJ	SEX	RT
AA	F	7	GG	F	8	MM	F	6
BB	M	9	HH	M	8	NN	F	9
CC	M	7	II	M	8	OO	M	9
DD	F	7	JJ	M	5	PP	F	8
EE	F	8	KK	M	9	QQ	F	9
FF	M	7	LL	M	8	RR	F	7

We begin by assigning the numbers "1", "2", and "3" to the levels, NON, LOW, and MOD respectively. To obtain a list of random numbers of "1", "2", and "3", use a random number table such as provided in Appendix E. A portion of that table is reproduced below as Table 7.3.1-2. Arbitrarily pick a starting place and direction. For this example, we start with the "2" in the first row and first column and proceed to the right. List each "1", "2" or "3" in a table such as Table 7.3.1-3 below. Each number used is highlighted in bold print. (When you do this, you will find it helpful to circle the number on your random number table.) At the end of the first row, we drop down one row and proceed back to the left. We continue back and forth until we have 18 random numbers - one for each subject. The subjects are then paired up with these numbers as shown in the table.

⁴ You could omit steps 5 and 6 and just enter the subjects' names directly into a table as the numbers are read off. However, it is easy to make a mistake or to lose your place and waste more time checking back on what you did.

Table 7.3.1-2
A Portion of the Random Number Table in Appendix F:
Starting at the 3rd Row and 1st Column.

2	4	1	3	0	4	8	3	6	0	2	2	5	2	7	9	7	2	6	5	7	6	3	9	3
4	2	1	6	7	9	3	0	9	3	0	6	2	4	3	6	1	6	8	0	0	7	8	5	6
3	7	5	7	0	3	9	9	7	5	8	1	8	3	7	1	6	6	5	6	0	6	1	2	1
7	7	9	2	1	0	6	9	0	7	1	1	0	0	8	4	2	7	5	1	2	7	7	5	6
9	9	5	6	2	7	2	9	0	5	5	6	4	2	0	6	9	9	9	4	9	8	8	7	2

Table 7.3.1-3
A List of Successive Numbers from the Above Random Number Table Paired with an Alphabetical List of the Subjects.

RAND #	SUBJ	RAND #	SUBJ	RAND #	SUBJ
2	AA	2	GG	2	MM
1	BB	2	HH	3	NN
3	CC	3	II	3	OO
3	DD	3	JJ	1	PP
2	EE	1	KK	2	QQ
2	FF	3	LL	3	RR

We complete the procedure by making a final table in which the subjects are organized according to the level represented by their random number. See Table 7.3.1-4.

Table 7.3.1-4
Results of the Random Assignment of the 18 Subjects to the Three Levels of the Independent Variable, Together with How Effectively the Groups Are Matched.

NON ("1")			LOW ("2")			MOD ("3")		
SUBJ	SEX	RT	SUBJ	SEX	RT	SUBJ	SEX	RT
BB	M	9	AA	F	6	CC	M	7
KK	M	9	EE	F	8	DD	F	6
PP	F	8	FF	M	7	II	M	8
			GG	F	8	JJ	M	5
			HH	M	8	LL	M	8
			MM	F	6	NN	F	9
			QQ	F	9	OO	M	9
MEAN:		8.67			7.43			7.38
S.D.:		0.47			1.05			1.32

How similar are the subjects assigned to the three levels? In Table 7.3.1-4 we can immediately see that the NON level has substantially fewer subjects than the other two levels.

Let's assume that the purpose of the experiment is to find out whether noise affects cognitive ability. Cognitive ability will be operationally defined in terms of how many shape discrimination problems are solved. Since the score obtained on this test also depends on how quickly a person can respond to the figures as they are presented on a computer, it would be desirable for the three groups of subjects to be approximately equal in this characteristic. It just so happens that in this case we "know" the reaction times of these subjects prior to the experiment. This information on each subject was presented in Table 7.3.1-1 above. While such information about subjects is not usually available when subjects are randomly assigned to levels, for this example we can use it to see how well the random assignment worked.

To check on group similarity, each subject's gender and reaction time were added to Table 7.3.1-1 along with the mean and standard deviation of the subjects in each level. The NON and MOD levels have mostly male subjects. The NON level also has a significantly longer mean reaction time than the LOW level and is the most uniform in its reaction times. The LOW level has the lowest mean reaction time and is the only level with the mostly female subjects. The MOD group is the least uniform. Are these groups sufficiently similar that you would be willing to say that any differences in the experimental results at each level might not be attributable to differences between the groups of subjects?

ADVANTAGE

- 1) It is simple and fast.

DISADVANTAGE

- 1) It usually produces groups of uneven size, which wastes power in statistical analyses, and may produce biased results when one level has relatively few subjects.

The random number technique is rarely used because it produces groups of uneven size. It is only used when large numbers of subjects are involved. However, working out an example yourself is a good way to get acquainted with the techniques used to assign subjects to levels.

B. BLOCK-RANDOMIZED ASSIGNMENT

This technique assures that an equal number of subjects are assigned to each group. (This, of course, requires that the number of available subjects is a multiple of the number of levels.) You can make your own block-random tables by flipping coins, rolling die⁵, or using a spreadsheet program. The following instructions are based on using a random number table.

⁵ Variations of the "Dungeons and Dragons" game have die with 4, 8, and other numbers of sides too.

HOW TO DO IT

Block-random assignment of subjects to levels is done as follows:

0. Start with a number of subjects that is a multiple of the number of levels to be tested. If a few more persons volunteer, you unfortunately won't be able to use them, since that would produce uneven groups.
1. Assign a number to each level.
2. Arbitrarily select a starting place and proceed along the random numbers as before.
3. Move along the numbers until you reach a number that is one of the numbers assigned to the levels. Ignore other numbers.
4. List that number.
5. Continue along the random number table until you reach a different level's number and add that to the list. Keep adding different numbers until each level's number has been used once. Ignore numbers of levels that have been used.
6. Repeat steps 3-5 to create multiple sets until you have as many numbers as subjects.
7. Pair up the first subject to the first number on the list; the second subject to the second number, and so forth.
8. Make a new table putting each subject in the column of the level represented by the subject's random number.

AN EXAMPLE

Let's again consider that experiment on how noise affects cognitive ability. We'll use the same subjects and the same section of the random number table as before. Starting with the first random number in the upper left corner and proceeding as previously, I highlighted each number that was used while following STEPS 1 to 8. Since a level's number is not repeated until all levels have been assigned an equal number of times, the numbers differ from those used for the pure random assignment.

Table 7.3.1-5
From the Random Number Table in Appendix F - Starting at Row 3, Column 1.

[2 4	1	3]	0	4	8	[3	6	0	2	2	5	2	7	9	7	2	6	5	7	6	3	9	3	
4	2]	[1	6	7	9	3	0	9	3	0	6	2	4	3]	6	[1	6	8	0	0	7	8	5	6
3	7	5	7	0	3	9	9	7	5	8	1]	8	[3	7	1	6	6	5	6	0	6	1	2]	[1
7	7	9	2	1	0	6	9	0	7	1	1	0	0	8	4	2	7	5	1	2	7	7	5	6
9	9	5	6	2	7	2	9	0	5	5	6	4	2	0	6	9	9	9	4	9	8	8	7	2
9	6	3	0	1	9	1	9	7	7	0	5	4	6	[3	0	7	9	7	2	1	8	8	7	6

Table 7.3.1-6
A List of the Successive Numbers from the Above Random Number Table Paired with an Alphabetical Listing of the Subjects.

RAND #	SUBJ	RAND #	SUBJ	RAND #	SUBJ
2	AA	3	GG	3	MM
1	BB	2	HH	1	NN
3	CC	1	II	2	OO
3	DD	2	JJ	1	PP
2	EE	3	KK	2	QQ
1	FF	1	LL	3	RR

Table 7.3.1-7
Results of the Block-Random Assignment to the Three Noise Levels.

NONE ("1")			LOW ("2")			MOD ("3")		
SUBJ	SEX	RT	SUBJ	SEX	RT	SUBJ	SEX	RT
BB	M	9	AA	F	6	CC	M	7
FF	M	7	EE	F	8	DD	F	6
II	M	8	HH	M	8	GG	F	8
LL	M	8	JJ	M	5	KK	M	9
NN	F	9	OO	M	9	MM	F	6
PP	F	8	QQ	F	8	RR	F	7
MEAN		8.17			7.50			7.17
S.D:		0.69			1.50			1.07

Well, at least we got the same number of subjects to be tested at each level. To see how well these groups are matched, we again tabulate the subjects' gender and calculate the mean value and standard deviation of the reaction times of the subjects assigned to each level. How does the similarity of the gender and reaction time means and standard deviations at the three levels compare with the results obtained previously using random assignment?

ADVANTAGES

- 1) Block-random assignment requires no knowledge of what subject characteristics are relevant to the performance being measured. (The significance of this advantage will become clear shortly.)
- 2) It's quick and easy to do.

DISADVANTAGE

- 1) Can produce groups whose subjects differ in their average ability. This can confound the results and leave any conclusion in doubt. This problem is worse with small groups of subjects.

7.3.2 MATCHING ON CATEGORICAL CHARACTERISTICS

Instead of hoping that randomization will provide you with similar groups of subjects at each level, you can try to match them deliberately. This requires information about one or more subject characteristics that are relevant to the dependent variable you'll be measuring. Sometimes this information is categorical and obtained by a questionnaire that asks about their occupation, handedness, attitudes or whatever else might be relevant. Other characteristics may involve test scores or require laboratory measurements such as reaction time. Somewhat different procedures are needed to match subjects for categorical versus numeric characteristics. They may appear complicated when you first try them, but neither is "rocket science". They are simply logical ways to go about avoiding systematic bias arising from the order in which the subjects are assigned to the levels. If you were faced with having to do so without any instruction, you would probably end up doing it the same way - eventually.

A. MATCHING ON A SINGLE CATEGORY

Examples of categorical characteristics are gender, occupation, personality type - any information about a subject that is not numerical.

HOW TO DO IT

1. List the prospective subjects by category. Each category must have a number of subjects that is some multiple of the number of groups to be formed. This may mean that you will not be able to use all available subjects or that you will have to do some more recruiting to fill in the gaps.
2. Assign a number to each level for which you are creating a group of subjects. Use the same block-randomization technique as described above to assign successive subjects in each list to one of the levels.

AN EXAMPLE:

Consider an experiment on how the effectiveness of information may vary depending on the age of the person who provides it. "Effectiveness" is operationally defined in terms of how much information the subjects remember after a short delay. The independent variable is four ages of information providers: elderly persons, middle aged, persons about 30 years old, and persons the same age as our subjects - introductory psychology students who are generally between 17 and 22. To avoid having subjects who differ in familiarity with the information provided, we want to ensure that the subjects assigned to each level have similar educational backgrounds. On a sign-up sheet passed around the introductory classes, the subjects were asked to provide their name, telephone number and whether they were enrolled in Arts, Education, or Science. The sheet also explained that since the experiment required subject matching, not everyone who signed up might be asked to participate. The following list shows the results of the recruitment effort.

NAME	MAJOR	NAME	MAJOR
AA _♀	arts	NN _♀	education
BB _♂	arts	OO _♂	education
CC _♂	arts	PP _♀	education
DD _♀	arts	QQ _♀	education
EE _♀	arts	RR _♀	education
FF _♂	arts	SS _♀	education
GG _♀	arts	TT _♀	science
HH _♂	arts	UU _♂	science
II _♂	arts	VV _♂	science
JJ _♂	arts	WW _♀	science
KK _♂	arts	XX _♀	science
LL _♂	arts	YY _♂	science
MM _♀	arts	ZZ _♂	science

First we arrange the students according to their major: (By astonishing chance, their initials just happened to be alphabetical order of their major.) This is indeed fortunate because it will help you see how the block random assignments were made.) Since the number of potential subjects in each category must equal some multiple of the number levels, we list the subjects in groups of four. Unfortunately, this means we must omit the last arts major, MM, and the last education major, RR.

Next we identify the levels of the independent variable by number (elderly = 1, mid-aged = 2, thirties = 3, peers = 4), and prepare a table with these headings:

ARTS		EDUCATION		SCIENCE	
SUBJ	RAND #	SUBJ	RAND #	SUBJ	RAND #
AA	2	NN	2	TT	3
BB	4	OO	4	UU	1
CC	1	PP	3	VV	2
DD	3	QQ	1	WW	4
EE	4	(RR)		XX	1
FF	3	(SS)		YY	2
GG	2			ZZ	4
HH	1			TT	3
II	3				
JJ	4				
KK	2				
LL	1				
(MM)					

Using a random number table such as the one below, go down the list of names in each category and assign each name to one of the levels by blocks - i.e. don't repeat assigning to any level until each level has been assigned an equal number of times. In the following example, I started with the first row and bolded each number used to make the above assignments. For the first arts student, AA, the first number in the first row, "2", is usable. Therefore AA is

assigned to the "mid-age" level. The next number "4" is also usable, so the next arts student, BB, is assigned to the "30's" level. And so forth.

We have three science students left over. Rather than drop them, we ask around and find that ΓΓ (evidently a Greek exchange student) is willing to participate.

[2 4 1 3] 0	[4 8 3 6 0	2 2 5 2 7	9 7 2 6 5	7 6 3 9 3
4 2] [1 6 7	9 3 0 9 3	0 6 2 4 3]	6 [1 6 8 0	0 7 8 5 6
3 7 5 7 0	3 9 9 7 5	8 1] 8 [3 7	1 6 6 5 6	0 6 1 2 1
7 7 9 2 1	0 6 9 0 7	1 1] 0 0 8	[4 2 7 5 1	2 7 7 5 6
9 9 5 6 2	7 2 9 0 5	5 6 4 2 0	6 9 9 9 4	9 8 8 7 2
9 6 3 0 1	9 1 9 7 7	0 5 4 6 [3	0 7 9 7 2	1 8 8 7 6

Finally, we organize the subjects into another table according to the level at which they are to be tested.

Table 7.3.2-1
The Results of Matching Subjects According to their Major in Assigning them to Four
Levels of a Between-Subjects Experiment.

ELDERLY (1)		MID-AGE (2)		≈30'S (3)		PEERS (4)	
SUBJ	MAJ	SUBJ	MAJ	SUBJ	MAJ	SUBJ	MAJ
CC _♂	ART	AA _♀	ART	DD _♀	ART	BB _♂	ART
HH _♂	ART	GG _♀	ART	FF _♂	ART	EE _♀	ART
LL _♂	ART	KK _♂	ART	II _♂	ART	JJ _♂	ART
QQ _♀	ED	NN _♀	ED	PP _♀	ED	OO _♂	ED
UU _♂	SCI	VV _♂	SCI	TT _♀	SCI	WW _♀	SCI
XX _♀	SCI	YY _♂	SCI	TT _♀	SCI	ZZ _♂	SCI

As you can see in the above table, this procedure produces an equal number of Arts, Education and Science majors to be tested at each level. However, a substantial imbalance is revealed if you look at the subjects' gender. The ELDERLY and MID-AGE levels have mostly males subjects while the ≈30's and PEER levels have mostly females. Next we'll see how to deal with that inequity.

B. MORE THAN ONE CATEGORY

In the above experiment on how age of an information provider affects recall, it would be a good idea to ensure that there were also an equal number of male and female subjects assigned to each level. To match on two (or more) categorical characteristics involves repeating the sorting and block-random assignment process as many times as there are characteristics. This does complicate the recruiting, because it requires that the number of

subjects with each combination of characteristics equal the number of levels of the independent variable. Just as you need not have the same number of education students as arts students in the first example, neither must you have the same number males and females - just the correct multiples.⁶

To illustrate how to do it, I'll use the same students and the same hypothetical experiment as before. We begin by sorting the subjects according to major and gender. Again, we'll group them into batches of four to remind us that we have to assign them to four levels. As Table 7.3.2-2 below shows, we have five females in Arts. This is only enough to assign one Arts female to each level. MM is left over and can't be used in the study. The eight male Arts majors are enough to assign two to each level.. With five female Education majors we must drop the last, SS. The lone male Education major, OO, also can't be used. We are unable to find a fourth female Science major so TT, WW, and XX have to be omitted from the study - pity. Neither can we use the fifth male science major, TT.

**Table 7.3.2-2
How the Subjects Were Sorted into Groups According to Both Gender and Major.**

ARTS		EDUCATION		SCIENCE	
SUBJ	RAND #	SUBJ	RAND #	SUBJ	RAND #
FEMALE					
AA	2	NN	2	(TT)	
DD	4	PP	4	(WW)	
EE	1	QQ	3	(XX)	
GG	3	RR	1		
(MM)		(SS)			
MALE					
BB	4	(NN)		UU	3
CC	3	(OO)		VV	1
FF	2			YY	2
HH	1			ZZ	4
II	3			(TT)	
JJ	4				
KK	2				
LL	1				

Using a random number table, block-random assignment is now done, first for Arts majors, then Education, and finally the Science majors. Below I use the same portion of the random number table as before, and the same bold numbers from 1 to 4 indicate the ones I used to fill in the above table. Starting with the female arts major, AA, the first number in the

⁶ However having the same number of subjects in all combinations of characteristics does become important if you want to use analysis of variance to treat subject characteristics as independent variables and thereby take into account their contribution to the variability of the data.

154 - BETWEEN

first row, "2", is usable. Therefore AA is assigned to the "mid-age" level. The next number is "4", so the next female arts major, DD, is assigned to the "peers" level; and so forth. After the first four female Arts majors have been assigned, we do the same with the male Arts majors. Then we continue with the other majors in sets of four.

[2 4 1 3] 0	[4 8 3 6 0	2 2 5 2 7	9 7 2 6 5	7 6 3 9 3
4 2] [1 6 7	9 3 0 9 3	0 6 2 4 3]	6 [1 6 8 0	0 7 8 5 6
3 7 5 7 0	3 9 9 7 5	8 1] 8 [3 7	1 6 6 5 6	0 6 1 2 1
7 7 9 2 1	0 6 9 0 7	1 1] 0 0 8	4 2 7 5 1	2 7 7 5 6
9 9 5 6 2	7 2 9 0 5	5 6 4 2 0	6 9 9 9 4	9 8 8 7 2
9 6 3 0 1	9 1 9 7 7	0 5 4 6 3	0 7 9 7 2	1 8 8 7 6

Finally, we prepare a table organized in terms of the levels of the independent variable and assign a number to each level. The subjects are then assigned to the level having the random number that they were paired with in the above procedure.

Table 7.3.2-3
Results of the Procedure for Matching Subjects on the Basis of Major and Gender for a Between-Subjects Experiment with Four Levels of the Independent Variable.

ELDERLY (1)			MID-AGE (2)			30'S (3)			PEERS (4)		
SUBJ	MAJ	SEX	SUBJ	MAJ	SEX	SUBJ	MAJ	SEX	SUBJ	MAJ	SEX
EE	ART	F	AA	ART	F	GG	ART	F	DD	ART	F
HH	ART	M	FF	ART	M	CC	ART	M	BB	ART	M
LL	ART	M	KK	ART	M	II	ART	M	JJ	ART	M
RR	ED	F	NN	ED	F	QQ	ED	F	PP	ED	F
VV	SCI	M	YY	SCI	M	UU	SCI	M	XX	SCI	M

ADVANTAGE

- 1) The more characteristics on which subjects can be matched, the less the chance of some characteristic confounding the results in a between-subjects experiment.

DISADVANTAGES

- 1) A major difficulty is knowing what subject characteristics are the important ones to match. This may not be known until more is understood about the research topic itself. It is not unusual for psychologists to find themselves in a "Catch 22" dilemma of not knowing how to do the experiment properly until they have done the experiment properly. This is one of the challenges that makes research in psychology fun to do.
- 2) Given some reasonable indications about which characteristics could be important to match, information about those characteristics needs to be obtained before the

volunteers can be selected as subjects. When those characteristics are not obvious, preliminary testing of all volunteers is required.

- 3) The more characteristics you match, the more volunteers you'll need available as potential subjects. To have the same number of subjects with any given combination of characteristics assigned to each level, probably means that you will not be able to use some subjects
- 4) It can be awkward to have to turn down persons who have volunteered. Yet recruiting a few students to fill in gaps at the last minute, may result in subjects who differ in some significant way from the other volunteers.
- 5) Since you can't anticipate who will volunteer, you will usually have to wait to start the experiment until you enough subjects with any given combinations of characteristics to assign at least one to each level.

7.3.3 MATCHING ON NUMERICAL CHARACTERISTICS

As researchers, we might wish that subjects knew their Minnesota Multiphasic Personality Inventory (MMPI) profile, IQ and various subscores on the Wechsler Adult Intelligence Scale (WAIS), Handedness Score, brain volume, etc. (Do you know any of these for yourself?) The WAIS, for one, requires professional administration and scoring and a couple of hours to complete. Moreover, many people feel that such information is private and do not want to disclose it.⁷ When information is not available about certain attitudes, aptitudes, and skills that could affect how a person responds, researchers often obtain some preliminary measurements from subjects prior to using them in between-subjects experiments. This enables groups of subjects to be matched on this characteristic before their assignment to the levels.

A. SINGLE NUMERICAL CATEGORY

HOW TO DO IT

1. Measure or obtain numerical information about subjects on some characteristic (eg. reaction time or average grade last semester) believed to affect the dependent variable.
2. Arrange subjects in ascending or descending order according to that numerical value.
3. Assign subjects alternately to each level - reversing the order of how the levels are assigned as each set is completed.
4. Try to complete an even number of sets so that the last set does not introduce a bias such as more higher values being assigned to the first level. (This "rule" may be broken if there is an odd number of sets with equal values assigned to each level.)

⁷ To use personal information about subjects and protect their anonymity, such information can be provided to a colleague who does not know the subjects. The colleague then assigns each subject a number and provides you with a list that only shows subjects' numbers and their characteristics. When you have done the matching, the list is given back to the colleague, who then provides a matched list with the subjects' names.

ADVANTAGES:

- 1) More likely than random assignments to prevent subject differences from confounding the results.
- 2) May provide a more uniform match than is possible with categorical characteristics.
- 3) Enables treating subject characteristics as another independent variable to reduce error variance using multivariate statistics..

DISADVANTAGES:

- 1) You need to know which characteristics count
- 2) It may be time consuming to obtain the information
- 3) Presents opportunity for experimenter bias - depending on⁸ the order in which the levels are addressed.
- 4) All subjects must be signed up before you can proceed with the experiment.

AN EXAMPLE:

For this example we'll return to the earlier experiment on noise and cognition. In the random assignment examples, we thought that a subject's reaction time could influence her/his performance on the measure of cognitive ability. We'll now assume that we have measured our subject's reaction times before starting the experiment. We'll use this information to deliberately make the subjects assigned to each level as similar as possible. We'll use the same 18 subjects as used for the random assignment techniques. We'll assign them to three noise levels, NON, LOW, and MOD, which will be represented by the numbers "1", "2", and "3" respectively.

We begin by arranging the subjects in ascending order on the basis of preliminary data on their reaction times (RT). For future reference, we'll also keep track of their gender and preliminary data on their error scores (ES).

The first subject is assigned to level 1, the next subject to level 2, the next to level 3, the next to level 3, the next to level 2, etc. Continue using this 1, 2, 3; 3, 2, 1; 1, 2, ... sequence. This is as good as possible using a simple rule.

⁸ To avoid the possibility of such experimenter bias is to add an additional step to the procedure. Instead of assigning subjects to levels directly, assign them to groups. Then use the block-random technique to assign the groups to the levels.

Table 7.3.3-1
Preliminary Data on Reaction Time (RT), Error Score (ES), and the Gender of Subjects
for the Noise Experiment; Their Arrangement into Ascending Order on Basis of RT; and
Their Systematic Assignment to the Noise Levels.

PRELIMINARY DATA			REARRANGED ACCORDING TO RT		ASSIGNED TO LEVEL
SUBJ	RT	(ES)	SUBJ	RT	
AA _♀	6	(2)	JJ	5	1
BB _♂	9	(3)	MM	5	2
CC _♂	7	(4)	CC	6	3
DD _♀	6	(2)	FF	6	3
EE _♀	8	(1)	RR	6	2
FF _♂	7	(5)	AA	7	1
GG _♀	8	(3)	DD	7	1
HH _♂	8	(3)	EE	7	2
II _♂	8	(2)	GG	7	3
JJ _♂	5	(5)	HH	7	3
KK _♂	9	(4)	II	8	2
LL _♂	8	(2)	LL	8	1
MM _♀	6	(6)	PP	8	1
NN _♀	9	(1)	QQ	8	2
OO _♂	9	(5)	BB	9	3
PP _♀	8	(3)	KK	9	3
QQ _♀	9	(2)	NN	9	2
RR _♀	7	(3)	OO	9	1

To see how this worked, we organize the subjects by level into a second table:

Table 7.3.3-2
Results of the Matched Assignment on Basis of Reaction Time
to Three Levels of Noise, and the Similarity of the Subjects at Each Level.

NON (1)			LOW (2)			MOD (3)		
SUBJ	RT	(ES)	SUBJ	RT	(ES)	SUBJ	RT	(ES)
JJ _♂	5	5	AA _♀	6	2	DD _♀	6	2
FF _♂	7	5	CC _♂	7	4	MM _♂	6	6
RR _♀	7	3	EE _♀	8	1	GG _♀	8	3
LL _♂	8	2	II _♂	8	2	HH _♂	8	3
PP _♀	8	3	BB _♂	9	3	KK _♂	9	4
QQ _♀	9	2	OO _♀	9	5	NN _♀	9	4
MEAN	7.33	3.33		7.83	2.83		7.67	3.17
S.D:	1.25	1.25		1.07	1.34		1.25	1.57

158 - BETWEEN

Look at the mean value and standard deviation of the reaction times of the subjects assigned to each group. How do the similarity of the reaction time means and standard deviations of the three groups compare with those obtained using block-random assignment?

Note that while the mean reaction times (RT) of the subjects at each level are quite close, their mean error scores (ES) are not well matched. More about this shortly. Note also that males outnumber females 4 to 2 in the LOW level while females outnumber males 4 to 2 in the MOD level.

B. ONE NUMERIC & ONE CATEGORICAL CHARACTERISTIC

The following method can be used when both a categorical characteristic and a numerical characteristic may influence subject performance. There should be an equal number of subjects in each category assigned to each level of the independent variable. Therefore, the total number of subjects should be a multiple of the number of subject categories and the number of levels. Furthermore, to also obtain a balanced assignment in terms of the numerical characteristic, it is best to have even multiples. (That latter point is easier to understand by looking at an example.)

HOW TO DO IT:

1. Divide the subjects into subgroups based on their categorical characteristics.
2. For each subgroup, follow the steps for matched numerical assignment of subjects.

AN EXAMPLE

Again we'll use the example of an experiment with 18 subjects to study the effects of three levels of noise on cognition. We'll use the same subject pool as was used in the previous example except that a two females will be replaced by two males to obtain an equal number of males and females. While 18 is a multiple of 6 (the 2 categories times the 3 levels), note that it is not an even multiple. We'll see the consequence of this decision shortly.

1. Make a table and arrange the subjects into male and female groups.
2. Arrange each gender group into ascending order of reaction time.
3. Use the 1, 2, 3, 3, 2, 1, 1, ... assignment method separately for each gender group to assign each subject in that group to one of three experimental sub-groups for that gender. Fortunately we have an even number of multiples of the three levels so the back-and-forth assignments balance.
4. Make a new table. Organize the subjects according to levels and calculate the mean reaction time for the subjects at each level.

Table 7.3.3-3
The Subjects Have Been Sorted by Sex and then Arranged in Ascending Order
on the Basis of Their Preliminary Reaction Times.

INFORMATION			ASCENDING ORDER		ASSIGNED TO LEVEL
SEX	SUBJ	RT	SUBJ	RT	
FEMALE	AA	6	AA	6	1
	DD	6	DD	6	2
	EE	8	MM	6	3
	GG	8	RR	7	3
	MM	6	EE	8	2
	NN	9	GG	8	1
	PP	8	PP	8	1
	QQ	9	NN	9	2
	RR	7	QQ	9	3
MALE	BB	9	JJ	5	3
	CC	7	CC	7	2
	FF	7	FF	7	1
	HH	8	HH	8	1
	II	8	II	8	2
	JJ	5	LL	8	3
	KK	9	BB	9	3
	LL	8	KK	9	2
	OO	9	OO	9	1

The results of this procedure for matching subjects by both gender and reaction time are shown in Table 7.3.3-4. There are an equal number of males and females assigned to each level. In this example the mean reaction times at each level are just as similar to each other as they were when matching by reaction time alone. (Usually the when you match on two characteristics neither is matched as closely as when done alone.)

Table 7.3.3-4
The Results of the Matched Assignment to Three Levels of Noise
Based on Reaction Time and Gender.

NON (1)			LOW (2)			MOD (3)		
SUBJ	SEX	RT	SUBJ	SEX	RT	SUBJ	SEX	RT
AA	F	6	DD	F	6	MM	F	6
GG	F	8	EE	F	8	RR	F	7
PP	F	8	NN	F	9	QQ	F	9
FF	M	7	CC	M	7	JJ	M	5
HH	M	8	II	M	8	LL	M	8
OO	M	9	KK	M	9	BB	M	9
MEAN: 7.67			7.83			7.33		
SD: 0.94			1.07			1.49		

C. MATCHING ON TWO NUMERICAL CHARACTERISTICS

When we matched subjects on the basis of just reaction time in the previous section, we found that the subjects were not well matched in terms another numerical characteristic, their error scores (ES). As you probably suspect by now, it is possible to match subjects on the basis of two numerical characteristics. In fact, the following method will work for any number of numerical characteristics.

HOW TO DO IT

The basic idea is to convert each set of numerical characteristics to Z values. This conversion puts all the characteristics on a numerically equal basis whether they be reaction times in hundreds of milliseconds with a relative small variability or they be ratings scaled from 1 to 4 with a large variability. The Z values are then averaged and arranged in ascending or descending order. After that it's the same old 1, 2, 3, 3, 2, 1, 1 ... technique of assigning the ordered subjects to the levels.

AN EXAMPLE

We'll use the same subjects as before and again assign them to be tested at three noise levels (NON, LOW, and MOD).

1. Convert the reaction times and the error scores for each subject to their Z-value equivalents. (This done by finding the mean and standard deviation of each set of values. The subtract the mean of each from ever value and divide the remainder by the standard deviation.)
2. Average the reaction time Z-values and the error score Z-values.
3. Arrange the subjects in ascending order in terms of these average Z-values.
4. Use the 1, 2, 3, 3, .. technique to assign the subjects to the levels.
5. Organize the assignments in a table according to noise level.

Table 7.3.3-5
The Mean Z-Values of the Subjects' Reaction Times and Error Scores.
The Subjects Are then Arranged in Ascending Order of Their Mean Z-Values,
and then Systematically Assigned to the Noise Levels.

NUMERICAL CHARACTERISTICS						ASCENDING ORDER		ASSIGNED TO LEVEL	
SUBJ	RT	Z	ES	Z	avg Z	SUBJ	avg Z		
AA _♀	6	-1.33	2	-0.79	-1.06	AA	-1.06	1	
BB _♂	9	1.15	3	-0.08	0.54	DD	-1.06	2	
CC _♂	7	-0.51	4	0.63	0.06	EE	-0.59	3	
DD _♀	6	-1.33	2	-0.79	-1.06	JJ	-0.41	3	
EE _♀	8	0.32	1	-1.50	-0.59	RR	-0.29	2	
FF _♂	7	-0.51	5	1.34	0.42	II	-0.23	1	
GG _♀	8	0.32	3	-0.08	0.12	LL	-0.23	1	
HH _♂	8	0.32	3	-0.08	0.12	NN	-0.17	2	
II _♂	8	0.32	2	-0.79	-0.23	CC	0.06	3	
JJ _♂	5	-2.16	5	1.34	-0.41	GG	0.12	3	
KK _♂	9	1.15	4	0.63	0.89	HH	0.12	2	
LL _♂	8	0.32	2	-0.79	-0.23	PP	0.12	1	
MM _♀	6	-1.33	6	2.05	0.36	QQ	0.18	1	
NN _♀	9	1.15	1	-1.50	-0.17	MM	0.36	2	
OO _♂	9	1.15	5	1.34	1.24	FF	0.42	3	
PP _♀	8	0.32	3	-0.08	0.12	BB	0.54	3	
QQ _♀	9	1.15	2	-0.79	0.18	KK	0.89	2	
RR _♀	7	-0.51	3	-0.08	-0.29	OO	1.24	1	
<i>mean</i> 7.61		<i>3.11</i>		<i>s.d.</i> 1.21		<i>1.41</i>			

Table 7.3.3-6
Results of the Matched Assignment to Three Levels of Noise
Based on Reaction Time (RT) and Error Score (ES).

NON (1)			LOW (2)			MOD (3)		
SUBJ	RT	ES	SUBJ	RT	ES	SUBJ	RT	ES
AA _♀	6	2	DD _♀	6	2	EE _♀	8	1
II _♂	8	2	RR _♀	7	5	JJ _♂	5	5
LL _♂	8	2	NN _♀	9	1	CC _♂	7	4
PP _♀	8	3	HH _♂	8	3	GG _♀	8	3
QQ _♀	9	2	MM _♀	6	6	FF _♂	7	5
OO _♂	9	5	KK _♂	9	4	BB _♂	9	3
MEAN:	8.00	2.67		7.50	3.17		7.33	3.50
SD:	1.00	1.11		1.26	1.57		1.25	1.38

The mean reaction times of the subjects in each level are no longer as similar as they were when the assignment was based only on reaction time. This is to be expected. As more characteristics are averaged, any given characteristic is likely to be less well matched. It is a tradeoff. The benefit is that the error scores are certainly better matched than they were when the matching was based only on reaction time. Note also that the LOW and MOD levels are not gender balanced. Don't despair. There is still another method which can often produce even better matches based on multiple subject characteristics - if you're willing to be challenged.

7.3.4 TRIAL AND ERROR MATCHING

Imagine a computer program that would assign the subjects in all possible combinations to the various levels of the independent variable. For each combination, it calculates the mean values and standard deviations of each characteristic at each level. While there are some 30 million combinations of assigning 18 subjects to 3 levels, this by itself is quite feasible with today's computers. The tricky part would be to select the best combinations, since you certainly don't want to pick them out by reading them all. Undoubtedly there are programs that do this - perhaps employing available "discriminant analysis" techniques.⁹ Alternatively, the human mind is quite capable of sensing trends and spotting possibilities that makes it possible to achieve reasonable matches by less tedious techniques. Unfortunately, I can't tell you exactly how to do that, but I can show the way,

To illustrate what is possible, I started with the results of the earlier assignment of subjects to levels based on just reaction time as was summarized in Table 7.3.2.3. Those subjects were arranged with levels in ascending order of reaction time. Now I further sort the subjects in terms of ascending error score (ES) within each reaction time value. (In this example that additional degree of sorting made only a small difference, but additional sorting helps substantially when there are more subjects.

Table 7.3.4-1
Results of Matched Assignment of Subjects to Three Groups on Basis of Reaction Time (RT) and then Further Sorting them within RT's on the Basis of Error Scores

GROUP 1			GROUP 2			GROUP 3		
SUBJ	RT	ES	SUBJ	RT	ES	SUBJ	RT	ES
JJ _♂	5	5	AA _♀	6	2	DD _♀	6	2
RR _♀	7	3	CC _♂	7	4	MM _♂	6	6
FF _♂	7	5	EE _♀	8	1	GG _♀	8	3
LL _♂	8	2	II _♂	8	2	HH _♂	8	3
PP _♀	8	3	BB _♂	9	3	NN _♀	9	1
QQ _♀	9	2	OO _♀	9	5	KK _♂	9	4
MEAN	7.33	3.33		7.83	2.83		7.67	3.17
S.D:	1.25	1.25		1.07	1.34		1.25	1.57

⁹ Developing a general purpose program to do matched assignments of X subjects to Y levels based on 1 to 6 numerical and categorical characteristics might make a good thesis for an enterprising student with programming abilities.

Note that the major headings which identified the NON, LOW, and MOD noise levels in Table 7.3.4-1 have been changed to GROUP 1, GROUP 2, and GROUP 3. This is important. Trial-and-error matching must be done "blind" without the researcher knowing the levels to which level the subjects are being assigned. This prevents researcher bias from inadvertently influencing the assignments. When the groups have been matched as closely as possible, the block-random method is used to assign groups to levels.

The task is now to make the groups more equal. To avoid endless trials, involves a combination of being both systematic and relying on "insight". Begin by focusing on one characteristic and matching it as well as possible. Picking the characteristic with the most number of different values or categories may help. Proceed with additional characteristics in order of decreasing number of categories while trying to avoid undoing the earlier matching.

I began with the reaction times. GROUP 2 has the highest mean reaction time (RT), and GROUP 1 has the lowest. Several combinations of subjects could be exchanged between groups 1 and 2 to raise the mean RT in GROUP 1 and lower it in GROUP 2. However, only a few of those will also lower the mean error score (ES) in GROUP 1 while raising it in GROUP 2. Switching JJ in GROUP 1 with CC in GROUP 2 accomplishes both effects without upsetting the gender balance in GROUP 1:

Table 7.3.4-2
The Results When Subject JJ who was formerly in Group 1
 was exchanged with CC who was formerly in GROUP 2.

GROUP 1			GROUP 2			GROUP 3		
SUBJ	RT	ES	SUBJ	RT	ES	SUBJ	RT	ES
CC _♂	7	4	AA _♀	6	2	DD _♀	6	2
FF _♂	7	5	JJ _♂	5	5	MM _♀	6	6
RR _♀	7	3	EE _♀	8	1	GG _♀	8	3
LL _♂	8	2	II _♂	8	2	HH _♂	8	3
PP _♀	8	3	BB _♂	9	3	NN _♀	9	1
QQ _♀	9	2	OO _♂	9	5	KK _♂	9	4
MEAN	7.67	3.17		7.50	3.00		7.67	3.17
S.D:	0.75	1.07		1.50	1.53		1.25	1.57

The mean reaction times and the mean error scores in Table 7.3.2.6-2 are as closely matched as is possible. As for gender, GROUP 1 is balanced, but groups 2 and 3 are not. Any of the four males in GROUP 2 could be switched with any of the four females in GROUP 3 to make them equal. Most of these exchanges would unbalance either the reaction time or error score means. However, since the mean error score is lower in GROUP 2 than in GROUP 3, switching II with GG will work:

Table 7.3.4-3
Results of Trial-and-Error Assignment to Three Levels of Noise.

MOD (2)			LOW (1)			MOD (3)		
SUBJ	RT	ES	SUBJ	RT	ES	SUBJ	RT	ES
CC _♂	7	4	AA _♀	6	2	DD _♀	6	2
FF _♂	7	5	MM _♀	6	6	JJ _♂	5	5
RR _♀	7	3	EE _♀	8	1	II _♂	8	2
LL _♂	8	2	GG _♀	8	3	HH _♂	8	3
PP _♀	8	3	BB _♂	9	3	NN _♀	9	1
QQ _♀	9	2	OO _♂	9	5	KK _♂	9	4
MEAN	7.67	3.17		7.50	3.17		7.67	3.00
S.D:	0.75	1.07		1.50	1.46		1.25	1.63

Did you notice in Table 7.3.4-3 the change in the major headings from "GROUP" to the various levels? I used the same section of the random number table as shown below to block-randomly assign the groups to the levels of the independent variable - and just didn't make another table.

Table 7.3.4-4
From the Random Number Table in Appendix F - Starting at Row 3, Column 1.

[2 4 1 3] 0	4	8	3	6	0	2	2	5	2	7	9	7	2	6	5	7	6	3	9	3
4 2 1 6 7	9	3	0	9	3	0	6	2	4	3	6	1	6	8	0	0	7	8	5	6

7.3.5 PROBLEMS WITH MATCHING

Regardless of which method you use to assign subjects to levels, be aware of the following problems:

1. "NATURAL GROUPS" assigned to different levels - It may be convenient to assign levels to certain pre-existing groups (such as school classes, morning and afternoon volunteers), but this leaves the experiment vulnerable to the confounding effect of factors that formed these groups. (When you must use natural groups, it is no longer a true experiment. See Chapter 10.)
2. SELECTIVE LOSS OF SUBJECTS at some levels but not others. Since different levels may vary in demand placed on subjects, a selective loss of subjects may result in no apparent effect of the independent variable or could create the false appearance of an effect. (Note: This also applies when the groups are formed by random and block-random assignment.)

7.4 FACTORIAL DESIGNS

Up till now we have been using research examples that involved only a single independent variable. As mentioned earlier, it is also possible to do research with several independent variables - each having two or more levels. Most readers have taken an introductory statistics course which included Analysis of Variance (ANOVA). Not only can ANOVA techniques handle any number of levels of an independent variable, they can also handle any number of independent variables - all with two or more levels.

A. COMBINED DESIGNS

You could study two independent variables in the same between-subjects experiment as follows: Test some subjects at all the levels of one independent variable while holding the second independent variable constant. Then test more subjects at different levels of the second independent variable while holding the first independent variable constant. This combined design would be more efficient than running two separate experiments. You would not have to test the two constant levels together twice. Consider the following example.

Which has the greater effect on fatigue while doing sedentary work, room temperature or lighting? To answer this question, we'll measure the fatigue of subjects working at three room temperatures 34^0 , 25^0 , and 16^0 C and at 100, 300 and 900 lux illumination. To study temperature and illumination together, we could hold one variable constant at its normal, mid-range value, while testing the three levels of the other variable; then switch over without repeating the 25^0 - 300 lux combination. We've decided to test 6 subjects at each temperature and each illuminance level. (Testing all conditions once is called a "replication". This design involves 6 replications and requires a total of 36 subjects.) Any of the previous methods can be used to assign 18 subjects so 6 of them are tested at each temperature. Then we repeat the process with 12 subjects to test 6 more at the 900 and 100 lux luminance levels. Without going through the whole assignment process, the overall design would look like this with S1, S2, etc representing subjects that have been assigned using one of the previous methods:

Table 7.4-1.
Design of a Non-Factorial, Between-Subjects Experiment with Two Independent Variables, Temperature and Luminance, Each with Three Levels.

CONDITIONS		SUBJECTS			
TEMP (0 C)	ILLUM (lux)	REP 1	REP 2	...	REP 6
34	300	S1	S6		S26
25	300	S2	S7		S27
16	300	S3	S8		S28
25	900	S4	S9		S29
(25	300)	(don't need to repeat this combination)			
25	100	S5	S10		S30

B. FACTORIAL DESIGNS

Alternatively, you could study the effects of two independent variables in a slightly different way. Test all levels of one variable at all levels of the second. In terms of the above example, that would mean also testing the 900 and 300 lux conditions at 34° and at 16°C. In total there are 9 combinations of temperatures and illuminances. To test 6 subjects at every combination of conditions would require a total of 54 subjects. Again without going into an actual process of assigning subjects to the various conditions, the overall design would look like this:

Table 7.4-2.
Design of a Factorial, Between-Subjects Experiment with Two Independent Variables, Temperature and Luminance, Each with Three Levels.

CONDITIONS		SUBJECTS			
TEMP (°C)	ILLUM (lux)	REP 1	REP 2	...	REP 6
34	900	S1	S10		
25	900	S2	S11		
16	900	S3	12		
34	300	S4	S13		
25	300	S5	S14		
16	300	S6	"		
34	100	S7	"	"	
25	100	S8			S53
16	100	S9			S54

When all combinations of levels of two or more independent variables are tested, the design is called "factorial". As we have just seen factorial designs require substantially more subjects and more work than running separate or combined experiments for each independent variable. Why would psychologists consider doing research this way? Read on.

7.4.1 INTERACTION EFFECTS

Testing two or more independent variables in a single factorial-designed study (as compared to a different study for each variable) enables looking for interaction effects between the variables. When the effect that different levels of one variable have on your measurements are changed by different levels of another variable. An example of an interaction effect encountered by psychologists include is how arousal affects performance. Increasing arousal increases work performance when the task is simple, but decreases performance when the task is difficult. This interaction is illustrated by the following figure.

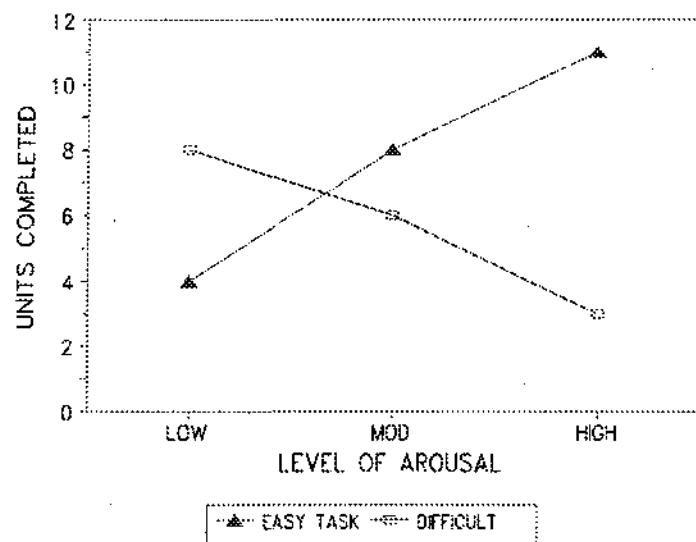


Figure 7.4. How arousal affects performance of easy and difficult tasks.

If task difficulty were not taken into account by using a factorial design, different experiments on arousal could come up with puzzling and conflicting results. When different psychologists obtain different results in similar experiments, it is more likely that some unknown variable is producing an interaction effect, than that one or the other has made some mistake in the procedure. The complexity of human is due to both extraordinary innate processes and to extraordinary sensitivity to changes in the testing conditions and conditions preceding that testing. Factorial experiments are the best way to find out about interactive effects that usually occur to some degree between most of the variable that interest psychologists. The lack of more research employing factorial designs continues to hinder understanding ourselves. So why are factorial designs not always used? The previous example of research on temperature and lighting tells you why. Factorial experiments to test several

levels of more than one independent variable require a lot of subjects and more work than one variable experiments.

7.4.2 ASSIGNING SUBJECTS IN FACTORIAL DESIGNS

You might think that assigning subjects to several independent variables each with two or more levels would be complicated. As you'll see that need not be the case if you just proceed systematically. The first thing to keep in mind, however, is that a true between-subjects factorial design requires a lot of subjects. The total number should a multiple of the product of the number of levels in independent variable. For an experiment with three independent variables such as: 1) three temperatures, cool (16^0), comfortable (25^0), and warm (34^0); 2) three levels of illumination, 30, 90, and 270 lux; and 3) three types of lighting, incandescent, fluorescent, and daylight - you will need some multiple of $3 \times 3 \times 3$ or 27 subjects. Testing 6 subjects at each combination of conditions requires 162 subjects. Yet, all the previous methods work the same as before if we take one simple step. String out all combinations of temperature-illumination-type as if they were simply 27 levels of one independent variable. Then proceed as before. I'll illustrate this for the block-random assignment. That will probably suffice to give you the general idea.

A. BLOCK-RANDOM

This is a situation where the use of block-random assignment has an important advantage over the matching methods. You can start testing subjects right away as they sign up, without waiting until you have all 162. (What if you are concerned that some characteristic which requires laboratory measurement might confound the results? You could do such testing at the same time, and analyze the results afterwards to see if there was a problem.)

To get your random numbers from 1 to 27, you can use a random number table like the one in the Appendix. Just use the numbers as two-digit clusters - we'll do that for this example. For a complete study, you'll go through a lot of random number pairs, but they're cheap. Advanced methods books and references can provide thousands. However, if you are tackling a project of any size, you will probably want to use a computer program anyway to organize and analyze the data. Therefore you might as well use a spreadsheet program to do the assigning using its random number generator customized to give you just numbers from 1 to 27. Afterwards, you can use the same sheet for the analysis.

To begin we go to a random number table and obtain a sequence of numbers between 1 and 27 for every set of subjects to be tested. For block random selection, remember not to select a number again until you have a full set. To illustrate the process I'll begin by taking them from the same portion of the random number table as before and highlight each one that is usable. We list these in sequence and assign subjects to a random number as they show up. For this example we'll assume that the subjects happen to show up in alphabetical order. That will make it easy to see how the random assignment works.

Table 7.4.2-1
From the Random Number Table in Appendix F - Starting at Row 3, Column 1.

24	13	04	8	3	6	0	22	5	2	7	9	7	26	5	7	6	3	9	3				
4	2	1	6	7	9	3	0	9	3	06	24	3	6	16	8	0	07	8	5	6			
3	7	5	7	03	9	9	7	5	8	1	8	3	7	1	6	6	5	6	0	6	12	1	
7	7	9	2	10	6	9	0	7	11	0	0	8	4	2	7	5	1	27	7	5	6		
9	9	5	6	2	7	2	9	05	5	6	4	2	0	6	9	9	9	4	9	8	8	7	2
9	6	3	0	1	9	19	7	7	0	5	4	6	3	0	7	9	7	2	18	8	7	6	

Table 7.4.2-2
Assigning Subjects to the Above Random Numbers.

RANDOM #	24	13	04	22	26	07	16	06	03	12	27	11	10	05	
SUBJECT:	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN	
<hr/>															
RANDOM #	26	18	19	
SUBJECT:	AO	AP	AQ	AR	GF
<hr/>															

Then we organize all combinations of the independent variables and their levels into a list as shown in Table 7.4.2-3. Each combination of levels is given a code number from 1 to 27. The subjects are assigned to the combination of levels whose code equals their random number. I haven't completed the assignments - just done enough to give you the idea,

Table 7.4.2-3.
How Subjects Were Assigned in a Between-Subjects, Factorial Experiment with
 3 Levels of Temperature, 3 Levels of Illuminance, and 3 Types of Lighting.

TEMP	ILLUM	TYPE	CODE#	REP 1	REP 2	...	REP 6
34	60	INC	1				
34	60	FLU	2				
34	60	DAY	3	AI			
34	180	INC	4	AC			
34	180	FLU	5	AN			
34	180	DAY	6	AH			
34	540	INC	7	AF			
34	540	FLU	8				
34	540	DAY	9				
25	60	INC	10	AM			
25	60	FLU	11	AL			
25	60	DAY	12	AJ			
25	180	INC	13	AB			
16	540	DAY	27	AF			

B. MATCHING

The procedures for matching on the basis of subject categories or some numerical characteristic are also similar to those used for single independent variable designs. Once the levels are arranged in a linear fashion as was done above in Table 7.4.2-3, the sorting and/or systematic assignment to levels are the same.

EXERCISES

1. For each of the following between-subjects designs, find a recent journal article which has used that design. Use APA format to provide a reference for each article below. Copy the relevant portion of the METHODS section and circle or high-light the words that identify how subjects were assigned to the levels of the independent variable.

pure or blocked random assignment: _____

matched assignment by category: _____

matched assignment by a numerical value: _____

2. Ask a colleague to quickly write down 50 numbers from 1 to 6 chosen at random. Omit the last 2 numbers. Perform a Chi-Square analysis of the frequency of occurrence and calculate an R value for each number category. Did any numbers occur significantly more or less often than expected by chance?

3. Below is a list and some relevant information on subjects who have volunteered for a between-subjects experiment that has three levels (ZERO, MIN, and MOD) of the independent variable. Assign the subjects to the levels as using the following methods: a) randomly; b) block randomly; c) according to sex and 3 age categories (teens = < 20, peers = 20 to 25, mature >25); d) according to sex and numerical age; e) in terms of both Freundlich Score, and assertiveness rating; f) using trial error starting with the results of any of the above methods. Summarize and compare the results obtained with each method by filling out the following table:

Table 7Q3-1.

Characteristics of 24 Subjects for Between-Subjects Experiment in Chapter 7, Question 3

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
SEX	F	F	M	F	M	M	F	M	F	F	M	F	M	M	M	F	F	F	M	M	F	M	F	N
A.C.	t	t	p	m	p	t	p	m	t	p	p	p	t	m	m	t	m	p	m	t	m	t	m	
YRS.	16	18	21	27	17	29	23	30	19	21	18	22	19	23	27	17	30	24	28	22	28	24	29	1
F-S.	3	4	6	8	5	9	6	9	5	6	5	5	5	6	7	4	8	7	7	5	8	7	8	
A.R.	3	4	2	1	3	2	2	4	4	3	3	2	3	1	3	2	3	1	2	3	3	2	4	

Table 7Q3-2.

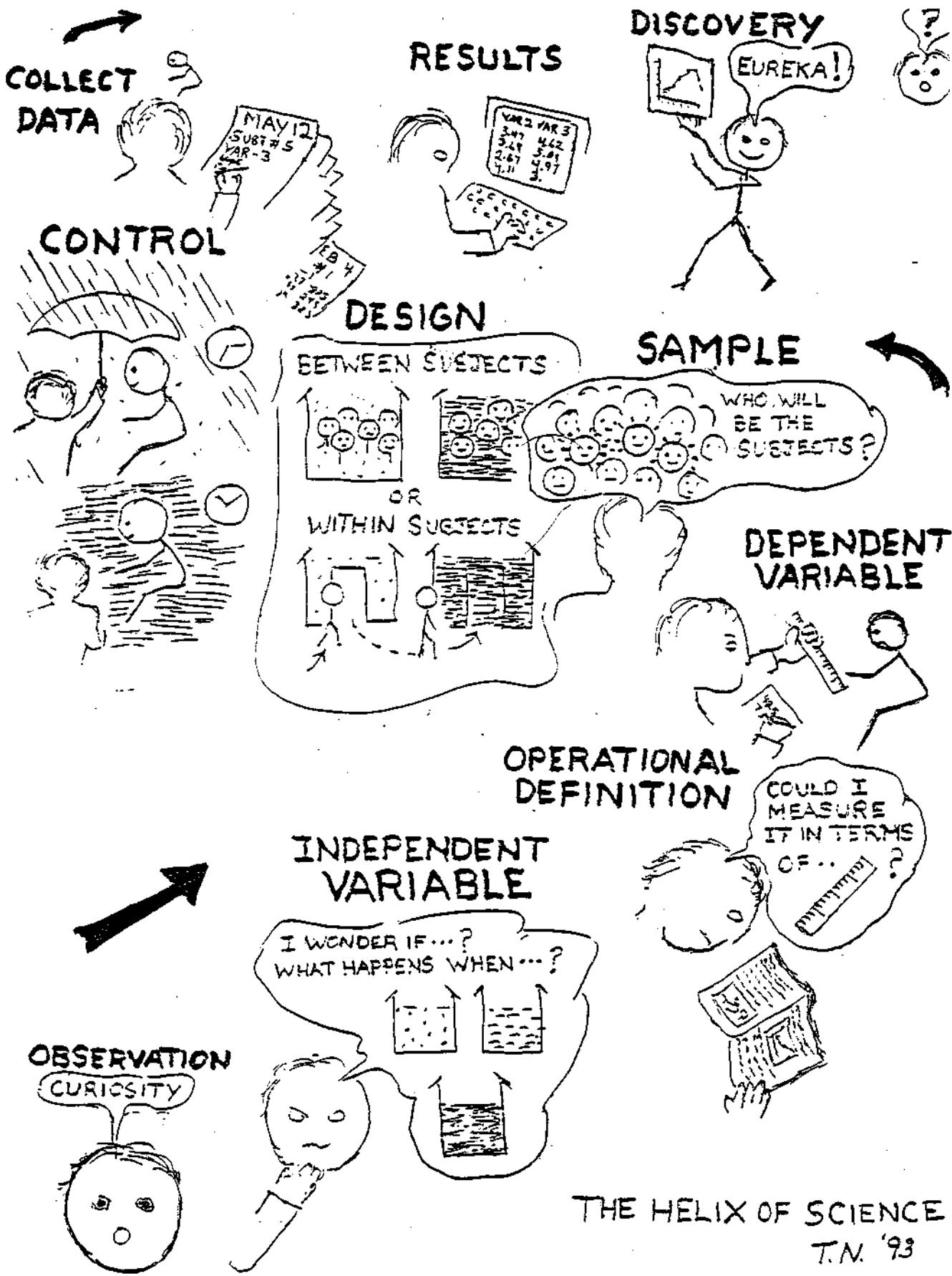
Number or Means of Various Characteristics of Subjects Assigned to Various Levels Using Various Methods to Assign the Subjects.

LEVEL	CHAR.	RANDOM	BLOCK RAND.	AGE CAT. & SEX	YEARS & SEX	F-SCORE & A RAT.	TRIAL & ERROR
ZERO	# fem						
	# male						
	# teen						
	# peer						
	# mat						
	mean age.:						
	F-Score:						
	A Rating:						
MIN	# fem						
	# male						
	# teen						
	# peer						
	# mat						
	mean age.:						
	F-Score:						
	A Rating:						
MOD	# fem						
	# male						
	# teen						
	# peer						
	# mat						
	mean age.:						
	F-Score:						
	A Rating:						

Table 7Q3-3.

Answers to Number and Means of Various Characteristics of Subjects Assigned to Various Levels Using Various Methods to Assign the Subjects.

LEVEL	CHAR.	RANDOM	BLOCK RAND.	AGE CAT. & SEX	YEARS & SEX	F-SCORE & A RAT.	TRIAL & ERROR
ZERO	# fem	7	6	3	4	3	4
	# male	9	3	3	4	5	4
	# teen	5	5	3	2	3	matched
	# peer	6	2	2	3	3	by avg
	# mat	5	3	2	2	3	age
	mean age.:	21.7	23.4	23.0	22.8	22.4	22.8
	F-Score:	5.8	6.5	6.2	5.8	6.1	6.1
	A Rating:	2.8	2.4	2.8	2.8	2.8	2.6
MIN	# fem	4	3	3	4	4	4
	# male	1	5	3	4	4	4
	# teen	1	3	2	1	2	
	# peer	2	4	2	4	3	
	# mat	2	1	2	3	3	
	mean age.:	23.2	22.9	20.7	22.8	24.8	23.0
	F-Score:	6.0	6.1	5.5	6.5	7.0	6.1
	A Rating:	2.2	2.6	2.7	2.4	2.1	2.8
MOD	# fem	1	3	3	4	5	4
	# male	2	5	3	4	3	4
	# teen	2	2	2	4	2	
	# peer	0	1	2	1	2	
	# mat	1	4	2	3	4	
	mean age.:	28.3	22.3	24.7	56.0	21.4	22.8
	F-Score:	8.0	5.8	6.7	6.1	5.3	6.1
	A Rating:	3.0	3.0	2.7	2.9	6.1	2.6



CHAPTER 8

WITHIN-SUBJECTS EXPERIMENTS

These research designs are experiments because the INDEPENDENT VARIABLE is controlled by the researcher. They differ from between-subjects experiments in that each subject is tested at each level of the independent variable. This prevents subject differences from being confounded with the effects of the independent variable's levels. Unfortunately, testing the same subject two or more times creates a host of other problems, which require know-how to overcome.

8.1 HOW THEY ARE DESIGNED

Let's say you wanted to test the effects of three conditions "red", "green", and "blue" on some response such as the number of correct answers. If you were using a BETWEEN-SUBJECTS design, you could find 18 subjects and assign 1/3rd of them to the "red" condition, 1/3rd to the "green", and 1/3rd to the "blue" condition by one of the methods explained in the previous chapter. To use a WITHIN-SUBJECTS design instead and obtain the same amount of data, you would only need 6 subjects and test each in the "red", "green", and "blue" conditions. The basic organization of the two designs is shown in the following table:

Table 8.1

How Subjects Are Tested at Three Levels, Red, Green, and Blue, of an Independent Variable Using Between-Subjects and Within-Subjects Designs

BETWEEN-SUBJECTS			WITHIN-SUBJECTS		
RED	GREEN	BLUE	RED	GREEN	BLUE
Ali	Bo	Chris	Ali	Ali	Ali
Don	Evan	Fran	Bo	Bo	Bo
Gale	Hap	Izz	Chris	Chris	Chris
Jan	Kim	Lee	Don	Don	Don
Max	Nate	Pat	Evan	Evan	Evan
Ron	Sam	Terry	Fran	Fran	Fran

8.2 ADVANTAGES

8.2.1 DON'T NEED AS MANY SUBJECTS

It can take considerable time to recruit and train subjects to participate in research. In a between-subjects design, the number of subjects needed gets multiplied by the number of levels of the independent variable. With several levels or more than one independent variable (see the end of this chapter) the number of subjects needed for a between design may become unattainable. Within-subjects designs do not generally require more subjects as the number of levels increases.

8.2.2 DON'T NEED TO MATCH SUBJECTS

This can be time consuming. Consider for example the difficulties of administering to every subject in a cognitive experiment a Wechsler Adult Intelligence test, which takes several hours and requires a trained administrator. Even more challenging are experiments where you do not understand enough about the behavior being studied to know on what basis subjects should be matched. For example, how would one go about equating subjects for aesthetic sensitivity?

8.2.3 LESS RANDOM ERROR -> MORE POWER

Consequently you are more likely to obtain statistically significant differences between means. Within-subjects experiments are the most powerful research method known. This sort of power is demonstrated by two widely-used statistical tests:

INDEPENDENT VERSUS PAIRED T-TESTS

Two means that don't differ with an independent test, may differ significantly with a dependent test - even though the same numbers are used. There is nothing "magical" here. Subjects tend to vary in sensitivity, ability or some other characteristic that affects how they respond to any independent variable that you are likely to test. A subject that is more sensitive than average at one level tends to be more sensitive at another level as well. Consequently the differences of the subjects' responses tend to be less variable than the response themselves.

Consider the data in Table 8.2.2 which represent hypothetical results of measurements obtained from subjects in two conditions, RED and GREEN. The t-test on the left is the one to use for between-subjects designs when different subjects are tested in the two conditions. Though the means of the responses in the RED and GREEN conditions appear to differ substantially, 5 and 9 respectively, there is no statistically significant effect here. The reason is too much variability in the subjects' responses.

A different conclusion is warranted if these data had been obtained from only five subjects, each of whom was tested in both the RED and the GREEN condition. Then you should use the "paired t-test" on the right. Here each row represents the results from a single subject. Notice how the first two subjects' responses are lower than the mean, whereas the last two subjects' are higher than the mean. Since these differences between the subjects occur in both conditions, we can infer that the first two subjects are generally somewhat slower or more accurate than the last two subjects. That variability is a characteristic of the subjects not of variability in the measurements.

To take that variability into account, one could look at the relative effects that the two conditions had on each subject. This can be done simply by subtracting each subject's result in condition RED from his/her result in condition GREEN. That is done in the right hand section of Table 8.2.3. The question of a significant effect now depends only on how consistently did all subjects show the same relative effect. That is the same as asking, "How variable are those differences compared to the size of the mean difference?" The paired t-test answers this question systematically, and a significant effect is found.

Table 8.2.3-1
t-Tests on Subjects' Responses in Conditions "Red" and "Green":
A Comparison of Between and Within Designs

BETWEEN						WITHIN			
RED	diff from mean	diff sqr'd	GREEN	diff from mean	diff sqr'd	GREEN	diff minus RED	diff from mean	diff sqr'd
2	3	9	1	8	64	-1	5	25	
3	2	4	8	1	1	5	1	1	
5	0	0	9	0	0	4	0	0	
5	0	0	11	2	4	6	2	4	
7	2	4	12	3	9	5	1	1	
8	3	9	13	4	16	5	1	1	
Sum:	30	26	54	94		24		32	
Mean:	5	4.33	9	15.7		4		5.33	
Standard Deviation:		2.08		3.96				2.48	
Standard Error:		0.931		1.77				1.033	

$$\text{SE combined} = \sqrt{(0.93^2 + 1.77^2)} \\ = 2.00$$

$$t = (9 - 5) / 2.00 \\ = 2.00 \text{ NOT SIGNIF.}$$

$$t = 4 / 1.033 \\ = 3.873 \text{ SIGNIF!}$$

$$\text{at } (6-1) + (6-1) = 10 \text{ df,} \\ \text{need } t = 2.228$$

$$\text{at } (6-1) = 5 \text{ df,} \\ \text{need } t = 2.571$$

Note: These t-test calculations use the conceptual method explained in my introductory statistics text, Statistics: Number Work for Psychologists.

BETWEEN VERSUS WITHIN ANOVA's

Table 8.2.3-2 presents some hypothetical data from three subjects in an within-subjects experiment with thee levels, X, Y, and Z.

Table 11.4.1-1
Hypothetical Effect of Three Levels of an Independent Variable on Responses by Three Subjects.

SUBJECT	X	Y	Z	SUBJ. MEAN
Lu	1	2	3	2
Mo	1	5	6	4
Ny	4	5	9	6
MEAN:	2	4	6	

If these same results had been obtained with a between-subjects design, the usual calculations would produce the following ANOVA table:

SOURCE OF VARIANCE	SUM OF SQUARES	degrees of freedom	MEAN SQUARE	F
XYZ LEVELS	24	2	12	2.4
ERROR	30	6	5	
TOTAL	54	8		

At 2 and 6 degrees of freedom, F needs to exceed 5.14. There is no significant effect due to the XYZ levels by this analysis.

In the above analysis, all of the variability not attributable to the levels was attributed to random error. Some of that so-called error variability is indeed due to random effects on the measurements, but some is due to subject differences. Testing each subject at each level makes it possible to calculate the variability attributable to subject differences. The subjects variability can be calculated by considering how the results would look if the only effect that occurred was subject differences. The results would then look like this:¹

SUBJECT	X	(diff from grand mean) ²	Y	(diff from grand mean) ²	Z	(diff from grand mean) ²
Lu	2	4	2	4	2	4
Mo	4	0	4	0	4	0
Ny	6	4	6	4	6	4
SUM:		<u>8</u>		<u>8</u>		<u>8</u>

All the subject means in these ideal data were subtracted from the grand mean of 4 and the differences were squared. The sum of these squares, $8 + 8 + 8 = 24$, is the variance attributable to subject differences and nothing else. A proper ANOVA table for the within-subjects design can now be made:

SOURCE OF VARIANCE	SUM OF SQUARES	degrees of freedom	MEAN SQUARE	F
XYZ LEVELS	24	2	12	8
MEASUREMENTS	30	6	5	
SUBJECTS	24	2	12	
ERROR	6	4	1.5	
TOTAL	54	8		

¹ This conceptual method of doing ANOVA is fully explained in my statistics book, "Statistics: A Number Work Introduction". If you find doing $\sum X^2 - (\sum X)^2/n$ calculations incomprehensible, you may like the conceptual method.

Subtracting the subjects' variance from what is now the measurements variance leaves a rather small variance of "6" to be attributed to random error. The new F-value of "8" is significant since at 2 and 4 degrees of freedom the critical value for F is 6.94.

Repeated measurements of each subject permits calculating the variability of the subjects. This accounts for some of the overall variability in the data. Once recognized, this subject variability can be removed from the overall variability of the data. This leaves a smaller variance attributable to random error. (The same thing is actually accomplished in the paired "t" test, when taking the differences.)

8.2.4 WHAT TO DO WITH THE EXTRA POWER

DO A MORE SENSITIVE EXPERIMENT WITH SAME AMOUNT OF DATA

(A within-subjects design may enable you to find a significant effect that would not otherwise not have been proven to be significant using a between-subjects design)

OR

REDUCE THE AMOUNT OF DATA COLLECTED

(This can save you time, effort, and money.)

Deciding amongst these two alternatives is one of the pleasant chores accompanying within-subjects designs. The greater ability to reveal an effect was demonstrated in the previous section and the consequences of that are self evident. However, whether to collect less data requires further consideration. You could either reduce the number of repeated measurements obtained with each subject at each level, or reduce the number of subjects. For example, the within-subjects results in Table 8.2.3 are still significant if the data from any one subject had not been collected. If subjects are difficult to find, you may have no alternative. Otherwise it is a matter of judgement whether to go for more power or efficient use of resources. Some statistics text books recommend power analysis techniques to determine how much data are needed. True you may have some idea of how variable the data will be based on previous research of a similar nature - either your own or published by others. I've found such analysis of little value: If I'm doing something similar to a previous study that I have done, I will know about how many measurements are needed. However, the most fun research is when you are doing something quite unlike what has been done previously. Then you may have no idea what the variability will be until you have the data. There is no way to avoid the fact that all research is a gamble - the more innovative, the bigger the gamble.

8.3 DISADVANTAGES

With so many advantages, you may wonder why psychologists use anything else. Unfortunately, this silver lining does have its cloud.

8.3.1 CAN'T BE USED FOR SOME KINDS OF RESEARCH

Human and animal subjects do not come equipped with an "reset buttons" for their memories. Consequently after a subject has been tested at one level of certain independent variables, it is not possible for the subject to respond to other levels without the influence of the first level confounding the results. Such awkward independent variables include:

1. Methods of Teaching - (has something to do with teaching old dogs the same trick in different ways)
2. Variables Involving Novelty - (even a three-month old's attention wanes by the 7th time you've said "BOO!")
3. Subject Characteristics such as sex, aptitude, personality, age, etc. - (these require the use of "quasi-experimental designs, which we'll get to in the last chapter)

There are certain types of questions and other dependent variables that do not work well in within-subjects designs because they are not readily altered by any independent variables. These include:

1. Attitudes - (true attitudes are not readily altered by changing some independent variable)
2. Long-term memory - (you either remember it or you don't)
3. Performance measurements intended to represent skill - (skill generally only improves)
4. Spontaneous responses - (the second time around is unlikely to be really spontaneous)

There are also certain types of subjects who are not generally suitable for within-subjects experiments because it takes considerably more time to explain and test several levels of an independent variable:

1. Members of the general public, such as you may meet "on the street" - (People are usually busy. They may have time for a single question or a quick task, but you may be pushing your luck to try changing an independent variable and have the subjects do the experiment again.)
2. Limited time availability of some professionals, the very elderly, infants, or ill persons.

8.3.2 NEEDS MORE TIME AND EFFORT FROM EACH SUBJECT

The more time it takes to do an experiment, the more difficult it will be to find person willing to participate. If more than 45 minutes is required of each subject, it is fair to compensate subjects. The professors for some psychology courses may consider that certain types of research participation by class members contributes to their understanding of the course or the discipline. Accordingly many offer 1 or 2 grade points to participants. This is so common in introductory psychology courses half of all psychology data on human subjects is based on Psychology 100 students. If more than an hour or two is needed, it may be necessary to pay subjects. An honorarium of \$5 or more depending on the required time (perhaps officially specified as compensation for travel expenses to avoid tax hassles) is one way. It is not unusual for some experiments to require daily testing sessions over several weeks. Then an hourly wage may be the most appropriate compensation. Funds to pay research subjects is one of the expected costs in many areas of psychology.

8.3.3 MAY REQUIRE AUTOMATION

Any thing you can do to reduce the amount of time required of your subjects is worth considering for two reasons: 1. The longer it takes to do an experiment, the more difficult it is to recruit subjects. In some cases, automation may make the difference between being able to test 40 subjects as free volunteers versus trying to recruit and pay for some minimal number of subjects. 2. Even more important may be the effects of quicker measurement procedures on maintaining the subject's interest, avoiding fatigue, and generally reducing random error in the measurements. To illustrate these points, consider the following example:

Let's assume that we want to measure the "believability" of advice. One independent variable is to be the age of the person offering the advice. Four age levels are to be tested: a high school student, a freshman undergraduate - the subjects' peer group, middle aged, and an elderly. The second independent variable is the sex of the person offering the advise: either the same or different from the subject. A third independent variable is the difficulty of the instructions: easy, moderate, or difficult. The fourth independent variable is the value of the outcome for following the instructions: no particular value, feedback of how the subject's score compared with others, and a \$1.00 for every point earned. testing all four independent variables at every level amounts to $4 \times 2 \times 3 \times 3$ or 72 conditions in all to be tested. Based on preliminary testing, various steps in the procedure of running the experiment were found to take following mean times:

ACTIVITY	PERFORMED BY	TIME (sec)
select and present descriptions of advisors	experimenter	11
read description of advisor and look at picture	subject	8
select and present the advice	experimenter	7
read the advice	subject	12
decide what to do and make the response	subject	9
record the response	experimenter	6

The total time to test each combination of levels is 54 seconds. Multiplying that by the 72 combinations means a total testing time of 65 minutes. In addition we should allow 10 minutes for introductions, explanation of the experiment, and a bit of practice. At the end, another 5 minutes should be available for debriefing, answering any questions and thanking the subject. Therefore, it will take about 80 minutes to test the average subject. So we ought to allow close to 1 1/2 hours per subject. This is rather long time to expect subjects to volunteer their time, and may exceed what course credit guidelines for students in the introductory courses.

If we automate the above experiment by having a computer present the descriptions, activate a slide projector to show a picture of the advisor, offer the advice, and record the subject's response, the testing duration shortens substantially. It takes the computer essentially no significant time to perform these tasks. Eliminating the experimenter times from the above chart, results in a test time of 28 seconds per condition and a total time of less than 34 minutes to test the average subject. With an additional 15 minutes at the beginning and end of the session, we should be able to test most subjects within an hour.

This will make it considerably easier to obtain subjects for the experiment.

Automating psychology experiments requires some special knowledge, skill, and/or resources such as technicians and equipment. It may also take considerably longer to automate a particular experiment. (The additional preparation time is often made up in time saved to prepare subsequent experiments of a similar type.) Even with technicians available, you still have to understand something about what is involved in order to know what to request. However, automation of most psychology experiments is hardly rocket science. Check out Appendix E - A Primer on Research Automation to see what it's about.

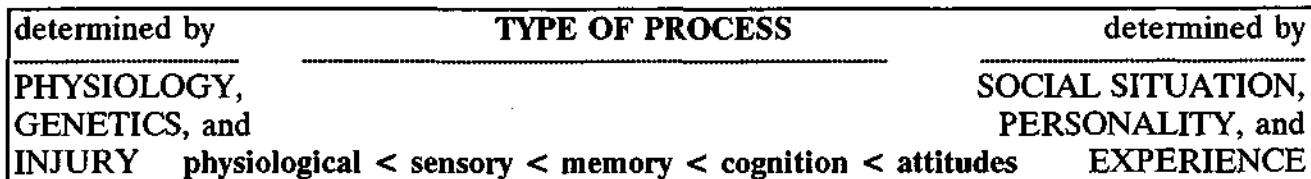
8.3.4 MORE VULNERABLE TO SUBJECT IDIOSYNCRASIES

As explained in the previous chapter, experiments do not usually attempt to test a representative sample of some general population. However, when data are obtained from only a few subjects, it is particularly important to avoid testing person's who have a certain special skill or disability in performing the task. For example, if you were doing an experiment where color discrimination played a role, a single subject with a mild color weakness or who misunderstood some aspect of the response task could substantially alter the results when he/she provides one-third of the data. One thing to do is to put more emphasis on training the subject by doing practice runs and providing feedback. Just be careful not to bias the subject in the process. Another thing is to test potential subjects' abilities that are important to the independent or dependent variables. Those whose performance does not meet some normal criterion specified for a test are not used. This is called "screening" the subjects, and should be reported in the METHODS section of your report - either under "Subjects" or "Procedure". The following table lists some of the commonly used screening tests. (Between-subjects designs can benefit from using these techniques too to reduce their error variance.)

Table 8.3.4
Various Screening Tests

TYPE	NAME	HOW IT WORKS
visual acuity	Snellen Chart	read different size letters at 20 feet
color vision	Dvorine or Ishahara Tests	color vision - patterns formed by colored dots
hearing	Audiometry Test	spoken numbers at different loudness levels
smell	Sensormetrics	olfaction - uses "scratch and sniff" cards
intelligence	Wechsler Adult Intelligence Scale (WAIS)	various subtests for verbal reasoning, math, and spatial abilities - often used individually
thinking	Miller Analogies Test	ability to understand various relations
reading	Metropolitan Readiness	recognition of letters and letter combinations
coordination	Pursuit Rotor Test	eye-hand skill at following a moving hole

Why not always screen your subjects using all these tests? It would certainly reduce the error variance in your data. However, it simply is not practical. Some of these tests' like the WAIS cost \$20 or more each; some of them take 20 minutes or more to do. Certain tests like intelligence tests may be considered by the subjects as an invasion of privacy as well as being fairly demanding to do. You should also realize that the effects of individual differences depends on what is being measured. Physiological and sensory processes tend to be quite consistent across individuals despite differences in personality experience and attitudes. These processes are largely determined by genetics, though they may also be injured. Intelligence and personality characteristics make little difference except perhaps in the degree of cooperation. On the other hand, physiological and sensory difference have little effect on thinking and attitudes except to the extend that they may interfere with how you make the measurements. The following diagram summarizes this perspective:



The following two problems with Within-Subjects designs arise directly as a result of the repeated measuring needed to test a subject at each level of the independent variable. They represent the most serious problems in using within-subjects designs. They can be reduced by attending to the sequence in which the levels are presented to the subject. How to do this is the most important part of the chapter.

8.3.5 SERIAL-POSITION EFFECTS

These are due to changes in the subject as the subject makes repeated responses and time passes in the experiment:

1. Practice Effects - Most experiments involve presenting a subject with some type of information and asking them to make a response. As subjects do this repeatedly, they improve somewhat in acquiring and interpreting the information as well as in how efficiently they respond. The result can be a greater sensitivity to stimuli, and faster or more accurate responses.

2. Fatigue Effects - One characteristic of well controlled experiments is that they are boring. Ideally, nothing should change except the precise changes introduced by the different levels of the independent variable. As subjects continue their participation, they are likely to get bored, have difficulty paying attention, and feel increasingly tired. This tends to result in less sensitivity to stimuli and slower or less accurate responses.

3. Relief Effects - When subjects anticipate the that the session is nearly over, they often experience a rekindling of attention and loss of tiredness. Consequently there may be a paradoxical increase in sensitivity and improved performance for those levels tested near the end of a session.

Figure 8.3.5-1 illustrates three hypothetical examples of the above serial-position effects as they develop over the course of an experimental session. The practice effect shows rapid improvement in performance during the first 20 minutes.; then performance becomes fairly stable though there is some modest improvement towards the end of the session. For the first hour of testing there is no appreciable fatigue; it then develops at an accelerating rate of declining performance till the session ends. A relief effect produces a bit of improved performance during the last 15 minutes of testing as the subject knows it will be soon be over.

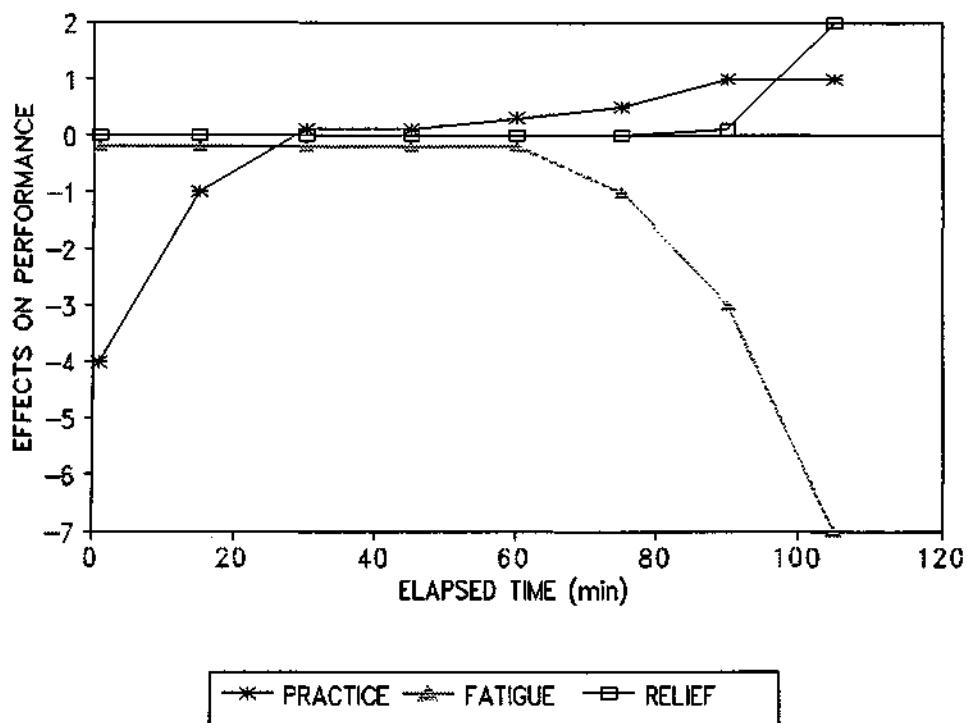


Figure 8.3.5. How three types of serial-position effects develop over the course of an experimental session.

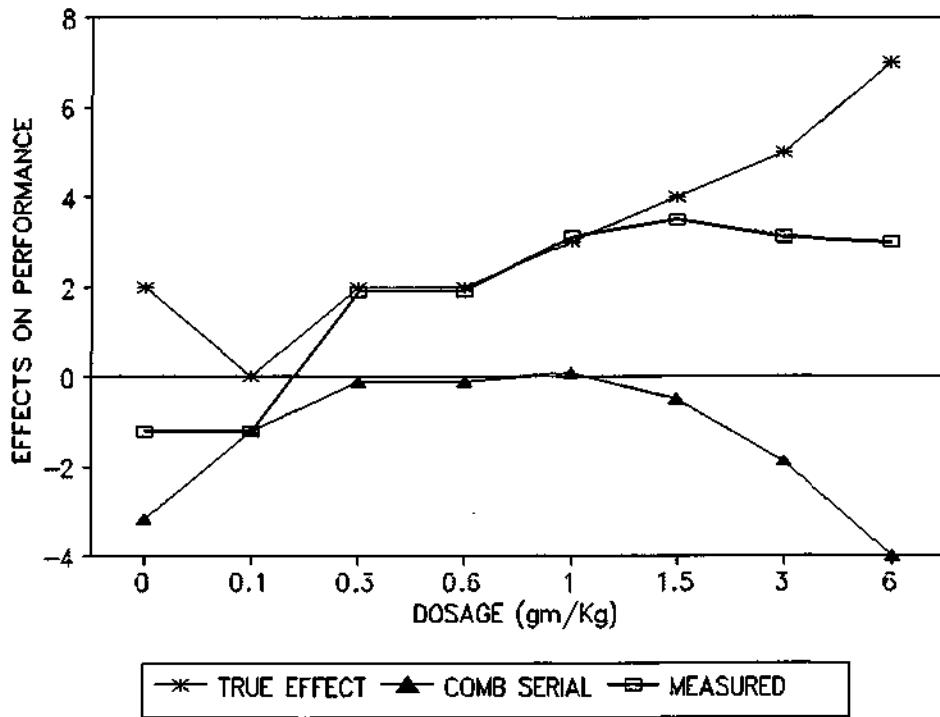


Figure 8.3.5-2. The combined effect of the three serial-position effects shown in Figure 8.3.5-1 interact with the true effect of 8 successively higher drug doses tested over a 105 minute session to produce the measurements obtained in this hypothetical experiment.

When effects of different levels of an independent variable are measured over the course of an experimental session, serial position effects combine to distort those

measurements. What can occur is illustrated in Figure 8.3.5-2. In it the three serial-position effects in Figure 8.3.5-1 are combined. Their combined effect interacts with a hypothetical true effect of 8 drug dosages that were tested at successively higher levels over the course of a session that lasted as long as the session in the Figure 8.3.5-1. Note how the measurements indicate maximum performance at the 1.5 gm/Kg dosage with declining performance at higher doses - quite different from the true effect. In actual research, one never knows either the true effect of the levels nor the serial position or other effects that may be distorting ones measurements. You can only suspect that some distortions may be present and design the experiment so as to minimize them. That what most of this chapter is all about.

8.3.6 CARRY-OVER EFFECTS

Sometimes a stimulus has an effect which alters the subject's sensitivity or response to subsequent stimuli. These effects become more prominent the greater the difference between successive stimuli. Three types illustrate these effects:

1. Changes in sensory sensitivity - especially when a **STRONG STIMULUS** precedes a weak one.
2. Expectancies arising from information inherent in levels that are presented prior to other levels.
3. Attitudes evoked by initial stimuli and which continue to affect the subject during subsequent stimuli.

NB: ORDER AND CARRY-OVER EFFECTS can also occur in RESEARCH APPARATUS due to:

1. Equipment warm-up, fading and wear.
2. Hysteresis arising from desensitization or sensitization of detectors or arising from inertia or rebound in output components.

These can affect both within- and between-subject designs. They are NOT specifically a disadvantage of within designs.

8.4 REDUCING ORDER AND CARRY-OVER EFFECTS

The required tactics are really not new. You cope with these effects in the same basic way as you have learned to cope with other variables which you can't avoid or hold constant. Namely, you prevent them from having a SYSTEMATIC effect on the measurements at various levels by assuring the effects occur randomly or by deliberately balancing them so they affect all levels equally. Balancing the order in which levels are tested is a lot like matching the subjects assigned to various groups for between-subjects experiments and like balancing environmental variables which you are unable to hold constant.

For each of the following methods, we shall consider the sequence in which to test an independent variable which has 5 levels. For a practical example consider the following research scenario. To operate the Canada Arm by remote control, an astronaut uses a camera mounted on the arm. To improve visibility, five image enhancement techniques are available: averaging (avg), brightening, (bri), contrast increase (cont), density equalization (dens), and edge finding (edg). An experiment is performed to determine which technique is best for accurately positioning the arm. The dependent variable will be distance between a target and the end of the arm after 30 seconds of operation. The experimenter must now decide on the order in which the five imaging techniques are to be tested.

8.4.1 RANDOM PROCEDURES

First a few words about "random numbers". They are not simply numbers you make up as you go along. People are not good at making up numbers that are truly random. When they try to do so, they tend to have too many odd numbers and repeat numbers less often than happens in truly random sets. You can not claim to have used random numbers unless you have obtained the numbers from a random number table, such as found in statistics books or a method of obtaining them that is recognized as valid such as drawing numbered slips from a hat or using a random number generator available in some computer programs.

8.4.1.1 TESTING LEVELS IN RANDOM ORDER

This method is fast, simple, and poor. It is sometimes used in computer controlled experiments where many measurements are made rapidly. We'll start with this method because it's easy to see how it works and why it's not usually the method of choice.

A. EXAMPLE OF HOW IT'S DONE

1. Decide how many times you will test each subject, and make a list with that many rows. For this example, we'll test each subject 12 times.
2. Assign a number to each of the levels. For this example: avg = 1, bri = 2, cont = 3, dens = 4, edg = 5.
3. Pick a location on a random number table. Going either up, down, left or right, read the numbers. Note the first number that is either a 1, 2, 3, 4, or 5.
4. Write the level assigned to that number at the top of the list for the first subject.
5. Continue in the same direction getting more 1's, 2's, 3's, 4's or 5's, adding them to the list for that subject, and entering the designated levels.
6. When you reach the end of a row or column go to the next. (For this example, I shall drop down to the right end of the next row go to the left.)
7. Continue until you have provided for enough measurements.
8. Continue along the random number table for each subject.

Table 8.4.1.1
A Listing of the Sequences in Which the Various Image Enhancements
Were Randomly Selected for Testing with Each Subject.

TEST NUMBER	SUBJECT											
	ALI		LES		PAT		SAM					
	RAND	LEVEL	RAND	LEVEL	RAND	LEVEL	RAND	LEVEL				
1	1	avg	3	cont	3	= ?						
2	4	dens	2	bri	1							
3	1	avg	2	bri	?							
4	5	edg	2	bri								
5	1	avg	4	dens								
6	1	avg	1	avg								
7	1	avg	3	cont								
8	5	edg	4	dens								
9	2	bri	3	cont								
10	3	cont	2	bri								
11	5	edg	2	bri								
12	4	dens	3	cont								

Below is a portion of the first groups of rows and columns in upper left portion of the random number table in Appendix F. For clarity, I began filling in the above table by starting with the number in the first column of the first row. That happened to be a "1", which determines that "avg" is the first technique that Ali will use. To determine the next technique, I proceeded to the right along the first row and used each suitable number in succession. On reaching the last number in that row, I went down directly to the right end of next row and proceeded left. The numbers I used in Table 8.4.1-1 above are indicated in bold print on the portion of the random number table below. Complete the testing sequence for Pat, then do Sam's. Compare your results to those of a colleague to find out if you are doing it correctly/ (In practice you may start at any location and proceed in any systematic manner that strikes your fancy - including south-east to north-west diagonals by starting from the lower right hand corner of the random number table.)

Random Numbers from the Table in Appendix F

[1	0	4	8	0	1	5	0	1	1	0	1
2	2	3	6	8]	[4	6	5	7	3	2	5
2	4	1	3	0	4	8	3	6	0	2	2
4	2	1	6	7	9	3	0	9]	[3	0	6
3	7	5	7	0	3	9	9	7	5	8	1
7	7	9	2	1	0	6	9	0	7	1	1
9	9	5	6	2	7	2	9	0	5	5	6

B. PROBLEMS:

- 1) This method usually results in some levels being tested more often than others.

In the above example, the number "5" never came up for Ali, so that subject was never tested with the edge finding enhancement.

- 2) Requires a lot of measurements to avoid random bias - If Ali was exceptionally skilled at operating the arm, the fact that Ali never worked with the edge finding enhancement could lower its average accuracy score relative to the other techniques. Ali would have to be tested 19 times to obtain at least two measurements at each level.
- 3) Frequent juxtaposition of very different levels leaves the design vulnerable to carry-over effects. While that does not happen to be a problem in the above example, it could be if the levels involved increasingly intense or complex stimuli. If a "1" represented a high intensity and "5" a low intensity, there could be problem with the above sequences. Ali is always tested at relatively low intensity of "4" immediately after a relatively high intensity of "2". Les is tested at the lowest level of "5" immediately after being tested at a "1".

8.4.1.2 BLOCK-RANDOMIZATION

The idea here is to select the levels "sort of randomly", while ensuring that every level is tested once before any level is tested again. As long as the number of measurements is a multiple of the number of levels, all levels will be tested an equal number of times. By the way, a sequence of levels that has one of each level is called a "replication". To be considered a within-subjects experiment, each subject must be tested on each level at least once. In the following example, we'll test each subject four times at each level.²

Block-randomization was applied to the hypothetical example of serial-position effects illustrated back in Section 8.3.5 (p. 186). Figure 8.4.1 shows the average measurements obtained as each of the eight dosage levels are tested in block random-orders with 6, 18, and 72 replications. Note how the shape of average effect comes closer to that of the true effect as the number of replications are increased. Yet even with 36 block-random replications, the average performance at every level is consistently lower than that of the true effect. No matter what design you use to minimize serial position and order effects, you can not remove them from your measurements. The best you can do is to avoid having such effects introduce a bias that affects some levels more than others.

²One of the first things I ever did on a digital computer was to make random sequences of 1's, 2's and 3's that had equal numbers of each for a within-subjects experiment. I wrote a little trial-and-error program that worked after a few minutes delay. Then I asked for several much larger sets. Nothing happened after I pressed "enter", but neither did it give an error message. Not knowing how long computers took to do such things, I went away after a while and checked the terminal several times over the course of the day. It was the one of first university systems using a new Amdahl super computer. I was one of its first student users. Time on it was leased to oil companies for thousands of dollars a second but free for students. Coming back the next morning, I was pleasantly surprised to find a printout of hundreds of numbers that were just what I wanted. Months later I heard of an immensely expensive incident that occurred shortly after the system had been put on-line. It seems that the computer suddenly got hung up for 19 hours doing some unfathomable calculations. No one could stop it. Then it suddenly resumed normal operation. Just coincidence of course.

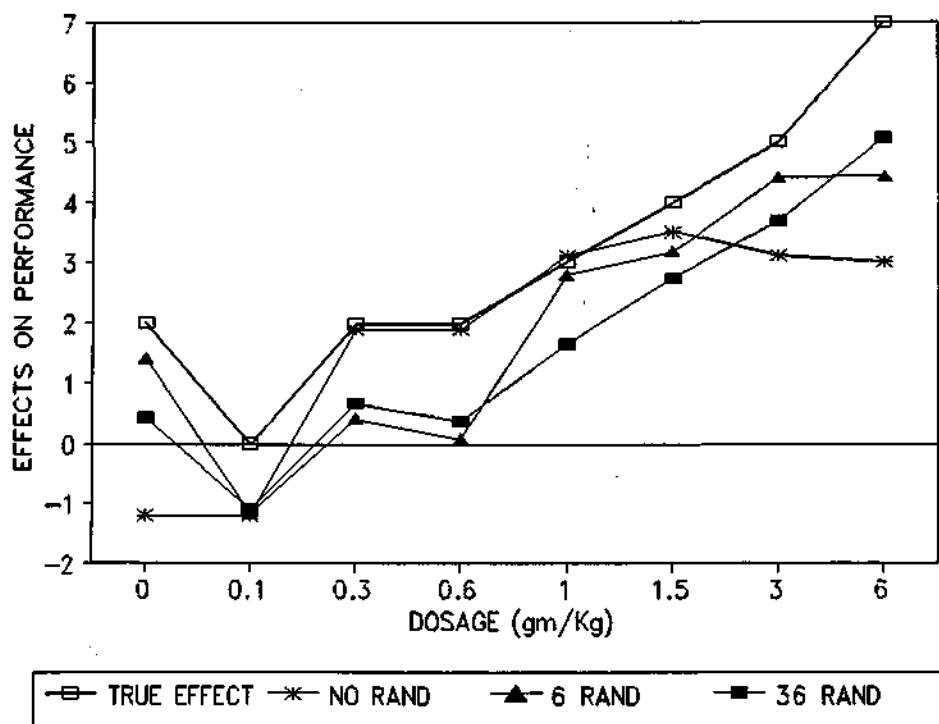


Figure 8.4.1-1. How block randomization reduces distort in measurements that arise from serial-position effects.

A. HOW IT'S DONE

1. Decide how many replications each subject will do. For this example, we'll have four replications. For each subject, list the series of replications in a left column.
2. Assign a number to each of the five levels: avg = 1, bri = 2, con = 3, den = 4, edg = 5.
3. Pick a location on a random number table. Going either up, down, left or right, read the numbers. Write down the first number that is either a 1, 2, 3, 4, or 5.
4. Write the name of the level assigned to that number in the adjacent space of the 1st replication section
5. Continue in the same direction the random number table till you get another 1, 2, 3, 4 or 5, but skip any repetitions of a number you have already used. Write down each usable number and its level.
6. Repeat steps 5 and 6 until you have listed each level of the independent variable.
7. Start the selection of levels for the next replication with the next number for any level. Repeat steps 3 through 6 for the 2nd replication.
8. Repeat steps 3 through 7 until all replications are completed.
9. Repeat steps 3 - 8 for each subject.

192 - WITHIN

The instructions look much harder than the actual doing. The best way to get the idea is to do it yourself. I have completed a set for Ali and started Les', using the portion of the random number table shown below. As in the previous section, I began selecting numbers from the upper left corner of the first group of rows and proceeded right along the first row. When I got to the end of that row, I dropped directly down to the second row and proceed left. (I had to go back to the Appendix table to find a "4" to complete Ali's 4th replication.) To complete the table in way that will enable you to compare your results with others, start doing Les at the left end of the tenth row of the random table in Appendix F.

Random Numbers											
1	0	4	8	0	1	5	0	1	1	0	1
2	2	3	6	8	4	6	5]	7	[3	2	5
2	4	1]	[3	0	4	8	3	6	0	2	2
4	2	1	6	7	9	3	0	9	3	0	6
3	7	5]	7	0	[3	9	9	7	5	8	1
7	7	9	2	1	0	6	9	0	7	1	1
9	9	5	6	2	7	2	9	0	5	5	6

Table 8.4.1.2-1
A Listing of the Sequences in Which the Various Image Enhancements
Were Randomly Selected for Testing with Each Subject.

REPLIC NUMBER	SUBJECT					
	ALI		LES		PAT	
	RAND	LEVEL	RAND	LEVEL	RAND	LEVEL
1	1	avg				
	4	dens				
	5	edg				
	2	bri				
	3	cont				
2	5	edg				
	4	dens				
	3	cont				
	2	bri				
	1	avg				
3	3	cont				
	4	dens				
	2	bri				
	1	avg				
	5	edg				
4	3	cont				
	5	edg				
	1	avg				
	2	bri				
	4	dens				

Now that you (hopefully) understand how to do this, for the rest of this chapter, I'll use a more compact table format to illustrate the sequences in which levels are tested. Table 8.4.1.2-2 has the same information as the one above except that the random numbers have been omitted.

Table 8.4.1.1-2
A More Compact Format to Show the Sequence of Levels
Obtain Using Block-Random Selection.

SUBJECT	REPLIC	SEQUENCE OF LEVELS				
Ali	1	avg	dens	edg	bri	cont
	2	edg	dens	cont	bri	avg
	3	cont	dens	bri	avg	edg
	4	cont	edg	avg	bri	dens
Les	1	bri	avg	edg	cont	dens
	2	edg	avg	bri	cont	dens
	3	bri	avg	dens		

B. PROBLEMS:

1. May need a lot of replications to avoid order-effects such as one level being tested mostly at the beginning or the end of the sequences - This did not happen with Ali, but may for some of the other subjects.
2. Frequent juxtaposition of very different levels leaves the design vulnerable to carry-over effects.

Block-randomized sequences are often used for within-subjects designs, but that is because the levels often do not have any evident differences in strength or complexity - as in the above example. However, they sometimes get used by researchers who do not appreciate the confounding effects that can arise from order and carry-over effects. Unless simplicity of devising testing sequences is critical, why rely on chance to avoid bias?

8.4.2. BALANCING PROCEDURES

The basic idea of these is to eliminate order effects by testing each level an equal number of times at each position in the testing sequence. This spreads-out order effects evenly amongst all levels. Therefore, any order-effect will be the same, or "balanced" for each level and not bias the results.

8.4.2.1 TEST ALL POSSIBLE SEQUENCES

This is almost an ideal method. It tests every level at every position in sequence to balance order effects. It also tests every level preceded and followed by every other level. This tends to balance interactions that carry-over from one level to the next. I say "tends to" because it will not eliminate nonlinear carry-over effects such as the effect of a strong stimulus on a subsequent weak stimulus. A loud sound's effect on perceiving a subsequent soft sound is not balanced when the soft sound is presented first because soft sounds have little effect on loud ones. However, that is not usually the reason why this method is rarely used.

A. HOW TO DO IT

If there were only two levels, such as averaged and brightened images, this procedure is easy. There are only two possible orders of two levels.

With four subjects, the testing sequences would be as follows:

SUBJECT SEQUENCES

REPLICATION	SAM	LES	PAT	ALI
1st	Avg, Bri	Bri, Avg	Avg, Bri	Bri, Avg
2nd	Bri, Avg	Avg, Bri	Bri, Avg	Avg, Bri

Note how an added degree of balancing has been introduced between subjects by reversing the order of the sequences for successive subjects. That way half the subjects are tested first with the averaged images and half are tested first with the brightened images.

With three levels, six replications are needed, and the sequences become a bit trickier:

SUBJECT SEQUENCES

REPLICATION	SAM	LES	PAT	ALI
1st	Avg, Bri, Con	Bri, Avg, Con	Con, Avg, Bri	Avg, Con, Bri
2nd	Avg, Con, Bri	Bri, Con, Avg	Con, Bri, Avg	Avg, Bri, Con
3rd	Bri, Avg, Con	Con, Avg, Bri	Avg, Bri, Con	Con, Bri, Avg
4th	Bri, Con, Avg	Con, Bri, Avg	Avg, Con, Bri	Con, Avg, Bri
5th	Con, Avg, Bri	Avg, Bri, Con	Bri, Avg, Con	Bri, Con, Avg
6th	Con, Bri, Avg	Avg, Con, Bri	Bri Con, Avg	Bri, Avg, Con

In the above example, note how the order of the sequences was changed for various subjects so that each subject started with a different type of enhancement. When possible, a good approach would be to have as many subjects as replications. Then every sequence could be tested first and last.

Now let's see what happens when there are 5 levels to test. To test all possible orders of five levels, requires 125 replications: (To save space, I'll indicate each enhancement technique by only its first letter, and show enough replications to give you the idea of what is needed.)

SUBJECT SEQUENCES

REPLICATION	SAM	LES	PAT	ALI
1st	A, B, C, D, E	B, C, D, E, A	D, E, A, B, C	E, A, B, C, D
2nd	A, B, C, E, D	B, C, D, A, E	D, E, A, C, B	
3rd	A, B, D, C, E	B, C, E, D, A		
4th	A, B, D, E, C	B, C, E, A, D		
5th	A, B, E, C, D			
125th	E, D, C, B, A	A, E, D, C, B		

I think you see the problem: Testing all possible orders of just 5 levels requires 125 replications! For most experiments with more than 3 levels of an independent variable, testing each subject with all possible orders would take far too long. An alternative is to use an "incomplete within-subjects design" - see Section 8.5. In the above five-levels case, an incomplete design could have Sam do replications 1 - 30, Les do replications 31 - 62, Pat do 63 - 94, and Ali do 95 - 125. If that is still too many replications per subject, get more subjects.

B. PROBLEMS

- 1) Testing all possible sequences becomes difficult with more than 3 levels and nearly impossible with more than 4 levels.
- 2) Frequent juxtaposition of very different levels leaves the design vulnerable to nonlinear carry-over effect.

8.4.2.2 LATIN-SQUARE SEQUENCES

The basic idea is to balance out order-effects by testing each level once in each position. This can be accomplished with as many replications as there are levels. There are several ways to construct Latin-Square sequences. One advanced method has each level preceded once

and only once by each of the other levels (Lewis, 1993). That helps to balance certain carry-over effects. For an introduction, we'll use a simpler method which only balances order effects.

A. HOW TO MAKE THEM

The simplest way to make a Latin Square is just to rotate the levels one step for each replication:

REPLICATION	SUBJECT SEQUENCES			
	SAM	LES	PAT	ALI
1st	<u>A, B, C, D, E</u>	B, C, D, E, A	C, D, E, A, B	D, E, A, B, C
2nd	E, A, B, C, D	<u>A, B, C, D, E</u>	B, C, D, E, A	C, D, E, A, B
3rd	D, E, A, B, C	E, A, B, C, D	<u>A, B, C, D, E</u>	B, C, D, E, A
4th	C, D, E, A, B	D, E, A, B, C	E, A, B, C, D	<u>A, B, C, D, E</u>
5th	B, C, D, E, A	C, D, E, A, B	D, E, A, B, C	E, A, B, C, D

Note that in the above example with four subjects, the sequences themselves have been rotated through the replications for each successive subject. This is a fine touch that further balances orders in which the levels are tested. Between the four subjects, each level gets tested first, second, ..., and in last place over the four replications. Another "fine touch" is to have as many subjects as levels. Why is that?

*You forgot to answer the above question. - Editor
I think readers can answer it themselves by looking at the above table - TN*

B. HOW IT WORKS

To show how the balancing works, let's assume that the various enhancement techniques actually had the following effects on the distance in centimeters at which images could be recognized:

ENHANCEMENT TECHNIQUE	TRUE EFFECT
Avg	320
Bri	340
Con	340
Den	350
Edg	370

These "true effects" are something that researchers never actually know. (If they did, there would be no reason to do the experiment.) Researchers can only try to learn true effects by making measurements. Unfortunately all measurements, even those in physics and chemistry, are subject to errors introduced by the measurement process itself. The whole enterprise of doing research in any field can be summarized as having two facets: 1) devising

an method or "instrument" which enables making the measurement. 2) designing a procedure for using that instrument in a manner that minimizes the error introduced by the measuring process. This whole book can be said to deal with the latter facet in psychological research.

Balancing techniques would not be needed if there were no order effects or carry-over effects influencing the measurements. Therefore I'll make up a simple order effect such as a fatigue that reduces the subject's sensitivity on successive sets of measurements. In this case that results in the subject bringing the images 10 centimeters closer to make them even easier to see on successive measurements. (This is a "linear" order effect because it progresses uniformly over the entire sequence as the various enhancement techniques are successively tested with the same subject.) Consequently on successive sets of measurements, the fatigue order effect will reduce the true recognition distance by, 10, 20, etc. centimeters regardless of which enhancement is being tested as shown below:

POSITION IN ORDER OF TESTING	FATIGUE EFFECT
1 ST	0
2 ND	-10
3 RD	-20
4 TH	-30
5 TH	-40

If the five enhancement techniques were tested in alphabetical order with one subject, only the first set of measurements using the Averaging techniques would represent the "true" effect of the enhancement. Subsequent measurements using the other techniques would have their true effects corrupted by the increasing amounts of fatigue as shown in Figure 8.4.2.2-1.

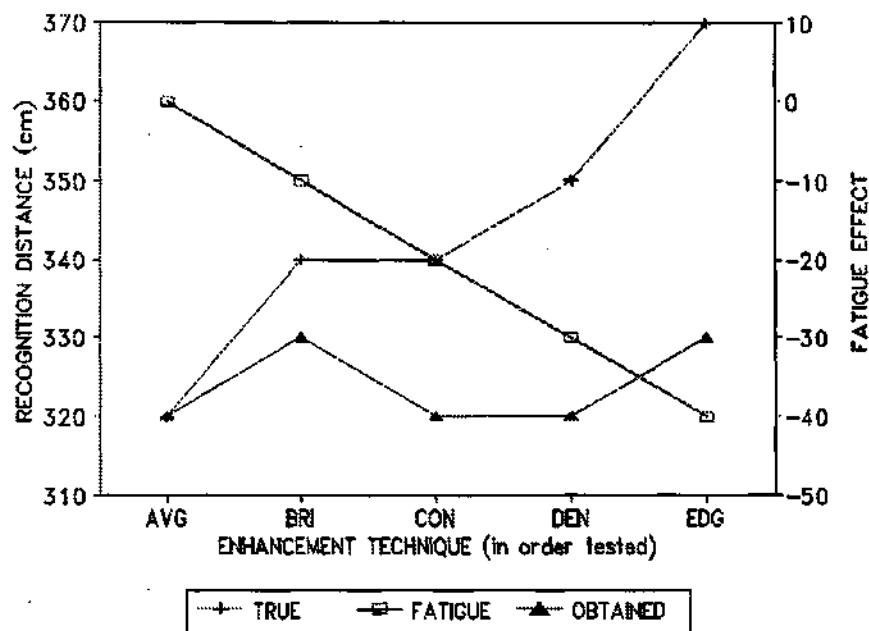


Figure 8.4.2.2-1. The true effect of the five enhancement methods is confounded by a linear fatigue effect which reduces recognition distance by 10 centimeters on successive sets of measurements. The consequences of this fatigue on the recognition distance measurements are shown when the enhancement levels are tested in alphabetical order.

It should be evident from Figure 8.4.2.2-1 that the fatigue order effect would substantially corrupt measurements of the effectiveness of the enhancement techniques if they were simply tested in alphabetical order. Based on the obtained measurements, a researcher would have to conclude that the various enhancements differed little in their effect on recognition distance, and that the Bri (brightness increase) and Edg (edge finding) techniques were equally good. (If this is not evident to you, re-read the above section. If that doesn't help ask a colleague, or ask the instructor because this is important stuff. The rest of this chapter and the next won't make much sense if you do not understand the need for balancing and counterbalancing in psychological research.)

Now let's see how Latin-Square balancing copes with an order effect that fatigue has on the measurements. We'll assume that Sam is a perfect subject whose responses would represent the "true" effects of the enhancements on recognition distance - were it not for the fatigue effect. To avoid having fatigue distort the measurements, Sam is tested using one Latin-square sequence a day over five days to complete the five replications required for this design. It is assumed that by the next day Sam has recovered from the fatigue effects of the previous day's successive measurements. (The order in which the enhancements are tested for Latin Square balancing was described above in Section A.) Table 8.4.2.2-1 shows what would happen.

Table 8.4.2.2-1

How Fatigue Affected Sam's Results Obtained with a Latin Square Design During Daily Replications Measuring Recognition Distance of Enhanced Images

REPLICATION	TYPE OF ENHANCE	TRUE EFFECT	FATIGUE EFFECT	MEASURED RESULT
1	Avg	320	0	320
	Bri	340	-10	330
	Con	340	-20	320
	Den	350	-30	320
	Edg	370	-40	330
2	Bri	340	0	340
	Con	340	-10	330
	Den	350	-20	330
	Edg	370	-30	340
	Avg	320	-40	280
3	Con	340	0	340
	Den	350	-10	340
	Edg	370	-20	350
	Avg	320	-30	290
	Bri	340	-40	300
4	Den	350	0	340
	Edg	370	-10	360
	Avg	320	-20	300
	Bri	340	-30	310
	Con	340	-40	300
5	Edg	370	0	370
	Avg	320	-10	310
	Bri	340	-20	320
	Con	340	-30	310
	Den	350	-40	310

Notes: Avg = averaged, Bri = brightened, Con = contrast enhanced, Den = density equalization, Edg = edge finding. All values are in centimeters.

The above results look like a lot of mumbo jumbo until they are summarized by averaging the measurements for each enhancement technique across replications as shown in Table 8.4.2.2-2

Table 8.4.2.2-2
A Summary of the Measurements of Sam's Recognition Distances
of Images that Were Enhanced Using Five Methods

TYPE	REPLICATION					MEAN
	1ST	2ND	3RD	4TH	5TH	
Avg	320	280	290	300	310	300
Bri	330	340	300	310	320	320
Con	320	330	340	300	320	320
Den	320	330	340	310	330	330
Edg	330	340	350	370	350	350

Figure 8.4.2.2-2 compares the mean results obtained using a Latin-square balanced design with the "true" results that are listed in the first table. Yes, the means are different, from the true effects, but look at the overall trend. The relative effectiveness of the enhancements is accurately represented by the means. The reductions in recognition distances that were introduced by fatigue have been spread out evenly across the five types of enhancement, and consequently no bias has been introduced by the fatigue. Nevertheless, the overall effects of fatigue have reduced all means by 20 centimeters. This is the average value of the fatigue order effect. It illustrates a general truth: **"Most psychological measurements never represent the absolute effects of an independent variable.** Order and carry-over effects (or subject differences or confounding effects of other variables for that matter) can never be avoided.³ The best one can hope for is a lack of bias on the levels."

³ If you could figure out a way to estimate the order effect in a Latin-square design, you could subtract its mean value from each result and get a closer approximation of the true effect of the independent variable. The trick, of course, is how to measure the order effect by itself.

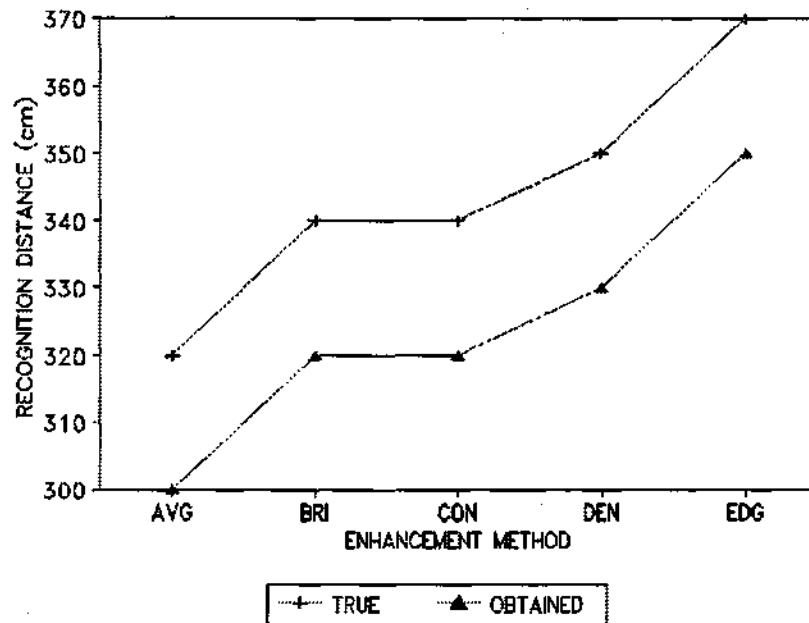


Figure 8.4.2.2-2. The results obtained with a Latin-square design, which has balanced the confounding effects of a fatigue order effect, are compared with the true effects of five image enhancement methods on recognition distance.

C. ADVANTAGES

- 1) Usually requires many fewer replications than testing all possible sequences.
- 2) May be preferable to ABBA Counterbalancing (see below) when order effects are nonlinear and strong.

D. PROBLEMS

- 1) Experiments with a large number of levels of the independent variable may find that Latin-Square designs require too many replications.
- 2) The inherent juxtaposition of very different levels leaves the design vulnerable to carry-over effects.

8.4.3 COUNTERBALANCING PROCEDURES

Counterbalancing is a shortcut alternative to balancing the sequences in which levels of an independent variable are tested. As long as the order effects are gradual, these procedures are as good as balanced ones for preventing bias.

8.4.3.1 ABBA COUNTERBALANCING

No, it's not named after my favorite popular music group. It is A, B; then B, A. This simple, forwards-and-backwards procedure is adequate for most within subjects experiments. By a large margin, it is the most frequently used design.

A. HOW TO DO IT

Begin by arranging the levels in some natural order such as increasing (or decreasing) strength, complexity, aesthetic appeal, etc. Two replications are required. For the first replication, present the levels successively in that natural order. For the second replication, reverse that order. Here is how it would be done for the four imaging methods with the four subjects we have been using to illustrate the random and balancing methods: It so happens that their Avg, Bri, Con, .. alphabetical order is also an ascending order of the amount of computation required to produce each enhancement. (Sometimes the "natural order" of a set of levels may not be obvious, and the experimenter simply has to decide as best as she/he is able or else do some preliminary testing.)

LEVELS IN NATURAL ORDER OF COMPLEXITY: Averaging (A) , Brightened (B), Contrast Enhanced (C), Density Equalization (D), Edge Finding (E)

- 1) For the first replication, test each level once in their natural order.
- 2) For the second replication, reverse the sequence.

SUBJECT SEQUENCES

REPLICATION	SAM	LES	PAT	ALI
1st	A, B, C, D, E	E, D, C, B, A	A, B, C, D, E	E, D, C, B, A
2nd	E, D, C, B, A	A, B, C, D, E	E, D, C, B, A	A, B, C, D, E

That's all there is to it. Note that above as in the previous balancing methods, an extra degree of balancing has been introduced by switching the sequences used for the 1st and 2nd replications for successive subjects.

B. HOW IT WORKS

To see how ABBA works, we'll assume the same true effects of the various enhancement types and that fatigue affects successive measurements the same as in the Latin-square example in the previous section.

Table 8.4.3.1-1

How Fatigue Affected Sam's Results Obtained with an ABBA Counterbalanced Design
During Daily Replications Measuring Recognition Distance of Enhanced Images

REPLICATION	TYPE OF ENHANCE	TRUE EFFECT	FATIGUE EFFECT	MEASURED RESULT
1	Avg	320	0	320
	Bri	340	-10	330
	Con	340	-20	320
	Den	350	-30	320
	Edg	370	-40	330
2	Edg	370	0	370
	Den	350	-10	340
	Con	340	-20	320
	Bri	340	-30	310
	Avg	320	-40	280

Notes: Avg = averaged, Bri = brightened, Con = contrast enhanced, Den = density equalization, Edg = edge finding. All values are in centimeters.

The measurements obtained from Sam's first replication show how much fatigue can distort the results. Had no balancing been used, the various enhancements would have appeared to be quite similar in effectiveness, which is not true. (This was also illustrated in Figure 8.4.2.2-1 in the previous Latin-square section.) When the two counterbalanced replications are summarized by averaging the measurements for each enhancement technique across replications, the results appear as shown in Table 8.4.3.1-2

Table 8.4.3.1-2

A Summary of the Measurements of Sam's Recognition Distances
of Images that Were Enhanced Using Five Methods

TYPE	REPLICATION		
	1ST	2ND	MEAN
Avg	320	280	300
Bri	330	310	320
Con	320	320	320
Den	320	340	330
Edg	330	370	350

The mean results that would be obtained for the other subjects would again be the same as Sam's when averaged across both replications. Compare these means with the "true" results as shown in Figure 8.4.3.1.

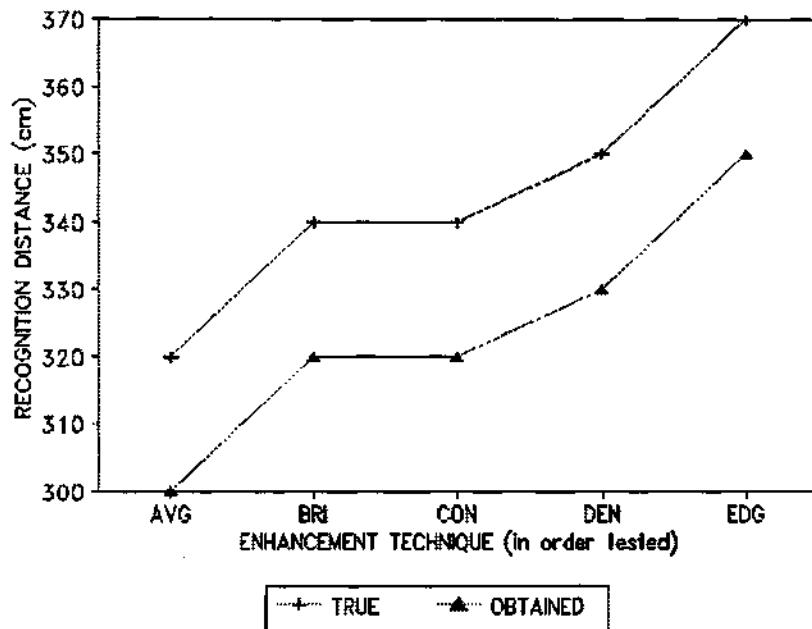


Figure 8.4.3.1 The results obtained with ABBA counterbalancing are compared with the true effects.

ABBA counterbalancing has removed the distortion introduced by a fatigue order effect. As was the case with means obtained using the Latin-Square design, the means are all smaller than the true effects by the 20 cm average of the fatigue effect, but the relative effectiveness of the enhancements is accurately represented. The reductions in recognition distances that are introduced by fatigue have been spread out evenly across the five types of enhancement, and consequently no bias has been introduced by the fatigue. Note that ABBA has accomplished this with only two replications compared to the five replications required by the Latin-square. However, we shall soon see there is a limit to ABBA's superiority.

C. ADVANTAGES

- 1) Requires only 2 replications no matter how many levels.
- 2) Minimizes carry-over effects. That's because the difference between successive levels is always kept as small as possible.

D. DISADVANTAGE

Can not cope with non-linear order effects. Unfortunately, order effects such as practice, fatigue, and end effects are usually non-linear. Subjects tend to learn the skills for doing an experiment quickly and after that do not improve further. Conversely, if you have made the subjects comfortable and the task is not overly demanding, subjects tend not start becoming fatigued until after 30 minutes or more of testing. These effects are illustrated in Figure 8.4.3.1-1.

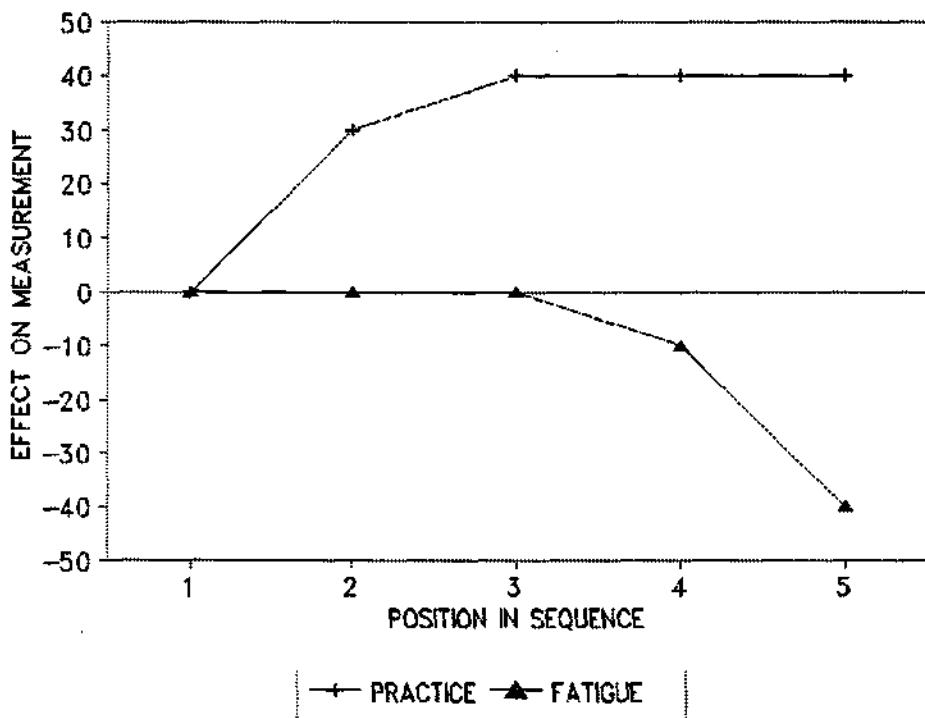


Figure 8.4.3.1-1. Examples of two non-linear order effects that are often encountered in psychological research.

Table 8.4.3.1-1
How Non-Linear Fatigue Affected Sam's Results Obtained with an ABBA Counterbalanced Design During Daily Replications Measuring Recognition Distance of Enhanced Images

REPLICATION	TYPE OF ENHANCE	TRUE EFFECT	FATIGUE EFFECT	MEASURED RESULT
1	Avg	320	0	320
	Bri	340	0	330
	Con	340	0	320
	Den	350	-10	320
	Edg	370	-40	330
2	Edg	370	0	370
	Den	350	0	340
	Con	340	0	320
	Bri	340	-10	310
	Avg	320	-40	280

Notes: Avg = averaged, Bri = brightened, Con = contrast enhanced, Den = density equalization, Edg = edge finding. All values are in centimeters.

Again the measurements obtained from Sam's first replication show how much fatigue can distort the results. Had no balancing been used, the various enhancements would have appeared to be quite similar in effectiveness, which is not true. Table 8.4.3.1-2 summarizes the

two replications would be by averaging the measurements for each enhancement technique across replications. The results are shown in Figure 8.4.3.1-2 and compared with the true effects.

Table 8.4.3.1-2
A Summary of the Measurements of Sam's Recognition Distances
of Images that Were Enhanced Using Five Methods

TYPE	REPLICATION		
	1ST	2ND	MEAN
Avg	320	280	300
Bri	330	310	320
Con	320	320	320
Den	320	340	330
Edg	330	370	350

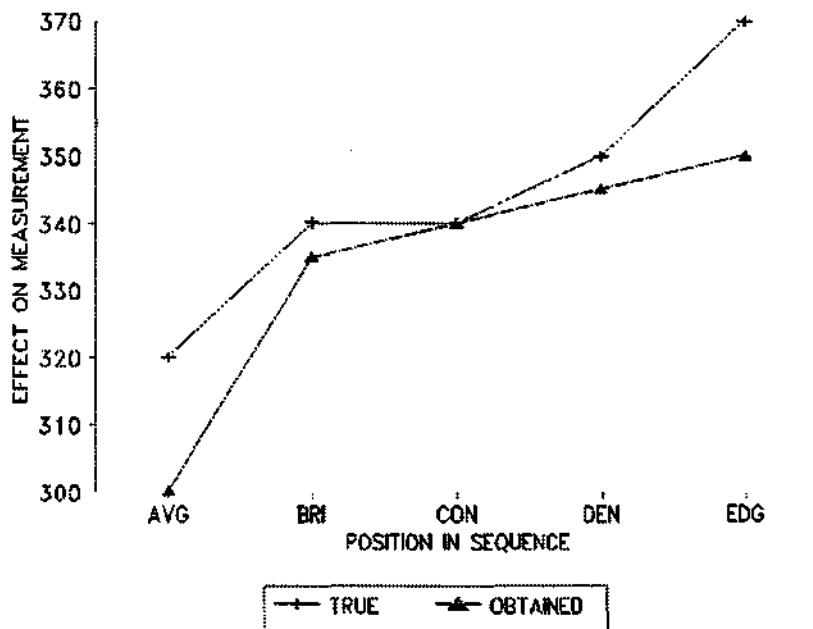


Figure 8.4.3.1-2. The results of an ABBA counterbalanced experiment when a non linear fatigue effect has occurred compared to the true effects of five enhancement techniques on recognition distance.

Compared to no counterbalancing at all (i.e. the results of Sam's first replication in Table 8.4.3.1), ABBA counterbalancing is clearly better than no counterbalancing. At least ABBA indicates that edge finding is the best technique. However, Figure 8.4.3.1-2 shows how markedly the non-linear fatigue effect has distorted the true effects of the various enhancement techniques. While Edg is best, it appears to be only slightly better than the Den or Con techniques. The Avg technique looks worse than it really is compared to the others.

What can you do when there are too many levels for Latin-Square balancing to be practical and yet you are concerned that a non-linear order effect could distort the results? Read on.

8.4.3.2 QUADRATIC COUNTERBALANCING

It has been said that "necessity is the mother of invention." That is certainly how quadratic counterbalancing was invented. In 1975 I set to replicate and more precisely define the temporal tuning characteristics of characteristics of visual channels. This required testing many more frequencies of flickering light than the original experiment (Nilsson, Richmond & Nelson, 1975). The method used to reveal the tuning was particularly time consuming, and more levels of the independent variable made it a very boring task for the subjects despite the use of automation. It took six months to complete the experiment. The results showed little evidence of tuning. We had taken some shortcuts in the procedure and concluded that this was the problem. So we repeated the experiment using the original procedure. This required longer testing sessions and took nine months. Again the results were inconclusive. Should we publish a retraction?

The data did show signs of tuning, but the tuning got scrambled as ascending and descending replications were averaged. A non-linear fatigue effect was distorting the results. With 22 frequencies to test, a Latin-Square design would take years. How non-linear trends could be coaxed out of data suggested a way of counterbalancing the sequence in which the levels were tested to overcome non-linear order effects. It required four replications instead of ABBA's two. We resolved to try again using this new design. It took a year to do the experiment, but this time the evidence of tuning was beyond doubt (Nilsson, Richmond & Nelson, 1978). I decided to call the design "quadratic" counterbalancing since it was based on quadratic trend analysis. Modelling has shown that it is 16X better at removing any nonlinear order effect than ABBA counterbalancing Nilsson & Donald, 1987).

A. HOW TO DO IT

Consider it a higher order of ABBA counterbalancing. Four replications are required. As with ABBA, you begin the same way by arranging the levels in some "natural" order. Using the enhancement testing as an example, the techniques were arranged in order of complexity:

LEVELS IN NATURAL ORDER: Avg, Bri, Con, Den, and Edg

The sequence in which to test the levels for the first replication is obtained by selecting the first level in that natural order, and selecting every other level till you get to the end of the order. Continue the sequence by reversing direction and selecting those levels that were skipped the first time through:

1ST REPLICATION: Avg, Con, Edg, Den, Bri

For the second replication, reverse the sequence of the first replication:

2ND REPLICATION: Bri, Den, Edg, Con, Avg

For the third replication, start at the far end of the natural order and select every other level. When you get to the beginning, reverse direction and pick up the levels that were skipped the first time through:

3RD REPLICATION: Edg, Con, Avg, Bri, Den

The fourth replication is the reverse of the third:

4TH REPLICATION: Den, Bri, Avg, Con, Edg

Here are the quadratic counterbalanced sequences for four subjects. Note that an added degree of balancing has been added by shifting the order of the replications for each successive subject:

SUBJECT SEQUENCES				
REPLICATION	SAM	LES	PAT	ALI
1st	<u>A, C, E, D, B</u>	D, B, A, C, E	E, C, A, B, D	B, D, E, C, A
2nd	B, D, E, C, A	<u>A, C, E, D, B</u>	D, B, A, C, E	E, C, A, B, D
3rd	E, C, A, B, D	B, D, E, C, A	<u>A, C, E, D, B</u>	D, B, A, C, E
4th	D, B, A, C, E	E, C, A, B, D	B, D, E, C, A	<u>A, C, E, D, B</u>

B. HOW IT WORKS

To show how quadratic counterbalancing works, I'll use the same basic data on how various image enhancement techniques affect recognition distance and the same nonlinear fatigue effect for one subject, Sam. as done previously.

Table 8.4.3.2-1
How Fatigue Affected Sam's Results Obtained with Quadratic Counterbalancing of Levels
During Daily Replications Measuring Recognition Distance of Enhanced Images

REPLICATION	TYPE OF ENHANCE	TRUE EFFECT	FATIGUE EFFECT	MEASURED RESULT
1	Avg	320	0	320
	Con	340	0	340
	Edg	370	0	370
	Den	350	-10	340
	Bri	340	-40	300
2	Bri	340	0	340
	Den	350	-0	350
	Edg	370	0	370
	Con	340	-10	330
	Avg	320	-40	280
3	Edg	370	0	370
	Con	340	0	340
	Avg	320	0	320
	Bri	340	-10	330
	Den	350	-40	310
4	Den	350	0	350
	Bri	340	-10	330
	Avg	320	-20	300
	Con	340	-30	310
	Edg	370	-40	330

Notes: Avg = averaged, Bri = brightened, Con = contrast enhanced, Den = density equalization, Edg = edge finding. All values are in centimeters.

Averages of the subjects' data are shown in Table 8.4.3.2-2. Figure 8.4.3.2 compares these averages with the true effect of the five enhancement techniques.

Table 8.4.3.2-2
A Summary of the Measurements of Sam's Recognition Distances
of Images that Were Enhanced Using Five Methods

REPLICATION					
TYPE	1ST	2ND	3RD	4TH	MEAN
Avg	320	280	300	300	310
Bri	300	340	310	330	327.5
Con	330	310	330	310	335
Den	320	340	310	350	337.5
Edg	330	340	350	370	360

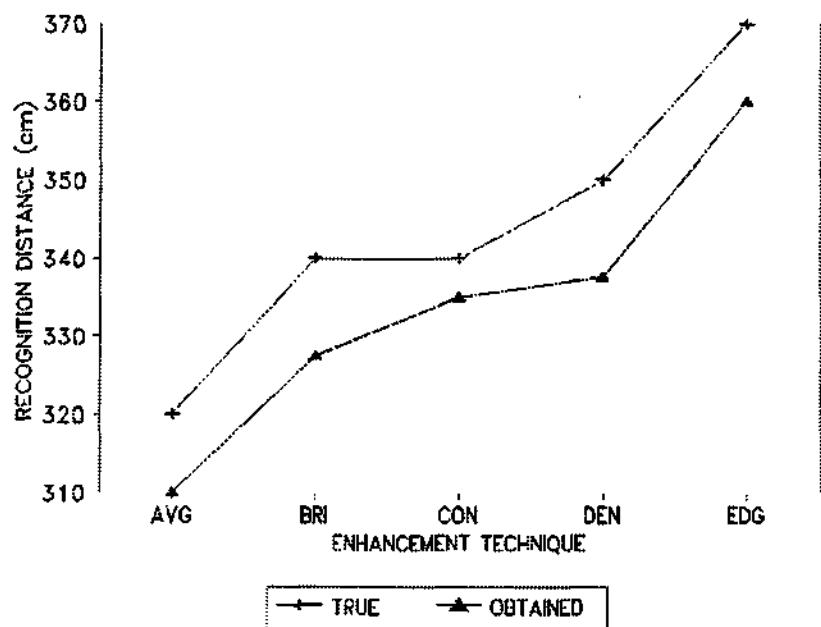


Figure 8.4.3.2. The average results of a quadratic counterbalanced experiment compared to the "true" effects of the independent variable.

It's not perfect, but it's a pretty good representation of the true effects.(Remember, an experimenter never knows the true effects, only the results of his/her measurements.) It is considerably more accurate than the ABBA results. Yet, it did it without doing as many replications as there are levels. I admit that with only the five levels in this example, the saving is not great. The\is advantage of quadratic counterbalancing over Latin-square balancing becomes important only as the number of levels of the independent variable exceeds six.

C. ADVANTAGES

- 1) Reduces the effects of nonlinear order effects to a least 1/16th the magnitude that those effects would have in an ABBA counterbalanced design.
- 2) Requires only 4 replications, no matter how many levels.
- 3) Keeps carry-over effects smaller than all methods except ABBA. Though the steps between successive levels are usually twice as large as the steps in ABBA, it still avoids abrupt changes in level.

D. DISADVANTAGES

- 1) Requires twice as many replications as ABBA counterbalancing
- 2) More prone to carry-over effects than ABBA.
- 3) There is some residual distortion of the results compared to results obtained with Latin-square or completely balanced designs.

8.5 "INCOMPLETE" WITHIN-SUBJECTS DESIGNS**HOW TO DO THEM**

These designs test each subject on all levels of an independent variable, but does not have the subject replicate those observations in other sequences to balance out order and carry-over effects. Rather, different subjects are used for each of the replications. Incomplete designs are used when it would take too long for any one subject to do all the replications. The basic "incomplete" design is done as follows:

1. Each Subject is Tested Only Once at Each Level
2. Different Subjects are Tested for Each Required Replication
3. The Number of Subjects Needed = $n * \text{Number of Required Replications}$
4. Can be Used with Randomized-Block, Latin Square, ABBA, or Quadratic methods

To make up for the loss of statistical power, consider using more subjects when using an incomplete design. An *extended*, incomplete within-subjects version of the quadratic counterbalanced example of testing image enhancement techniques would look like this:

SUBJECT	SEQUENCE TESTED
SAM	averaged, contrast, edge, density, brightened
LES	brightened, density, edge, contrast, averaged
PAT	edge, contrast, averaged, brightened, density
ALI	density, brightened, averaged, contrast, edge
TER	averaged, contrast, edge, density, brightened
KIM	brightened, density, edge, contrast, averaged
VAN	edge, contrast, averaged, brightened, density
BIL	density, brightened, averaged, contrast, edge

ADVANTAGES

- 1) Requires Less Time per Subject
- 2) Useful when testing each level more than once is not suitable due to enduring effects on subject characteristics such as memory or attitude.

DISADVANTAGES

- 1) If subjects differ in their serial-position or carry-over effects, the counterbalancing will only partially succeed in reducing these effects. The consequence can be a confounding effect that could have been avoided with a complete design.
- 2) More Variability -> Less Statistical Power
3. Requires More Subjects

AN ADVANCED EXAMPLE

To show you what is possible, I'll describe how a incomplete within-subjects experiment with three independent variables was counterbalanced. The purpose of this experiment was to measure how the color of a figure and the color of the background in a sign such as traffic sign affected the distance at which the figure could be recognized. Six primary colors, black, white, red, yellow, green, and blue, were used for both the figure and the background - except we did not test the same color in both the figure and the background. Ten different figures, children, slippery road, rocks falling, etc, were used in the signs. This made a total of 300 combinations of levels to be tested. Not for a moment did we consider using All-Possible-Orders or Latin-Square. We used an incomplete, ABBA counterbalanced design.

Fifteen sequences of stimulus levels were created that had the color of the figure, the color of the background, and the figure all change for each successive measurement. Half of the subjects were presented the 300 figure X figure-color X background-color combinations in 15 sequences of 20 symbols each that started with the Rower and ended with the Fragile symbol. Table 8.5 shows how we produced the first sequence, which started with a black rower on white background, K/W - Rower 1₁, and ended with a white on blue Fragile symbol, W/B-Fragile 1₂₀. Sequence 2 started with a black rower on a red background, K/R - Rower 2₁. Successive sequences also started in the first column, each beneath the previous for a total of 15 sequences with 20 levels in each. The other subjects saw these sequences in reverse order starting with the black/green Fragile symbol, K/G - Fragile 15₂₀.

While this research for the Canadian Space Agency may look complicated, creating the sequences was hardly "rocket science". It just involved simple logic and a lot of step by step tabulation. It took more than a day to work out the sequences, enter them into a spreadsheet, and check that all combinations had been used and none repeated. Don't worry. If you are asked about an incomplete within-subjects design on a test for this course, it will not be quite this complicated.

Table 8.5
 Sequence 1 of Method for Counterbalanced Testing of
 10 Symbols Printed in the 6 Primary Colors on 5 Primary Backgrounds
 (Table 4.2-2 from Clements-Smith, Nilsson, Connolly & Ireland; 1993)

Letter/Bkgrd	Rower	Children	Slippery	Rocks	Picnic	Lighthouse	Dentist	Launch	Workers	Fragile
K/W	1 ₁	10 ₁₂	4 ₁₃	13 ₁₄	12 ₅	6 ₆	5 ₇	9 ₁₈	3 ₁₉	12 ₂₀
K/R	2 ₁	6 ₁₂	5 ₁₃	14 ₁₄	13 ₅	7 ₆	1 ₇	10 ₁₈	4 ₁₉	13 ₂₀
K/Y	3 ₁	7 ₁₂	1 ₁₃	15 ₁₄	14 ₅	8 ₆	2 ₇	6 ₁₈	5 ₁₉	14 ₂₀
K/G	4 ₁	8 ₁₂	2 ₁₃	11 ₁₄	15 ₅	9 ₆	3 ₇	7 ₁₈	1 ₁₉	15 ₂₀
K/B	5 ₁	9 ₁₂	3 ₁₃	12 ₁₄	11 ₅	10 ₆	4 ₇	8 ₁₈	2 ₁₉	11 ₂₀
W/K	6 ₁	5 ₂	9 ₁₃	3 ₁₄	12 ₁₅	11 ₆	10 ₇	4 ₈	8 ₁₉	2 ₂₀
W/R	7 ₁	1 ₂	10 ₁₃	4 ₁₄	13 ₁₅	12 ₆	6 ₇	5 ₈	9 ₁₉	3 ₂₀
W/Y	8 ₁	2 ₂	6 ₁₃	5 ₁₄	14 ₁₅	13 ₆	7 ₇	1 ₈	10 ₁₉	4 ₂₀
W/G	9 ₁	3 ₂	7 ₁₃	1 ₁₄	15 ₁₅	14 ₆	8 ₇	2 ₈	6 ₁₉	5 ₂₀
W/B	10 ₁	4 ₂	8 ₁₃	2 ₁₄	11 ₁₅	15 ₆	9 ₇	3 ₈	7 ₁₉	1 ₂₀
R/K	11 ₁	10 ₂	4 ₃	8 ₁₄	2 ₁₅	11 ₁₆	15 ₇	9 ₈	3 ₉	7 ₂₀
R/W	12 ₁	6 ₂	5 ₃	9 ₁₄	3 ₁₅	12 ₁₆	11 ₇	10 ₈	4 ₉	8 ₂₀
R/Y	13 ₁	7 ₂	1 ₃	10 ₁₄	4 ₁₅	13 ₁₆	12 ₇	6 ₈	5 ₉	9 ₂₀
R/G	14 ₁	8 ₂	2 ₃	6 ₁₄	5 ₁₅	14 ₁₆	13 ₇	7 ₈	1 ₉	10 ₂₀
R/B	15 ₁	9 ₂	3 ₃	7 ₁₄	1 ₁₅	15 ₁₆	14 ₇	8 ₈	2 ₉	6 ₂₀
Y/K	11 ₁₁	15 ₂	9 ₃	3 ₄	7 ₁₅	1 ₁₆	15 ₁₇	14 ₈	8 ₉	2 ₁₀
Y/W	12 ₁₁	11 ₂	10 ₃	4 ₄	8 ₁₅	2 ₁₆	11 ₁₇	15 ₈	9 ₉	3 ₁₀
Y/R	13 ₁₁	12 ₂	6 ₃	5 ₄	9 ₁₅	3 ₁₆	12 ₁₇	11 ₈	10 ₉	4 ₁₀
Y/G	14 ₁₁	13 ₂	7 ₃	1 ₄	10 ₁₅	4 ₁₆	13 ₁₇	12 ₈	6 ₉	5 ₁₀
Y/B	15 ₁₁	14 ₂	8 ₃	2 ₄	6 ₁₅	5 ₁₆	14 ₁₇	13 ₈	7 ₉	1 ₁₀
G/K	1 ₁₁	15 ₁₂	14 ₃	8 ₄	2 ₅	6 ₁₆	5 ₁₇	14 ₁₈	13 ₉	7 ₁₀
G/W	2 ₁₁	11 ₁₂	15 ₃	9 ₄	3 ₅	7 ₁₆	1 ₁₇	15 ₁₈	14 ₉	8 ₁₀
G/R	3 ₁₁	12 ₁₂	11 ₃	10 ₄	4 ₅	8 ₁₆	2 ₁₇	11 ₁₈	15 ₉	9 ₁₀
G/Y	4 ₁₁	13 ₁₂	12 ₃	6 ₄	5 ₆	9 ₁₆	3 ₁₇	12 ₁₈	11 ₉	10 ₁₀
G/B	5 ₁₁	14 ₁₂	13 ₃	7 ₄	1 ₅	10 ₁₆	4 ₁₇	13 ₁₈	12 ₉	6 ₁₀
B/K	6 ₁₁	5 ₁₂	14 ₁₃	13 ₄	7 ₅	1 ₆	10 ₁₇	4 ₁₈	13 ₁₉	12 ₂₀
B/W	7 ₁₁	1 ₁₂	15 ₁₃	14 ₄	8 ₆	2 ₇	6 ₁₇	5 ₁₈	14 ₁₉	13 ₂₀
B/R	8 ₁₁	2 ₁₂	11 ₁₃	15 ₄	9 ₅	3 ₆	7 ₁₇	1 ₁₈	15 ₁₉	14 ₂₀
B/Y	9 ₁₁	3 ₁₂	12 ₁₃	11 ₄	10 ₅	4 ₆	8 ₁₇	2 ₁₈	11 ₁₉	15 ₂₀
B/G	10 ₁₁	4 ₁₂	13 ₁₃	12 ₄	6 ₅	5 ₆	9 ₁₇	3 ₁₈	12 ₁₉	11 ₂₀

Note: K = black, W = white, R = red, Y = yellow, G = green, B = blue.

EXERCISES

1. Look through some research journals concentrating on the "Procedure" sections of their reports on experiments. For each of the following within-subjects methods of determining the sequences in which the levels are tested, find a report which uses that design; provide a proper APA reference; identify the independent variable(s), the levels, the dependent variable(s); and describe how the subjects were obtained.
 - a) random or randomized blocks
 - b) Latin-Squares
 - c) ABBA counterbalanced
2. Look through some journals, concentrating on the "Procedure" sections. For each of the following types of within-subjects designs, find an experiment of that type, provide a proper APA reference; identify the independent variable(s), the levels, the dependent variable(s); and describe how the subjects were obtained.
 - a) complete within-subjects design
 - b) an incomplete design
 - c) a factorial or mixed design
3. Look through some journals, concentrating on the "Procedure" sections. See if you can find a report that used a randomized block design, when a balanced or counterbalanced procedure might have been better to avoid order and carry-over effects. Explain what order or carry-over effect might have threatened the results of this experiment.
- 4-a. Assume that you want to do an experiment which has 4 levels (A, B, C, D) of an independent variable. For each of the design method presented in this chapter, produce the sequences in which you would test 16 subjects.

Assume that it takes 16 minutes the first time a subject is tested to explain the experiment, handle questions, obtain consent and practice the subject's task. It also takes another 4 minutes at the end of a session to debrief and thank the subject. To prevent fatigue or boredom you limit the sessions to about an hour.
- 4-b If each trial takes 10 minutes and each subject will come only once, which designs can you use?
- 4-c. If each trial takes 10 minutes and each subject can come up 6 times. which designs can you use?
- 4-d. If each trial takes a minute and each subject will only come once, which designs can you use?

214 - WITHIN

4-e. If each trial takes a minute and each subject can come up to 6 times, which designs can you use?

5. You wonder if infants are more interested by many items than a few. The independent variable will be the number of items presented. You are going to study this using the preferential viewing and measuring the latency (dependent variable) of the infant's orienting response to look at the items. Since infants of similar age vary considerably in their performance, a within-subjects design is desirable. You present various numbers of items such as dolls, toys, plush animals, flowers, doll furniture on one side of the infant's field of view and just a single example of a similar item on the other. (To avoid preferences, you randomly alternate which side has the single item, and the kind of the items presented on successive trials. For this example we'll ignore these details and just consider how the number of items affected the latency of the preferential viewing response.)

5-a. Assume that you have 8 infants available. Prepare the sequences in which you would present different numbers of items using randomized blocks, Latin-square, ABBA, and quadratic designs.

5-b. Assume that the true effects of the number of items on viewing latency is presented in the table below. Also assume that after 3 trials in any session the infants start getting bored and their latency increases by 2 seconds on every subsequent trial so that by the 8th trial their latency is 10 seconds longer than the true latency regardless of the number of items. Prepare some tables to show how this boredom effect biases the average results obtained with each of the designs you created in 4-a.

NUMBER OF ITEMS	TRUE LATENCY (sec)
2	15
3	12
4	10
5	8
6	7
7	6
8	5
9	5

5-c. For each method make a graph comparing its mean results with the "true effect". Considering the difficulties of doing research with infants (see Chapter 5) which design would you use?

CHAPTER 9

FIELD RESEARCH

Many psychologists do research in laboratories, use exotic equipment, and some even wear white lab coats like movie scientists. However, not all behavioral scientists work in laboratories as illustrated by archaeologists, ethologists, and sociologists. Psychologists go into the community to do research shown by examples using observational, trace, and survey methods of measurement. Field research is no less "scientific" than that done in labs and in many ways is more challenging. It requires careful attention to procedure to obtain satisfactory data in situations that can not be as carefully controlled as laboratories.

Before venturing into the fields let's look back briefly, take a broader view of the two "forests" we have just come through, and set our bearings. The last two chapters were laden with a lot of 'how to do it' methods that may be blurring into a multitude of trees all alike. To do between-subjects experiments, we focused on the various sequences in which the subjects could be tested to minimize bias from individual differences. To do within-subjects experiments, we focused on various sequences for presenting levels of the independent variable to minimize order and carry-over effects. So what's next?

Whether they involve between- or with-subjects designs, when you do research in the field, you meet an additional set of factors that can bias your results. In the laboratory the place was always the same and the surrounding conditions constant. Venturing into the 'real world' you encounter all sorts of changes that could bias the results. In the field you usually have little control over these changes except to select where and when to make your measurements. That will be the focus of this chapter.

9.1 ADVANTAGES OF FIELD RESEARCH

CHECK VALIDITY OF LAB RESULTS Laboratories make it possible to hold many factors constant to ensure that they do not confound the effects of the independent variable and to minimize variability. However, we do not want to learn only how people behave in these special environments. Basic functions of the central nervous system such as sensation, memory or reaction time are probably only less variable in the controlled confines of a laboratory. On the other hand, attitudes, interpretations of social situations, and expectancies may differ considerably from those outside a laboratory. The only way to be certain about the generality of laboratory results is to see if similar results are obtained in the field.

WHEN FIELD CONDITIONS CAN'T BE REPLICATED IN LAB Sometimes psychologists go to great lengths to reproduce field conditions such as a class room, work place or vehicle cab in the laboratory. However, some situations such as Arctic weather, traffic, and social incidents can prove extremely difficult, expensive or impossible to reproduce - especially when subjects know they are in a laboratory. The alternative to high technology simulators is, of course, to do conduct the research on site. Field research can be a better solution when the costs of transportation, access time, and portable testing apparatus are substantially less than the simulation costs.

ACCESS TO SUBJECTS WHO WON'T COME TO YOUR LAB One reason so much psychological research is done using undergraduate students is that they are readily

available in large numbers to come to psychology laboratories, which usually are located on university campuses. In terms of age, education and socio-economic status, they are hardly representative of the general population when it comes to matters of cognitive ability, emotion and attitudes. Yet it is not easy to recruit persons off campus to come and be subjects in your laboratory. An ad in the local paper, for example, will probably attract only a few persons from the community. These persons are likely to be unemployed, and somewhat adventurous - hardly representative of the general population either. So, "If the mountain won't come to Mohammed - , you have to go into the community to get members of the general public to participate in your research. Laboratories built into trailers and vans can be brought to market parking lots. Portable testing booths can be set up in shopping malls. Even stalls set up on the sidewalk have been used to gain access to the community. Access to certain special populations such as the very elderly, young children, or physically challenged persons may require that you go to their home, school or institutional residence.

SPECIFIC NEEDS OF APPLIED RESEARCH Psychologists are sometimes asked to find answers to specific questions such as the best lighting conditions for a certain factory, how to improve the training of equipment operators, or better signage for a museum. Doing the research on site simultaneously provides access to the same persons who will be using the results of your research, avoids simulation costs, and helps ensure that the results apply to the intended situation.

9.2 VARIABILITY IN THE FIELD

With all these advantages, you might wonder why psychologists don't do more field research. After all, with a world full of people perceiving, thinking, behaving and interacting, why do we need laboratories for psychological research?

Though artificial, laboratory situation makes it easier to control a lot of factors that can otherwise increase the variability to the measurements. Distracting events, variations in the surrounding conditions such as lighting, noise level, and temperature, changes in the social milieu of spectators and bypassers are just some of the possible influences on how subjects respond. Collecting more data is an effective way of coping with random variability. You probably saw examples of how this works with "t" tests in introductory statistics. Yet it may help to look at it again from the perspective of research design. For this example, we'll look at a statistical test that tells whether two sets of measurements differ significantly in their variability - the "F" test for two standard deviations.

Consider a preliminary experiment on touch perception of numerosity - the number of contact points on the skin that can be recognized without counting. To prevent counting when measuring numerosity perception, stimuli are presented for only a second. Les, wondered whether the accuracy would be improved if the contact points produced a temperature sensation as well as pressure. (The skin has separate sensory pathways for pressure, warmth, and cold. Since they differ substantially in conduction time, confusion is a plausible result.) The contact points in this experiment were produced by the ends of wires held in various arrangements. A thermostatically controlled waterbath enabled Les to vary the temperature. For just contact pressure stimulation, a 32°C bath temperature matched surface temperature of the skin. To add warmth to pressure, a 42°C bath was used. Les recruited 6 students from a perception class to serve as subjects. In a quiet laboratory setting, each subject practiced touching and estimating various numbers of contact points in both the pressure-only

practiced touching and estimating various numbers of contact points in both the pressure-only and pressure-plus-warmth conditions. Data were then obtained by asking the subjects to estimate the number of contacts when ten points were presented twice - once without warmth and once with warmth. The order of testing without-warmth and with-warmth conditions was counterbalanced across subjects.¹ Les' results are presented in Table 9.2-1

Table 9.2-1.
Number of Points Perceived When Briefly Touching
10 Points at Skin Temperature and at 42° C

SUBJ	PRESSURE ONLY	PRESSURE & WARMTH
AA	10	11
BB	6	9
CC	10	9
DD	12	11
EE	15	11
FF	7	9
mean:	10	10
sd:	3	1

Under both conditions, the subjects were correct on average in their estimates of the number of rods. What is different is that their estimates were more variable when there was only contact pressure information about the number of rods. To determine if the standard deviations differ, Les did an F-test as follows:

$$\begin{aligned} F &= (\text{SD}_{\text{press only}})^2 / (\text{SD}_{\text{pres \& warm}})^2 \\ &= 3^2 / 1^2 = 9 - \text{a significant difference} \\ &\text{(at 5 \& 5 degrees of freedom, need } F > 5.05) \end{aligned}$$

Judgements of the number of items was significantly less variable when information about the array was carried by two sensory systems instead of just one. During debriefing, Les realized that these subjects all knew about the limitations of short-term memory to 7 + or - 2 items. Though not clear from their comments, it was possible that at least some expected that they were being tested on such a matter. Would naive subjects show the same effect? Since a most students had an introductory psychology course, Les decided to go public and arranged to set up the experiment in the public library foyer. Eagerly analyzing the results from the six subjects as shown in Table 9.2-2, Less found the same means as previously, but the difference in variability was no longer significant.

¹ The proper name for this design would be "an incomplete within-subjects design with two levels of the independent variable." The design is overly simplified to illustrate a principle. A more realistic experiment would use arrays containing 4 to 14 rods with complete counterbalancing. Such an experiment does not appear to have been published.

Table 9.2-2.
Number of Items Perceived When Touching a 10 Rod Array
at Skin Temperature and at 42° C - Community Sample

SUBJ	PRESSURE ONLY	PRESSURE & WARMTH
JJ	10	10
KK	5	8
LL	13	14
MM	5	8
NN	16	10
OO	11	10
mean:	10	10
sd:	4	2

$$F = 4^2 / 2^2 = 4$$

Looking at the "F" Table, Les noted that a ratio of "4" was significant at 7 & 7 degrees of freedom. If the public measurements continued as they had, 8 subjects would show a significant effect. Knowing that the variability of means tends to increase as the number of measurements are increased, Les decided to continue the experiment until 12 persons had been tested. The complete results are shown in Table 9.2-3.

Table 9.2-3.
Number of Items Perceived When Touching 10 Rod Array at
Skin Temperature and at 42° C - Community Sample of 12 Persons

SUBJ	PRESSURE ONLY	PRESSURE & WARMTH
JJ	10	10
KK	5	8
LL	13	14
MM	5	8
NN	16	10
OO	11	10
PP	9	11
QQ	7	9
RR	6	6
SS	7	11
TT	16	12
UU	15	11
mean:	10	10
sd:	4	2

$$F = 4^2 / 2^2 = 4$$

at 11 and 11 degrees of freedom, need $F > 2.82$

Les now had evidence that information about the number of items conveyed by the warmth receptors in the skin led to more accurate judgements for subjects who were unlikely to be acquainted with the expected limits of short-term memory. Though the results were the same as those obtained in the laboratory, additional measurements were needed to overcome the effects of the greater variability of measurements made in the field. Due to greater variability, field research usually requires more data to achieve the same degree of significance as laboratory research.

9.3 CONFOUNDING EFFECTS

There is a worst threat than statistical insignificance posed by uncontrolled variables that influence your measurements. They may introduce significant effects that have nothing to do with your independent variable.² This can occur when uncontrolled variables do not happen to be random with respect to the levels of your independent variable. The presence of any uncontrolled variables make field designs more vulnerable to the possibility that one or more of these change at the same time that you change the independent variable. For example, research done in a public place must contend with changes in traffic that are not random but vary systematically with the time of day and day of the week. Research done outdoors must contend with weather which varies in complex cycles.

To illustrate what can happen, consider the following field study on social conformity and age: Posing as a roving reporter in a large shopping mall, Pat selected persons who appeared to be either young adults between the ages of 18 to 24, persons who appear to be middle aged, and persons who appear to be over 65. From classmates, relatives and neighbors, Pat recruited a male and a female accomplice at each age level to participate on half of the interviews. The accomplices informed Pat when they were in position near a likely subject. Pat then came over and asked the group their opinion on a fictitious bill under consideration by the legislature. The accomplices spoke out strongly for the bill. Then the "real" subject was asked how strongly he/she felt for or against the bill using a differential -5 to +5 rating scale. Afterwards the subject was debriefed, thanked, asked not to divulge the study to others. For the other half of the interviews, there were no accomplices - the subject was interviewed alone.

This is an example of a between-subjects design that has two independent variables. One independent variable is the social situation in which people are interviewed. It has two levels: alone, and in a social situation where others favored the bill. Using just the first independent variable, one could do an experiment that measured whether people were influenced by the opinions of others. Such an experiment might be a clever way of gauging public attitudes about that particular bill, but otherwise would not be particularly interesting. (It is well known that people are generally influenced by the opinions of others.) The second independent variable is the age of the subjects. One could simply ask persons alone from different age groups what they thought about the bill. This too would be of little general interest, unless it was the bill itself that was the issue.

² Confounding effects can also counteract an appreciable effect of the independent variable. In this case a significant effect would not be obtained by collecting more data.

But since it is not a real bill, people wouldn't have any opinion about it. It seems that Pat could measure conformity simply by asking people's opinion in that contrived social setting. - Editor

Yes, that's correct in principle, but it would be a risky design. People may have certain opinions about a bill they have never heard of because they confuse it with some other bill or because just the name of the bill creates certain expectations. By measuring the difference in opinion between those asked alone and those asked in a group, Pat can be more certain that the results represent conformity. - Nilsson

What about if half the time Pat had the accomplices argue for the bill and the other half against the bill? - Editor

That's a clever idea! It would also improve Pat's design. - Nilsson

It happened that Thursday was best for the seniors accomplices to help Pat, Tuesday for the middle aged, and Saturday for the young adult accomplices. Therefore, Pat did the alone interviews on Monday, Wednesday, and Friday. An equal number of persons from each age group were interviewed on each of those days. The whole schedule presents a problem because malls tend to be progressively more crowded as the week progresses. Assume that crowding has a negative affect on how agreeable people are. We'll represent "agreeableness" on scale of -5 (tendency to strongly disagree with anything) to +5 (tendency to heartily agree with anything). Assume that someone had done a preliminary study in the mall on different days of the week. In that study lots of persons had been asked about a fictitious bill before the legislature. "Many people seem to favor the MacDonald bill. On a scale of -5 to +5 how strongly do you disagree or agree with it?" From this we know that on Mondays and Tuesdays when the mall is not at all crowded the average agreeableness was 0 (uncertain), but by Saturday it had dropped to -2. Furthermore, can you recall stretches of weeks on end when it always rained on the weekend, but was nice all week? Let's assume this happened when Pat did this study, and that rainy days made people less agreeable by a -1 on the agreeableness scale. The combined effects of crowding and weather on agreeableness are presented in Table 9.3-1.

Table 9.3-1
How Crowding and the Weather Influence
the Tendency of People to Agree with the Opinions of Others

DAY	CROWDING	WEATHER	COMBINED
MONDAY	0	0	0
TUESDAY	0	0	0
WEDNESDAY	-0.5	0	-0.5
THURSDAY	-1	0	-1
FRIDAY	-1.5	0	-1.5
SATURDAY	-2	-1	-3

Because the young, mid-aged, and seniors were measured on different days (Saturday, Thursday, and Monday, respectively), the effects of age are confounded with the effects of crowding and weather. Because 2/3 of the persons interviewed alone were early in the week (Monday, Wednesday, and Friday) and 2/3 of the persons in a group were later in the week (Tuesday, Thursday, and Saturday), the effects of a social influence are also confounded by these effects. To illustrate how the confounding occurs, let's assume that the "true" effect of age is as follows: 1) young persons are likely to agree with others to an extent represented by a +2 in agreeableness; 2) mid-aged are likely to agree with others to a lesser extent of a +1 in agreeableness; 3) seniors are least influenced by the others, which we'll represent as a "0". Table 9.3-2 shows what happens to these true effects when the measurements are influenced by the above confounding influences.

Table 9.3-2
How the Combined Effects of Crowding and Weather on Agreeableness
Changed Measurements of How Much Persons of Various Ages
Favored a Certain Bill When Alone and in a Group

AGE	ALONE			SOCIAL		
	TRUE EFFECT	CONFFOUND EFFECT ¹	MEASURED EFFECT	TRUE EFFECT	CONFFOUND EFFECT ²	MEASURED EFFECT
YOUNG	0	-0.7	-0.7	2	-3	-1
MID-AGE	0	-0.7	-0.7	1	0	1
SENIOR	0	-0.7	-0.7	0	-1	-1

Notes: 1) The ALONE confounded effects are based on the combined effects of crowding and weather averaged for Monday, Wednesday, and Friday. 2) The SOCIAL confounded effects are based on the combined effect of crowding and weather on Saturday for the young, Tuesday for mid-aged, and Thursday for seniors.

When alone, it appears that most people regardless of age tended to be a bit negative about the bill. When in a group, it appears that middle aged persons were most influenced by the opinions of others to favor the bill, while young people and seniors reacted against the opinion of others to be less in favor of the bill. To see the effects of the social influence more clearly, a researcher would treat the results from persons interviewed alone as baseline data. The difference between the opinions expressed alone and the opinions expressed in a group indicates the size of the social influence. Figure 9.3.1 shows the results of this further analysis of the data and also shows the true effect for comparison.

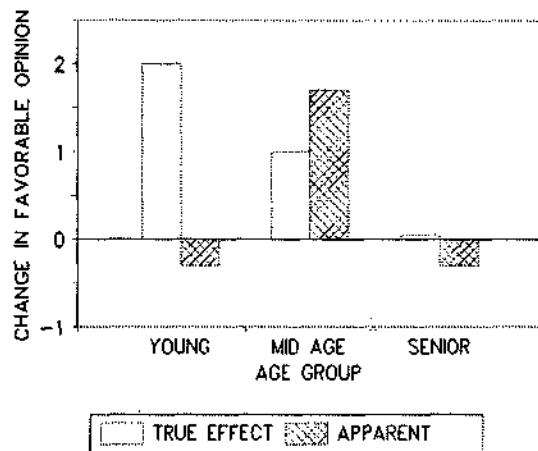


Figure 9.3. A comparison of the true and the measured effects of social influence on the opinions of three age groups.

Taking into account the somewhat negative opinion measured when persons were alone, the above analysis suggests that mid-aged persons were more changed by a group of peers to favor the bill than were young and senior persons. However, in this hypothetical example, we know two things Pat could not know. 1) People had no particular opinion about the obscure bill. We can attribute the measured negativity to the influence of the crowding on Wednesday (-.5) and Friday (-1.5) averaged over the three days when subjects were interviewed alone. 2) Seniors were actually the least influenced by others, while young adults were most influenced by others. We can attribute the apparent maximal social influence on the mid-aged group to the fact that they happened to be least negatively influenced by crowding and weather because they happened to be tested on Tuesday.

9.4 HOW TO DEAL WITH CONFOUNDING EFFECTS

In the above field research example, two things went wrong. One, the accomplices' preferred days to help introduced a substantial confounding effect, because crowding and weather varied at the same time as the age levels were changed. Pat should have noticed the difference in crowding and sought other accomplices. Let's assume that Pat can get any team of accomplices out to the mall on any given day. How should the interviews be organized to prevent those differences from confounding the measurements?

9.4.1 HOLD POTENTIAL VARIABLES CONSTANT

Pat could **hold constant** the testing day - i.e. only test subjects on Tuesday. Then, obviously, day of the week can not be a confounding variable. While effective, this may create another problem with respect to the **representativeness** of the results. Certain people may go to malls on Tuesday to avoid others because they are less (or more) influenced by the presence of others than are people in general. To solve the representativeness problem would require testing each age group every day - something that may not feasible.

There is another approach to holding things constant. Above I said **two** things went wrong in Pat's study. The second thing is the influence of the weather - nice all week, except it rained on Saturday. Typical of sporadic factors that can affect the results of any

research, field research is particularly vulnerable since there is more opportunity for them to affect measurements outside the laboratory. The effect of a rainy day in a mall tends to be magnified by the numerous social interactions. In doing field research you have to alert for such sporadic factors. When encountered, the researcher should consider postponing measurements until another time. Pat should have noted the lousy weather, realized that it could affect the results, and postponed testing till the following Saturday.

This raises another matter. What if a certain day started out marginal and then got worse? With research team in place, it may seem best continue the interviews and "hope for the best". Unfortunately, you do not know how things turned out until the results are analyzed. Then you face a dilemma. To "throw out" the results at that point would represent a selective repression of data that does not meet your expectations. Yet to keep the data, might distort the results. This probably is the most common threat to the "honesty" of data in any science. To avoid the dilemma, you must decide whether to keep data obtained under dubious conditions before they are analyzed.

9.4.2 BALANCING - THE LATIN-SQUARE METHOD

In experiments, balancing and matching were the ways to cope with order effects and subject characteristics that could not be held constant. The same is true when it comes to field research. Yet differences between laboratory experiments and typical field research designs need to be considered. In within-subjects experiments, balancing and counterbalancing are used to organize the sequence in which levels of the independent variable are presented to individual subjects. While within-subjects designs can be used in field settings (as in the touch experiment discussed in Section 9.2), the more typical field research involves many subjects - each tested briefly in a between-subjects design. In between-subjects experiments, balancing by matching was used to assign subjects to levels. In field research, balancing is used to select when, where, and ultimately which subjects get tested. Since the Latin-Square method is the only balancing method generally suitable for field research, we'll focus our attention on it. First, I'll illustrate how Latin Squares can be used to balance field conditions. Then we'll consider Pat's research problem.

A SIMPLE EXAMPLE

In within-subjects experiments, Latin Squares were used to test a set of levels of the independent variable once in each position of the sequence in which they were presented to the subject. Let's say we want to find out whether socio-economic status influences people's respect for law and order. We'll operationally define respect for law in terms of the minimum speed a driver reaches at a STOP sign. We'll operationally define status in terms of average home size in three residential neighborhoods that vary from small apartments to spacious two storey homes with two garages. Ideally, we would like to send a teams of researchers to each neighborhood at the same time. However, we don't have that many assistants nor three sets of speed measuring devices.

You might simply shrug your shoulders and test each neighborhood on successive Saturdays in a random order. After all there are no obvious reasons why a relationship between status and respect would change over three weeks. Yet a number of other factors could alter how people respond to a STOP sign: The weather could change. A tragic auto accident could happen during the study. The police might decide on a stricter enforcement policy. All of these could confound the results of the neighborhood that was studied last.

Therefore, you decide to test each neighborhood on a single day - one in the morning, one around midday, and one late afternoon. However, now you are faced with possible differences in driving at various times of the day. The solution is use a Latin Square balancing and test each neighborhood three times - once at each time of day. Table 9.4.2-1 shows how this could be done.

Table 9.4.2-1
Testing Schedule for Three Neighborhoods
at Three Times of Day Over Three Saturdays.

TIME	MAR 12	MAR 19	MAR 26
8-11	LOW	HIGH	MED
12-3	MED	LOW	HIGH
4-7	HIGH	MED	LOW

BACK TO THE MALL

Rather than hold the day of the week constant in the above example, Pat could make measurements every day, as long as the days were **balanced** across age groups - i.e. test persons from each age group on each type of day. (Such a design would resemble the Latin Square approach used to balance order effects in within-subjects experiments.) Using one team per day, the study would require twice as many weeks as the types of days. Of course, Pat doesn't have to test every weekday, Monday, Wednesday, and Friday might suffice to represent various degrees of crowding. Accordingly, Pat could make up a testing schedule like the one in Table 9.4.2-1

Table 9.4.2-1
How Three Groups of Subjects, Could be Tested Alone or with Social Influence
at One Group per Day, for Three Days of the Week
with a Schedule Balanced Across Days Over a Period of Three Weeks

WEEK	MONDAY	WEDNESDAY	FRIDAY
1	SENIORS (ALONE)	MID-AGE (ALONE)	YOUNG (ALONE)
2	YOUNG (SOCIAL)	SENIORS (SOCIAL)	MID-AGE (SOCIAL)
3	MID-AGE (ALONE)	YOUNG (ALONE)	SENIORS (ALONE)
4	SENIORS (SOCIAL)	MID-AGE (SOCIAL)	YOUNG (SOCIAL)
5	YOUNG (ALONE)	SENIORS (ALONE)	MID-AGE (ALONE)
6	MID-AGE (SOCIAL)	YOUNG (SOCIAL)	(SENIORS (SOCIAL))

To see how this avoids having crowding confound the results, we'll look at what happens to the real social conformity effects as they interact with crowding over this 3-week period. See Table 9.4.2-2.

Table 9.4.2-2.
How the Effects of Crowding Would Influence the Average Opinions
of Three Age Groups on Three Days of the Week.

DAY	SENIOR			MID-AGED			YOUNG		
	TRUE EFFECT	CROWD EFFECT	RESULT	TRUE EFFECT	CROWD EFFECT	RESULT	TRUE EFFECT	CROWD EFFECT	RESULT
ALONE									
MON	0	0	0	0	0	0	0	0	0
WED	0	-0.5	-0.5	0	-0.5	-0.5	0	-0.5	-0.5
FRI	0	-1.5	-1.5	0	-1.5	-1.5	0	-1.5	-1.5
	MEAN:	-0.67		MEAN:	-0.67		MEAN:	-.67	
WITH SOCIAL INFLUENCE									
MON	0	0	0	+1	0	1	+2	0	2
WED	0	-0.5	0	+1	-0.5	+0.5	+2	-0.5	+1.5
FRI	0	-1.5	0	+1	-1.5	-0.5	+2	-1.5	+0.5
	MEAN:	-0.67		MEAN:	+0.33		MEAN:	+1.33	

The generally negative effects of crowding on all subjects' willingness to agree with the bill when alone are now the same, mean = -0.67, for all age groups. To reveal the magnitude of the social conformity effect, the opinions expressed while alone are subtracted from those subject to social influence. Table 9.4.2-3 shows the results of these subtractions.

Table 9.4.2-3
Estimates of Social Conformity Derived by Subtracting the Mean Rating Obtained Alone
From the Mean Rating Obtained in a Social Context Amongst Three Age Groups

	SENIOR	MID-AGE	YOUNG
SOCIAL	-0.67	+0.33	+1.33
ALONE	-0.67	-0.67	-0.67
CONFORMITY	0	+1	+2

Note: Subtracting a negative number adds its value.

The measured conformity now equals the hypothetical "true" conformity. Balancing has removed the confounding effect introduced by different degrees of crowding on different days. Lesson: When doing field research, try to balance your measurements

across time or place or whatever, whenever possible. You never know what confounding variable may be lurking to distort your measurements. This is also good advice for other research designs.

Those of you who have career goals in psychology, are likely to take an advanced statistics course. There you will encounter another reasons for using completely balanced designs like Latin-Squares to deal with potential confounding variables. When you make measurements at all levels of any variable, you can treat that variable as an independent variable - even if it was not intended to be an independent variable in your research. Analysis of variance, ANOVA, can be used for data from research designs that involve more than one independent variable. In treating a potential confounding variable as another independent variable, the ANOVA calculations in effect separate the effects of that variable from the effects of the independent variable that interests you. ANOVA also enables you to gage the extent to which a potential confounding variable has affected your intended independent variable in terms of what are called "interaction effects". Despite all these high faltering calculations, you still end up calculating averages across the various levels of the potential confounding variable just as we did above.³

9.4.3 COUNTERBALANCING - ABBA

Latin-Square balancing along with its ANOVA possibilities, may sound like the best way to design field research, but they have a serious drawback that can particularly prevent their use outside (as well as inside) the laboratory. They require that you collect a lot more data in the form of replicating your measurements at each level of every possible confounding variable. When there are just four levels such the four schools in the above example, this may not be serious drawback. Yet while teachers may permit you to disrupt their classrooms once or twice, many may object to your doing so four or more times. When a potential confounding variable such as the different schools in our example has four or more levels, you may be forced to consider using an alternative design that does not require as many replications as there are factor levels.

When an uncontrolled variable has an approximately linear, progressive effect on field measurements, counterbalancing is more efficient than balancing. Examples of such progressive variables include the following:

- 1) time - Over the course of a work day, people gradually become more fatigued and less alert.
- 2) temperature - Increases above the comfort level tend to reduce metabolic rate, increase the rate of fatigue. Within a limited range, decreases below optimum have the opposite effect.
- 3) seasonal - As the days get more pleasant from winter to spring, there tends to be gradual changes in metabolic rate, mood, amount of outdoor activity, clothing, etc.

³ Don't let these impressive claims about ANOVA lead you think that it's something really complicated. My introductory statistics book, Statistics for Psychology - A Number Work Approach, explains a method so simple that it's boring.

- 4) location of an audience member from the center of activity - This may reflect the person's degree of interest in the proceedings.
- 5) market value of automobile - This tends to reflect the drivers' economic status.
- 6) age of the subjects - This can reflect increasing cognitive skills or degrees of judgement.

*Note that these examples would have a progressive affect only on certain measurements. Increasing age may tend to be accompanied by increasing skill in solving anagrams, but it is unlikely to have an systematic influence on attitudes on abortion. A subject's level of income may correlate with their years of education, but not with spatial abilities.

How can you tell whether an uncontrolled variable would have a confounding effect on your measurements? For psychologists, a broad understanding of human behavior gained from a variety of life experiences and/or studies in a variety of areas such as sociology, history, literature can be as valuable as studies in the biological and physical sciences. A more immediate approach is to use a "pound of prevention" and always balance or counterbalance subjects and testing situations whenever possible.

AN EXAMPLE - COUNTERBALANCING TIME OF DAY

To illustrate how counterbalancing can work in field research, we'll go to a large factory. The job is to find out which brand of ergonomic workstation produces the least fatigue. Six brands (Ergot, Green, HiLo, N & M, Seato, and Workal) are to be evaluated. The company prefers that each worker participate only once. This precludes a within-subjects design, but avoids problems of order effects in the testing. To keep the work task similar, we will introduce the various workstations at one production location. Therefore each brand has to be tested at a different time. Over the work day, there are six 60-minute periods available which do not interact with scheduled work breaks. To test each brand at each period of the day would require six days of testing for a balanced design. The company, however, feels that would be too disruptive.

If worker fatigue increased approximately linearly over the course of a day, we could use ABBA counterbalancing and do the testing in two days - preferably mid-week. We obtain permission to conduct a brief survey to measure fatigue over the course of the work. A paper-and-pencil fatigue check list that takes less than a minute to do is given to a group of workers at 75 minute intervals over one day. These workers were all using the same current work station. The average results are plotted below.

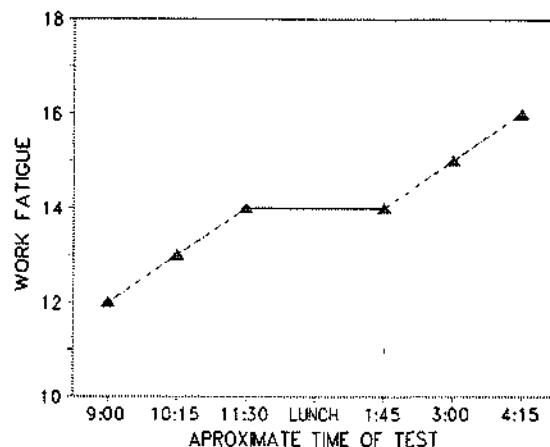


Figure 9.4.3-1. Mean fatigue scores over the course of the work day obtained from workers using the present workstations.

Except for the halt over the lunch period, fatigue amongst these work does increase linearly over the day. Therefore, we proceed with an ABBA counterbalanced design.

For levels of the independent variable, we have the six workstations: . Since the workstations do not vary along any known manner, the their order of testing is arbitrary. We select a random order (Green, Workal, Ergo, N&M, HiLo, Seato) to preclude any bias on our part. On the first day, we'll test the brands in that order; on the second day, in reverse order. We'll use a configuration of each workstation that accommodates four workers. Testing the six workstations twice therefore requires 48 workers. Without further information on the workers years of experience or skill, we block-randomly schedule their participation. Table 9.4.3-1 shows the testing schedule.

Table 9.4.3-1.
Schedule for Testing Six Workstations with Forty-Eight Workers Over Two Days

TIME	BRAND	DAY 1				DAY 2			
		WORKER #		WORKER #		WORKER #		WORKER #	
8:15-9:15	Green	25	24	35	16	Seato	1	41	48
9:30-10:30	Workal	10	30	14	44	HiLo	32	46	9
10:45-11:45 (lunch)	Ergot	18	15	12	37	N & M	31	26	5
1:00-2:00	N & M	23	36	17	19	Ergot	7	29	27
2:15-3:15	HiLo	13	28	40	39	Workal	11	47	22
3:30-4:30	Seato	43	20	34	21	Green	6	8	2

In each time period, the four selected workers are shown how to adjust their workstation, asked to continue the normal assembly work till the last ten minutes, and then tested. For the dependent variable, we'll use a battery of performance, physiological and subjective tests, which have a summary score that ranges from 1 (no fatigue) to 20 (high fatigue). Table 9.4.3-2 shows the results obtained on the first day.

Table 9.4.3-2
Fatigue Scores of Workers Using
Various Brands of Workstations - Day 1.

BRAND	DAY 1	mean
Green	11 8 9 12	10
Workal	8 9 10 9	9
Ergot	11 14 14 13	13
HiLo	13 13 11 12	12
H&M	7 9 13 11	10
Seato	14 15 9 14	13

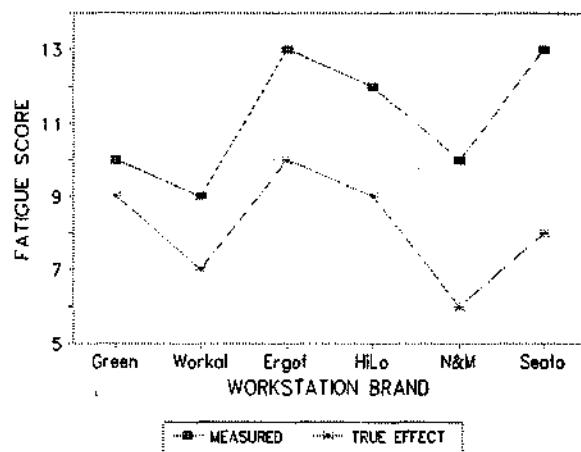


Figure 9.4.3-2. Mean results from Day 1 along with the "true fatiguing effect" of each brand.

These results by themselves show what would have been found had we only tested the workstations once in random order. The Workall brand appears to be the best, followed by the Green and H&M tied for 2nd best, with the Ergot and Seato brands being the worst. To show you how wrong these results would be, Figure 9.4.3-2 shows you something that researchers never actually know - the "true fatigue effect" of the various brands. (I can do this here because I made up the data based on these characteristics.) In reality, the H&M brand was best, Workal 2nd, and the Green brand worst. The distortion in the random tested results is due to the confounding effect of fatigue (the previous Figure 9.4.3-1) because each brand had to be tested at a different time of day.

The next day, the workstations are tested in the opposite order. Here are those results and a graph of the average fatigue produced by each brand.

Table 9.4.3-2
Fatigue Scores of Workers Using
Various Brands of Workstations - Day 2.

BRAND	DAY 2	mean
Seato	9 10 13 12	11
N&M	10 7 10 9	9
HiLo	13 11 11 13	12
Ergot	11 14 13 14	13
Workal	14 13 9 8	11
Green	15 13 14 14	14

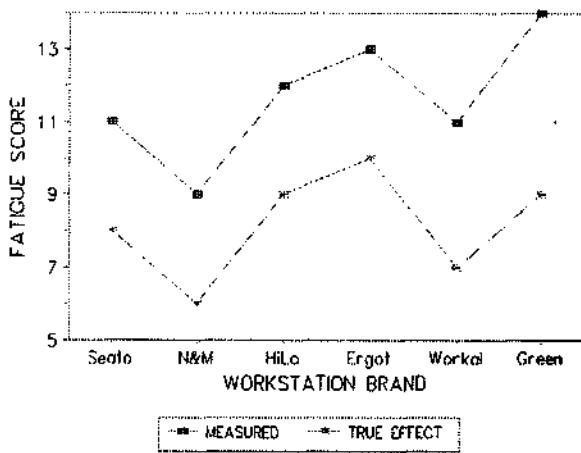


Figure 9.4.3-3. Mean results from Day 2 along with the "true fatiguing effect" of each brand.

These are quite different from the first Day's results. The N&M is clearly best and the Green is worst. The differences in the results is probably unlikely due to these 24 randomly

assigned workers having very different experiences with the various brands - though this can never be completely ruled out. The difference appear to be consistent with the time a brand was tested and the preliminary data on how fatigue increased over the day. Since that fatigue effect was approximately linear, the bias it introduces should be cancelled by averaging the results from the two sets of measurements obtained in counterbalanced order. Those averages are shown in Figure 9.4.3-4 again along with the true fatigue effects of the various brands.

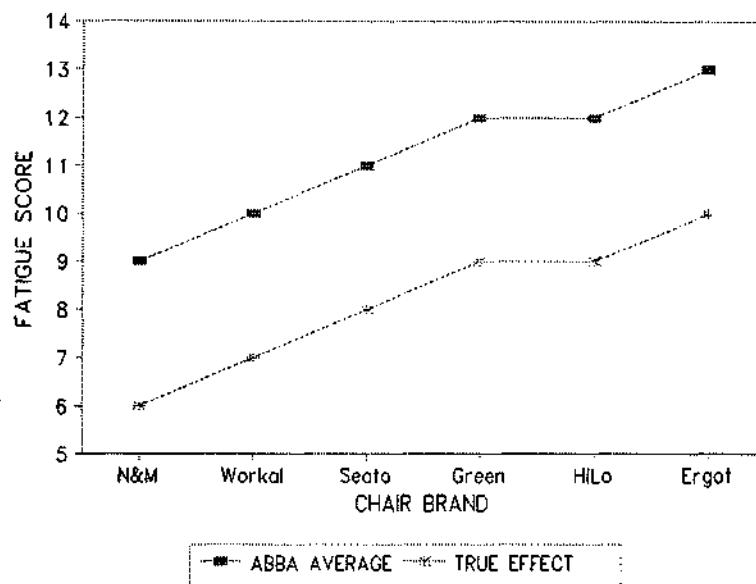


Figure 9.4.3-4. Fatigue scores averaged over two days as function of Workstation Brands, along with the true fatigue effect produced by each brand.

The above figure shows that The N&M workstation produced the least fatigue, followed by the Workal, and that the Green brand was worst. Since the scores for these brands are averages from testing at various times, they should comparable to the average of the fatigue scores obtained the whole day in the preliminary testing - see Figure 9.4.3-1. At 10 and 11 respectively, the average scores of the N&M and the Workal brands compare favorably with the 12 average of the present workstations.

More significant than these specific results. Note that the ABBA counterbalancing led to mean results for each brand that are similar the true fatiguing effects of these chairs. However, note also that the measured means are consistently larger than the TRUE means for these chairs. That increment is due to an inevitable effect of any counterbalancing method. Counterbalancing does not remove confounding effects from the results. Rather, it distributes those effects approximately equally across all levels of the independent variable. Though all of the results are raised by counterbalancing, that even distribution of the confounding variable prevents it from biasing the results.

Were you surprised that ABBA counterbalanced worked as well as it did in this between-subjects field experiment? Probably not. After all the principle is no different than the use of ABBA counterbalancing of the order in which various kevels of an independent variable are tested with a single subject in a with-subjects experiment. The fact that the above design involved the time of day at which the various brands were tested may help in making this connection.

ANOTHER EXAMPLE - COUNTERBALANCING LOCATIONS

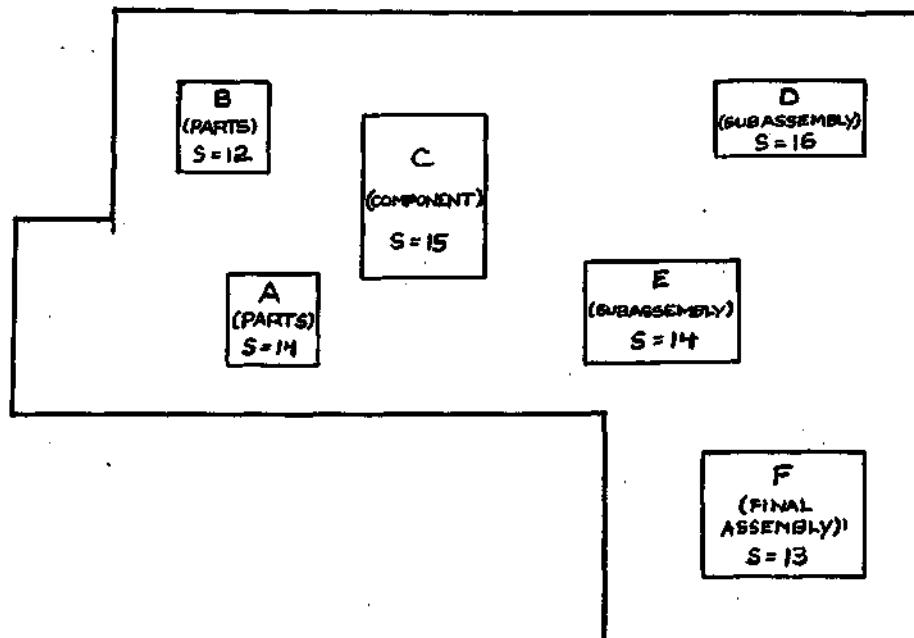


Figure 9.4.3-5. Layout of sections at the Wright Corporation's Factory in Cornwall.

Now consider how to proceed if the company had wanted each workstation tested for an entire day. Yet they want this done during a brief two day period when they are switching over their production line. Now, the workstations will have to be tested at different locations instead of different times. There is not enough time for a Latin Square design. If the production line itself varies in some systematic manner such as progressing from parts construction to component construction to sub assembly stages, to final assembly; you could assign the work stations to be tested twice in counterbalanced order across the locations. Better yet, you could do some preliminary testing to measure fatigue at these various locations. Arrange the locations in according to fatigue level in the research plan (not on the factory floor), and then counterbalance the brands across that sequence as shown in Table 9.4.3-5.

Table 9.4.3-3.

Factory Sections in Ascending Order of Preliminary Test Scores and How Various Brands of Workstations Were Tested over Two Days.

SECTION	PRELIMINARY TEST SCORE	BRAND TESTED	
		DAY 1	DAY 2
B (PARTS)	12	Green	Seato
F (FINAL ASSEM)	13	Workal	HiLo
A (PARTS)	14	Ergot	N & M
E (SUB-ASSEM)	14	N & M	Ergot
C (COMPONENT)	15	HiLo	Workal
D (SUB-ASSEM)	16	Seato	Green

Finally, what could you do if the preliminary testing revealed a fatigue effect which was definitely not linear over time or did not enable arranging the workstations in a linear manner according to fatigue? (Hint - Try to convince the company to give you four days.)

9.4.4 COMPARING COUNTERBALANCING AND BALANCING

When a potential confounding variable being addressed has effects that differ in various ways from one field situation to the next, counterbalancing will not effectively remove the confounding effect. Examples of such categorical variables include: 1) neighborhoods that differ in ethnic proportions; 2) persons having different occupations; 3) days of the week; 4) choice of eating establishment. Different examples of these variables have such different effects, that you had better test all levels of the independent variable in each category. **For these situations, the counterbalancing "shortcut" is not suitable.** Remember, in the typical, between-subjects, field study, you are not simply trying to remove the influence of order and carry-over effects. Rather you are trying to cope with differences between subjects and differences produced by the field conditions either of which can influence what you measure.

To make sure you don't forget, I'll show you what happens when the confounding variable is non-linear. For this demonstration I'll use a field situation that it involves testing different levels of an independent variable at different locations. (Amongst other real-world considerations, there are not enough subjects available at any one location to enable holding location constant.) The objective of this hypothetical study is to evaluate the effectiveness of three new methods plus the current method of teaching the alphabet to kindergarten children for whom English is a second language. There are only 32 children with English as a second language in each of the nearby schools. However, by testing the methods in four school, we could obtain data from 32 children for each of the four methods. To assign one method to each school would obviously invite differences between schools as a confounding variable. Therefore, an ABBA design counterbalanced across schools was used. Half of the children in each school were tested with one method the first two weeks; the other half was tested with a different method the second two weeks. The design is shown in the table below.

Table 9.4.4-1.
How Four Methods of Alphabet Teaching Were Tested
in Counterbalanced Order at Four Schools Over Two Two-Week Periods

SCHOOL	March 2-13	March 16-27
West	Phonolog	Semantic
River	current	Graftec
Park	Graftec	current
North	Semantic	Phonolog

To simplify matters, we'll consider only the mean results obtained at each school. The results that were obtained using this ABBA counterbalanced design are shown in Table 9.4.4-2. For instructional purposes, this table also shows something that researchers never know - the true effects of the independent variable.

Table 9.4.4-2.
Alphabet Proficiency Scores Obtained Using Four Teaching Methods:
Average of 16 Different Children in Each Two-Week Training Session.
Also Shown Are the True Effects of Each Teaching Method

METHOD	March 2-13	March 16-27	Mean	(True Effect)
Semantic	13.0	16.9	15.0	(13)
Graftec	10.1	15.0	12.5	(15)
Phonolog	11.7	9.0	10.4	(9)
current	11.0	7.4	9.19	(11)

The results as shown in Table 9.4.4-2 indicate a clear advantage of using the Semantic teaching method. However, we know that this is a misleading conclusion - true results would show that the Graftec method was best. We are being misled by the above results because the ABBA counterbalancing has not worked as well as we hoped. This problem arises because the schools differ in the overall reading skills of their students in a manner that was not balanced out by the ABBA counterbalancing. The schools are simply categorical levels of a variable in this experiment. We lacked the information needed to properly counterbalance them.

When the results were presented to the local school board in April, a member recalled that a provincial study of reading readiness skills had been done three years previously. She seemed to recall that it had found some regional differences in skills. A report on that study was obtained and revealed the following results for the four schools.

Table 9.4.4-3.
Mean Reading Readiness In the Four Schools

SCHOOL	MEAN READINESS SCORE	RELATIVE SKILL FACTOR
West	32	1.3
River	24	1.0
Park	16	0.6
North	24	1.0

Table 9.4.4-3 reveals that the four schools differed substantially in reading skill of their children. Assuming that these differences are characteristic of the neighborhoods and not just the attendance that year, the order in which we happened to distribute the counterbalanced tested did not represent even an approximately linear progression of these skills. The methods that were used in the West and North schools had the advantage in both the initial and counterbalanced order of involving children with more reading skills than the methods that were tested in the River and Park schools.

THE LATIN-SQUARE ALTERNATIVE

The following illustrates the results that would have been obtained in the above experiment using a Latin-Square design to balance any possible differences in the reading

skills of the children at the various schools in which the four teaching methods were tested. Since Latin-Square balancing requires as many replications as there are levels, we now have visit each school four times (instead of twice as with ABBA counterbalancing). Assuming that we still have only a total of 32 children available at each school, we will test 8 eight children during each two-week period at each school. To do the testing at the same time of the year, we'll start two weeks earlier and continue till two weeks later. To begin, we'll look at how the four teaching methods could be tested in the four schools:

Table 9.4.4-4.
How Four Methods of Alphabet Teaching Were Tested in
Latin-Square Balanced Order at Four Schools Over Two-Week Periods

SCHOOL	(rel. skill)	Feb 15-26	March 2-13	March 16-27	Apr 10-21
West	(1.3)	Phonolog	Semantic	Graftec	current
River	(1)	current	Phonolog	Semantic	Graftec
Park	(0.67)	Graftec	current	Phonolog	Semantic
North	(1)	Semantic	Graftec	current	Phonolog

The next table shows the results obtained with each teaching method during each of the four testing periods. (These simulated results were obtained by multiplying each school's mean skill level by the true effect of the teaching methods. In reality, we never know the true effect of what we are measuring.)

Table 9.4.4-5.
Alphabet proficiency Scores After Two Weeks with Each of Four Methods
of Teaching Used in Each of Four School

METHOD	true effect	Feb. 15-26		March 2-13		March 16-27		Apr 10-21		OBT'D MEAN
		rel school	mean obt'd result	rel school	mean obt'd result	rel school	mean obt'd result	rel school	mean obt'd result	
		skill	skill	skill	skill	skill	skill	skill	skill	
Semantic	13	1	13	1.3	16.9	1	13	0.67	8.7	12.9
Graftec	15	0.67	10.1	1	15	1.3	19.5	1	15	14.9
Phonolog	9	1.3	11.7	1	9	0.67	6	1	9	8.9
current	11	1	11	0.67	7.4	1	11	1.3	14.3	10.9

When averaged over the four periods of testing, Table 9.4.4-5 shows that the obtained mean proficiency scores for each of the teaching methods do resemble the "true" effects of these methods. This was obtained without any additional knowledge about the general levels of reading skill in these schools. Perhaps now you will believe me when I say, "When there is no evident basis for deciding on a particular counterbalancing order, it is better to do more replications and use a Latin-Square design."

9.4.5 USE PRIOR DATA TO DESIGN COUNTERBALANCING SEQUENCES

Had we known about the reading skills data before doing the above experiment, we could have used a better counterbalancing method. We would then arrange the schools in ascending (or descending) order of reading presumed reading skill, and counterbalance that order for the second set of measurements.

Table 9.4.5-1.

Counterbalanced Testing of Four Teaching Methods at Four Schools with the Schools arranged in Descending Order of Reading Skill Based on Other Data

SCHOOL	relative skill	MARCH 2 - 13		MARCH 16 - 27	
		METHOD	mean score	METHOD	mean score
West	1.3	Phonolog	11.7	Semantic	16.9
River	1	current	11	Graftec	15
North	1	Graftec	15	current	11
Park	0.67	Semantic	8.7	Phonolog	6.0

We would then unscramble these results to find the mean for each teaching method as follows:

Table 9.4.5-2.

Alphabet Proficiency Scores Obtained Using Four Teaching Methods: Average of 16 Different Children in Each Two-Week Training Session

METHOD	March 2-13	March 16-27	Mean
Semantic	8.7	16.9	12.8
Graftec	15	15	15
Phonolog	11.7	6	8.85
current	11	11	11.0

This improved counterbalancing sequence produces results that are similar to the true results (which we only know because this is a hypothetical example) and to the results obtained using a Latin-Square balancing. Unfortunately, in field research you generally can't rely on simply coming across other data to assist you in finding a good counterbalancing order. The alternative is to seek such data before starting the project or make some preliminary measurements yourself.

9.4.6 "CORRECTING" DATA IN TERMS OF ADDITIONAL INFORMATION

Some instructors may feel the following should not be included in an introductory research course because methods like this could be used inappropriately. On the other hand, I believe that it is valuable to know how to use all information available to avoid coming to erroneous conclusions. As long as you provide original data and clearly explain further calculations like these, the reader can decide its appropriateness.

Given additional information about the general reading skills of children in the four schools, we could use it to "correct" our results in terms of the relative skill of each school. To equate the schools in terms of their mean reading skill, we divide the mean alphabet proficiency scores obtained at each school by that school's relative skill factor, as shown in Table 9.4.6-1.

Table 9.4.6-1.
Correcting the Obtained Mean Results at Each School in
 Terms of the Relative Reading Skill at that School

SCHOOL	relative skill	MARCH 2 - 13			MARCH 16 - 27		
		METHOD	obtained mean	corrected mean	METHOD	obtained mean	corrected mean
West	1.3	Phonolog	11.7	9.0	Semantic	16.9	13
River	1	current	11	11	Graftec	15	15
Park	0.67	Graftec	10.1	15.07	current	7.37	11
North	1	Semantic	13	13	Phonolog	9	9

These corrected results can now be averaged to obtain a mean result for each method of teaching:

Table 9.4.6-2.
Alphabet Proficiency Scores Obtained Using Four Teaching Methods - Average of 16
 Different Children in Each Two-Week Training Session with Scores
 Corrected in Terms of the Mean Relative Reading Skill at Each School

METHOD	March 2-13	March 16-27	Mean
Semantic	13	13	13.0
Graftec	15.07	15	15.04
Phonolog	9	9	9.0
current	11	11	11.0

Even though a flawed ABBA counterbalancing method was used, the corrected results are the same as the true effects of the teaching methods. Indeed, the above results show that the correction factor produced correct results without any balancing or counterbalancing when only one of the methods was tested in each school. However, this is an ideal example. In the real world you have no way of knowing if a 'correction factor' based on other data is actually correct. Counterbalancing can hedge that bet. Some readers may be skeptical about correction factors based on other data that are three years old, but as far as the research design goes, the above example may represent the best that was possible in a field research situation.

9.4.7 Dealing With Confounding Effects In Other Designs

Confounding effects represent the most serious threat to research of any design. We have just looked at four ways to cope with such threats: 1) holding potential confounding variables constant, 2) balancing - repeating your measurements at each level of a potential

confounding variable, 3) counterbalancing - which involves fewer replications than balancing but which is also riskier if you do not know enough about the confounding effect, and 4) removing a confounding effect based on other information about the confound. All this was illustrated with respect to doing field research, where the threat of confounding variables is greatest. However, do not let those examples limit your perspective. These methods of dealing with confounding effects are equally suitable and desirable for use in descriptive research, between-subjects experiments, within-subjects experiments, and (next chapter) quasi-experiments.

9.5 OUT OF THE LABORATORY AND INTO THE FRYING PAN

While there are systematic solutions to the problems of greater variability and the threat of confounding effects, there are no general guidelines for how to deal with numerous other problems one can, and usually does, encounter in doing field research. I can point out some of these problems,

SPACE, TRANSPORTATION, TIMING A lot of things taken for granted in a laboratory such as a secure space, availability of equipment and assistants, and when subjects are scheduled to arrive just don't come easily in the field. Space is often at a premium in schools, work places, and commercial establishments. Equipment and furnishings may have to be moved to the research site. Arrangements for transporting assistants may be necessary. All this requires planning, time, and expenses not incurred in laboratory research.

ACCESSING SUBJECTS People in the real world are busy with their jobs, families and other necessities of life. Don't expect them to provide more than a few minutes without prior arrangements. At job sites, employers may be unwilling to have work interrupted for research. I once spent months negotiating with trucking companies for permission to study driving on long distance routes in order to learn more about driver fatigue. When the arrangements had finally been made with management and individual drivers, union objections were raised that cancelled the whole project. Not even retired persons are typically sitting around and available to participate in research. Between luncheons, club activities , exercise and volunteer work, many have difficulty fitting you into their crowded schedule.

[CARTOON OF MAN AT DOOR HOLDING YELLING INFANT AND WEARING HEADPHONES AND EYE-TRACKER IN FRONT OF COMPUTER BALANCED ON RESEARCHER'S BACK. RESEARCHER HOLDS AN UMBRELLA WITH ONE HAND, A CLIP BOARD IN THE OTHER. THROUGH WINDOW IS SEEN DENTIST WIFE BENT OVER PATIENT CALLING OUT, "IS THAT AMALGAM READY DEAR?"]

PUBLIC REACTIONS TO RESEARCH It is remarkable that people will often interrupt their activity to participate in research because they recognize the importance of learning about ourselves. During the Canadian Study of Health and Aging (1994), interviewers went to the homes of seniors, asked for details about their family circumstances and health, and even administered a test of cognitive ability. In the main study, a letter of introduction was sent to the selected participants. They were then contacted by telephone to schedule the

interview. Finally, they were phoned again just prior to the scheduled time to confirm it was still suitable for them - and remind them as well. However, a preliminary study had examined participation without any prior notice. We were amazed at the low refusal rates when interviewers showed up unannounced - even in Toronto. Notwithstanding such support for research, most members of the public have little understanding of how research is done. Consequently they may try hard to "help" the research by providing what may be construed as the "right" answers. Furthermore, despite assurances that the results are kept anonymous, they may still have reactive effects such as wanting to look good to the researcher. This presents a dilemma. To the extent that your informality establishes a certain relationship with the subjects to make their participation pleasant, you also tend to increase the latter type of reactivity.

CAN'T MANIPULATE CERTAIN VARIABLES Traffic jams, the weather, social unrest, or election results are examples of conditions that can't be manipulated. It can be unethical to manipulate certain independent variables such as therapy, education or dangerous incidents. Having to wait for such conditions to occur naturally precludes their use in experiments as independent variables. Yet you may have read about psychological research that looked at how weather affects mood or cognitive strategies used by men and women. Such research may use many of the same methods as experiments yet they are not true experiments. What are they? The answer brings us to the last chapter.

EXERCISES

1. Go to a public place such as a shopping center, street corner, or library foyer, where you might conduct field research. Make a list of the random factors that are present and which might affect the responses of subjects. Return several hours later and now look for factors that have changed systematically since before. Return to the same place at the same time on another day and note what factors have changed since you were last there.
2. Assume that you are doing a field research project that involves an independent variable with 3 levels (I, II, & III). Since it takes about forty minutes to set up each level, you can only make measurements at one level within a given testing session. To compare your results with that of some similar published research, you want to test at least 40 subjects at each level. You do some preliminary field testing and note the time it takes to recruit each volunteer, explain the task, obtain the measurements, and debrief the subject. You conclude that you can test about four subjects in an hour. Your academic schedule allows you to conduct the research from 9 to 10 in the morning and from 3:30 to 4:30 in the afternoon during weekdays. You decide to test only weekdays because you note that on weekends the potential subjects differ substantially in age, manner of dress, and level of activity. To control for time-of-day and day-of-week effects, it would be ideal to hold both of these factors constant, but the semester would be long over and you'd be into a different season before you were done. Design a testing schedule that counterbalances the time of day and the day of the week for the levels.
3. Kim wanted to compare memory of pictures that were described with emotional terms with memory of pictures that were described factually. Subjects were asked to remember 20 simple line drawings depicting common social situations. One example showed a boy and girl walking hand in hand. An emotional description of this was, "Young lovers head down the path of life." A factual description, "boy and girl walk hand in hand." In a laboratory setting, subjects were shown the pictures, read the descriptions, and 20 minutes later shown a set of 60 such pictures - including the 20 they had seen before. Seven subjects received the emotional descriptions; seven received the factual. A mean of 13 emotional pictures and 10 factual pictures were recognized.. With standard deviations were 3 and 1 respectively, Kim concluded that the significantly more emotional pictures were recognized. Seeking to replicate these results off campus, Kim tested 14 people in a local donut shop. The means for the two groups were the same as before, emotional mean = 13, factual mean = 10. However, the standard deviations increased to 4 and 2 respectively, so those differences were no longer significant. Assuming that the means and standard deviations in the field research remain the same, how many subjects in each group would Kim need to test for that difference in means to be significant?

Chapter 10

QUASI-EXPERIMENTS: Maximizing Power

WHEN YOU CAN'T EXPERIMENT

Recall that a critical requirement for an experiment is that the independent variable is controlled. Only by deliberately changing the independent variable or by assigning subjects to its various levels can we ensure that the results are due to what we have done and not something else. With certain independent variables, this type of control may not be practical or possible. The alternative is to use a quasi-experimental design. Unfortunately these are not as powerful nor convincing as true experiments. Therefore the researcher must rely on other techniques to compensate for the inability to control the independent variable. Quasi-experiments tend to have only two levels of the independent variable: one level that involves no treatment - often called the CONTROL; the second level being some TREATMENT. It is possible to have more levels, but this modest approach is safer given the variability from uncontrolled variables.

There are 3 types of quasi-experiments: BETWEEN-GROUPS, WITHIN-SUBJECTS, and those involving SUBJECT CHARACTERISTICS as independent variables. They are distinguished by how the independent variable is handled. Recognizing the type helps one to understand what sorts of doubts about their results are appropriate. It is also the first step for considering what techniques may be used to compensate for the weakness that results from the inability to manipulate the independent variable.

10.1 BETWEEN-GROUPS QUASI-EXPERIMENTS

This is probably the most common type of quasi-experimental design. They resemble BETWEEN-SUBJECTS experiments, but have assigned pre-existing groups of subjects to the levels of the independent variable. Therefore, the subjects are neither randomized nor matched in their assignment. Consequently there may be many reasons for any differences in the dependent variable other than the different levels of the independent variable.

Why would you do between-groups quasi-experiments instead of between-subjects experiments? The main reason is the availability of subjects. For certain independent variables such as various teaching methods or types of social stress, the only practical way to administer the levels of the independent variable may be to various pre-existing groups of subjects. Examples of such pre-existing groups include: 1) students in various classes, 2) customers in a cafeteria at various time periods, 3) people in a library foyer on various days, 4) workers at various factory sites, 5) patients at different clinics. The researcher has little choice over the particular subjects in any of these situations - they certainly can't be "assigned" to be at one location or another. Yet their presence in one group versus another can not be considered random. Any number of variables such as economic status, skills, attitudes, or occupation could affect your measurements. Therefore, differences in the results obtained at different levels of the independent variable could be due to these other variables instead.

10.1.1 BASIC EXAMPLE

Is a new design for a SCHOOL ZONE road sign better than the present sign? In a driving simulation laboratory, we could have subjects "drive" along a road where they occasionally encountered school zones marked by one or the other type of sign in a within-subjects experiment. We could monitor their speed and braking behavior to see if one type

of sign produced better compliance. In a field experiment, we could arrange for some sixteen towns of similar size and situation to participate. Block-random assignment or possibly some matching procedure could be used to form two groups of towns, one of which would receive the new sign.¹ In this field experiment, our dependent variable could be the number of speeding incidents in each town. But what if we have neither a driving simulator nor enough towns willing to participate? That's when quasi-experimental designs come into their own.

Let's say we manage to convince the council of one town, O'Leary, to replace their signs with the new sign. Getting a second town, Montague, to act as the control was not difficult since that involved no change. Though there are some 7,000 residents in each town, it is not a between-subjects experiment, because we have not used random or matched assignment to assign who lives in each town. Therefore there could be many reasons why the residents of one town may drive more carefully - reasons that have nothing to do with our signs. For example, perhaps the O'Leary town council agreed to participate because such violations had not been a problem there, and they knew there was little risk.

Having arranged for two levels of an independent variable as best as possible, we must now come up with a dependent variable that can tell us whether the new sign is better or not. In other words, we need an operational definition of "better" to devise a measurement. In quasi-experiments, you often can't do much to improve the independent variable, but you can try to make up for that inherent weakness by putting more effort into the dependent variable. I'll point out some possibilities for this example and illustrate some of the finer points about how psychologists make measurements.

10.1.2 HOW TO MAKE BETTER MEASUREMENTS

The easiest approach is probably to request access to municipal records of by-law infringements over a period of, say, 6 months in both towns, starting with the date the new signs were put in O'Leary. So we wait six months and then find that there were 8 traffic citations related to school zone speeding in O'Leary and 14 in Montague.

What can we do with this data? Not much by themselves. However, one could take into account the population of the two towns and do a Chi-Square contingency analysis of the obtained and expected number of violations. Montague has a population of 8,995 and O'Leary has 7,220. To use these data, we have to assume that each violation in a town was by a resident and that there were no repeat offenders. (This, of course, may not be accurate, but it's the best that can done if we have no other information.) According to population, we categorize the residents as either offenders or non-offenders. On this basis the Chi-Square calculations lead us to expect 10 offenders for O'Leary and 12 offenders for the slightly larger Montague if the signs had no effect. The calculations outlined in Table 10.1.2-1 show that the effect is not even close to being significant. (At 1 degree of freedom, a Chi-Square value of 3.84 is needed.) There simply are not enough violations in these towns to show much of an effect in these general terms.²

¹ In this case we would be treating whole towns as "subjects".

² Students have asked why a simple Chi-Square could not be done on the observed number of offenses and use an equal number of offenses as the "expected" values. This would only be appropriate if the offenders had chosen whether to speed in Montague or O'Leary.

Table 10.1.2-1
Chi-Square Analysis of School Zone Traffic Violations Over Six Months in
O'Leary and Montague Taking into Account the Town Populations

OBTAINED	O'LEARY	MONTAGUE	SUM
OFFENDERS	8	14	22
NON-OFFENDERS.	<u>7212</u>	<u>8981</u>	<u>16193</u>
SUM	7220	8995	16215
EXPECTED			
OFFENDERS	9.8	12.2	
NON-OFFENDERS.	7210	8983	
CHI-SQUARE	$(O - E)^2 / E$		
OFFENDERS	0.33	0.26	0.59
NON-OFFENDERS	0	0	0
		CHI-SQUARE TOTAL =	0.59

Further discussion with the RCMP reveals that a development project on the outskirts of O'Leary led to a substantial increase in traffic three months into the study, and furthermore this involved many persons from "away". Therefore we need information that reflects the amount of traffic. This could be obtained by counting all traffic violations in the two towns over the same period. The records reveal a total of 80 violations for O'Leary and 78 for Montague. Once again we try a Chi-Square analysis as shown in Table 10.1.2-2. This time we come closer to demonstrating some significant effect.

Table 10.1.2-2
Chi-Square Analysis of School Zone Traffic Violations in O'Leary and Montague
Taking into Account the Number of Traffic Violations During the Same Period

OBTAINED	O'LEARY	MONTAGUE	SUM
SCHOOL ZONE	8	14	22
ALL VIOLATIONS	<u>74</u>	<u>64</u>	<u>138</u>
SUM	82	78	160
EXPECTED			
SCHOOL ZONE	11.1	10.9	
ALL VIOLATIONS	68.9	67.1	
CHI-SQUARE			
SCHOOL ZONE	0.88	0.91	1.79
ALL VIOLATIONS	0.14	0.15	0.29
		CHI-SQUARE TOTAL =	2.08

To understand why we still are not obtaining any significant (assuming that the new sign really works better), we need to look at what is being measured from a broader perspective: **When you combine a weak research design with a dependent variable that permits only a weak, non-parametric statistical analysis, there has to be a very strong effect before you can claim it is unlikely due to chance.**

We need a dependent variable that can provide enough repeated measurements to enable the use of a metric statistical test such a t-test. First we'll consider what could be done using survey measurements. We wait a month after the new signs are up and then interview persons who live in the school zone in each town. There are a number of questions that could give us some quantitative data on their perceptions of drivers' reactions to the signs.³ For example:

SURVEY ON VEHICLE SPEEDS

This is a study being conducted by --- etc.

Please estimate the general speed of cars passing by here on weekdays over the past month:

QUITE	QUITE
SLOW	-----
20	30
40	50
60	70
FAST	
(KILOMETERS PER HOUR)	

Here are some hypothetical data:..

Table 10.1.2-3
Residents' Estimates of Vehicle Speed in a School Zone during Weekdays

O'LEARY		MONTAGUE	
SUBJECT	SPEED	SUBJECT	SPEED
AA	40	LL	50
BB	50	MM	30
CC	30	NN	60
DD	30	OO	70
EE	60	PP	30
FF	40	QQ	50
GG	30	RR	50
HH	50	SS	60
JJ	40	--	--
KK	30	--	--
<hr/>		<hr/>	
MEAN	40		50
SD =	10	SD =	175

³ Among other possibilities we could interview drivers in each town. These however might show a certain, understandable reactivity when it comes to answering how fast they drive past the school in their town. Given the legal aspects, even assurance of anonymity might not suffice in a small town. However, given some thought, there likely are some indirect ways to approach this matter by asking questions.

The above data produced a t-test value of 1.66. However, at $9 + 6 = 14$ degrees of freedom, t must exceed 2.13 for the means to differ significantly. These results are closer to showing a significant effect than could be obtained with the data on traffic violations. However, even though the mean estimate of vehicle speed is 10 km/hr faster in Montague than O'Leary, we can not conclude this difference is unlikely due to chance.

Two aspects of this field research may have prevented our measurements from revealing a significant effect. One, there were large individual differences in people's estimates of the speed of passing vehicles and a lot of other things as well. That variability can often be managed by asking more people. Which brings us to the second aspect. In this situation there were not enough persons living along the road within the school zone to provide enough such measurements. Comparable difficulties are not uncommon in field research.

When one encounters such difficulties in real life, one has two choices: Give up, and seek some less challenging research - like nuclear magnetic brain splicing, or figure out a still better a way to make the measurements. In this case "better" requires a dependent variable that will provide both more measurements and more accurate measurements of how school zone signs affect driving behavior. A direct measure of driving behavior would be how fast cars travel in the school zone. A radar ranging speed monitor is borrowed from the RCMP to try out this idea. One morning we go to O'Leary and make some preliminary measurements then drive to Montague to take some. The hypothetical results are presented above.

Table 10.1.2-4
Speed of Vehicles in School Zones During a Weekday Morning

O'LEARY		MONTAGUE	
VEHICLE	MEASURED	VEHICLE	MEASURED
1	44	1	48
2	40	2	52
3	48	3	50
4	44	4	46
5	50	5	54
6	38	6	48
7	44	7	52
8	48	8	50
9	40	9	54
10	38	10	46
11	50	11	50
12	44	12	42
13	40	13	58
14	48	14	50
15	40	15	44
16	44	16	56
17	48	17	50
MEAN	44		50
SD	4		4

*Note: Speed is shown in km/hr

The value of "t" for these two sets of data equals 4.22. At $16 + 16 = 32$ degrees of freedom, t must exceed 2.05 for the means to differ significantly. Clearly the mean vehicle speed in O'Leary is significantly less than that in Montague. Note also that these results are significant even though the measured speed difference is substantially smaller than the estimated speed difference. This illustrates how more thought and effort can lead to significant findings in psychology.

Hey, I thought the purpose of this book was to show potential students that psychological research is easy. - Editor

I never said it would be "easy". I said it could fun - provided that you know what you're doing and are willing to work. - Nilsson

In general terms the design used in the above basic example is called a "treatment-control, post-test only, between-groups quasi-experiment." Quite a mouthful, but the terminology may be useful for remembering the design as we go on to consider other variations. If you are a pictorial thinker, the following schematic representation may be easier to understand:

	<u>INDEP. VAR.</u> LEVEL	<u>DEPEND. VAR.</u>
GROUP 1:	CONTROL	MEASUREMENT
GROUP 2:	TREATMENT	MEASUREMENT

More specifically, in terms of the above example the schematic would look like this:

	<u>INDEP. VAR.</u> LEVEL	<u>DEPEND. VAR.</u>
MONTAGUE:	OLD SIGN	VEHICLE SPEED
O'LEARY:	NEW SIGN	VEHICLE SPEED

Yet even statistically significant results from such designs are not all that convincing. Why? When an each level of an independent variable is tested on separate predetermined groups, you can never be certain whether any measured differences are due to levels or the groups. In the above example, several of differences between O'Leary and Montague that morning might be responsible for the difference in mean driving speed: 1) There was a time difference of over an hour between the two sets of measurements with Montague's being the later. Later drivers may feel more alert and known that all the children are inside the school. 2) There could have been some special event in Montague that day which made people hurry to work. 3) People in Montague may generally drive faster than those

in O'Leary. As a student from the western end of this 140 km island put it, "They don't know how to drive right out east."

Clearly more evidence is needed to convince an astute person that a difference in speed is really attributable to the new sign. Perhaps there is archival data that shows the number of school zone violations in each town over several years past. A systematic review of local newspapers might provide a basis for comparing the driving habits in each town. Municipal budgets for road maintenance might reveal long term traffic patterns. While the right sort of data you need is often not available or non-existent, you can always gather additional data yourself.

10.1.3 STRENGTHENING BETWEEN-GROUPS DESIGNS

10.1.3.1 ADD A PRE-TEST "A two-group, pre-post test design"

This is the most widely used technique to strengthen quasi-experiments. It is also be used to strengthen real experiments.

a) Principle

Using the same dependent variable, measure both (all) groups before giving each their level of the independent variable. Why duplicate measurements before anything has happened? There are two good reasons: Such "baseline" data may provide convincing evidence that the groups did not differ with the dependent variable prior to treatment. When this the case, one can have more confidence that subsequent differences in the dependent variable are due to treatment effects and not to group differences. Furthermore, in designs where the subjects are identified, pre and post treatment measurements enable the use of the more powerful paired t-tests and within-subjects analysis of variance. Two or more measurements per subject make it possible to distinguish subject variability from random variability.

b) Schematic

	<u>DEPEND. VAR.</u> BEFORE	<u>INDEP. VAR.</u> LEVEL	<u>DEP. VAR.</u> AFTER
GROUP 1:	M ₁	CONTROL	M ₂
GROUP 2:	M ₁	TREATMENT	M ₂

c) An Example

By counting the number of school zone violations in both towns before introducing the new sign in one, you can determine how similar (or different) the towns are generally. For a Chi-Square analysis, the pre-treatment measurements will likely be of a similar magnitude to the post-treatment measurements. This will result in expected values that are more proportional to the post-treatment measurements and thus a larger Chi-Square.

Table 10.1.3.1-1
Chi-Square Analysis of School Zone Traffic Violations
in O'Leary and Montague Before and After New Road Signs in O'Leary

OBTAINED	O'LEARY	MONTAGUE	SUM
BEFORE	18	13	31
AFTER	8	14	22
SUM	26	27	53
EXPECTED			
BEFORE	15.2	15.8	
AFTER	10.8	11.2	
CHI-SQUARE			
BEFORE	0.51	0.49	1.00
AFTER	0.72	0.70	1.41
	CHI-SQUARE TOTAL =		2.42

Well, we got closer to finding a significant effect, but still no banana for the monkey. It just goes to show how big an effect is needed before a Chi-Square analysis will acknowledge it as being unlikely due to chance.

Let's try the survey measurements by interviewing the school zone residents before we put up the new signs. Here are some hypothetical results:

Table 10.1.3.1-2
Estimated Speed of Vehicles in School Zones
Before and After New Road Signs Were Put in O'Leary

SUBJECT	O'LEARY			MONTAGUE		
	SPEED*	SPEED*	SPEED CHANGE	SUBJECT	SPEED*	SPEED*
BEFORE	AFTER		BEFORE	AFTER	CHANGE	
AA	50	40	-10	LL	50	50
BB	50	50	0	MM	30	30
CC	40	30	-10	NN	70	60
DD	30	30	0	OO	70	70
EE	60	60	0	PP	30	30
FF	50	40	-10	QQ	40	50
GG	40	30	-10	RR	50	50
HH	50	50	0	SS	60	60
JJ	50	40	-10	-	-	-
KK	40	30	-10	-	-	-
MEAN	46	40	-6	50	50	0
SD			4.9			1.89

*Note: Speed is shown in km/hr

A t-test on the two sets of change in estimated speed, yields a value of "t" = 2.4, which indicates a significant difference. More data have overcome the variability of subjects differences and the less precise nature of these measurements compared to measurements of actual speed. This pre-post design could also have been used with the instrument

measures of vehicle speed, though it was needed to demonstrate a significant effect in this example. Still, for two reasons it worthwhile to consider "strengthening" that design too. 1) When you started the research, you could not know how much of a speed difference might be obtained or the variability of such measurements. Therefore, it is safer to go with the most powerful design you can obtain. If you find the results are clear you could also always stop data collection sooner and save your resources for another project. 2) Yet there is a more important reason. Statistical significance is not the only thing that convinces others of the value of your research. Though a significant difference in speed was obtained using the basic between-groups quasi design, a knowledgeable person would still ask whether those results may not simply have reflected that people in O'Leary generally drive somewhat more slowly - for reasons that have nothing to do with your new sign. Since psychologists do not have access to time travel technology (and never will), this example illustrates the importance of knowing how to strengthen research designs before starting a research project - which further explains why courses in research design are required for psychology students.

c) Problem

However there are still problems in interpreting the results even of pre-post designs. Lack of opportunity to train subjects and to control the test environment in real-world research leaves the results vulnerable to other explanations. The fewer violations measured over a six month period in O'Leary may simply have resulted from a temporary social effect (possibly a tragic accident shortly before the study began) or been due to the novelty of the new signs (any other new SCHOOL ZONE, SLOW sign whether better or worse might produce a similar effect). This possibility is illustrated by Figure 10.1.3.1-1.

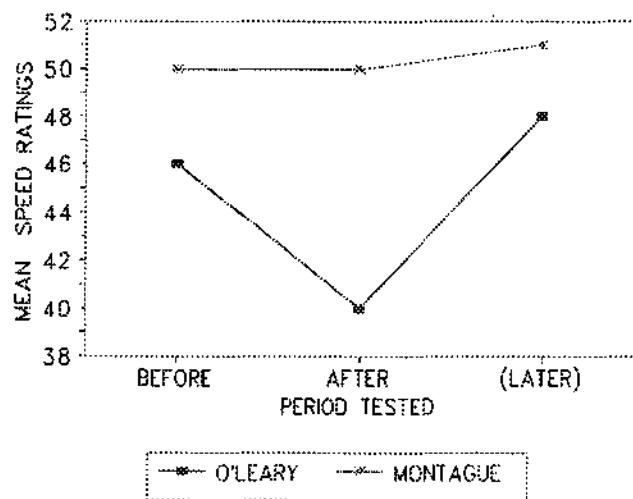


Figure 10.1.3.1-1. What might have happened had another set of measurements been taken somewhat later after the new sign was put in O'Leary.

It is also possible, but perhaps less obvious, that the speed reduction in O'Leary appeared because a construction project in O'Leary at the time of the "before" testing had temporarily produced a lot traffic. Had the testing been done prior to that construction project, the "improvement" after the new sign may not have been as impressive. This possibility is illustrated in Figure 10.1.3.1-2.

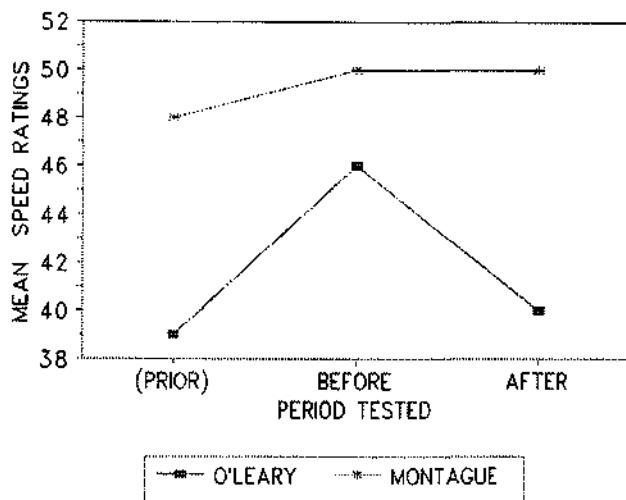


Figure 10.3.1-2. What might have been found had a still earlier set of measurements been made.

How might you convince your readers that neither of these possibilities occurred in your between-groups quasi-experiment? Read on.

10.1.3.2 USE MULTIPLE PRE-TESTS AND/OR POST-TESTS

a) Principle

More pre-treatment measures would help convince readers that the groups really were similar originally, or that the "baseline" data were correct. More post-treatment measurements would help assure that any effects of the treatment were not due to novelty or a random change. Taken together, multiple pre- and post treatment measurements can help distinguish a true effect of the independent variable from ongoing fluctuations, which should be anticipated in complex, real-world situations.

b) Schematic

	<u>DEPEND. VAR.</u> BEFORE			<u>INDEP. VAR.</u> LEVEL	<u>DEPEND. VAR.</u> AFTER		
GROUP 1:	M ₁	M ₂	M ₃	CONTROL	M ₄	M ₅	M ₆
GROUP 2:	M ₁	M ₂	M ₃	TREATMENT	M ₄	M ₅	M ₆

Such data can be analyzed using analysis of variance (ANOVA) followed by multiple means tests such as Tukey's or Duncan's tests. Combined with a graphic presentation, these can be quite convincing. Trend analysis techniques may wring out slightly more statistical power, but these are advanced techniques.

c) An Example

Hopefully, data from additional measurements will dispel doubts such as the possibilities illustrated in the previous two figures. If the BEFORE and AFTER

measurements truly represent the speed of drivers at each period, an even earlier measurement and a still later measurement should produce much the same results as the BEFORE and AFTER measurements. This ideal outcome from multiple before and after measurements is illustrated in Figure 10.1.3.2.

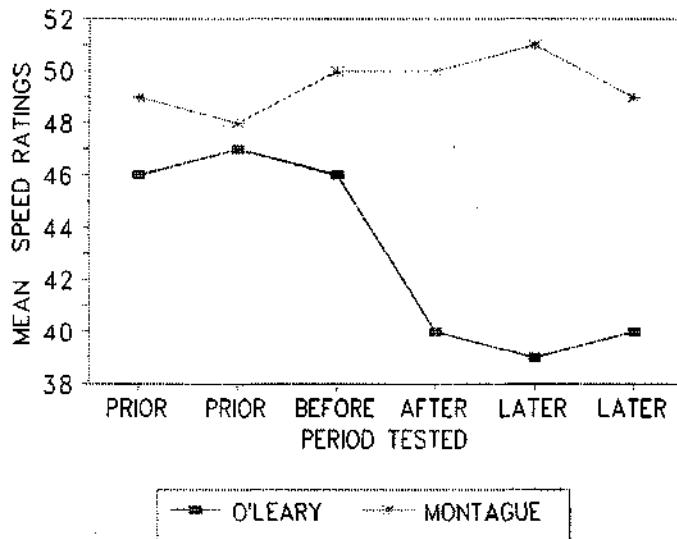


Figure 10.1.3.2. The results obtained with multiple measurements made at various times before and after a new traffic sign was installed in O'Leary.

10.1.3.3 REPEAT THE LEVELS OF THE INDEPENDENT VARIABLE

a) Principle

This technique can be considered when the "treatment" does not have a permanent effect or when it can be removed as in the case of a traffic sign. No matter how many pre- or post-tests are conducted, the possibility remains that any difference between the groups may have been a chance effect that occurred when the independent variable was present. To test this possibility after the effect of a treatment has been measured, remove the treatment and make another set of measurements. When these show no treatment effect, re-introduce the independent variable and repeat the measurements. Obtaining the same difference between the groups two or more times is definitely more convincing than just showing that it happened once.

b) Schematic

	<u>INDEP.</u> VAR. LEVEL	<u>DEP.</u> VAR. M #	TREATMENT REMOVED		<u>INDEP.</u> VAR. LEVEL	<u>DEP.</u> VAR. M #	TREAT REM'D		<u>INDEP.</u> VAR. LEVEL	<u>DEP.</u> VAR. M #
GROUP 1: CONTROL	M ₁		CONTROL	M ₂	CONTROL	M ₃	CONT.	M ₄	CONT	M ₅
GROUP 2: TREAT.	M ₁	(control)		M ₂	TREAT.	M ₃	(cont.)	M ₄	TREAT	M ₅

How does repeated application of the treatment differ from just doing the quasi-experiment again? Actually there is not much difference. If you were to repeat the study, you would probably use different subjects. However, sometimes you may lack the resources to recruit two more groups or it is otherwise not possible. The use of the same subjects is what keeps the repeated treatments enhancement a quasi-experiment. It may occur to you that when the treatment is repeated, it would be a good idea to provide the treatment to the other group instead of the same group. That is good reasoning. The only thing "wrong" with that is it would no longer be a quasi-experiment. Instead, you would have a properly counterbalanced within-subjects experiment with two groups. That is great when it can be done. However, this chapter is about doing research when experiments are not possible. That is the case when you are only able to provide the treatment to one group of subjects.

c) An Example

To illustrate how repeated introductions and removal of a treatment could work, we'll return once again to O'Leary and Montague. This time we obtained permission from the O'Leary town council to switch the traffic sign back to the old sign and switch again every six months. (Note that to conduct the design indicated in the above schematic with two removals and two re-introductions would take 2.5 years!) We shall also follow the basic format of this enhancement by **not** using the pre-test enhancement. simply to make it easier to compare the different enhancements. It is of course entirely possible to combine multiple introductions of the treatment with several pre and post treatment measurements.

10.2 WITHIN-SUBJECTS QUASI-EXPERIMENTS

In real-world research the independent variable often involves events that can't be controlled by the researcher. Since most quasi-experiments involve only two levels, a treatment and a no treatment or control condition, we'll restrict our discussion to these, though the principles hold as well for quasi experiments with more levels. When the same persons tested treatment and control conditions, you have a within-subjects design. However, when it is not possible to run replications using different sequences of the levels, any differences that are found may be due to order and carry-over effects rather than the independent variable.

To consider these problems, it helps to distinguish two basic types of "treatments" that occur in the real world:

- 1) ACCIDENTAL-TEMPORARY EVENTS In this type, some "special condition" occurs before an experiment can be planned. Consequently measurements can only be obtained during that condition and again after the situation has supposedly returned to normal. An example of this type would be research on the attitudes, social behavior, and health of persons who have experienced a natural disaster.
- 2) EVENTS WITH PERMANENT EFFECTS - A comparable problem is faced by researchers anticipating a "treatment" that will permanently affect the lives of the subjects. In this type of situation, measurements without the treatment can only be obtained beforehand. Once again replications to cope with order and carry-over effects are not possible. This type of problem would be faced by researchers wanting to study the effects of a hefty tax increase on the use of tobacco - nothing is more permanent than taxes, except death.

Quasi-experiments dealing with type of effect can be strengthened by using much the same techniques. Namely, obtain additional measurements which may convince readers that difference in results obtained at the various levels of the independent variable are not attributable to order or carry-over effects. Two techniques should be considered:

10.2.1 GET MORE PRE- OR POST-TREATMENT MEASUREMENTS

You can't repeat accidental independent variables, and you can't go back in time to get measurements before the "accident occurred, but you can make more measurements afterwards. Similarly, given enough warning about an independent variable that will produce a permanent change, you can take more measurements before the change occurs as well as afterwards. If this sounds a bit like the drunk who only looked near a lamp post for his/her lost keys, gave yourself an "A" for critical thinking.⁴ However, there is method to this apparent madness.

Accidental Studies

Due to carry-over effects of conditions preceding the accidental treatment, the effects of the treatment may not have developed when the first measurements are made. A difference between these and the second set of measurements may indicate an effect that is opposite to what really happened. If the effects of the treatment developed slowly, this would be revealed by subsequent measurements that gradually came to resemble the initial measurements.

If the effects of the treatment persist into the time when there is supposed to be no treatment, the second measurements may fail to show a difference when there was a difference. In this case additional subsequent measurements may reveal a change as the treatment effects diminish.

Order effects that arise simply from testing the same subjects more than once can produce the appearance of a difference when none existed.⁵ If an order effect was responsible for a measured difference, one could reasonably expect the additional measurements to also show order effects. If they don't, one can argue that order effects were not large enough to explain the observed difference.

Here is a schematic representation of an accidental within-subjects quasi-experiment that has been strengthen by additional post treatment measurements:

INDEPENDENT VARIABLE

ACCIDENTAL EVENT (treatment)	AFTERWARDS (control)	SUBSEQUENT MEASUREMENTS (additional data from control situation)		
M ₁	M ₂	M ₃	M ₄	M ₅

⁴ As the story goes, a police officer found a drunk crawling round and round a lamp post late one night. Asked what he was doing, the drunk explained, "Jus' looking for me keys, ossifer." "When did you loose your keys?" asked the officer. "A lill while ago when I was down the block." replied the drunk. "Then why are you looking here?" "Cause it's dark back there." replied the drunk.

⁵ It is also possible for an order effect to mask the effects of a treatment. While it is unlikely that they would actually cancel, the possibility should be kept in mind.

The subsequent control measurements accomplish two things: 1) They may convince the reader that the immediate control measurement M_2 was not due to chance. 2) They may reveal a difference from the treatment measurement M_1 , when the effect of the accidental event has worn off when the first control measurement M_2 was made. (Consider how you could take this approach to strengthen a quasi-experiment on how people react to a natural disaster like an ice-storm that cuts all electrical power for six days.)

Studies of Permanent Effects

Anticipation of a treatment may produce changes prior to the treatment itself. Consequently, there may be no difference in measurements made before and during a treatment. The effects of anticipation may be avoided by obtaining still earlier measurements prior to the treatment.

If the effects of the treatment develop more slowly than expected, no difference may have developed when the after-treatment measurements are made. In this case, additional measurements after the treatment has occurred would help show the effect.

As with accidental studies, any two sets of measurements from the same subjects could differ simply due to the subjects being tested twice. Additional pre-treatment or post treatment measurements can be used to reveal the magnitude of such an order effect. If such measured order effects are significantly smaller than the difference between the control and treatment conditions, one can argue that the effects of the treatment are real.

Here is a schematic representation of a within-subjects quasi-experiment on an event that had a permanent effect. It has been strengthened by additional post treatment measurements:

PERMANENT EFFECT					
ADDED PRIOR	BEFORE		AFTER		SUBSEQUENT
MEASUREMENTS	(more control data)	(control data)	(treatment data)		MEASUREMENTS
	M_1	M_2	M_3	M_4	M_5

An Island Example

Not so long ago a permanent change occurred to the geographic and social characteristics of Prince Edward Island. The federal government built a bridge from the mainland that replaced a forty minute ferry link. Prior to the decision to the build the bridge, many Islanders were concerned that it would alter the "Island Way of Life". Some feared it would reducing our sense of distinctness. Others looked forward to a better economy. In either case, both anticipatory and delayed reaction effects of the bridge could be hypothesized. A proper within-subjects quasi-experiment would have begun surveying Islanders when the proposal was first presented. Provision should also have been made to continue the study on a yearly basis until at least three years after the bridge was completed. Unfortunately, neither the provincial government nor most social scientists on the Island had any interest in evaluating this great social experiment.

10.2.3 TIME-SERIES DESIGNS

Of the many independent variables which give rise to the use of quasi-experimental designs, not all are accidental or permanent. Some can be controlled by the researcher, but they can not be restricted to some subjects and not others. Therefore all subjects are tested with the same sequence of levels. That restriction prevents the use of balancing or counterbalancing techniques to reduce the effects of order and carry-over effects. However, time-series designs can get around this problem to a large extent by repeated testing of the levels so that the effects of the sequence become "lost" or can be separated from the effects of the independent variable. As long as only a two levels involved, time series designs are not complicated.

The basic idea is simple enough. Instead of introducing each level just once to the subjects, the levels are presented several times in alternation. Here's a schematic a design for two levels consisting of a CONTROL and a TREATMENT:

CONTROL₁ TREATMENT₁ CONTROL₂ TREATMENT₂ CONTROL₃ TREATMENT₃

M_1 M_2 M_3 M_4 M_5 M_6

As you can see, by the time the 4th, 5th and 6th measurements are made, it is no longer evident whether the control level preceded or followed the treatment level. The multiple measurements of the sets of levels can also be considered as replications. Data that show a similar difference between the effect of the control and treatment every time they are changes are likely to be convincing. One factor to bear in mind with time-series designs is how soon after one level has been tested can you test another. This will vary greatly depending on the independent variable. Some insight into this question as well as further strengthening could be gained by combining time-series and multiple measurements designs as follows:

LEVELS OF INDEPENDENT VARIABLE

CONTROL₁	TREATMENT₁	CONTROL₂	TREATMENT₂	CONTROL₃	TREATMENT₃
↓	↓	↓	↓	↓	↓
M ₁ M ₂ M ₃	M ₄ M ₅ M ₆	M ₇ M ₈ M ₉	M ₁₀ M ₁₁ M ₁₂	M ₁₃ M ₁₄ M ₁₅	M ₁₆ M ₁₇ M ₁₈
TIME -->	MEASUREMENTS				

A HYPOTHETICAL EXAMPLE

There is growing recognition of the effects that ventilation have on human comfort and performance. The manager of the Rutherford Center was uncertain of how to set the ventilation rate of the building. While more ventilation is generally considered better, greater air flow is also accompanied by greater building noise due to low frequency turbulence and motor sounds. After some haphazard changes back and forth failed to reveal any consensus, a psychologist at the Center was asked for advice. The suggestion was to use a time-series design altering the conditions each week day over a two week period. Daily measurements of productivity and comfort would be made during a 3 o'clock coffee break.

Productivity was operationally defined in terms of how early or late people showed up for the break. Comfort was measured using the following 7-point rating scale.

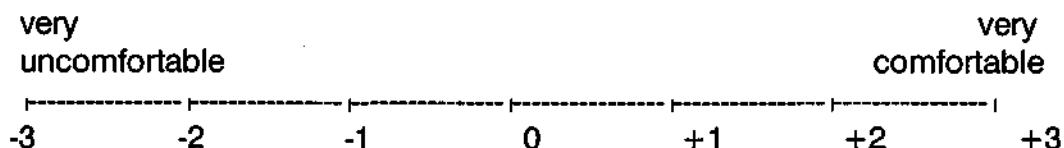


Figure 10.2.2 shows the hypothetical results.

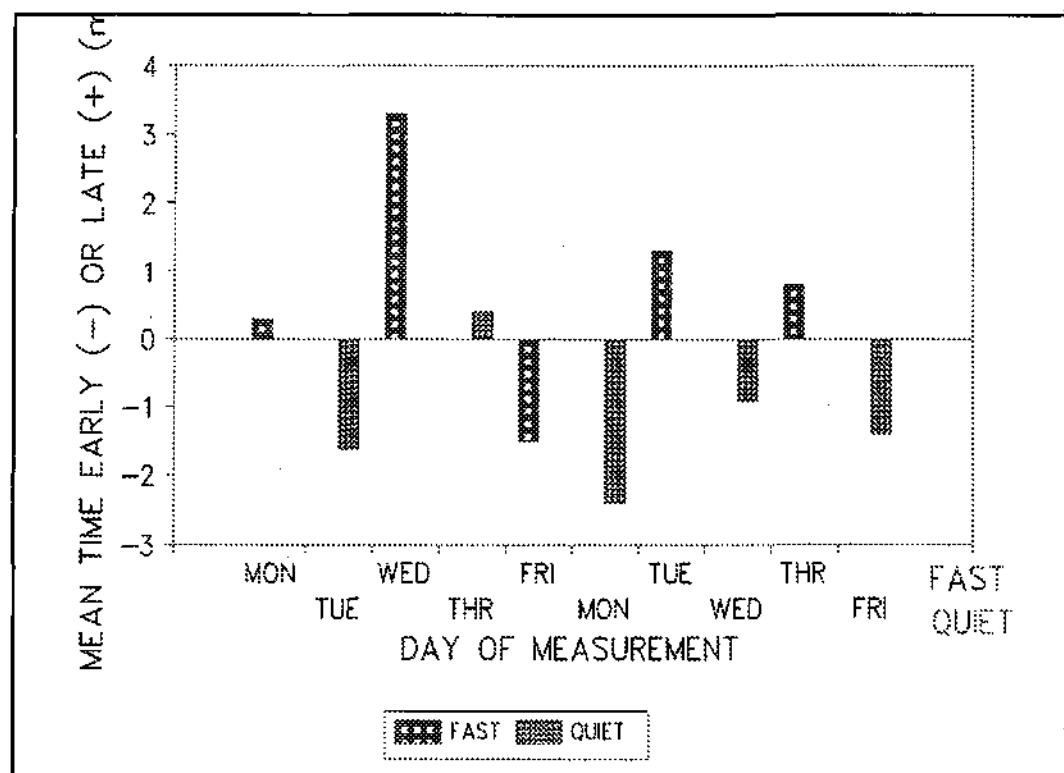


Figure 10.2.2. Mean number of minutes that workers were early (-) or late in coming for a mid-afternoon coffee break as a function of two types of ventilation.

Had these measurements been made only on Thursday and Friday of the first week, the conclusion would have been rather different from the overall trend in these data. In general, the faster ventilation was accompanied by workers showing up later for the break than when the ventilation was quieter. Transitions from faster to quieter were always accompanied by a greater delay. With exception, transitions from quieter to faster were accompanied by workers tending to take a break bit early.

10.2.3 TEST A SURROGATE GROUP

You may wear out your welcome by seeking lots of pre- and/or post-treatment measurements from the same subjects to strengthen a within-subjects quasi. There is an alternate that involves testing altogether different groups of subjects.

How It's Done

To use this technique in either an accidental or a permanent effects study, find a second group of subjects who will not experience the effects of the treatment being studied. Test these subjects in the same way and at the same times that the subjects in the main study are tested. What can be gained from the data of these surrogate subjects? 1) The data can provide an estimate of the size of the order effect that arises when these particular measurements are made twice with the same subjects. 2) One-time studies of accidental independent variables are particularly vulnerable to being confounded by the effects of other events that accompanied the accidental even being studied. Testing a second group of subjects at the same time does not avoid this threat, but the results can indicate whether a confounding variable could explain the results of the main study. One hopes that the two sets of surrogate measurements will not differ significantly or differ by much less than the difference between the control and treatment measurements.

The following schematic summarizes how the technique is used in an accidental study. A schematic for a studies of a permanent effect would be similar except that the no-treatment level would precede the treatment level.

INDEPENDENT VARIABLE			
	TREATMENT	NO TREATMENT	
MAIN STUDY:	M ₁	M ₂	
SURROGATE:	M ₃	M ₄	NO TREATMENT

An Example

Consider a study to find out whether a hefty cigarette tax reduces smoking. While the legislation is prepared, the health minister authorizes the study. A random sample of 400 persons is obtained from telephone listings. Three months before the tax is introduced an initial interview found that 25% of the sample smoked and their average was 48 cigarettes per week. Three months after the tax was levied, the same persons are interviewed again. The two sets of data show that after the tax 20 percent reported that they still smoked, and the average number of cigarettes decreased by 30 percent.

While these figures suggest that the tax works, the results can be questioned because they constitute a single set of measurements obtained from the same subjects. It can be reasonably argued that participating in the survey the first time, made the subjects think further about smoking and this led some subjects to reduce their smoking without regard to the tax. The first survey may also have made some subjects aware that smoking is less socially acceptable. Therefore, to look better on the second survey, they underestimate their smoking. It might also be reasonably argued that other events in the world at this time contributed to a decline in smoking that had nothing to do with taxes. In announcing the tax, the government and news media likely added information about the harmful effects of smoking. Furthermore, the government's action may well have been prodded by certain recent medical studies with convincing new evidence of the harm caused by smoking. Such information and not the tax may have sufficed to produce the observed reduction.

Anticipating these questions the researchers also conducted simultaneously a surrogate study with a random sample of 400 persons in a neighbouring province, which

had no immediate plans to introduce a similar tax. The initial interview found that there 30% of the sample smoked an average of 52 cigarettes per week. Six months later, the number of smokers in the neighbouring province was reduced to 27% and their average consumption was 54 cigarettes per week.

While either an order effect or increased awareness (one cannot tell which) may be responsible for a reduction in the number of persons who reported smoking, this effect is smaller than the reduction that accompanied the tax. The fact that the average number smoked increased slightly indicates that order effects or some other change in the world are not likely explanations for the reduction found in the main study. In this case the surrogate data helped answer doubts that were raised because the main study was not an experiment.

Back to the Island

The previous section had a short discussion of how a multiple pre-post design could have been used to study the social and economic impact of a bridge to Prince Edward Island. However over the eight year period of events leading up to the completion of the bridge and afterwards, many changes took place in Canada that had nothing to do with the bridge. A mid-90's depression came and went. The Eastern fishery was largely shut down. Quebec nearly separated from the rest of Canada. Electronic commerce burgeoned into being. All of these changes could have influenced a study that sought to evaluate the effects of the bridge. How could these changes be separated from any effects the bridge might have? The answer would have been to also use one or more surrogate groups, such as samples from New Brunswick and Newfoundland. One would then look for differences between the Island results and results from these other provinces.

10.3 GROUP CHARACTERISTICS AS INDEPENDENT VARIABLES

Pertaining to the Subject Groups

Particular differences between groups of subjects might be sought for the benefit of these groups: For example, measurements of how math problems are solved might reveal differences in how boys and girls think about numbers and help develop gender specific teaching approaches that would be more effective for both sexes.

Basic Research

Subject characteristics might also be considered as naturally occurring levels of a certain independent variable. Differences in handedness and gender, for example, are related to differences in brain organization. Accordingly, a comparison of cognitive, behavioral or physiological response between persons in these groups can provide insight into how the brain works. In this context, quasi-experiments might be done to learn about some basic aspect of brain function which might arise Measurements of field dependency using a rod-and-frame apparatus comparing fishermen with potato farmers might help answer questions of whether field dependency depends on vestibular experience.

10.3.1 CASE-CONTROL DESIGNS

These are like between-subjects quasi-experiments with pre-existing groups, but with an important difference: the subjects themselves form the levels of the independent variable. There is no "treatment". Such quasi-experiments are done to learn about more subtle differences between subjects which have been categorized on the basis of some evident differences such as: male/female, IQ level, introversion/extroversion, occupation. They are frequently used in clinical research to study persons suffering from some disorder by

comparing the data from these "cases" with data from "normal" people, who act as a control group.

Schematic

INDEP. VAR.	DEP. VAR.
GROUP 1: CASE	M_1
GROUP 2: CONTROL	M_2

Problems

Because subject characteristics can not be randomly assigned or assigned in some matched manner, other differences between subject groups could be responsible for any differences in the measurements. For example, let's assume that Groups 1 and 2 above represented boys and girls, and the dependent variable involved measurements of mathematical skill. Since the subjects have not been matched for confidence (or any of a number of characteristics that may arise as result of differences in how our society treats boys and girls) in their assignment to gender, differences in confidence might produce performance differences between boys and girls that have nothing to do with how they think about numbers. Similarly, consider the example above in which fisherman were compared with farmers as a means of learning about the effects of experiencing frequent vestibular stimulation and using vestibular information to remain standing. Since persons have not been block random assigned to being fisherman or farmers, it is possible that persons with less adequate vestibular functioning tend to become farmers rather than fishers. Consequently any difference that are found the measurements may not be attributable to experience, but rather are due to physiological differences in vestibular acuity.

Strengthening Case-Control Designs

A. Match the Subjects

In case-control studies, you usually have to take whatever case subjects are available. The control group, however, can often be individually selected by the researcher. Select subjects for the control group that are as similar as possible to the case subjects - except that they do not have the particular characteristics of that case. Typical factors include sex, age, ethnic background, education, income level, area of residence, occupation. Sometimes clever means can be found to equate subjects for factors such as institutional experience, comparable, but different, medical conditions.

B. Use Dependent Variables to Match Subjects Afterwards

Principle

Additional measurements can enable a sort of "post-hoc" matching of the subjects. Such data may enable you to demonstrate that the groups are similar with respect to other variables that could confound the results. In the case of subtype 1 above, a "confidence test" might establish that shyness in the testing situation was unlikely to be the cause of any differences in math performance.

Schematic

INDEP. VAR.	DEP. VAR.	SUPPLEMENTARY DEP. VAR. #1	SUPPLEMENTARY DEP. VAR. #2
GROUP 1:	M	M_1	M_2
GROUP 2:	M	M_1	M_2

Problems

When such similarity does not exist, the additional data nevertheless may provide a basis for some relative comparisons. For example, both fishermen and potato farmers could also be tested for ability to discriminate line orientation. If the fishermen did better than the farmers, the data on the rod-frame task could be rescaled in terms of orientation difference thresholds for each subject. If a difference on the rod-frame task is still significant, the difference could not be attributable to differences in discrimination.

More sophisticated analyses of data together with supplementary measurements calls for the application of analysis of covariance and for multiple regression analysis which are covered in advanced statistics and research design courses.

C. Go "Fishing"

If one dependent variable does not show a difference between the cases and the controls, another dependent variable might. This is often used in clinical studies where the researcher is seeking some factor which may explain the cases' abnormality. For example in seeking to understand why schizophrenia has developed, one might obtain from schizophrenic and normal persons a variety of measurements such as visual temporal acuity, auditory discrimination of complex patterns, spatial memory, the ability to organize structures as revealed by the ability to draw a bicycle, tests of basic logic, or the ability to recognize novelty as revealed by the P300 response in recordings of brain activity.

10.3.2 COHORT DESIGNS**10.3.2.1 Principles**

It is not possible to obtain "pre-treatment" measurements for subject characteristics like sex, IQ, and some personality traits because these inherited characteristics are present from birth. Yet other characteristics such as career achievement, life satisfaction, criminal behavior, and many pathologies develop later in life depending on a person's experiences. When the purpose of the research is to find experiential factors which cause such differences to develop, the best possible research method is to obtain measurements on several dependent variables before the particular characteristics develop and again later in life after the characteristics being studied have developed in many of the subjects. The expectation is that some of these measurements will have changed among persons that developed the characteristic but not changed among persons who did not develop the characteristic. Therefore, those changes indicate factors that may have contributed to the development of that characteristic.

Schematic

	DEP. VARS.	LATER THE INDEP. VAR. DEVELOPS	DEP. VARS.
INITIAL GROUP	M_1, M_2, \dots, M_X	GROUP 1	$M_{X+1}, M_{X+2}, \dots, M_{X+x}$
		GROUP 2	$M_{X+1}, M_{X+2}, \dots, M_{X+x}$

Problems

Cohort studies are more powerful than case-control studies, but that power is achieved at a substantially higher financial cost. The higher cost is due to the larger number of subjects that must be studied since you do not know which or how many of the subjects in the initial group will develop the particular characteristics that represent the independent variable. Because people move, die, or become unable or unwilling to continue their participation over a period of years, the initial sample must be increased further to make up for such loses. However, When the loses are substantial, selective survival becomes a potential confounding variable much the same as in archival and trace research.

What does it cost to contact persons in the general community and send a person to their home to conduct a one-hour interview? About \$100/subject in direct costs. Adding in the costs of for supervising a team of team interviewers, organizing the interview lists, preparing the interview materials, checking the data entries and producing a tabulation of the data tends to double the direct cost. Accordingly to do basic study for a cohort of 400 persons by interviewing once initially and once again after some period of time costs about \$120,000.

10.3.2.4 Strengthening Cohort Studies

There are two main ways to strengthen this design: 1) Increase the size of the initial group. 2) Make more measurements.

The Canadian Study of Health & Aging is a good example of a strong cohort study. It initially surveyed some 8000 healthy seniors over the age of 65 and surveyed them again after five years. The survey included some 100 questions about their life style, occupation and nutrition. During that period about 250 of those persons developed Alzheimer's disease. (Correcting for the number of seniors who died or were unable to be recontacted during that period led to an estimated incidence rate of a bit more than 1% per year.) The answers of the victims of this disease were compared with the answers of those who did not develop the disease.

10.4 COMPARING THE QUASI-EXPERIMENTAL METHODS

Consider the problem of trying to determine whether drinking tea increases the risk of developing Alzheimer's disease. (Oriental tea leaves contain relatively large amounts of aluminum, and traces of aluminum have been found in the neural plaques in brain tissue that accompany the development of the disease. By itself, this finding does not prove that aluminum is a cause, since the aluminum could simply accumulate as a consequence of the pathology that causes the plaques.) The only way to prove a causal connection would be a between-subjects experiment: Many seniors would be randomly assigned to drink tea or coffee over a period of four years. Then they would be tested for the presence of Alzheimer's disease. If significantly more subjects in the tea group developed the disease,

the cause would be proven. This experiment can not be done in the "real world" for two reasons: 1. It would be unethical to do such research if tea was a suspected cause. 2. People are generally not willing to be told what they can and can not drink.

Now consider how the above question could be addressed with a case-control quasi-experiment. Since you can not assign persons to drink tea or coffee, but you could compare persons who have chosen of their own free will to drink tea. You would contact persons who have been diagnosed with Alzheimer's disease and make a number measurements about factors in their life that might be related to the disease including how often they drank tea. Then contact a number of similar persons with no signs of the disease and obtain the same measurements from them. If tea drinking contributed to the disease, one would expect to find more tea drinkers amongst those who had Alzheimer's disease than amongst those who did not have the disease. However, the case-control study can not rule out alternative explanations for why persons with the disease drank more tea. Some plausible explanations include: 1) Tea drinking is more characteristic of persons with British ancestry while coffee drinking is more characteristic of persons from a European background. Genetic differences between the two groups may make persons of British descent more vulnerable, and tea drinking just happens to be a social tradition of the group with the genetic susceptibility. 2) In its nascent stages, Alzheimer's disease may produce changes in chemical sensitivity of the taste or olfactory systems so that tea acquires a more pleasant taste than coffee. 3) It is even possible that tea prolongs life the life of persons with Alzheimer's disease, there fore one would find more tea drinkers with the disease simply because they live longer.

10.6 TRICKS OF THE TRADE

I'll now let you in on a little secret. There is **no** rule which says that the techniques for strengthening quasi-experiments can only be used in quasi experiments! Think about it. The implications are staggering. Most of the various ways of doing research can be strengthen by 1) additional measurements before and after the effects of the independent variable have been tested; 2) testing another group of subjects at the same time without the presence of the independent variable; 3) presenting the various levels of the independent variable several times in succession; 4) including another independent variable with somewhat similar properties. Depending on the purpose of the research, its design, and the available resources, these techniques can be used in observational studies, surveys, some types of trace research, between- and within-subjects experiments. Together with combinations the basic methods, such as a between-subjects observational study using intervention supplemented by on-site trace measurements and followed up with a survey, the possibilities are only limited by your imagination.

EXERCISES

1. When New Brunswick was considering whether to require the use of seat belts in automobiles, they decided to do a study. A provisional law effective for two years was passed to enable determining whether this would reduce traffic fatalities. A Department of Transport research officer with a degree in psychology was assigned to compare the fatality statistics for those two years with fatality statistics from previous years when seat belts were not required. A t-test showed significantly fewer traffic fatalities during the 24 months when the legislation was in effect compared to 24 months prior to the legislation.
 - a. Why this study did not prove that seat belts reduce accidents?
 - b. What are some alternative explanations?
 - c. What means could the researcher have used to strengthen the study?
 - d. Originally the seat belt was announced as "temporary". When the research showed a significant reduction in fatalities, the regulation was made permanent. This prevented the use of a time-series analysis technique that was originally intended for evaluating the effects of requiring seat belts. Discuss the ethical issues involved in this legislative change.

EPILOGUE

Measurement and design - these are the basis of how psychologists do research. It is easy to foresee that new technologies will provide new ways for psychologists to make measurements. CAT, PET, and magnetic scan technologies are just beginning to be applied to probing the brain for functions of the mind. Yet do not let yourself think that all the rocket science stuff is just for physiological research. For example, use of the internet for survey research has begun, and other uses of this technology for social and personality and cognitive research will continue to be discovered. Nor must we forget that as night-vision, interferometric range microphones, and electropharmacological odor detectors present new opportunities for unobtrusive observations of behavior, these technologies also create a potential problems for ethical protection of privacy.

Nor should you think that research design is now a closed book. Quadratic counter-balancing for example, is still unknown to most researchers. Other ways to organize measurements and view data surely remain to be discovered, and these may be the greater challenge than measurement technology. That we are far from understanding what can be achieved by organizing information is clear when we study the how the mind works. We can make decisions about truth, beauty, or a good guess in the twinkling of an eye with no understanding yet of that is possible. As simple example of the power of the human mind over the best presently foreseeable technology, answer the following question. "What is the Prime Minister's telephone number?" What, you don't know! Without an exhaustive search of your entire memory, how do you know that you don't know? New uses of technology for psychological measurements and new research designs to unravel ever more complex behaviors await your discovery.

APPENDIX A USING THE LIBRARY

A.1 WHERE THE BOOKS ARE

At UPEI and most other university libraries across the world, you will find books organized the same way according to a topical code called the Library of Congress alphanumeric system. Adopted in the early 1900's, its alpha-numeric divisions provided a place for all areas of knowledge and literature. Since that time, some divisions such as those related to technology have expanded far beyond the expectations of that time, but the decimal aspect of the system has been able to accommodate these enlargements while retaining the overall code. Other areas of knowledge, such as phrenology, have not proven to be as important as was expected back then.

BF (The main location for psychology books)

30 - ANNUAL REVIEWS OF PSYCHOLOGY

(good for a relatively recent perspective and major references on various areas)

39 - STATISTICS

76 - RESEARCH METHODS

81 - HISTORY OF PSYCHOLOGY

121 - GENERAL & INTRODUCTORY

173 - FREUD

181 - EXPERIMENTATION

199 - BEHAVIORISM

201 - HUMAN EXPERIENCES

233 - SENSATION & PERCEPTION

295 - MOVEMENT & MOTOR SKILLS

309 - COGNITION & PROBLEM SOLVING

323 - ATTITUDES

335 - ADJUSTMENT

353 - ENVIRONMENTAL PSYCHOLOGY

371 - MEMORY

408 - CONSCIOUSNESS, CREATIVITY

431 - PSYCHOLOGICAL TESTING

455 - LINGUISTICS

481 - WORK, FATIGUE & ERGONOMICS

515 - EMOTION

575 - LOVE

610 - VOLITION & FREE WILL

637 - COUNSELLING

683 - MOTIVATION

697 - PERSONALITY

700 - DEVELOPMENT

789 - DEATH

891 - PHRENOLOGY, HANDWRITING

1023 - EXTRA SENSORY PERCEPTION

1078 - SLEEP & DREAMS

1148 - HYPNOSIS

1321 - TELEPATHY

LIBRARY A2

LB

1051 - LEARNING

1650 - READING

LC

4000 - LEARNING DISABILITIES

Q

100 - GENERAL SCIENCE

340 - ARTIFICIAL INTELLIGENCE

QA

76 - COMPUTER PROGRAMMING

QC

355 - OPTICS

495 - COLOR

QH

100 - BIOLOGY

400 - GENETICS

QL

790 - ETHOLOGY & ANIMAL BEHAVIOR

950 - EMBRYOLOGY

QM

100 - ANATOMY

QP (a major secondary location in the medical area for physiological topics)

33 - GROWTH AND AGING

121 - PHYSIOLOGY OF THE BODY

188 - ENDOCRINOLOGY

360 - PHYSIOLOGICAL PSYCHOLOGY

363 - NEUROPHYSIOLOGY

376 - THE BRAIN

406 - MEMORY

425 - SLEEP

431 - PAIN

456 - SMELL & TASTE

460 - HEARING

480 - VISION

RJ

500 - THERAPY & COUNSELLING

T

66 - OPERATIONS RESEARCH

TA

165 - SCIENTIFIC INSTRUMENTS

166 - ERGONOMICS

168 - SYSTEMS THEORY

TK

7867 - ELECTRONICS

7877 - MICROCOMPUTERS

Books on how to do psychological research:

- Anastasi, A. *Psychological Testing*. New York: Macmillan.
(BF 176.A5)
- Burros Institute (1994) *Tests in Print IV*. Lincoln, Nebraska: University of Nebrasks Press. (Z5814.E9T472) REF.
- Fairweather, G.W. & Davidson, W.S. (1986) *An Introduction to Community Experimentation*. New York: McGraw-Hill.
- Kling, J.W. & Riggs, L.R. (1971) *Woodworth & Schlosberg's Experimental Psychology*. New York: Holt, Rinehart & Winston.
- Sidowski, J.B. (1966) *Experimental Methods and Instrumentation in Psychology*. New York: McGraw-Hill.
- Weimer, J. (1995) *Research Techniques in Human Engineering*. Enhlewood Cliffs, New Jersey: Prentice Hall.

A.2 INTERLIBRARY LOANS

If you want a book that our library does not have, the library often can borrow it for you from another library. While this service eases the conscience of administrators when they slash library budgets, how can you find out about books that the library does not have? One way is from other books cited in your text book or research papers. The journals American Psychologist, and Contemporary Psychology contain reviews of recently published books. Various other journals also contain occasional book reviews and advertisements for new books.

A.3 REFERENCE SECTION

"If you cite it, you are honour bound to have read the complete work or else cite the abstract as well."

*PSYCHOLOGICAL ABSTRACTS
INDEX MEDICUS
BIOLOGICAL ABSTRACTS
SOCIOLOGICAL ABSTRACTS
*DISSERTATION ABSTRACTS (MICRO)
CANADA STATISTICS INDEX
DEBATES IN HOUSE OF COMMONS
DEBATES IN PEI LEGISLATURE

CURRENT CONTENTS - Issued bi-weekly. Contains the Table of Contents of selected journals, i.e. a listing of titles and authors, but no abstracts. There are specialty issues for biology, medicine, social sciences, etc.

CITATION INDICES - These are now available as an electronic reference. They are a very useful to find recent journal articles that pertain to a specific research topic. They consist of an alphabetical list of all authors of published research articles in major

LIBRARY A4

journals. Next to each author is a chronological list of his/her publications. For each publication is a list of published papers that have cited that publication. It works like this: A particularly important paper that has been published about a certain topic is likely to be cited by subsequent researchers doing similar research. Therefore, by knowing about a certain publication that is of interest to your work, you can quickly find the most recent research done on that topic.

(SCIENCE CITATION INDEX)
(SOCIAL SCIENCE CITATION INDEX)

Newspapers and popular magazines are generally not suitable sources of authoritative information, except to the extent that they present current events or reflect social attitudes. The following will direct you to magazine articles by title, subject, and sometimes author:

READER'S GUIDE TO PERIODICAL LITERATURE (WILSON'S INDEX)

The references area of the library also has many professional handbooks that are also organized by topic using the Library of Congress System:

The American Psychological Association's Diagnostic and Statistical Manual, IV (DSM-4) - a guide to various mental disorders.

Boff's Handbook of Human Perception and Performance - widely used by ergonomists and engineers to help them design machines and factories that are suitable for people to use.

The Merck Manual (RC-55.M4) - a guide to medical disorders.

The Physician's Desk Reference (RS-75.P5) - a guide to prescription drugs.

ENCYCLOPEDIAS are not acceptable authoritative references for scholarly papers at the university level. They may be used as sources of definition.

A.3 FINDING ABSTRACTS WITH COMPUTER

You can access these from certain computers in the Robertson Library. They can also be accessed from other computers around the campus including Psychology's Computer Lab if you are registered with UPEI Computer Services (main floor, south wing of the Vet College). Other students, Mr. Gray will help you get started.

These abstracts have the same contents as the printed abstracts listed above. However, the ability to search for several different combinations of topics or "key words" simultaneously makes computer abstracts easier to use.

PSYCH LIT
MEDLINE
NURSING ABSTRACTS
ERIC (related to educational issues)
SCIENCE INDEX
SOCIAL SCIENCE INDEX
CURRENT CONTENTS?

A.4 UPEI JOURNALS RELEVANT TO PSYCHOLOGY

GENERAL

American Psychologist (APA)
 Behavioral Science
 Canadian Journal of Behavioral Sciences
 Canadian J. of Experimental Psychology (CPA)
 Canadian Psychologist (CPA)
 [Current Directions - Dr. Boudreau (APS)]
 Journal of Experimental Psychology: General (APA)
 Journal of the History of the Behavioral Sciences
 (Nature)
 (Perspectives in Biology and Medicine)
 Psychological Bulletin (APA)
 Psychological Report
 Psychological Review (APA)
 [Psychological Science - Dr. Boudreau (APS)]
 (Science)
 (Scientific American)
 (Social Research)
 !American Journal of Psychology
 !Psychologia

() = not Psych. budget
 [] = privately available
 { } = donated
 ! = recent drop, back issues
 APA = Amer. Psychol. Assoc.
 APS = American Psychol. Soc.
 CPA = Can. Psychol. Assoc.
 PS = Psychonomic Society

REFERENCE & POPULAR

Contemporary Psychology (APA)
 (Current Contents - Social & Behavioral Sciences; Life Science)
 (Discover)
 (Newscientist)
 Psychology Today
 (Science News)

ANIMAL LEARNING & BEHAVIOR

(Animal Behavior)
 (Behaviour)
 Animal Learning & Behavior (PS)
 Journal of Experimental Psychology: Animal Behavior Processes (APA)*
 Learning and Motivation
 !Journal of Comparative Psychology (APA)
 !Journal of Experimental Analysis of Behavior

COGNITIVE

[British Journal of Experimental and Clinical Hypnosis - Dr. St. Jean]
 Canadian Journal of Psychology (CPA)
 Cognition
 (Cognitive Brain Research)
 Cognitive Psychology
 International Journal of Clinical and Experimental Hypnosis

LIBRARY A6

Journal of Experimental Psychology: Human Learning and Memory (APA)
Language
Learning and Motivation
Memory and Cognition (PS)
(Reading Psychology)

CLINICAL & COUNSELLING

Archives of General Psychiatry
Canada's Mental Health
Canadian Journal of Behavioral Science (CPA)
Child Abuse & Neglect
Educational and Psychological Measurement
(European Journal of Child and Adolescent Psychiatry)
(Geriatric Nursing)
Journal of Abnormal Psychology
Journal of the American Medical Association
Journal of Child Psychology and Psychiatry
Journal of Clinical Psychology
Journal of Counselling Psychology (APA)
Journal of Consulting and Clinical Psychology (APA)
(Journal of Learning Disabilities)
Journal of Psychology and Theology
(Journal of Speech and Hearing Disorders)*
(Learning Disabilities Research)
Mental Retardation
Professional Psychology
Psychoanalytic Review
Psychotherapy (APA)
{Violence and Victims - Dr. Smith}
!Behavior Research and Therapy
!International Journal of Clinical and Experimental Hypnosis
!Journal of Applied Behavioral Science
!Journal of Nervous and Mental Disease

DEVELOPMENTAL

(American Education Research Journal)
(American Geriatrics Society Journal)
Adolescence
British Journal of Educational Psychology
Canadian Journal of Behavioral Science (APA)
Child Development
Child Development Abstracts
Child Development Monographs
Exceptional Children
(Developmental Brain Research)
Developmental Psychology
Gifted Children Quarterly
[Infant Behavior & Development - P. Boudreau]
Journal of Educational Psychology (APA)

LIBRARY A8

(Brain Research Reviews)
Canadian Medical Association Journal
Cognitive Brain Research
(Computers in Biology and Medicine)
(Developmental Brain Research)
(Discussions in Neuroscience)
(Experimental Neurology)
(Hormones and Behavior)
(Journal of the American Medical Association)
(Journal of Comparative Physiology - A: Sensory, Neural & Behavioral)
(Journal of Neuroscience)
(Journal of Neuroscience Research)
(Life Sciences)
(Neuropharmacology)
(Neuroscience Letters)
(New England Journal of Medicine)
(Physiology and Behavior)
(Physiological Reviews)
Psychobiology (PS)
(Trends in Neuroscience)
!(Pharmacology, Biochemistry and Behavior)

RESEARCH METHODS

[Behavior Research Methods, Instrumentation & Computers (PS) - Dr. Nilsson]
(Biometrics)
!(Byte)
(Computers in Biology and Medicine)
(Journal of Neuroscience Methods)
Psychological Bulletin (APA)

SOCIAL

Behavioral Science
Canadian Journal of Behavioral Science (CPA)
Changing Men
Journal of Applied Social Psychology
Journal of Experimental Social Psychology
Journal of Personality and Social Psychology (APA)
Journal of Psychology & Human Sexuality
Journal of Social Issues
Masculinities
Personality and Social Psychology Bulletin
Representative Research in Social Psychology
Sex Roles
Social Psychology Quarterly
!Small Group Behavior

OTHER PSYCHOLOGY RELATED JOURNALS AT THE VETERANS AFFAIRS LIBRARY - Mackenzie Building on Grafton St. Ask for directions at reception desk.

Aging
Behavioral Sciences Newsletter
Canadian Journal on Aging
Canadian Journal of Psychiatry
Canadian Occupational Safety
Clinical Gerontologist
Journal of Applied Psychology
Journal of Counselling and Development
Office Equipment and Methods
Personnel Journal
Psychology of Aging
Research on Aging
Scientific American Medicine
Social Casework
Social Work
The Gerontologist
Training and Development Journal

OTHER PSYCHOLOGY RELATED JOURNALS AT THE FOOD TECHNOLOGY CENTRE LIBRARY - Ask for directions at reception desk.

Chemical Senses
some journals related to nutrition

A.5 OBTAINING COPIES OF ARTICLES NOT AVAILABLE AT UPEI

All the technology needed to provide complete journal articles directly by electronic mail or the World-Wide Web are in place. Unfortunately, our society still lacks the social means and will to fully utilize this technology. Though authors gain no income from the journal articles they write or review, publishers want to maximize their profits. They are using every legal means to prevent the distribution of their publications other than by subscription sales. It may well take another decade to solve this major impediment to human progress.

A.5.1 FROM THE AUTHOR

Most journal publishers provide authors with a number of copies of a paper they are publishing - either free or for a modest cost per 100. Authors are usually glad to send these copies to anyone, including students, who requests them. This is another means for authors to spread word about their research and make contact with other researchers interested in the same topic. While free, there some unreliability: The author may be away, have moved, be too busy to answer your request promptly, or even be out of copies. If you need a copy for an assignment, get to work promptly on the assignment and **write early**.

A.5.2 THROUGH INTER-LIBRARY REQUESTS

Smaller universities can no longer afford to purchase many of the journals needed to support their teaching and research mission. Even large universities do not subscribe to all scientific journals that may be needed. While individual books can usually be borrowed, most libraries will not send recent or bound volumes of journals to provide a

LIBRARY A10

single article. Instead, a system has evolved whereby larger libraries will copy single articles and send them to the requesting library. This service is provided "at cost" and amounts to about \$XX per page. While such requests are sent electronically by your library, the photocopies are usually sent by post. Allow about 2-3 weeks for delivery.

A.5.3 PUBLICATION CLEARING HOUSES

These will send by fax copies of research articles within two days. The cost is about \$XX per page - depending on the journal. Enquire at the Information Desk in the library.

HOW TO BEGIN USING SPREADSHEET PROGRAMS

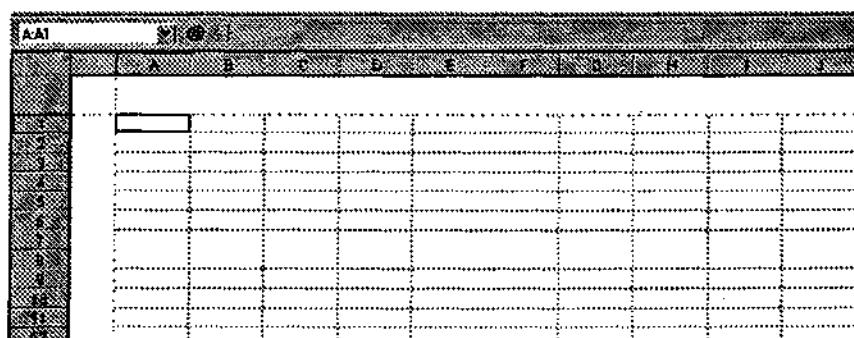
- with Yvonne Haskin

A spreadsheet is like a large blackboard marked out as boxes in rows and columns. You can use a spreadsheet to summarize, plot a graph, and do a statistical analysis on any set of data that can be listed as numbers in a table. To use the spreadsheet, you list your data in successive boxes and enter some calculations in other boxes. These calculation boxes do the number work for you. A significant aspect of spreadsheets is that they enable you to copy a calculation from a single location to many other locations in one step. Thus you may only need to enter the formula for a particular calculation once to work with hundreds of sets of numbers.

Locations

You keep track of the boxes by the number of their row and the letter of their column, which together are each box's address. All boxes are created equal. You can put either labels, data numbers, or calculations into any box. Despite their small size on the screen, boxes can hold several lines of words and lengthy calculations, but only one number.

The Notebook page



Columns are identified by letters at the top of each column

Rows are identified by numbers at the left hand side of each row.

Cells are the little white boxes in the notebook page (where you enter data)

Active cell is the cell ready for data; it is identified by a black border around the cell [Top Left (A:A1) in this example]

Cell Address is the notebook page name followed by the column letter and row number (A:A1)...you can tell what address you are on by looking at the 'pushing in' button along the columns and rows.

Input Line shows the data that is in the active cell; you can enter data here or directly into the active cell. This is the long rectangle shown above the column letters.

Menu Bar



The menu bar is where the majority of functions can be found. When you click your mouse on any item in the menu, you will be given a list of what functions that item can do. The Quickfunction icons are the commands that are used most frequently.

File: Opening, Saving, Closing, Printing

Edit: Cut, Copy, Paste, Undo, Convert to Variables, Move sheets, etc

View: Toolbars, formulas, etc

Insert: Inserting rows, columns, sheets, charts, graphics, etc

Format: Sets attributes of your document

Tools: Spellchecker, numeric tools, etc

Window: Change or eliminate window icons

Help: Help topics, and Help topic search engine to guide you through QuattroPro functions

Moving Around

Starting from scratch you will find an empty spreadsheet, with a flashing cursor in the upper left corner at address A1.

- a) **Scroll bar-** (far right of screen) allows you to move up and down in document
- b) **Edit-Go to-** (top left of screen) allows you to specify a cell to move to
- c) **Home key-** (upper right on keyboard) moves to beginning of the notebook page
- d) **Page Up Key-** (upper right on keyboard) moves up one screen
- e) **Page Down key-** (upper right on keyboard) moves down one screen
- f) **Arrow Up Key-** (lower right on keyboard) moves up one cell
- g) **Arrow Down Key-** (lower right on keyboard) moves down one cell
- h) **Arrow Right Key-** (lower right on keyboard) moves right one cell
- i) **Arrow Left Key-** (lower right on keyboard) moves left one cell
- j) **Ctrl Arrow Right Key-** moves to the right one page
- k) **Ctrl Arrow Left Key-** moves to the left one page
- l) **Page Tabs-** allow you to move from page to page



Number Work

Only type one data value into each box location- whole numbers, positive or negative numbers, and decimal numbers. As you type they appear in the input line on the top of the screen. When the number is typed, use the arrow key to move the cursor or press ENTER or point and click mouse in another cell. This is how data is entered.

Now you try it.....

In cell A1, type the number 4 and enter the data by one of the methods described above.

Numerical operations: add "+", subtract "-", multiply "*", divide "/", and powers and roots "^" are available right from the keyboard. To square 9, enter 9^2 . Taking the square root of 9 would be entered as $9^.5$ as decimal powers are roots, or you may also use the @SQRT function. The use of brackets "()" is recommended to ensure that the numerical operations are performed in the order you want.

Now lets try some basic mathematical functions.....

Point your mouse or use the arrow keys and go to cell A1 where you entered the number "4". When you reach the cell, the cell becomes active. Type in $4+2$ in cell, or edit the formula in the Input Line at the top of the screen. Enter the data by using arrow keys, mouse or ENTER key. Cell A1 should now have the number 6. Congratulations!!! You have just entered your first spreadsheet formula!!!! If you move your cursor back to A1, you will notice that the number 6 is in the cell, and the formula you used is shown in the Input Line. Now lets add to this formula. I want to subtract 0.5, then multiply this total by 5. Let's do this step by step.

- 1) Click your mouse arrow on the Input line to edit the formula.
- 2) Add brackets and subtraction to the formula. Formula should now look like this:

$$(+4+2-0.5)*5$$

The brackets tell the computer what formula to do first. If you do not use brackets, you will end up with a different answer. If you did this formula correctly, cell A1 will be 27.5 once you press ENTER, or move cursor.

Now lets make formula A1 more complex. Lets divide it's total by 2, Square it, then find the Square root. Let's give it a try....

- 1) Click your mouse arrow at cell address A2.
- 2) Type in formula: $(a1/2)$ This tells the computer that you want to take the total in cell A1 and divide it by 2. Enter the data. Cell A2 should now be 13.75.
- 3) Click your mouse arrow at cell address A3.
- 4) Type in formula: $(a2^2)$ This tells the computer that you want to take the total in cell A2 and square it. Enter the data. Cell A3 should now be 189.0625.
- 5) Now click to cell address A4 and type in formula: $(A3^.5)$ to get the square root. After entering the data, cell A4 should be 13.75.

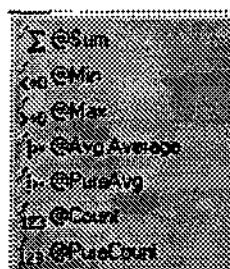
How to Create a Formula-Descriptive Statistics

You already know formulas can be basic math, but more complicated formulas can also be used. Quickfunction key on the function bar will allow you to save your formula for another time.



The spreadsheet program is user-friendly for all types of functions. The Quickfunction key: 

gives all kinds of quick calculations that are used frequently. Click on the arrow beside the sigma to see what other symbols are ready for use.....



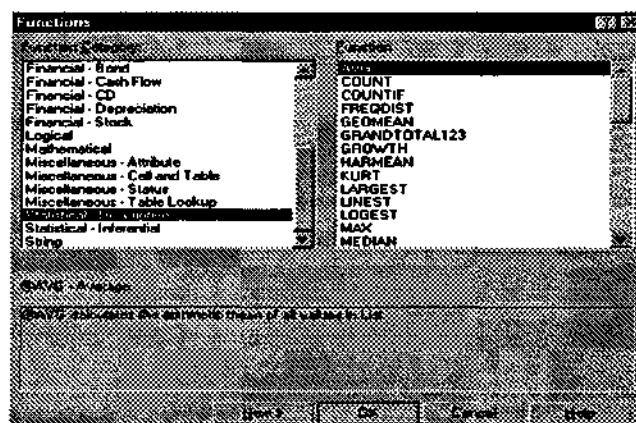
- (@Sum) click this directly under numbers you wish to add...and it does it for you
- (@Min) click this directly under data to get the smallest value
- (@Max) click this directly under data to get the largest value
- (@Avg) click this to get the mean
- (@PureAvg) click this to get average while ignoring labels, and blank cells
- (@Count) click this to get how many entries have been made in a column (good for calculating N)
- (@PureCount) click this to get number of entries while ignoring labels, and blank cells.

.....These are great tools, but what about other descriptive statistics?????

Quickfunction button

will give you an opportunity to see what type of formulas are available, including an entire section on descriptive statistics, which is most useful to the Psychology student.

This is the 'Pop up box' that appears when you click the Quickfunction button: 



You will likely be most interested in the category 'Statistical-Descriptive'. After you have chosen this category, a list of descriptive statistical functions appears on the right hand side.

Copy, Cut, Paste

These options are under the Edit menu at the top of the computer screen or you can use the Quickfunction buttons:

-  CUT...This will cut out any highlighted material
-  COPY...This will copy any highlighted material
-  PASTE...This will paste copied material on current active cell

You can also Right click on the mouse to get the 'cut', 'copy', and 'paste' functions. By highlighting data and choosing copy, the computer remembers this information. You can then choose cut if you want to remove the data from its location. You can then click on the area you wish the data to be instead, and choose paste.

****The computer only remembers that last thing you copied, so only do this to one set of data at a time.

When you copy boxes that contain formulas but display the results of that calculation, *you will copy only the formula-not the result*. The program automatically adjusts the address in the formula into new address relative to the new location.

Copying Numbers-Absolute Address

Convert to Values (under Edit in menu)

This allows you to convert your formulas to values. If you would like to copy results, not the formulas. Just convert the result to a value, then copy and paste it elsewhere.

Other Quickfunction buttons you may find useful:

- Create a Text Box...Makes a box that is good for text/explanations
- Sort function....Sorts highlighted data in ascending or descending order
- QuickFill function....Fills in pattern set. For example if you type JAN and highlight where you want the months and press this function button, it will fill in the rest of the months for you. This works the same with patterns of numbers
- Speedformat....Places your work into attractive tables that are preformatted
- JoinCells function...This will join a group of highlighted cells into one cell
- PerfectExpert button....This gives you step by step options to create a custom spreadsheet
- Change Text Colour button...This gives you colourful options for your text.
- Symbols Button...gives you assorted symbol options, including mathematical symbols

Using the quickfunction buttons are not necessary, as any function can be found on the menu bar, and any formula can be typed into the active cell or the Input line. Basic formulas are provided: sum is @SUM, mean is @AVG, standard deviation is @STD. These usually need to be followed by brackets that contain the range of addresses to which you want the formula to apply. For example, @avg (b3..b11) will produce the mean value of the nine numbers in locations B3, B4, B5,...B11.

Correlations can be done by calculating Z-scores and finding the mean value of their cross-products. There is also a built-in formula.

- 1) Go to Tools on the Menu Bar and Select 'Numeric Tools'
- 2) Select Advanced Regression and Press 'NEXT'
- 3) Enter dependent and independent variables....Y Range= cells containing Y values (dependent variables)...X Range= cells containing X values (independent variables)...Summary, Residual, Probability= the upper left cells where you want these tables to appear.
- 4) Select 'FINISH'

Graphs

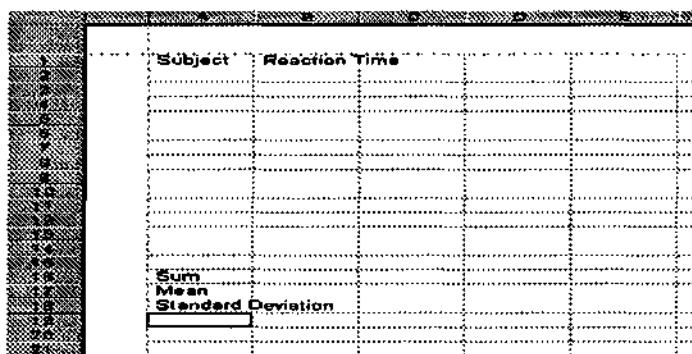
To make a graph or chart, you can highlight your data and press:

 Quickchart button

Right click mouse to get more options, clicking inside chart gives you even more options, and double clicking mouse inside chart gives you some neat pattern ideas. You can cut and paste chart to another sheet if you wish.

Another way to make a graph is to go to the Insert Menu and choose Chart. Enter your chart data and press NEXT. Choose your chart type and press NEXT. Now you can give your chart a title, subtitle, x-axis, y-axis and press NEXT. You then choose your colour scheme and press FINISH. You are then back to the spreadsheet. Find the location you wish to place your chart/graph and click mouse. Your graph will appear. You can manipulate the size by going to the edges of the chart where arrows appear. You can then click and drag the chart to the size you want.

Labels



The screenshot shows a Microsoft Excel spreadsheet. On the left, there is a vertical column of numbers from 1 to 10. To its right is a table with two columns: 'Subject' and 'Reaction Time'. The first few rows of the table are empty. Below the table, there is a summary row with three cells labeled 'Sum', 'Mean', and 'Standard Deviation'. The 'Sum' cell contains the value 10.0, the 'Mean' cell contains 1.0, and the 'Standard Deviation' cell contains 0.866.

Start by making some labels for the rows and columns to keep track of your data:

1. Using the arrow keys (or mouse) move the cursor to a suitable location and just type in a label like "Reaction Time". What you are typing goes into the Input line. (You can use whole sentences as 'labels'-the program won't know the difference. However, only the first word may show up on the spreadsheet, depending on the width of the column. The rest of the sentence you can only see by moving the cursor to that location and reading it in the special row above the spreadsheet).

2. When the label is typed, move the cursor to the next location, and the word appears in the row-column box where the cursor was.
3. If your label has both letters and numbers, the program may confuse it with an incorrect command or location, and you'll get an error message. Hit RETURN, which puts you back in control. You need to modify this type of label by putting a quotation mark in front of this sort of label to prevent confusion. You could simply retype the label preceded by a quotation mark.

Manipulating Rows or Columns

1) Inserting Rows or Columns

Click on row or column just after where you want the row or column to be inserted. Then

- click on the insert icon:
- Choose row or column from the pop up box.

2) Deleting Rows or Columns

Click on row or column that you want to delete. Then click on the delete icon:



3) Changing Column Widths

Click on any cell in the column that you want to change the width of, then click the width button:



.... and the column will automatically adjust to the size of the widest entry in the column...OR....place your cursor on the right edge of the column letter you want to adjust. When the cursor becomes a double headed arrow, you can then click and drag the double headed arrow until the column is the size you want.

Adding, Deleting and Naming Pages

*To insert or delete a page, you can do this by Right clicking the sheet tabs and then choose either 'insert' or 'delete'.

OR

To insert a new page click on Insert and then Sheet- this inserts a new page before the page you are working on.

To delete a page click on Edit and then delete, then choose sheet from the pop up options box.

*To name a page in your spreadsheet, double click on the sheet tab and type in the label you want for the page.

Page Setup

Under the File Menu....This is where you will be able **
to choose your page type, header/footers, print margins, scaling, etc. The File menu is also where

you can open, close, save and print documents.

Saving the Spreadsheet

Go to File on the Menu bar and select Save As. You then type in the name you wish to save it under and where (A drive if you have a disk, C drive if you want it on the hard drive; M drive if you want it saved on your student account). If you have already saved your material and named it, yet want to save it throughout your time working, choose the Save option. If you have not saved the file under a name and you choose Save, it will automatically go to Save As anyways.

The Quickfunction button for saving is the button in the shape of a computer disk. It works the same as going through the File Menu.

If your mouse freezes, or if you prefer to use the keyboard, press Ctrl 's' to save your material.

Printing

Go to File on the Menu bar and select Print Preview, or go to the Quick function button on the far right that looks like a blank piece of paper. Here you can manipulate how you want your spreadsheet to print (margins, headers etc). When the spreadsheet looks the way you would like it, click on the button that looks like a printer, and your spreadsheet should print.

**CANADIAN PSYCHOLOGICAL ASSOCIATION'S
CODE OF ETHICS**
(an unofficial, abridged version for research students)

PRINCIPLE I: RESPECT FOR THE DIGNITY OF PERSONS

Psychologists accept as fundamental the principle of respect for the dignity of persons, that is, the belief that each person should be treated as a person or an end in him/herself, not as an object or a means to an end. In so doing, psychologists acknowledge that all persons have a right to have their innate worth as human beings appreciated and that this worth is not enhanced or reduced by such differences as culture, ethnicity, colour, race, religion, gender, marital status, sexual preference, physical or mental abilities, age, socio-economic status, and/or any other preference or personal characteristic, condition or status.

Although psychologists have a responsibility to respect the dignity of all persons with whom they come in contact in their role as psychologists, the nature of their contract with society demands that their greatest responsibility normally be to those persons directly receiving or involved in the psychologist's activities and, therefore, in a more vulnerable position (e.g., clients, students, research participants). This responsibility is almost always greater than their responsibility to those indirectly involved (e.g., employers, third party payers, the general public).

Adherence to the concept of moral rights is an essential component of respect for the dignity of persons. Rights to privacy, self-determination, and autonomy are of particular importance to psychologists who have a responsibility to protect and promote these rights in their service, research, and teaching activities. As such, psychologists have a responsibility to provide services and develop procedures for informed consent and confidentiality that are consistent with those rights.

- | | |
|----------|---|
| General | <ol style="list-style-type: none">1. Demonstrate appropriate respect for the knowledge, insight, experience and areas of expertise of those persons with whom they come in contact in their role as psychologists.7. Respect the right of research participants, and students to safeguard their own dignity.8. Not practice, condone, facilitate or collaborate with any form of unjust discrimination. |
| Informed | <ol style="list-style-type: none">13. Obtain informed consent for all research activities which involve Consent obtrusive measures, invasion into the and private lives of research participants, risks to the participant, or any attempt to change the behaviour of research participants.16. In obtaining informed consent, provide as much information as a reasonable or prudent person would want to know before making a decision or consenting to participate in research. This information must be relayed in language which the person understands and take whatever reasonable steps are necessary to assure that the information was, in fact, understood.19. Take all reasonable steps to ensure that consent is not given under conditions of coercion.21. Respect the right of the research participant to discontinue at any time. |

ETHICS C2

- Vulnerability 24. If fully informed consent cannot be obtained due to age or serious handicap, obtain consent from persons who are legally appointed to give informed consent on behalf of the individual concerned.
25. Seek the willing participation of any child or other persons of diminished capacity to give informed consent, and proceed without this assent only if the research activity is of direct benefit to that person.
26. Be particularly cautious in establishing freedom of consent of a research participant who is in a dependent relationship (eg. student employee). This may include offering alternative activities to fulfil their educational or employment goals.
- Privacy 27. Explore and collect only information which is germane to the purpose for which consent has been obtained.
28. Store and handle all records, both written and unwritten (e.g. video-tapes), in a way that attends to the needs for privacy and security.
29. Take all reasonable steps to ensure that records which they have control remain personally identifiable only as long as is necessary in the interests of the research project for which they were collected, and rendered anonymous or destroy any records they control that no longer need to be personally identified.
- Confidentiality 30. Be careful not to relay information which they have gained about colleagues, students, and research participants gained in the process of their activities as psychologists and which is considered confidential by those persons.
- Extended Responsibility 31. Assume overall responsibility for the professional activities of their assistants, students; supervises employees with regard to ethical practice, all of whom, however incur similar obligations.

PRINCIPLE II: RESPONSIBLE CARING

One of the most basic ethical expectations of any profession in our society is that its activities benefit members of society or at least, do no harm. Therefore, ethical conduct by psychologists is characterized by an active concern for the welfare of any individual, family or group with whom they come into relationship in their role as psychologists. This concern includes both those directly and those indirectly involved in their research and teaching activities.

As individuals are the most likely persons to be concerned about their own welfare, obtaining informed consent (see Principle I) is one of the best methods for ensuring that their welfare will be protected. However, it is only when informed consent is combined with the responsible caring of the psychologist that there is considerable ethical protection of the welfare of the person(s) involved.

Psychologists define harm and benefit in terms of both the physical and psychological dimensions. They are concerned about such factors as feelings of self-worth, fear, humiliation, interpersonal trust, cynicism, and both self-knowledge and general knowledge, as well as such factors as physical safety, comfort, pain, and injury. They are concerned about immediate, short-term, and long-term effects.

Psychologists' use and treatment of animals in their research and teaching activities are also a component of responsible caring. Although animals do not have rights in the same way as persons have rights, psychologists consider it their responsibility to treat them humanely and to not expose them to unnecessary discomfort or pain.

- | | |
|---|--|
| General | <ol style="list-style-type: none"> 1. Protect and promote the welfare of clients, students, research participants, colleagues and others with whom they come in contact in their role as psychologists. 2. Avoid doing harm to clients, students, research participants, colleagues and others with whom they come in contact in their role as psychologists. |
| Risk/
Benefit
Analysis | <ol style="list-style-type: none"> 11. Be sufficiently sensitive to vulnerabilities and individual differences to discern what will benefit and not harm persons involved in their research and teaching activities. 12. Carry out pilot studies to determine the effects of all new procedures and techniques which might carry some risks, before considering their use on a broader scale. 13. Before making a decision to proceed, seek an independent and adequate ethical review of the balance of risks and potential benefits of all research which involves procedures of unknown consequence, or where pain, discomfort and harm are possible. |
| Maximize
Benefit
Minimize
Harm | <ol style="list-style-type: none"> 19. Debrief research participants so that their knowledge is enhanced and they have a sense of contribution to the enhancement of knowledge. 25. Be careful not to put incidentally involved individuals at risk. 31. Minimize the impact of their research on the participants' physical and mental integrity and personality. |
| Offset/
Correct
Harm | <ol style="list-style-type: none"> 36. Not place an individual, group, or family needing service at a serious disadvantage by offering them no service over an unreasonable period of time in order to fulfil the conditions of a control condition in a research study, and, where resources allow, would offer such person(s) the service found to be most effective after the research study is completed. 37. Debrief research participants in such a way that any harm caused can be discerned, and act to correct any resultant harm. |
| Care of
Animals | <ol style="list-style-type: none"> 38. Not use animals in their research unless there is reasonable expectation that the research will increase understanding of the structures and processes underlying behaviour, or increase understanding of that animal species, or result eventually in benefits to the health and welfare of humans or other animals. 39. Use a procedure subjecting animals to pain, stress, or privation only when an alternative procedure is unavailable and the goal is justified by its prospective scientific, educational or applied value. 40. Make every effort to minimize the discomfort, illness, and pain of animals. This would include performing surgical procedures only under appropriate anaesthesia, using techniques to avoid infection and minimize pain during and after surgery, and if disposing of experimental animals is carried out at the termination of the study, doing so in a humane way. <p>() Avoid using a more sentient species when a less sentient species is adequate. (a suggested addition)</p> |

PRINCIPLE III: INTEGRITY IN RELATIONSHIPS

The relationships formed by psychologists in the course of their work embody explicit and implicit mutual expectations of integrity. These expectations include: fairness; impartiality; straight-forwardness; avoidance of misrepresentation; avoidance of conflicts of interest; and, the provision of accurate information. Psychologists have a responsibility to meet these expectations and to encourage reciprocation. Psychologists who fail to do so undermine the trust and mutual respect upon which professional relationships are built. In addition, a lack of honesty, the presentation of inaccurate information, and bias in reporting can distort and even invalidate scientific progress, which rests on the accumulated work of many investigators.

Of special concern to psychologists is the use of deception in research or therapy (e.g., paradoxical intention). Although research which uses deception can lead to knowledge which is beneficial, and therapy which uses deception can lead to beneficial changes for the client, such benefits must be weighed against the individual's right to self determination and the importance of public and individual trust in psychology. For these reasons, psychologists have a serious obligation to consider the need for, the possible consequences of, and their responsibility to correct any harmful effects of deception.

- | | |
|-----------|---|
| Accuracy/ | <ul style="list-style-type: none"> 2. Accurately represent their own and their associates' professional Honesty qualifications, education, experience, competence and affiliations. 5. Accurately represent their activities, functions, and likely outcomes of their work. 7. Take credit only for the work that they have actually done and give credit for work done by others in proportion to their contribution. 8. Acknowledge the limitations of their knowledge, findings, interventions and views. 9. Not suppress disconfirming evidence of their findings and views, acknowledging alternative hypotheses and explanations. |
| Openness | <ul style="list-style-type: none"> 15. Conduct research in a way that is consistent with a commitment to honest, open inquiry, and to clear communication of sponsorship and research aims. 16. Submit their research in some accurate form to independent colleagues for their comments and evaluations |
| Bias | <ul style="list-style-type: none"> 20. Communicate their knowledge, findings and views as completely, accurately and fairly as possible, taking care to distinguish what is supported by objective evidence and what is personal interpretation or opinion. |
| Deception | <ul style="list-style-type: none"> 26 Not withhold information or use temporary deception if there are alternative procedures available and/or if the negative effects of the deception cannot be predicted or offset. 27. Not withhold information or use temporary deception if it would interfere with the clients' or research participants' understanding of facts which clearly might influence their decision to give informed consent. 29. If information has been withheld or temporary deception has occurred in research, provide participants during debriefing with a full clarification of the nature of the study and remove any misconceptions which might have arisen, assuring the participant that the withholding or deception was neither arbitrary nor capricious. |

- Reliance 34. If faced with difficult situations (ethical, or otherwise), unless in an emergency, seek consultation from on colleagues and/or appropriate professional groups and committees, and give due regard to their advice in arriving at a responsible decision.

PRINCIPLE IV: RESPONSIBILITY TO SOCIETY

Psychology exists as a profession within the context of human society.^{↑↑} Psychologists, both as professionals and as private citizens, have responsibilities to the societies in which they live and work, such as the neighbourhood or city, and to the welfare of all human beings in those societies.

Two of the legitimate expectations of psychology as a discipline are that it will increase knowledge and that it will conduct its affairs in such ways that it will promote the welfare of all human beings.

- Develop 1. Contribute to the profession of psychology and to society's Knowledge understanding of itself and humans generally through a free pursuit of knowledge, unless such pursuit of knowledge conflicts with other basic ethics.
2. Keep informed of progress in their area(s) of psychological activity, take this progress into account in their work, and try to make their own contributions to this progress.
6. Protect the skills, knowledge and interpretations of psychology from being misused, used incompetently, or made useless (e.g., loss of security of assessment techniques) by others.
- Respect 9. Acquire an adequate knowledge of the culture, social structure, and Society customs of a community before beginning for any major work there.
12. In any apparent conflict between keeping a law and following a professional ethical principle, unless in an emergency, consult with colleagues and seek consensus as to the most ethical course of action and the most responsible, knowledgeable effective and respectful way to carry it out.
- Improve Society 14. In research and service activities, be sensitive to the needs and problems of society if determining what questions will be asked or what kind of services will be developed, what information collected, and how results or findings will be interpreted.
18. Be aware of the current social and political climate and of previous and possible future societal misuses of psychological knowledge, and exercise due discretion in communicating psychological information (e.g., research results, theoretical knowledge) in order to discourage any further misuse.

HUMAN FACTORS SOCIETY CODE OF ETHICS

ADOPTED OCTOBER 14, 1989

Preamble

The Human Factors Society is dedicated to the betterment of humankind through the scientific inquiry into and application of those principles that relate to the interface of humans with their natural, residential, recreational, and vocational environments and the procedures, practices, and design considerations that increase a human's performance and safety at those interfaces. To promote and sustain the highest levels of professional and scientific performance by its members, the Human Factors Society has adopted this Code of Ethics. No special oath to these Articles is necessary; its provisions are incumbent on all classes of membership of the Society.

No such code can be expected to completely anticipate all of the various and complex arrangements that professionals create, nor can it fully explore the many ramifications of these arrangements. The following Articles, then, are a guide and serve to set the tenor of professional behavior. The details must be left to the conscience and goodwill of the elected and appointed officers of the Society who must administer adherence to this code.

Article I—Professional Qualifications

Human factors scientists and engineers have the responsibility of factually representing their professional qualifications as well as those of the institution they represent.

Principle 1

Members limit their practice to those areas of human factors wherein they maintain a competence by virtue of training and/or experience and not extend their endeavors beyond their realm of competence. They enter into additional areas of human factors practice and teaching only after sufficient professional preparation or with proper professional oversight.

Principle 2

Where a brief or summary statement of qualifications would be deceptive or misleading, members present their educational background in the detail and with the additional explanation necessary for an accurate interpretation of their area of study and the level of attainment achieved. Members do likewise with their representations of their work experience so that there is little chance for a misunderstanding of the extensiveness or intensiveness of their work achievements.

Principle 3

Members represent their employers' capabilities and interests accurately so as not to mislead their clients or potential clients or damage the business interests or reputation of their employers.

Principle 4

Members, when representing their professional affiliations, factually represent their current or past affiliations with any institution or organization as well as factually represent the aims and purposes of those institutions or organizations.

Principle 5

Members do not use their affiliation with the Human Factors Society or its Chapters for purposes not consonant with the stated purposes of the Society, nor do they announce their affiliation with the Human Factors Society in such a way as to falsely imply sponsorship or approval by that organization.

Article II—General Conduct

Human factors scientists and engineers have the responsibility of comporting themselves in a manner consistent with that generally expected of the professional community.

Principle 1

In the conduct of their professional activities, members do everything necessary to reflect personal integrity as well as to convey the integrity of their profession.

Principle 2

Members avoid sensationalism, exaggeration, and superficiality that constitutes deception, and must similarly avoid any misrepresentation in all public statements, presentations, and submissions to mass media.

Principle 3

Members avoid all situations that contain elements of conflict of interest or must provide full disclosure of those conflicts to all potentially affected parties.

Principle 4

Members do not use a position as a teacher, a granting or contracting official, an employer or employee, or any other position of influence to coerce or harass others.

Principle 5

Members do not use race, handicap, sex, sexual preference, age, religion, or national origin as a consideration in hiring, promotion, or training or in any research or application where such consideration is irrelevant to the situational demands for performance.

Principle 6

Members factually represent all aspects of an employment offer, fully disclosing the terms and conditions of work, the length of employment, research projects and facilities available, work assignments, and opportunities for advancement.

Principle 7

Where responsible for design, members include considerations for the safety of person and property, and, through the appropriate source, notify those concerned when a hazardous condition exists.

Principle 8

Members clearly present the adverse safety and health consequences to be expected from deviations proposed if their technical judgment is overruled by technical or administrative authority.

Article III—Publications

Human factors scientists and engineers generally have the obligation to report their work to the general scientific community and to give credit to those who have contributed on a professional level to that publication.

Principle 1

Members give credit, proportional to their contribution, to all those responsible for the formulation, experimental design, analysis, or other treatment of the material if their contribution was on a professional level. Such credit should be extended by a listing of all contributors' names in the publication. That listing can be in the form of joint authorship with the name of the most substantial contributor listed as senior author, or by a footnote or introductory statement when the contribution is minor. This Principle deals with credit for professional contributions only and in no way affects copyright ownership.

Principle 2

Members ensure that their work is reported factually, bearing professional responsibility for all elements of their reportage, including the accuracy of analysis, quotation from other works, and conclusions drawn. Members maintain the highest standards of scientific experimentation and analysis.

Principle 3

Members maintain a position of objectivity when editing publications and reviewing papers that reflect views other than their own, as well as papers that present data in conflict with those they themselves may have previously published.

Article IV—Subject Precautions

Human factors scientists and engineers have the responsibility of treating both human and animal subjects humanely and in accordance with federal, state, and local laws or regulations, as well as the generally accepted procedures within the scientific community.

Principle 1

Members determine, through consultation with colleagues or institutional review committees, that the exposure of human or animal research subjects to hazards, stress, divulgence of history or preferences, or tedium is commensurate with the significance of the problem being researched.

Principle 2

Members determine the degree of hazard present in the exposure of human or animal research subjects, avoiding any exposures to human subjects that may result in death, dismemberment, permanent dysfunction or extreme pain, and utilize the lowest levels of exposure to both human and animal subjects consistent with the phenomenon under consideration.

Principle 3

Members ensure the ethical treatment of human and animal research subjects by collaborators, assistants, students, and employees.

Principle 4

Members establish an informed consent with human research subjects when required by institutional, state, or federal codes or regulations, making explicit in plain language the terms of participation, particularly with respect to any elements of risk or stress involved, and adhere to those terms throughout the experiment. One of these terms must be that the subject has the right to terminate participation at any time without prejudice.

Principle 5

Members do not coerce potential human research subjects to participate as subjects, nor do they use undue monetary rewards to induce subjects to take risks they would not otherwise take.

Principle 6

Members preserve the confidentiality of any information obtained from human research subjects that, if divulged, may have harmful effects on those subjects.

Ethical Standards for Research with Children (Society for Research in Child Development, 1990)

Report from the Committee for Ethical Conduct in
Child Development Research

Principles

Principle 1. Non-harmful procedures: The investigator should use no research operation that may harm the child either physically or psychologically. The investigator is also obligated at all times to use the least stressful research operation whenever possible. Psychological harm in particular instances may be difficult to define; nevertheless its definition and means for reducing or eliminating it remain the responsibility of the investigator. When the investigator is in doubt about the possible harmful effects of the research operations, consultation should be sought from others. When harm seems inevitable, the investigator is obligated to find other means of obtaining the information or to abandon the research. Instances may, nevertheless, arise in which exposing the child to stressful conditions may be necessary if diagnostic or therapeutic benefits to the child are associated with the research. In such instances careful deliberation by an Institutional Review Board should be sought.

Principle 2. Informed consent: Before seeking consent or assent from the child, the investigator should inform the child of all features of the research that may affect his or her willingness to participate and should answer the child's questions in terms appropriate to the child's comprehension. The investigator should respect the child's freedom to choose to participate in the research or not by giving the child the opportunity to give or not give assent to participation as well as to choose to discontinue participation at any time. Assent means that the child shows some form of agreement to participate without necessarily comprehending the full significance of the research necessary to give informed consent. Investigators working with infants should take special effort to explain the research procedures to the parents and be especially sensitive to any indicators of discomfort in the infant.

In spite of the paramount importance of obtaining consent, instances can arise in which consent or any kind of contact with the participant would make the research impossible to carry out. Nonintrusive field research is a common example. Conceivably, such research can be carried out ethically if it is conducted in public places, participants' anonymity is totally protected, and there are no foreseeable negative consequences to the participant. However, judgments on whether such research is ethical in particular circumstances should be made in consultation with an Institutional Review Board.

Principle 3. Parental consent: The informed consent of parents, legal guardians or those who act in loco parentis (e.g., teachers, superintendents of institutions) similarly should be obtained, preferably in writing. Informed consent requires that parents or other responsible adults be informed of all the features of the research that may affect their willingness to allow the child to participate. This information should include the profession and institution affiliation of the investigator. Not only should the right of the responsible adults to refuse consent be respected, but they should be informed that they may refuse to participate without incurring any penalty to them or to the child.

Principle 4. Additional consent: The informed consent of any persons, such as school teachers for example, whose interaction with the child is the subject of the study should also be obtained. As with the child and parents or guardians informed consent requires that the persons interacting with the child during the study be informed of all features of the research which may affect their willingness to participate. All questions posed by such persons should be answered and the persons should be free to choose to participate or not, and to discontinue participation at any time.

Principle 5. Incentives: Incentives to participate in a research project must be fair and must not unduly exceed the range of incentives that the

child normally experiences. Whatever incentives are used, the investigator should always keep in mind that the greater the possible effects of the investigation on the child, the greater is the obligation to protect the child's welfare and freedom.

Principle 6. Deception: Although full disclosure of information during the procedure of obtaining consent is the ethical ideal, a particular study may necessitate withholding certain information or deception. Whenever withholding information or deception is judged to be essential to the conduct of the study, the investigator should satisfy research colleagues that such judgment is correct. If withholding information or deception is practiced, and there is reason to believe that the research participants will be negatively affected by it, adequate measures should be taken after the study to ensure the participant's understanding of the reasons for the deception. Investigators whose research is dependent upon deception should make an effort to employ deception methods that have no known negative effects on the child or the child's family.

Principle 7. Anonymity: To gain access to institutional records, the investigator should obtain permission from responsible authorities in charge of records. Anonymity of the information should be preserved and no information used other than that for which permission was obtained. It is the investigator's responsibility to ensure that responsible authorities do, in fact, have the confidence of the participant and that they bear some degree of responsibility in giving such permission.

Principle 8. Mutual responsibilities: From the beginning of each research investigation, there should be clear agreement between the investigator and the parents, guardians or those who act in loco parentis, and the child, when appropriate, that defines the responsibilities of each. The investigator has the obligation to honor all promises and commitments of the agreement.

Principle 9. Jeopardy: When, in the course of research, information comes to the investigator's attention that may jeopardize the child's well-being, the investigator has a responsibility to discuss the information with the parents or guar-

dians and with those expert in the field in order that they may arrange the necessary assistance for the child.

Principle 10. Unforeseen consequences: When research procedures result in undesirable consequences for the participant that were previously unforeseen, the investigator should immediately employ appropriate measures to correct these consequences, and should redesign the procedures if they are to be included in subsequent studies.

Principle 11. Confidentiality: The investigator should keep in confidence all information obtained about research participants. The participants' identities should be concealed in written and verbal reports of the results, as well as in informal discussion with students and colleagues. When a possibility exists that others may gain access to such information, this possibility, together with the plans for protecting confidentiality, should be explained to the participants as part of the procedure of obtaining informed consent.

Principle 12. Informing participants: Immediately after the data are collected, the investigator should clarify for the research participant any misconceptions that may have arisen. The investigator also recognizes a duty to report general findings to participants in terms appropriate to their understanding. Where scientific or humane values justify withholding information, every effort should be made so that withholding the information has no damaging consequences for the participant.

Principle 13. Reporting results: Because the investigator's words may carry unintended weight with parents and children, caution should be exercised in reporting results, making evaluative statements, or giving advice.

Principle 14. Implications of findings: Investigators should be mindful of the social, political and human implications of their research and should be especially careful in the presentation of findings from the research. This principle, however, in no way denies investigators the right to pursue any area of research or the right to observe proper standards of scientific reporting.

Department of Psychology
Human Ethics Committee
Request for Review of Research Proposal

1. Title of Project: _____

Course Number: _____

Start/Finish Dates: _____

2. Name Address Phone No.

Primary Researcher: _____

Supervisor: _____

Other Personnel: _____

3. ACADEMIC PURPOSE OF THE RESEARCH:

- a. Honours Thesis
- b. Special Study or Directed Research
- c. Independent Research
- d. Other (Explain) _____

4. RATIONAL AND/OR HYPOTHESIS

Provide a brief description of the reason for doing the research.

5. SUBJECTS

5.1 Projected number of subjects needed: _____

5.2 How was this number decided?

5.3 Are any Special Subject Characteristics Required?

(e.g. age, sex, normal or corrected vision, driver's license, etc)

5.4 Subject Recruitment

Briefly describe how you will recruit the subjects.

5.5 Instructions to Subjects

Attach the script of the basic instructions given to the subjects.

5.6 Subject Consent Form:

Here is the basic outline. Please complete it with a brief statement that identifies the research project.

I consent to participating in research on _____

I understand that I may discontinue my participation at any time.

(signature)

(date)

5.7 Does the Research Involve:

a) Full Disclosure of Aims and Expectations?

Yes ____ No ____ (If No, EXPLAIN)

b) Deception?

No ____ Yes ____ (If Yes, EXPLAIN)

c) Psychological Stress/Anxiety?

No ____ Yes ____ (If Yes, EXPLAIN)

d) Physical Stress, Fatigue, or Endangerment?

No ____ Yes ____ (If Yes, EXPLAIN)

e) Lingering Psychological Issues?

No ____ Yes ____

(If Yes, EXPLAIN and describe steps taken to deal with this issue).

5.8 Debriefing

Attach a script of the basic debriefing given to the subjects on completion of their participation.

5.9 Confidentiality

Is confidentiality a possible issue? NO YES

(If uncertain consult your supervisor, or a member of the Ethics Committee.)

If it is a possible issue, briefly describe how the confidentiality of subjects will be ensured:

6. APPARATUS AND/OR MATERIALS)

Describe the instrumentation and/or provide copies of questionnaires or surveys. (Let's follow the sequence of the APA format that we are striving to teach, and put this item next. To reduce duplication from what was previously Section 9. Put it all here.)

7 PROCEDURE

Describe how the research will be conducted.

8. CERTIFICATION

I certify that the information provided is accurate and complete.

Principal investigator:

Date

Approved: _____ **Date** _____
(Chair of Ethics Committee)

Date

A Primer on Research Automation

You can use computers for a lot more in your research than simply to analyze the data. You can automate many types of research by using just the monitor to present instructions and stimuli materials and the keyboard to record the subjects' responses. With little further ado than a word processing program, you can automate questionnaires. I want to introduce you to the further opportunities made possible by installing a general purpose interface board into a computer. It provides a basis for the computer to adjust variables such as the brightness of a light, turn stimuli on and off, sense when a subject had pushed a button, measure how far the subject has adjusted a dial. Add a bit of electromechanical ingenuity and the possibilities for fully automated research are endless. My undergraduate mentor, Gilray Kandel, called this the "electric trains" fascination of research, but there is more than fun at stake. Allow me to explain; then I'll summarize the basic principles and show you examples. Reading it won't make you a lab wizard, but it can make you aware of what is happening at leading laboratories.

1. WHY AUTOMATE

TO MEASURE WHAT IS OTHERWISE UNMEASURABLE

Research use of most physiological measurements such as eye-movements, heart rate, and electroencephalogram (EEG) requires a computer to store the records so they compared for different levels of the independent variable. Cortical evoked potentials can only be seen after a computer has averaged some thirty or more EEG signals whose recording was synchronized to an evoking stimulus.

TO AVOID RESEARCHER VARIABILITY IN MAKING MEASUREMENTS

Many responses happen too fast to be accurately counted or timed directly. Errors made during a tracking task or how quickly a person can recognize a stimulus are examples. Consider the situation in which a researcher uses a stop watch to measure how fast subjects recognize upright faces versus faces tilted 90°. The difference of some 30 milliseconds reflects the additional cognitive processing, which in turn might be used to measure familiarity, emotional state, etc. Let's say that measuring recognition time ten times for each type of face was accompanied by a standard deviation of 30 msec in the subjects' responses. A t-test indicates that the 30 millisecond difference in recognition times is significant despite the subject variability. However, when the researcher is operating a stop watch, the researcher's variability in reaction time adds to the subjects' variability.¹ If the researcher's variability is the same as the subjects', twice as much data will be needed to obtain a statistical significance. On the other hand, a using a computer to measure the recognition time adds essentially no variability to the measurement process. Consequently using a computer substantially reduces the number of measurements needed.

¹ Since the researcher's and subjects' variabilities are independent, the combined variability is their geometric sum - i.e. the square root of the sum of the standard deviations squared.

AUTOMATION D2

TO AVOID RESEARCHER BIAS

As mentioned in the chapter on ethics, it can be difficult for a researcher to remain totally unconcerned about how the data are turning out. Despite the best of intentions and without awareness, the researcher may introduce bias without awareness in terms of vocal cues to the subject or interpretation of some ambiguity in the response. For example, "Did you 16.3 or 63?" The computer is a totally objective, dispassionate operator.

TO SHORTEN THE PROCEDURE

A computer can look up the next setting of the independent variable, make the adjustment via some electromechanical device, activate the presentation, read an indicator of the subject's response, record the reading, and prompts the subject that everything is ready for the next trial in the twinkle of an eye. By comparison, a researcher may take a few seconds to do any of these tasks. "Is saving a few seconds really worth all that technology and the time and resources needed to get it all working?" In some cases, probably not. However consider an experiment that involves making 240 measurements with each subject. If a research could do each of the above tasks in 2 seconds, the preparation time for each trial becomes 12 seconds. Multiply that by 200 and you have the subject waiting for a total of 40 minutes. That is forty minutes of the subject loosing his/her attention and getting bored - which adds to the overall variability of the data. The time savings possible with automation can be particularly important to within-subjects designs, which often require many measurements to be made with each subject.

2. DRAWBACKS AND OPPORTUNITIES

Computers do "hang up", electromechanical devices jam, digital input/output lines can get misactivated by electronic noise. However, computers very rarely make an actual mistake such as switching a 7 for a 3. The usual problem with computers is the program, and that, of course, is the responsibility of the researcher. Good design and proper components can make electromechanical and other automation devices operate more reliably than manual control systems. In short, reliable automation requires expertise. If you lack that expertise, you have two choices if you want to pursue research of the calibre of major psychology laboratories: You acquire some of these skills yourself or you can hire technicians. Most researchers do a bit of both. Knowing something about what is possible is often necessary for developing realistic requests; yet there is more technology than most researchers have time to learn.

If you pursue your studies at one of the major universities, you will find that their Psychology departments have workshops and technicians to help with research automation. Consider taking a course in instrumentation which may be offered in Engineering, Physics, Chemistry, and in some places Psychology Departments. Electronics project kits such as sold by Radio Shack are an excellent way to develop some hands-on-experience on your own.

What follows summarizes the basics concepts and principles of interfacing computers for automated research. By no means will its study enable you to automate your next research project, but it may convince you to pursue the matter further and possibly help you overcome some conceptual hurdles. (When I took my first electronics course from a physics department, I was totally confused about how vacuum tubes worked. Fortunately, I came across an Army electronics training manual. Something about its explanation of how transistors worked suddenly cleared up my confusion and I had no difficulties in the course after that.)

3. FUNDAMENTAL STUFF ABOUT COMPUTERS

(OR WHAT EVERYONE WHO USES ONE SHOULD KNOW)

As every Introductory Psychology student knows, humans can grasp up to ten (7 + 3) things simultaneously. So why be intimidated by something too stupid to grasp a number larger than "1"? Asked to consider a number larger than "1", a computer does just what we do when we want a number larger than 9. Namely, go back to the beginning and add a zero. Having cleverly obtained a two-place number, one can continue counting by incrementing numbers in the second position and then the first.

When computers do this, at first it seems ridiculously simple: Starting with zero (for nothing), a computer counts one thing by incrementing the "0" to a "1". To count two things, it then adds a zero and comes up with a two place number, "10", which is what we would call "ten" but to a computer it means "two". To count another thing, the computer increments the second-place "0" and proudly comes up with "11", which means "three". To count a fourth item, it has reached the end of its intellectual tether, and again reverts to adding another place and resetting the existing places to "0". This gives it a "100", which means "four". Five becomes "101", and so on.²

Because this number system has only two numbers, "0" ad "1", it is called a "binary" number system. People did not design computers to use such simple arithmetic just to feel superior. Rather this simple-minded math, could be achieved by the simple mechanics of a switch or its automated equivalent the relay.

3.1 SWITCHES, RELAYS, AND FLIP-FLOPS

Your basic switch, such as a wall light switch, has two states - it is either "off" or "on". Since there is no in-between state, the switch is an example of a "digital" device. Since the switch has two and only two states, it is a "binary" digital device. (One could use a bunch of manual switches like an abacus to keep track of how many sheep are traded or the number of correct minus wrong responses of a subject. As long as they are left undisturbed, both an abacus and an array of switches have memory that is more reliable than my computer's memory).

The next stage of sophistication in binary digital devices incorporated an electromagnet built into a switch. Now instead of flipping the lever, you can push a button which sends some electricity through a coil and the magnet closes the switch for you. Now you have a "relay". By itself that would hardly be worth the bother. But relays have some redeeming potentials. Your push button could activate several relays at the same time. One of these relays could turn off a light; another turn on a buzzer; another could start a timer; a fourth could activate a lever for a rat to press in a Skinner Cage. An experimenter could do all those things quickly by hand, but not simultaneously. Subsequently, if the rat's lever press operated some more relays, the rat could turn off the buzzer, activate a feeder, and stop the timer. Just try teaching a rat to push that many switches.

The problem with basic relays is that they stay closed only as long as you (or the rat) hold down the button to keep the electromagnet pulling on the switch. (If you built the switch part of the relay to latch like a wall switch, you could activate with a momentary push

² If you follow me so far try collapsing your mental powers and see if you can come up with the computer equivalent of 8, 9, and 26.

AUTOMATION D4

of button, but you would then have to reset the switch manually. The answer to this little dilemma is, of course to put two electromagnets on each switch - one magnet to close the switch, a second to open it. Now you have reached the pinnacle of binary digital logic, a "flip-flop", which is the main component in all digital computers.

For some forty years from the 1930's to the 1970's, research in psychology grew exponentially as laboratories incorporated electromagnetic flip-flops to automate experiments. Cornsweet's (1968?) book was the bible for laboratory-wise graduate students who wanted to be at the cutting edge. His book shows how multiple flip-flops can be combined to perform all basic logic functions. "Programming" an experiment was literally a matter of wiring together dozens of dozens of flip-flops and switches to operate various devices like timers, shutters, list rollers, etc. Enough racks of flip-flops to fill a walk-in closet along with hundreds of connecting wires might be needed to automate a sophisticated experiment. It took a lot of ingenuity, time, and money. If a single wire came loose or an electromagnet wore down and jammed - - !!@#@**!. In the mid 1960's transistor flip-flops began to replace the electromechanical, with substantial improvements in reliability, speed, and compactness. However, in less than ten years, the whole laboratory automation situation changed to a completely new format.

Transistorized flip-flops were miniaturized to the point where thousands and then millions could be put into a shoe box. Rather than programming them by adding and removing wires, they were all preconnected with other flip-flops to determine whether each connection was open or closed. Thus it was no longer necessary rewire the flip-flops when it was time to try something different. The connects were malleable or "soft". That is how computers came into the psychology laboratory.

3.2 BINARY NUMBERS, CODES, AND LOGIC

Have you made the connection between binary numbers and the flip-flops which make up computers? Not only can just zeros and ones be represented by a single flip-flop they can also be stored as either a space or a hole in a certain location on a paper card, as neutral or magnetized spot on a film coated with tiny iron particles, as light reflecting or light absorbing spot on a compact disk. Computers use binary numbers in two ways. One is for counting and holding the results of all those other arithmetic operations we use to summarize and analyze data. This is done by using flip-flops in sets of 16, 32, or more to represent binary numbers to 16, 32 or more places. Sixteen flip-flops together can represent any number up to 2^{16} , which is 64,000.

The second function is that the binary numbers are also used as codes that tell the computer what to do. These codes set and unset various processor flip-flops to perform logical operations such as AND's and OR's; arithmetic operations such as adding or subtracting; internal operations such as "get a number from a certain location in memory" and "get the next instruction from a certain location in memory". There is no difference between binary numbers that are stored as numbers or as codes. Only the sequence of instructions that the computer is following determines how the binary numbers in memory are used.

*So that's why one little mistake in a program
can make a computer go crazy! - Editor*

*Yes, it may think the results of a long division
represents some completely inappropriate command - Nilsson*

3.3 DIGITAL INPUTS AND OUTPUTS

Lets say you want to let a computer know that a subject had pushed a switch - perhaps to indicate that she/he is ready for the next measurement. With no additional interfacing you could hand over the computer's keyboard and ask the subject to press a certain key. Your operating program would then keep checking if a key has been pressed. If it is the wrong key, it would keep checking. If it's the correct key, the program goes on to the next stage of the experiment. However, that might be inconvenient for you and unnecessarily complicated for the subject to select from 88 keys when one would do. The automation alternative is to install a digital I/O (input/output) board in your computer and connect a pair of wires from one of its flip-flops to a push button for the subject. If all the I/O board flips were in the "0" position initially, and you connected the button to the third flip-flop, the subject's response would change the array to "00000100" (with eight I/O flip-flops). To the computer looking at them, this would be the number "4". The operating program would have an instruction to "read the I/O flip-flops" then do a logic test to see if the result is a "4". If not, try again. If it is "4" go to the next step. With eight I/O flip-flops you could give the subject several more switches to indicate responses such as "yes", "no", "repeat", or even a multipoint rating scale.

Digital outputs work the opposite way. The computer sets an output flip-flop on the board by taking a number you have previously put in memory and connecting those memory flip-flops to the corresponding ones on the I/O. Thus if you had the number "5" (00000101) stored in memory, connecting those flip-flops to the I/O board would turn on the first and third flip-flops so they emit 5 volts. You could use those five volts to do things like power a light or buzzer directly. The 5 volts could also operate a relay which in turn could open a shutter or activate a tape recorder. To illustrate the advantages of the "soft" connections between flip-flops, imagine that you have just completed an experiment which involved a light onset followed by a buzzer. You now want to find out what happens if the buzzer is presented first. Instead of untiring the buzzer and light and then switching the connects to their respective flip-flops, you now simply switch which number you load into the I/O board.

E.3.4 COUNTERS AND TIMERS

4 FANCIER INTERFACES

4.1 ANALOG - DIGITAL CONVERTERS

4.2 MIDI CONTROLLERS

4.3 PROGRAMMABLE CLOCKS

4.4 MOTOR CONTROLLERS

AUTOMATICO D6

Practicum # 1

DOING ARCHIVAL RESEARCH

Ever since the earliest kingdoms of the middle east first enumerated their subjects for taxation, organized society continues to gather data on itself. In Canada most of that information is in the public domain - available to anyone willing to seek it out. For certain types of research, this knowledge that our society has gathered about itself may suffice without the need for the researcher to gather new data. To use archival data, one must first be aware of its existence and then know how to locate it - as well as how to analyze what one has found. This practicum is designed to introduce you to using this type of knowledge.

Archival information can be found in many places including newspapers, journals, catalogs and regular books, but the largest concentration exists in government publications. Like most libraries, the Robertson has a special section for government publications. These shelves are organized according to the originating government department. A Guide to Canadian Government Publications is indexed by author, subject and government department. Furthermore, Statistics Canada has its own monthly index which is organized according to title, subject and commodity. Of particular interest to psychologists may be the results of the 1996 census, which was the most thorough gathering to date of a variety of demographic information. Statistics about UPEI's enrollment, courses and grades can be obtained at our library's Information Desk.

Librarians are the experts at directing one to this stored knowledge. However, the researcher's job is rarely as simple as merely locating some information. Archival discoveries depend on finding new relationships within the available data. Such relationships can take one beyond the obvious facts to yield an understanding of why the world is the way it is. Statistics Canada does a certain amount of organization and comparisons of information for special areas of concern such as education, health and labor. Yet today's archival data has an almost limitless number of potential relationships that are worth exploring.

Your task is to think up a question that could be answered by making some comparison of two sets of archival data. Find some reference in print (eg. Psychology text, newspaper or magazine article) to this question, which you can cite in your Introduction. In the Method section, describe the materials from which you obtained your archival data and describe the procedure you used to determine the possible relationship. Use your knowledge of descriptive statistics to present the Results in a convincing graph and/or use an appropriate statistical test to make your case. In the Discussion, compare your findings with what was indicated in the reference source.

PRACTICA E2

You may choose your own variables or you may select from the following variables:

- Divorce rate
- Death rate
- Crime rate
- Unemployment rate

Historical Canadian data on these variables are available on reserve. Choose the two variables you wish to examine and then extract the necessary information from the appropriate files.

In addition, on the UPEI Library home page, you will find a section called the "Data Liberation Initiative" DLI which provides access to a great deal of useful data.

It is recommended that you use this source of information since it is up-to-date and readily accessible and allows you to select from a large variety of variables..

You might test the data for a relationship between two of these variables, or you might look for group differences, or you might consider looking for a relationship with some possible cause. For this practicum, you are asked to look for a correlation between the two variables you choose.

In any case, you should not rely on an isolated piece of data, which may simply be an artifact, but rather you should use data demonstrating the variables over a longer period of time. For this study you should examine about a 10 year period.

Practicum # 2

OBSERVATIONAL RESEARCH

You may or may not have noticed that there seems to be a tendency for male and female students to carry books in a different manner. Jenni (1976) performed an observational study which demonstrated that this difference is statistically significant. While differences between males and females in body shape and strength are likely important reasons for this difference, Jenni argued that social customs and personality variables also may be involved.

Jenni and Jenni (1976) showed that this behaviour is also dependent upon sexual maturation. Young children of both sexes tend to carry books in the typical 'male' manner. However, from the 2nd grade on, females increasingly switch over to using the typical 'female' pattern; this transition being completed by the 10th to 12th grade.

These studies provide an example of how the observational research method can be applied to apparently simple, everyday behaviour to yield insights into human nature. Though they (Jenni & Jenni; Jenni) examined only book-carrying behaviour, it is possible that similar differences exist for other types of carrying behaviour. Your assignment is to select some other type of carrying behaviour and perform an observational study to determine if it is performed differently by males and females.

It would probably be more interesting if you come up with a type of carrying that turns out to be statistically different in how it is performed by males and females, but the scholarly value of your report will not be diminished if matters turn out otherwise. The principal aim of the practicum is to have you conduct your observational research in a professional manner and report your research properly according to APA format.

Here are some guidelines on how to conduct your study:

1. From past experience and general observations, format a specific question about some kind of carrying behaviour that may be different for males and females.
2. Make some preliminary observations to estimate your observational categories.-- Since you probably can't take into account all possibilities for human variability, use an OTHER category as did Jenni (1976).
3. Restrict your categories and observations to behaviour that pertains to the question you are asking. (i.e. Don't try to report everything).
4. Prepare your data record sheet before going into the 'field' to collect data.
5. Take notes describing the situation in which you conducted your observations and how you proceeded.
6. Use the same statistical test Chi-Square as Jenni did to determine the significance of your findings.

References

- Jenni, D. A. and Jenni, M.A. (1976). Carrying behaviour in humans: Analysis of sex differences. Science, 194, 859-860.
- Jenni, M.A. (1976). Sex differences in carrying behaviour. Perceptual and Motor Skills, 43, 323-330.

Practicum # 3 SURVEY RESEARCH

Surveys using questionnaires are probably the most common type of research. Questionnaires yield a 'snapshot' assessment of the opinions, attitudes and/or characteristics of a group of people at a particular point in time.

There are a variety of purposes for administering questionnaires. The main purposes are as follows:

1. Descriptive: To determine the present state of the group. These types are known as One-shot or Cross-sectional surveys.
2. Before/After: As a means of assessing the effectiveness of some intervening variable.
3. Comparative: To establish a baseline or 'norm' group with which to assess changes over time. These may involve new independent samples tested with the same questionnaire or the same sample may be retested in a longitudinal survey.
4. As a Research tool to test the effect of an independent variable upon a dependent variable.

Questions may be either open-ended or closed-ended. Open-ended questions allow respondents to write in their own answers. Because the responses may be quite varied, these are usually more informative but more difficult to analyze. Closed-ended questions require the respondent to select a single response from a list of choices. These are also called forced-choice questions. The range of responses is more limited, but much simpler to analyze.

Questionnaires may be administered by mail, by phone, or to a person or group. For this practicum, you will administer eight questionnaires to individuals or small groups.

The questionnaire which you will be using consists of a few questions which should be of interest to the respondents and which will not require a great deal of time to complete.

In order to encourage the respondents to answer the questionnaire as conscientiously as possible, you should assure them that their responses will be kept confidential. The answers from each questionnaire will be combined with all the other questionnaires so that no individual's answers will be identifiable.

When you approach potential respondents, inform them that you are conducting the survey as part of a course on research methods and that the results will only be used to assist in learning about survey techniques. Indicate to the respondents that you will only need a few minutes of their time. Assure them of the confidentiality of their responses. If they agree to take part, simply give them the questionnaire form to complete. When they have finished, collect the questionnaire and thank them.

Return with your set of completed questionnaires to the lab meeting next week so that they may be combined with the others to generate the data for this practicum.

While designing a valid, reliable questionnaire is generally a process of drafting, editing, pretesting and revising questions, for this practicum we will design and administer a simple survey with no more than six questions.

Since this will be a research questionnaire, it needs at least one independent and one dependent variable. The independent variable may be any condition which produces separate groups, e.g.

Handedness: LEFT, AMBIDEXTROUS, RIGHT

Year of studies: 1st, 2nd, 3rd, 4th

Faculty of studies: arts, science, business, education, nursing.

Include at least two questions which will could be measurements of a dependent variable. These questions could directly assess some factual aspect of behavior, e.g.

The average number of hours I study each week is: _____

Or could assess the strength of some attitude, e.g.

To what extent do you disagree or agree with the following statement?

"Silence is important when I study."

STRONGLY DISAGREE

STRONGLY AGREE

-4 ----- -3 ----- -2 ----- -1 ----- 0 ----- +1 ----- +2 ----- +3 ----- +4

One question should be designed to involve a qualitative response that represents a choice from different categories, e.g.

When is the most productive time for you to study? (Circle one)

EARLY
MORNING

DURING
THE DAY

LATE
AFTERNOON

EVENING

LATE AT
NIGHT

JUST BEFORE
THE EXAM

PRACTICA E6

The following form, which differs somewhat from the checklist provided in Chapter 1, will be used to grade your report on the survey exercise:

EVALUATION FORM FOR SURVEY REPORT

title page

Title

Student's name

Institutional Affiliation

abstract page

Abstract

Summary of the survey's purpose, what was conducted, who were the respondents, what was found. It should be specific, concise, non-evaluative, stand on its own, and be less than 75 words.

(the introduction starts on new page but has no heading)

Explain the purpose in terms of what question was addressed why it is of interest; what has recently been reported about this issues in journal or news media; and how you went about answering the question. Explicitly identify the INDEPENDENT and DEPENDENT VARIABLES.

=====

=====

Method

Participants

Who was surveyed; how were they selected, and what instructions were they given?

=====

Materials

Brief description of the questionnaire. Provide a copy in the Appendix and refer to it in the text.

=====

Procedure

Where, when, and how the survey was done.

=====

Results

Describe what were the overall results and any statistical analysis.- Refer to a table and/or figure which summarizes the results.

=====

Show calculations in an Appendix. Are the statistics appropriate and correctly done?

=====

(possible table page)

Table #.

Title of Table Should Summarize the Answers to Each Question.

Organization, Labelling, Graphics, Appropriateness of Title All Count (

)

=====

(possible figure page)

Figure 1. Figure to illustrate apparatus or materials. The legend is placed below. Organization, graphic quality, labelling, appropriateness of legend, and neatness count.

()
=====

Discussion

What does your statistical summary and analysis imply? What was generally concluded from the questionnaire results? How does this compare with what was previously reported?

=====

REFERENCES (start on a new page)

Properly cited in the text

Use of APA format in alphabetical list of references

=====

Grammar, spelling, use of past tense, clarity of expression

=====

Evaluation of the questionnaire itself:

format (title, identification, instructions, general layout, neatness)

=====

appropriateness of questions and answer choices

=====

ingenuity or creativity of topic

=====

Practicum # 4

A BETWEEN-SUBJECTS EXPERIMENT ON THE EFFECTS OF XANTHINE ON ATTENTION, MEMORY AND MOTOR SKILLS

Psychoactive drugs affect our cognition and performance because they mimic the neurotransmitter substances and the neuromodulator substances which the brain produces to do its work. The brain uses many diverse neural circuits, and thus a combination of many more chemicals, to perform even the simplest of behaviors. Therefore the effects of any one substance may be both subtle and far reaching. Xanthines inhibit the effects of adenosine, which in turn is a behavioral depressant, and xanthines stimulate the heart, increase skeletal muscle tension and relax smooth muscles such as those used in the respiratory system (Groves & Rebec, 1988).

In this practicum you are one of the subjects assigned randomly to one of two groups. One group will consume a xanthine containing potion while the other group will consume a similar potion without this drug. While no experiment or any other human endeavor such as crossing the street is ever totally safe, and psychoactive drugs always involve some risk to the user, the dosage being used in this experiment is probably safe and commonly used by college students. To avoid contaminating the effect of whichever potion you will receive as well as to avoid possible drug interaction effects, which can be unpredictable, it is important that you abstain from the use of xanthines and any other psychoactive drugs for at least three hours PRIOR to your appointment to serve as a subject.

(If you have an allergic reaction to xanthines or for other reasons are a nonuser of this drug, please inform the experimenter.)

To participate as a subject, sign up for a time period - approx. one hour in total. This should be done as soon as possible and your participation completed by the deadline. Reports will only be accepted from those who have participated as subjects. At the appointed time, you will be given a potion which you must consume within five minutes. There will then be given a brief waiting period to allow the drug to enter the central nervous system. During this waiting period, you may pursue any light, non-strenuous activity - such as studying.

In the actual experiment, you will be placed under the control of a computer which will present stimuli and record your responses. The principal task is to follow and learn an invisible maze presented on the cathode ray screen. You will be given a number of trials to learn this maze which remains the same throughout the experiment. Work as quickly and as accurately as you can to solve the maze each time. An auxiliary task will measure your vigilance to detecting a signal while simultaneously moving through the maze. It is important that you do not miss responding to any of these signals.

After the experiment, all the data collected will be summarized for you. Examine the data on the various measurements which were made with respect to the main independent variable - the presence or absence of the drug in the potions drunk by the two groups of subjects. Perform an appropriate statistical test on one set of measurements to determine whether the two groups differed, then write your report using as a reference source the Loke & Melsika (1984) article on reserve which describes another experiment which examined the

effects of a xanthine drug on performance in college students. You should also locate a second reference as well (a PsycINFO article will do).

Note : Xanthines: One of a class of stimulant compounds. "The xanthines are the most widely used stimulants because they include the drug caffeine." (Groves & Rebec, 1988, p. 181)

References

- Groves, P. M. & Rebec, G. V. (1988). Introduction to biological psychology (3rd ed.). Dubuque, Iowa: Wm. C. Brown.
- Loke & Meliska (1984). Effects of caffeine use and ingestion on a protracted visual vigilance task.

Practicum # 5

A WITHIN-SUBJECTS EXPERIMENT: MEASURING BRIGHTNESS DIFFERENCE THRESHOLDS

This experiment involves using a precise psychological method to measure a relationship between physical stimulus and sensory experience. While scientists probably will never be able to directly measure the strength of conscious experience that is produced by a given physical amount of light, we can measure how much of a change in the amount of light is needed to see a difference in brightness. This latter measurement is called a brightness "difference threshold". The fact that difference thresholds for brightness change as a complex function of the amount of light suggests that the visual processes which produce brightness do not simply reflect the amount of energy received. Because the shape of difference threshold function reveals the characteristics of the underlying mechanisms, Shepard (1978) has suggested that we may come closest to our understanding of consciousness through these measurements, which he refers to as the "second order" isomorphism between the physical world and consciousness.

Tackling this momentous task requires you to attend to the calibration of an otherwise imprecise apparatus; requires the experimenter to carefully organize how the measurements are made; requires the subject to patiently adjust the apparatus and to attend to what his or her sensations reveal while ignoring the physical changes which one knows are occurring as the apparatus is adjusted.

We will use the METHOD OF ADJUSTMENT or METHOD OF AVERAGE ERROR as it's also known, which is the fastest and also one of the more sensitive of the psychophysical methods. (See your text for further discussion of this and the other psychophysical methods.) Using this method to measure brightness discrimination, the subject repeatedly attempts to match the brightness of one stimulus, the MATCH field, to that of a second stimulus, the STANDARD field: Starting with the MATCH obviously brighter than the STANDARD, the subject decreases the brightness of the MATCH until the fields look equally bright. The experimenter records the subjects dial setting and then adjusts the MATCH so that it is obviously dimmer than the STANDARD. The subject now increases the brightness of the MATCH until the two fields look equally bright. (Prior to collecting data, the experimenter determines the approximate "obviously brighter" and "obviously dimmer" initial settings of the MATCH by adjusting the MATCH for him or her self. This must be done for each energy level of the STANDARD. The subject can record these settings.)

This "up - down" procedure is repeated a total of 16 times at each energy level of the STANDARD. The experimenter should take care to make the initial, obviously mismatched settings of the MATCH somewhat different each time.

Also, don't do all 16 at once. COUNTERBALANCE the measurements as follows: Start testing with the lowest energy level of the STANDARD. Do half of the measurements, and then proceed to the next energy level of the STANDARD. When you reach the highest energy level, repeat it, then proceed to test successively lower levels.

Make these measurements with the STANDARD set to 16, 56, 96, and 136 candelas/meter². However, you can't set the STANDARD dial directly to these values. The numbers on the two dials represent only arbitrary ORDINAL values. You must interpret the dial numbers by referring to the provided CALIBRATION chart: For example, look up 80 cd/m²

on the Y axis. Determine the plot's corresponding dial value, 123, on the X axis. Set the dial to 123 to obtain 80 cd/m^2 . (You will need to reverse this procedure to convert your dial data numbers into meaningful energy values before plotting or otherwise interpreting them)

Give the subject a chance to become familiar with the apparatus and the matching procedure by having him or her make some practice matches at an intermediate energy level of the STANDARD. Before collecting data, have the subject dark adapt for 5 minutes. Keep the laboratory as dark as practical during testing.

To calculate a difference threshold at each energy level of the STANDARD: a) Calculate the standard deviation of all the observer's matches at that level. b) Multiply the standard deviation by 0.6745. c) Add the product to the dial setting of the STANDARD. d) Determine the energy value of the sum from the CALIBRATION chart. e) Subtract the energy level of the STANDARD from this calibrated value, and you have the DIFFERENCE THRESHOLD. (For example, assume that a standard deviation of 25.2 was obtained when the STANDARD was set to produce 80 cd/m^2 . Multiplied by 0.6745 you have 17. This added to the 123 dial setting for the STANDARD gives 140. The 140 in dial numbers corresponds to a calibrated value of 102 cd/m^2 . Subtracting the energy level of the STANDARD from this calibrated value gives 22 cd/m^2 , which is then the obtained DIFFERENCE THRESHOLD at an energy level of 80 cd/m^2 .)

Make a graph showing how the brightness difference threshold changes as a function of light energy level. Compare your results with those of a previous study by Bartlett (1942).

References

Bartlett, N. R. (1942). The discrimination of two simultaneously presented brightnesses. Journal of Experimental Psychology, 31, 380-392.

{Mainly use the introduction, summary, and - for comparison - Figure 2B}

Shepard (1978). The mental image. {Not needed for report}

{Also find a second article, possibly involving Weber's Law}

Table R
Random Numbers

10480	15011	01536	02011	81647	91646	69179	14194	62590	36207	20969	99570	91291	90700
22368	46573	25595	85393	30995	89198	27982	53402	93965	34095	52666	19174	39615	99505
24130	48360	22527	97265	76393	64809	15179	24830	49340	32081	30680	19655	63348	58629
42167	93093	06243	61680	07856	16376	39440	53537	71341	57004	00849	74917	97758	18379
37570	39975	81837	16656	06121	91782	60468	81305	49684	60672	14110	06927	01263	54613
77921	06907	11008	42751	27756	53498	18602	70659	90655	15053	21916	81825	44394	42880
99562	72905	56420	69994	98872	31016	71194	18738	44013	48840	63213	21069	10634	12952
96301	91977	05463	07972	18876	20922	94595	56869	69014	60045	18425	84903	42508	32307
89579	14342	63661	10281	17453	18103	57740	84378	25331	12566	58678	44947	05585	56941
85475	36857	53342	53988	53060	59533	38867	62300	08158	17983	16439	11458	18593	64952
28918	69578	88231	33276	70997	79936	56865	05859	90106	31595	01547	85590	91610	78188
63553	40961	48235	03427	49626	69445	18663	72695	52180	20847	12243	90511	33703	90322
09429	93969	52636	92737	88974	33488	36320	17617	30015	08272	84115	27156	30613	74952
10365	61129	87529	85689	48237	52267	67689	93394	01511	26358	85104	20285	29975	89868
07119	97336	71048	08178	77233	13916	47564	81056	97735	85977	29372	74461	28551	90707
51085	12765	51821	51259	77452	16308	60756	92144	49442	53900	70960	63990	75601	40719
02368	21382	52404	60268	89368	19885	55322	44819	01188	65255	64835	44919	05944	55157
01011	54092	33362	94904	31273	04146	18594	29852	71585	85030	51132	01915	92747	64951
52162	53916	46369	58586	23216	14513	83149	98736	23495	64350	94738	17752	35156	35749
07056	97628	33787	09998	42698	06691	76988	13602	51851	46104	88916	19509	25625	58104
48663	91245	85828	14346	09172	30168	90229	04734	59193	22178	30421	61666	99904	32812
54164	58492	22421	74103	47070	25306	76468	26384	58151	06646	21524	15227	96909	44592
32639	32363	05597	24200	13363	38005	94342	28728	35806	06912	17012	64161	18296	22851
29334	27001	87637	87308	58731	00256	45834	15398	46557	41135	10367	07684	36188	18510
02488	33062	28834	07351	19731	92420	60952	61280	50001	67658	32586	86679	50720	94953
81525	72295	04839	96423	24878	82651	66566	14778	76797	14780	13300	87074	79666	95725
29676	20591	68086	26432	46901	20849	89768	81536	86645	12659	92259	57102	80428	25280
00742	57392	39064	66432	84673	40027	32832	61362	98947	96067	64760	64584	96096	98253
05366	04213	25669	26422	44407	44048	37937	63904	45766	66134	75470	66520	34693	90449
91921	26418	64117	94305	26766	25940	39972	22209	71500	64568	91402	42416	07844	69618
00582	04711	87917	77341	42206	35126	74087	99547	81817	42607	43808	76655	62028	76630
00725	69884	62797	56170	86324	88072	76222	36086	84637	93161	76038	65855	77919	88006
69011	65795	95876	55293	18988	27354	26575	08615	40801	59920	29841	80150	12777	48501
25976	57948	29888	88604	67917	48708	18912	82271	65424	69774	33611	54262	85963	03547
09763	83473	73577	12908	30883	18317	28290	35797	05998	41688	34952	37888	38917	88050
91567	42595	27958	30134	04024	86385	29880	99730	55536	84855	29080	09250	79656	73211
17955	56349	90999	49127	20044	59931	06115	20542	18059	02008	73708	83517	36103	42791
46503	18584	18845	49618	02304	51038	20655	58727	28168	15475	56942	53389	20562	87338
92157	89634	94824	78171	84610	82834	09922	25417	44137	84813	25555	21246	35509	20468
14577	62765	35605	81263	39667	47358	56873	56307	61607	49518	89656	20103	77490	18062
98427	07523	33362	64270	01638	92477	66969	98420	04880	45585	46565	04102	46880	45709
34914	63978	88720	82765	34476	17032	87589	40836	32427	70002	70663	88863	77775	69348
70060	28277	39475	46473	23219	53416	94970	25832	69975	94884	19661	72828	00102	66794
53976	54914	06990	67245	68350	82948	11398	42878	80287	88267	47363	46634	06541	97809
76072	29515	40980	07391	58745	25774	22987	80059	39911	96189	41151	14222	60697	59583
90725	52210	83974	29992	65831	38857	50490	83765	55657	14361	31720	57375	56228	41546
64364	67412	33339	31926	14883	24413	59744	92351	97473	89286	35931	04110	23726	51900
08962	00358	31662	25388	61642	34072	81249	35648	56891	69352	48373	45578	78547	81788
95012	68379	93526	70765	10592	04542	76463	54328	02349	17247	28865	14777	62730	92277
15664	10493	20492	38391	91132	21999	59516	81652	27195	48223	46751	22923	32261	85653

