DOCUMENT RESUME

ED 184 854 SE 030 502

AUTHOR Hirshfeld, Stephen F.; Bart, William M.
TITLE Algebraic Systems: Applications in the Behavioral and Social Sciences.
INSTITUTION Army Research Inst. for the Behavioral and Social Sciences, Alexandria, Va.
REPORT NO ARI-TR-P-78-6
PUB DATE Feb 78
NOTE 123p.

EDRS PRICE MF01/PC05 Plus Postage.
DESCRIPTORS *Algebra; *Behavioral Sciences; *College Mathematics; *Mathematical Applications; *Mathematics; Mathematics Education; Matrices; Psychology; Researchers; Scientists; *Social Sciences
IDENTIFIERS *Algebraic Systems

ABSTRACT

A variety of uses of algebra in the behavioral and social sciences is provided along with descriptions of several algebraic systems. This volume is intended to be a sourcebook for theoretical conceptualizations for professionals in the behavioral and social sciences. This publication with its emphasis on description, application, and utility should be a valuable aid to the behavioral and social science researcher. This book is presented in eight chapters. The first four chapters present the foundation material on algebraic concepts and should be read before attempting to examine the remaining chapters. Chapter 5 is concerned with the application of groups to psychology. Chapter 6 introduces rings and fields. Chapter 7 introduces another major algebraic system, a vector space. Chapter 8 is directed at the concept of matrix. (Author/MK)

ED 180 502 [side margin, rotated]

EU 18485 4 [side margin, rotated]

ARI TECHNICAL REPORT
P-78-6

ALGEBRAIC SYSTEMS: APPLICATIONS IN THE BEHAVIORAL AND SOCIAL SCIENCES

by

Stephen F. Hirshfeld
Army Research Institute for the Behavioral and Social Sciences

and

William M. Bart
University of Minnesota

February 1978

Prepared for

U.S. ARMY RESEARCH INSTITUTE
for the BEHAVIORAL and SOCIAL SCIENCES
5001 Eisenhower Avenue
Alexandria, Virginia 22333

Approved for public release; distribution unlimited.

U. S. ARMY RESEARCH INSTITUTE
FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency under the Jurisdiction of the Deputy Chief of Staff for Personnel

JOSEPH ZEIDNER
Technical Director

WILLIAM L. HAUSER
Colonel, US Army
Commander

NOTICES

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER P-78-6	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) ALGEBRAIC SYSTEMS: APPLICATIONS IN THE BEHAVIORAL AND SOCIAL SCIENCES		5. TYPE OF REPORT & PERIOD COVERED -- 6. PERFORMING ORG. REPORT NUMBER --
7. AUTHOR(s) Stephen F. Hirshfeld (ARI) and William M. Bart (University of Minnesota)		8. CONTRACT OR GRANT NUMBER(s) --
9. PERFORMING ORGANIZATION NAME AND ADDRESS U.S. Army Research Institute for the Behavioral and Social Sciences 5001 Eisenhower Avenue, Alexandria, VA 22333		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 2Q761102B74F
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Deputy Chief of Staff for Personnel Washington, DC 20310		12. REPORT DATE February 1978 13. NUMBER OF PAGES 119
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) --		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE --

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

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

--

18. SUPPLEMENTARY NOTES

--

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

Algebraic systems
Set theory

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

This publication is an introduction to the uses of algebraic systems in the behavioral and social sciences. Algebra has already manifested its utility in the physical sciences. For example, the algebraic theory of groups has been very useful to physicists in their formulation of quantum mechanics, and Boolean algebra has been crucial to computer science theorists in their exposition of digital circuit theory.

(Continued)

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

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Item 20 (Continued)

Behavioral and social scientists have tended to rely heavily on statistical analyses, because of their substantial applicability to behavioral and social science problems. However, there are certain basic limitations in applying statistical methods to these problem areas. Statistics cannot be used to describe formally the system of relationships within a class of phenomena. Statistical techniques can indicate levels of interactions among variables, but they cannot be used to depict the form or quality of these interactions.

Algebraic theory contains concepts and principles which can be used to articulate the structural properties of classes of behavioral phenomena. It refers to the study of classes of behavioral rule systems, each of which has a set of elements, operation(s) defined on the set, and rules determining certain interrelationships among elements and operations.

Algebra provides a language which is precise, intuitive, and formal. Algebraic systems have been used to synthesize separate models and theories. Synthesis of the proliferation of seemingly disparate and expanding bodies of behavioral science knowledge is greatly needed. Algebra as a field can become as useful to the behavioral and social scientist as statistics. Its utility will be most evident in the activities of description and conceptualization. As in the case with statistics, the use of algebra does not require any substantive theoretical commitments.

In this book, a variety of uses of algebra in the behavioral and social sciences is provided along with descriptions of several algebraic systems. This volume is intended to be a sourcebook for theoretical conceptualizations for professionals in the behavioral and social sciences. This publication with its emphasis on description, application, and utility should be a valuable aid to the behavioral and social science researcher.

This book is presented in eight chapters. The first four chapters present the foundational material on algebraic concepts and should be read before attempting to examine the remaining chapters. The following paragraphs provide a brief summary of the content of each chapter.

In chapter 1 the basic terminology and elementary concepts of set theory are introduced. The discussion presupposes no knowledge of mathematics; the explanations are presented in a quite thorough, yet highly intuitive manner. Ample examples are presented, many of them having direct psychological relevance.

We all have an intuitive idea of what is meant by a "relation." A relation reflects some type of association or connection between two entities. In order to be more precise in describing this vague idea of a bond between entities, a mathematical formulation of a relation is needed. Chapter 2 serves this function.

One of the most important ideas in all of mathematics is that of a function or mapping. This term is so fundamental that it is commonly used in most disciplines. Chapter 3 defines and discusses the role of functions in algebraic systems.

(Continued)

Item 20 (Continued)

A class of algebraic entities useful in psychology is groups. The presentation on groups will be made in chapters 4 and 5. Chapter 4 includes a discussion of the definition of a group and other related terms. Other key terms such as subgroups, generators, homorphisms, isomorphisms, and semigroups are introduced. The chapter concludes with examples of several important types of groups.

Chapter 5 is concerned with the application of groups to psychology. Examples are given from Piagetian theory, the theory of kinship relations, studies of measurement, perception, language, and automata theory.

Chapter 6 introduces rings and fields. It is a relatively short chapter, because presently there are very few applications of these concepts to psychology. Their applicability has not really been tested yet. In this chapter important terminology is defined and illustrated through example(s).

Chapter 7 introduces another major algebraic system. A vector space has structural similarities to the other systems already considered, but introduces a new operation. The value of particular vector spaces in statistical and measurement analysis of psychological phenomena has been recognized. Many of these techniques are based on vector space theory. The examination of vector spaces proceeds in two parts. Chapter 7 introduces the concept and discusses linear combinations, linear independence and dependence and bases.

Chapter 8 is directed at the concept of a matrix. The matrix is an excellent concept to conclude the book with, because it will be proved that the set of matrices may be used in defining a group, or ring, or a vector space, or under certain special conditions, in defining a field. This will serve as a review of the key structures introduced in the book. Matrices also are valuable to discuss because they have a wide range of applications outside of mathematics.

PREFACE

In dealing with Army problems over the years the Army Research Institute for the Behavioral and Social Sciences has always insisted on bringing to bear a scientific point of view. This point of view includes objectivity, use of theoretical models and their resulting hypotheses, reliance on empirical data rather than armchair estimates, and use of mathematical and statistical methods of analysis. In particular the Institute has drawn heavily upon the formal systems and methods found in the disciplines of psychometrics, statistics, linear algebra, probability theory, and operations research.

The current volume presents for behavioral scientists, both inside and outside of the Army, an introduction to another set of mathematical systems with potentially interesting applications. These systems, often referred to as "modern algebra" and here called "algebraic systems," have potential, not so much for purposes of data analysis, but rather for describing formally the system of relationships within a class of phenomena. As is typical of mathematical systems, the ideas and structures presented here have great power and generality. They could well be useful in constructing models of social and behavioral phenomena.

J. J. Mellinger

J. J. MELLINGER
Chief, Research Statistics
and Computer Science Office

J. E. Uhlaner

J. E. UHLANER
Technical Director, ARI and
Chief Psychologist, U.S. Army

ALGEBRAIC SYSTEMS: APPLICATIONS IN THE BEHAVIORAL AND SOCIAL SCIENCES

CONTENTS

ALGEBRAIC SYSTEMS: APPLICATIONS IN THE BEHAVIORAL AND SOCIAL SCIENCES

INTRODUCTION

This book is an introduction to the uses of the branch of mathematics called algebra in the behavioral sciences. Basically, there are three branches of mathematics--geometry, analysis, and algebra. Geometry is the field concerned with the properties and relationships of points, lines, angles, surfaces, and solids. Analysis is the field concerned with functions and limits and includes the calculus. Algebra is the field concerned with sets that have sums and/or products defined on their elements and includes arithmetic and set theory. As opposed to the preconceived views of many behavioral scientists, algebra is not merely the study of polynomials as a remembrance of high school algebra could effect. Of the three branches of mathematics, algebra is the most abstract and foundational branch.

Each of those branches has been found to have substantial utility. Geometry is a very useful field to architects and civil engineers. Analysis is probably the branch of mathematics that is used the most and this is reflected in the fact that training in the calculus is required for an education in practically every scientific and engineering area. Algebra, though being the most abstract branch of mathematics, has manifested its utility to the sciences in a variety of ways. For example, the algebraic theory of groups has been very useful to theoretical physicists in their formulation of quantum mechanics and Boolean algebra has been crucial to computer science theorists in their exposition of digital circuit theory. What is to be indicated is that algebra has a host of important uses in the behavioral sciences.

It may seem strange to one interested in the behavioral sciences that being acquainted with algebra would be an aid in his work. He may think that he studied algebra in high school and it was found to be useful in determining roots of quadratic equations and the like but it certainly is not the methodological tool chest that statistics is for research in the behavioral sciences. So why study, of all things, algebra?

Well, statistics is a field which has substantial applicability to the behavioral sciences. However, it does have limitations. Statistics cannot be used to describe formally the system of relationships within a class of phenomena in a manner that is as exacting and rich as algebra can. Though statistical techniques can be used to indicate the level of interaction of two or more variables, they cannot be used to depict the form or quality of that interaction. On the other hand, within the corpus of algebra there is a rich reservoir of concepts and principles which can be used to articulate the structural properties of classes of behavioral phenomena as it refers to the study of a wide class of rule-systems, each of which has a set of elements, operation(s) defined on the set, and rules determining certain interrelationships among elements and operations. Also this branch of mathematics is laden with concepts and principles as it is centuries old and has grown at an extraordinary rate in this century.

A property of algebra that is often overlooked is that it is quite natural. Much of our everyday thinking is in conformance with algebraic principles. To a great extent, algebra is a rigorous articulation and logical extension of patterns of reasoning that are common to people. For example, much of set theory is merely a formal exposition of modes of mental organization that are evidenced in everyday life. Thus, behavioral scientists may rightly view algebra not as an exotic, arbitrary, abstruse field but as a field which provides a meaningful discussion of patterns that are very immediate, common, and even obvious.

There is another quality of algebra that should be of interest to behavioral science devotees. Algebra, especially with the development of algebraic logic, provides a language which is very precise, primitive, and rich, and nearly perfect in its lucidity. Such a precise language should be of use to behavioral scientists.

Another property of algebra relates to one of its primary uses in mathematics. Elements of algebra such as its constituent systems and structures have been used to tie parts of mathematics together and to show how different entities in mathematics are interconnected and related. In other words, algebra has had and will continue to have a decisive synthesizing effect on the proliferating corpus of mathematics. Presently, the algebraic theory of categories is being used in this regard to depict the forms of integration amidst mathematical systems. It is contended that algebra, when applied properly, would have a similar influence in the behavioral sciences. Needless to say, synthesis is greatly needed in the behavioral sciences as most of the research in the behavioral sciences is directed to experimental analyses of theories and models and this emphasis on analysis has resulted in a proliferation of seemingly disparate and expanding bodies of behavioral science knowledge. For example, there is a variety of psychologies of school learning that have resulted in a multitude of empirical studies, many of which remain trivial and disconnected. With synthesis, more direction will be provided to allow for more research in the behavioral sciences.

Algebra is a field that should become as useful a field as statistics is to the behavioral scientist. Its greatest utility will be evident in the activities of description and conceptualization in the behavioral sciences. As is the case with statistics, the use of algebra does not require any substantive theoretical commitments. Thus, in the area of psychology, algebra should be as potentially useful in the areas of operant conditioning or associationistic psychology as it would be in the areas of cognitive developmental psychology from a Piagetian viewpoint.

Already important uses of algebra in the behavioral sciences have been made. For example, Jean Piaget, a noted pioneer in developmental psychology, has employed the algebraic theory of lattices to describe the system of cognitive processes proper to adolescence. Also, Noam Chomsky, seminal thinker in the psychology of language, has used algebraic concepts and principles to articulate the structural properties of grammars which refer to the systems underlying human linguistic capabilities.

In this book, a variety of uses of algebra in the behavioral sciences is provided along with descriptions of several algebraic systems. This volume is intended to be a sourcebook for theoretical conceptualizations for students and professionals in the behavioral sciences.

With the use of algebra, the physical sciences have made considerable progress--much more than the behavioral sciences. It is likely that the behavioral sciences can also make profound progress if it makes greater use of algebra. This volume with its emphasis on description and utility should be an aid in that endeavor to behavioral science students and professionals.

CHAPTER 1

SET THEORY

In this chapter the basic terminology and elementary notions of set theory are introduced. The discussion presupposes no knowledge of mathematics; the explanations will be presented in a quite thorough, yet highly intuitive manner. This discussion is not a rigorous study of axiomatic set theory, but rather a concise overview of a very elegant theory. There will be an ample number of examples, many of them psychologically relevant, to assist the reader in his understanding of what may at first be rather abstract material.

The idea basic to the entire text will be that of a set. The notion of a set will not be formally defined, but will be taken to mean any collection of entities, objects, or stimuli. These objects may have some common property, such as each object in a collection of objects is red, or there may be no apparent mutuality among the items. The individual objects belonging to the given set will be called elements. For example, a red triangle would be an element in a collection of red objects. A convention that will be adhered to throughout the book is to denote sets by capital letters, and use lower case letters to represent elements. If an element, x, belongs to a set A, we write $x \in A$. If x does not belong to A, then we write $x \notin A$. Suppose r represents a red triangle, s denotes a silver circle, and R is the set of all red objects, then $r \in R$, but $s \notin R$. The set consisting of no elements is called the null set and is denoted by ϕ. An example would be the set of all triangles with 360°.

We may indicate a set by listing all of its elements. In the case of infinite sets this is impossible, and often it is also inconvenient to list all the elements in a large finite set. In this case we use what is called the set builder notation. The following examples illustrate the situation.

Examples

1. If A consists of the numbers 1,2,3,4, and 5, then A may be written as $A = \{1,2,3,4,5\}$.

2. If B equals all the counting (natural) numbers from one to one hundred, then B may be denoted as $B = \{1,2,3,\ldots,100\}$, or equivalently, $B = \{n | n \text{ is a natural number and } 1 \leq n \leq 100\}$, which is read "B equals the set of all n, such that n is a natural number and 1 is less than or equal to n and also n is less than or equal to 100."

3. If C = {Connecticut, Rhode Island, Massachusetts, Vermont, New Hampshire, Maine}, then C may be more concisely represented as C = {x|x is a New England state}.

4. Suppose D = {signal learning, S-R learning, chaining, verbal association, multiple discrimination, concept learning, principle (rule) learning, problem solving}. A person familiar with Gagné's work would describe D by saying D = {x|x is one of the eight types of learning described by Gagné}.

In order to make comparisons between sets, we must first define the equality of two sets. Two sets A and B are <u>equal</u> if and only if whenever $x \in A$, then $x \in B$, and conversely whenever $x \in B$, then $x \in A$, i.e., when the two sets consist of the same elements. The set consisting of Hubert Humphrey and Walter Mondale is equal to the set of United States Senators from Minnesota, because both sets have exactly the same members. A is a <u>subset</u> of B if every element in A is also an element of B. This is denoted by $A \subset B$ or equivalently $B \supset A$. A is a <u>proper subset</u> of B if every element in A is in B and there exist additional elements in B not in A. This is equivalent to $A \subset B$ and $A \neq B$. We write this as $A \subsetneq B$. The set of states consisting of Vermont and Maine is a proper subset of the New England states. The reader should note that a second form of notation is also widely used. We could write $A \subseteq B$ to represent A is a subset of B, and write $A \subset B$ to represent A is a proper subset of B. Therefore, it is important to check which notation is being used in the text which is being read. Returning to our discussion, we see that we have an alternative definition for the equality of sets A and B. We may define $A = B$ if and only if $A \subset B$ and $B \subset A$.

Examples

1. Let E = The set of fifty states in the United States
   = {x|x is a state in the United States};

   let F = {all states in the United States having a location with an elevation of at least 3000 feet}.

   Then we may conclude that $F \subset E$, and more specifically that $F \subsetneq E$, because there exist states with highest elevation less than 3000 feet, e.g., New Jersey. We have Colorado $\in$ F, New York $\in$ F, and California $\in$ F. Hence, we could define a set G as G = {Colorado, New York, California}, where $G \subsetneq F$. But, for example, if H = {Colorado, New Jersey}, then $H \not\subset F$ even though $H \subsetneq E$. One final related example is designed to illustrate that a set of one element is not identical with that element. We can speak of Colorado in two ways, Colorado $\in$ E, or {Colorado} $\subsetneq$ E. In the first case we are talking about Colorado as an element of the set E; in the second case we are talking about Colorado as a set of one element which is included in but not equal to the set E.

2. A more mathematical example is the following:

Define $B = \{1/2\}$;
$R = \{\text{positive multiples of } 3\} = \{3,6,9,\ldots\}$;
$A = \{\text{positive multiples of } 2\} = \{2,4,6,8,\ldots\}$;
$I = \{\text{even natural numbers}\} = \{2,4,6,8,\ldots\}$; and
$N = \{\text{natural numbers}\}$.

Clearly, R, A, and I are proper subsets of N. Also, observe that $A = I$ since both sets have the same elements. $A \not\subset R$ because, for example, $4 \notin R$, also $R \not\subset A$ since $3 \notin A$. Therefore, it is often the case that when considering two sets, neither set is a subset of the other. It is also interesting to notice that $B \not\subset N$. This is an illustration where a set with infinitely many elements, such as A, is a subset of N and where sets with one element, such as B, are not a subset of N.

Basic Operations

Set theory would be of little worth if there were no ways of forming new sets from the given ones. We will define several operations on sets. The definitions will be for two sets, but they can be easily generalized to any finite number or an infinite number of sets.

Definition 1. The intersection of two sets, A and B, denoted $A \cap B$, is the set consisting of those elements belonging to both A and B. Symbolically, this may be expressed as

$$A \cap B = \{x \mid x \in A \text{ and } x \in B\}.$$

Two sets, A and B, are said to be disjoint or mutually exclusive if $A \cap B = \phi$.

We may generalize this definition. For three sets, C, D, E,

$$C \cap D \cap E = \{x \mid x \in C \text{ and } x \in D \text{ and } x \in E\}; \text{ and}$$

for N sets, $A_1, A_2, \ldots, A_N$,

$$\bigcap_{i=1}^{N} A_i = \{x \mid x \in A_i \text{ for every } i;\ i = 1,2,\ldots,N\}$$
$$= A_1 \cap A_2 \cap \ldots \cap A_N.$$

The intersection of two sets A and B may be pictorially represented by Venn diagrams as illustrated in Figure 1.

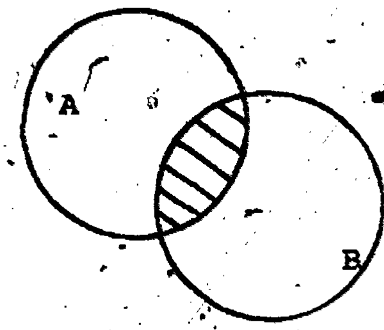

Figure 1

Examples

1. In discussions of psychological space, two stimuli are often considered to have a psychological distance between them. If the dimension of color, shape, and size are involved, where color is black or white, shape is triangle or square, and size is small or large, then if

$$\Large\square \normalsize = \{\text{white, square, large}\};$$

$$\blacksquare = \{\text{black, square, small}\};\ \text{and}$$

$$\square = \{\text{white, square, small}\};$$

it may be observed that $\Large\square$ and $\square$ are closer than $\Large\square$ and $\blacksquare$ in terms of psychological space, because they differ on only the dimension of size, i.e., their intersection shows a common color and shape. Therefore, any discussion of psychological distance between stimuli implies an understanding of the intersection operation.

2. Another illustration is in considering similarity between words. Suppose in a free association test, the subject is told to give five associations to words A, B, and C. If A and B have four common words, B and C have two common words, and A and C have one common word, then this would be one index of claiming there is greatest similarity between A and B. Notice that the consideration of commonality implicitly requires the use of the intersection operation.

<u>Definition 2</u>. The <u>union</u> of two sets, A and B, is the set consisting of elements belonging to A or to B or to both A and B. It is denoted by $A \cup B$, with $A \cup B = \{x | x \in A \text{ or } x \in B \text{ or } x \in A \text{ and } B\}$. The word "or" will be taken to include the possibility of membership in both sets. Thus, "or" will be interpreted in an inclusive manner. The union operation is pictorially described in Figure 2.

$A \cup B$

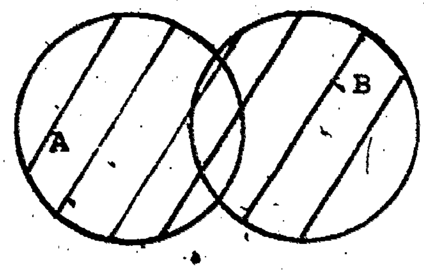

Figure 2

The definition may be generalized to N sets:

$$\bigcup_{i=1}^{N} A_i = A_1 \cup A_2 \cup \ldots \cup A_N = \{x \mid x \in A_i \text{ for some } i;\ i = 1, \ldots, N\}.$$

Example

1. If in one issue of psychological journal A, the contributors are Bruner, Gagné, Mandler, Piaget, and Simon, and in one issue of psychological journal B, the contributors are Berlyne, Elkind, Gagné, Jenkins, and Simon, then the set of contributors to the two journals would be Bruner, Gagné, Mandler, Piaget, Simon, Berlyne, Elkind, and Jenkins. This is precisely the union of the two table of contents in that it includes all those individuals in journal A, in journal B, or in both journals A and B.

Definition 3. The universal set $\mathcal{U}$ consists of all those elements under consideration. Then the complement of a set A, denoted $\tilde{A}$ or $\sim A$, consists of those elements in the universe that are not elements of A. The complement is represented pictorially in Figure 3.

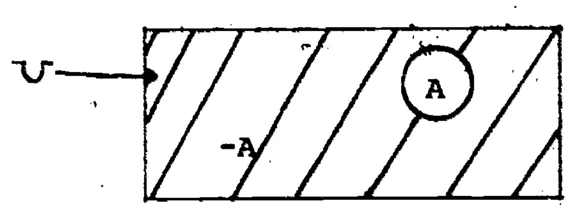

Figure 3

The set notational definition for the complement is

$$\tilde{A} = \{x \mid x \in \mathcal{U} \text{ and } x \notin A\};$$

but usually we just write

$$\tilde{A} = \{x \mid x \notin A\}.$$

Example

1. In an experiment with 100 subjects, 50 individuals receive treatment A, and the remaining 50 subjects form the control group. We may think of the 100 subjects as being the universal set $\mathcal{U}$, the 50 individuals receiving treatment A as set A, and those in the control group as $\tilde{A}$.

Definition 4. The difference of A and B, denoted A − B, is the set consisting of those elements belonging to A and not belonging to B. This operation is pictorially described in Figure 4.

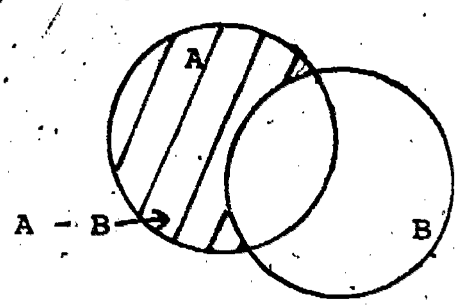

Figure 4

$$A - B = \{x \mid x \in A \text{ and } x \notin B\}$$

is the set notational definition for the difference operation. Clearly, we may say that $A - B = A \cap \tilde{B}$.

Example

1. The newspaper carries an advertisement that there is a job available for a person with a B.A. in psychology and specifies that the person must be under 30. If P represents all those individuals with a B.A. in psychology, and T denotes all those people 30 years old and over, then P − T consists of those individuals who meet the minimal qualifications for employment.

Definition 5. The symmetric difference of A and B is defined to be those elements in A, but not in B or those elements in B, but not in A. Figure 5 is a Venn diagram for the symmetric difference of A and B.

17

$A \Delta B$

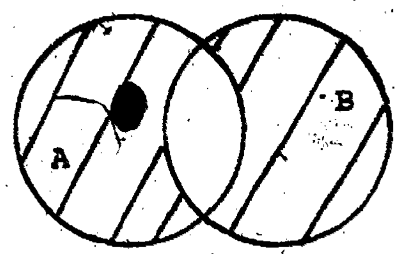

Figure 5

$$A \Delta B = \{x \mid x \in A \text{ and } x \notin B \text{ or } x \in B \text{ and } x \notin A\}.$$

Therefore,

$$A \Delta B = (A - B) \cup (B - A).$$

Example

1. In conditioning experiments a pigeon is rewarded if he pecks a key and is punished if he does not. Therefore, key pecking and punishment never go together. If K denotes the times a pigeon pecked the key, and P represents the times the pigeon was punished, then $K \Delta P$ would describe the principle involved in conditioning, i.e., if the pigeon pecks the key he is not punished and if he is punished he did not peck the key.

As a means of reviewing and interrelating the five definitions just given, an example with sets of numbers is included.

Examples

1. Let $\mathcal{U} = \{1,2,3,4,5,6,7,8,9,10,11,12,13,14,15\}$;
   $A = \{2,4,6,8,10,12,14\}$;
   $B = \{1,6,11\}$;
   $C = \{3,6,9,12,15\}$.

   Then, we have

   $A \cap B = \{6\}$;
   $B \cup C = \{1,3,6,9,11,12,15\}$;
   $\tilde{A} = \{1,3,5,7,9,11,13,15\}$;
   $A \cap (B \cup C) = \{2,3,6,8,10,12,14\} \cap \{1,3,6,9,11,12,15\} = \{6,12\}$;
   $A - B = A \cap \tilde{B} = \{2,4,6,8,10,12,14\} \cap \{2,3,4,5,7,8,9,10,12,13,14,15\}$
   $= \{2,4,8,10,12,14\}$;
   $B \Delta C = (B - C) \cup (C - B) = \{1,11\} \cup \{3,9,12,15\} = \{1,3,9,11,12,15\}$.

2. With the 1976 presidential election approaching there is much conjecture as to whom the Democratic Party will nominate. In determining who the nominee will be, each candidate will be weighed as to his strengths and weaknesses on various personal qualities, political views, electability, etc. It would be an interesting exercise to define or compile a list of criteria most desirable to you. Then derive a rating system involving union, intersection, and complementation to evaluate each contender, and see if your personal choice and your highest rated individual are the same person.

In order to aid the reader in his understanding of the set theoretic terminology, several elementary proofs using the new abstract language are included. The analogy to learning a language is a meaningful one, because for a person to really understand a word, he must be able to use it in appropriate situations. The following proofs serve a similar purpose for the words, intersection, union, complementation, difference, and symmetric difference.

Proofs

1. $A \subset B$ if and only if $A \cap B = A$.

Proof: One must first realize that an "if and only if" proof requires two proofs. We must show that $A \subset B$ implies that $A \cap B = A$, and also that if $A \cap B = A$, then $A \subset B$. As an aid in the proof, draw a Venn diagram similar to the one in Figure 6.

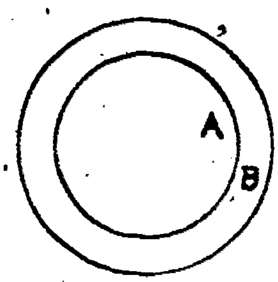

Figure 6

In proving results such as this one, accompanying pictures may aid in visualizing the problem, but one must realize that even if a picture's intuitive worth may be a thousand words, it is not a formal proof. With this thought in mind we begin the proof.

a. Suppose $A \subset B$.

(i) Let $x \in A$, then by hypothesis $x \in B$, and we have $x \in A$ and $x \in B$. Therefore, we have $x \in A \cap B$, but $x \in A$ implying $x \in A \cap B$ is precisely $A \subset A \cap B$.

(ii) Let $x \in A \cap B$, then we have $x \in A$ and $x \in B$, but in particular $x \in A$. Hence, $x \in A \cap B$ implies $x \in A$, or equivalently $A \cap B \subset A$.

Combining (i) and (ii), $A \subset A \cap B$ and $A \cap B \subset A$, yield that $A \cap B = A$, which is the desired result.

b. Suppose $A \cap B = A$.

If $A \cap B = A$, we may say that $A \subset A \cap B$. Further, by the definition of intersection $A \cap B \subset B$, and thus we have $A \subset A \cap B \subset B$, which implies $A \subset B$.

2. $-(A \cup B) = \tilde{A} \cap \tilde{B}$.

Proof: a. If $x \in -(A \cup B)$, then $x \notin (A \cup B)$, which means that $x \notin A$ and $x \notin B$, since $x \in A \cup B$ if x belongs to either A or B or both of them. But $x \notin A$ and $x \notin B$ is equivalent to $x \in \tilde{A}$ and $x \in \tilde{B}$, which by definition is $x \in \tilde{A} \cap \tilde{B}$.

b. If $x \in \tilde{A} \cap \tilde{B}$, then $x \in \tilde{A}$ and $x \in \tilde{B}$, which implies $x \notin A$ and $x \notin B$, from which it follows that x cannot belong to the union of A and B, i.e., $x \notin (A \cup B)$ or $x \notin -(A \cup B)$. A picture of this is presented in Figure 7.

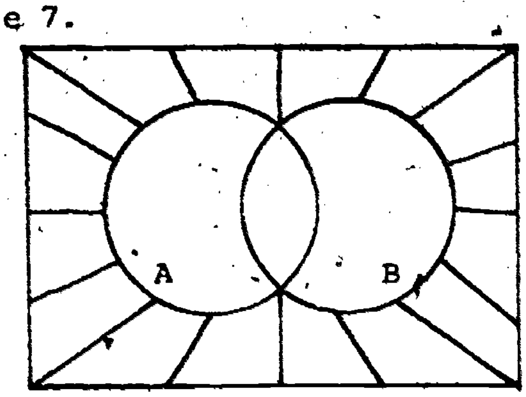

Figure 7

3. $A \Delta B = B \Delta A$.

Proof: By definition $A \Delta B = (A - B) \cup (B - A)$, which equals $(B - A) \cup (A - B)$ which equals $B \Delta A$. Therefore, $A \Delta B = B \Delta A$. Notice that for sets P, Q, $P \cup Q = Q \cup P$. The reader should prove this, because in the proof of $A \Delta B = B \Delta A$ it was needed. In this proof P corresponds to $(A - B)$ and Q corresponds to $(B - A)$. Figure 8 demonstrates this pictorially.

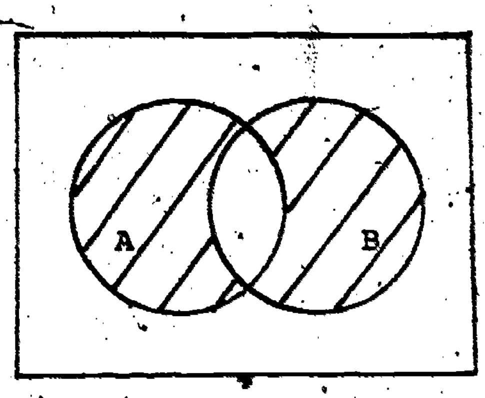

Figure 8

4. $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.

Proof: a. Let $x \in A \cup (B \cap C)$, then $x \in A$ or $x \in (B \cap C)$.

(i) If $x \in A$ and $x \in B \cap C$, this implies $x \in A$, $x \in B$, $x \in C$, and clearly $x \in A \cup B$ and $x \in B \cup C$, i.e., $x \in (A \cup B) \cap (A \cup C)$.

(ii) If $x \in A$ and $x \notin (B \cap C)$, then $x \in (A \cup B)$ and $x \in (A \cup C)$, regardless of whether $x \in B$ or $x \in C$, and therefore, $x \in (A \cup B) \cap (A \cup C)$.

(iii) If $x \notin A$ and $x \in (B \cap C)$, then $x \notin A$, $x \in B$, $x \in C$, so again $x \in (A \cup B)$ and $x \in (A \cup C)$, or equivalently $x \in (A \cup B) \cap (A \cup C)$.

b. Let $x \in (A \cup B) \cap (A \cup C)$, then $x \in (A \cup B)$ and $x \in (A \cup C)$; $x \in A \cup B$ implies that $x \in A$ or $x \in B$ and similarly, $x \in A \cup C$ implies that $x \in A$ or $x \in C$.

(i) If $x \in A$, $x \in B$, $x \in C$, clearly $x \in A \cup (B \cap C)$. In fact, if $x \in A$, then $x \in A \cup (B \cap C)$, regardless of whether $x \in B$ or $x \in C$.

(ii) Suppose $x \notin A$, then we must have $x \in B$ and $x \in C$, since $x \in A \cup B$ and $x \in A \cup C$. Therefore, $x \in B \cap C$, and finally, $x \in A \cup (B \cap C)$. A picture of this is presented in Figure 9.

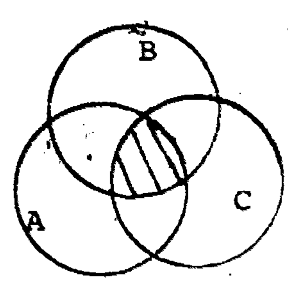

Figure 9

The first chapter introduced the fundamental idea of a set, its terminology, and the types of operations that may be performed on sets. The fundamental nature of a set should be clear from the many and varied uses of it in this chapter. Sets of numbers, red objects, states, types of learning, United States Senators, states with elevations above 3000 feet, words, noted psychologists, people over 30, etc., were considered. The rich diversity of areas covered is an illustration of the generalizability of the term. The operations on sets, such as union and intersection, allow us to generate new sets or describe sets with more specific properties.

An appropriate way of concluding the chapter is to review the set theoretic terminology in relation to the problem of concept learning. In concept learning, an individual who knows a given concept, say redness, can be shown a collection of stimulus objects, i.e., a set of stimuli, and can determine which stimuli are exemplars of the concept redness. He will select only those objects that are red in color. He is manifesting an understanding of the operation of intersection, because each of these objects is individually red. Those objects that are not red are nonexemplars, and require the application of complementation. If a second concept is introduced, say triangle, and the individual is asked to choose all objects that are red or triangles, then he will select those stimuli that are red, are triangles, or are both red and triangles. This would refer to a grasp of the operation of union. To find all the objects that are red, but not triangles utilizes the difference operation. Finally, in choosing objects that are red, but not triangles, or objects that are triangles, but not red, the operations of symmetric difference is referred to. Interesting research is being carried out to determine if there exists a hierarchy of difficulty among operations such as those just described in this chapter.

CHAPTER 2

RELATIONS

There is one further operation involving sets that we would like to consider. The notion of a Cartesian product of two sets will be fundamental to this chapter. We will need to introduce the notion of an ordered pair. We will take an ordered pair to be two objects given in a fixed order. Therefore, (a,b) is generally not equal to (b,a). If the first position represents the number of ten dollar bills in your wallet, and the second position the number of one dollar bills in your wallet, then Bob has (7,2) and Dave (2,7), this means that Bob has $72 and Dave has $27, which certainly are not the same. We could define an ordered pair in a more formal manner, but an intuitive idea of the concept will suffice for our purposes. The ordered pairs (a,b) and (c,d) will be equal if and only if $a = c$ and $b = d$.

Definition 6. The Cartesian product of two sets A and B is defined to be the set of all ordered pairs, (a,b), such that $a \in A$ and $b \in B$, and is written A X B. $A \times B = \{(a,b) \mid a \in A \text{ and } b \in B\}$.

Examples

1. Let $A = \{1,2,3\}$ and $B = \{0,5,10,-2\}$, then

$$A \times B = \{(a,b) \mid a \in A \text{ and } b \in B\}$$

$$= \{(1,0), (1,5), (1,10), (1,-2), (2,0), (2,5), (2,10), (2,-2), (3,0), (3,5), (3,10), (3,-2)\};$$

$$A \times A = \{(1,1), (1,2), (1,3), (2,1), (2,2), (2,3), (3,1), (3,2), (3,3)\}.$$

2. Any graphical data from a psychological experiment may be interpreted in terms of ordered pairs. For example, in an intelligence test, each individual has a particular score, or in a discussion of S-R theory, the theory is described in terms of stimulus-response pairs called associations.

We all have an intuitive idea of what we mean by a "relation." A "relation" reflects some type of association or connection between two entities. In order to be more precise in describing this vague idea of a bond between entities, we want a mathematical formulation of a relation. Two objects either have this defined bond or they do not. Therefore, we can enumerate the set of all ordered pairs of entities having this bond. Thus, we may think of a relation as a collection of ordered pairs.

Definition 7. Let A and B be sets, then a relation R on the Cartesian product $A \times B$ is any subset of $A \times B$, i.e., $R \subset A \times B$. If (a,b) is an element of the collection of ordered pairs determining R, then we may either write $(a,b) \in R$ or $a\,R\,b$.

Examples

1. If $A = \{1,2,3,4,5,6,7\}$ and $B = \{3,6,8,11,13,14,19,22\}$, then the Cartesian product $A \times B$ has 56 ordered pairs. Examples of relations would be,

$$R_1 = \{(1,3),\ (5,19),\ (7,6)\}$$

$$R_2 = \{(3,6),\ (4,8),\ (7,14)\}$$

$$R_3 = \{(1,3),\ (2,4),\ (4,6),\ (6,8)\}$$

$$R_4 = \{(1,8),\ (2,13),\ (5,8),\ (1,6)\}.$$

   As may be observed, not all relations have a clear connection between elements in the ordered pairs. Often it is impossible to come up with a rule defining the relation. In R_2 we may observe that the ordered pair satisfies $b = 2a$, but R_1 does not have any such well-defined bond.

2. The notion of a relation has wide applicability. For example, any verb phrase in a sentence indicates a relation. Consider the set A to be composed of the cow, the moon, and the Pied Piper. Let our relation R be designated by "jumped over." Now, only the cow jumped over the moon, and no other elements in A are related by "jumped over"; thus, the ordered pair (the cow, the moon) in $A \times A$ determines our relation R. Notice that (the cow, the moon) is not the same as (the moon, the cow), the latter being the moon jumped over the cow.

Properties of Relations

We will discuss various important properties of relations on $A \times A$, i.e., $A \times A = \{(a_1,a_2) \mid a_1 \in A \text{ and } a_2 \in A\}$.

24

Definition 8.

(i) Let A be a set and R a relation, i.e., $R \subset A \times A$, then R is reflexive if for every $a \in A$, $(a,a) \in R$.

(ii) R is irreflexive if for every $a \in A$, $(a,a) \notin R$.

(iii) If R is neither reflexive or irreflexive, then R is called nonreflexive.

Examples

1. Equality "=" in the discussion of numbers is reflexive, because for every number, it is equal to itself.

2. The relation "is less than," "<," is irreflexive, since for every number, it is not less than itself. A more concrete example is the relation "weighs less than." Even though many dieters wish it were true, no one weighs less than himself, so "weighs less than" is irreflexive.

3. However, the relation "is less than or equal to," "≤," is reflexive, since for instance, every number is less than or equal to itself.

4. Another irreflexive relation is being a mother, because no one is her own mother.

5. In comparisons such as "is as intelligent as," "is as kind as," "is as tall as," etc., we have examples of reflexive relations from everyday language.

6. There exist relations that are neither reflexive nor irreflexive. Let $A = \{x,y\}$, hence $A \times A = \{(x,x), (x,y), (y,x), (y,y)\}$. Define $R = \{(x,x), (x,y)\}$, then we may observe that R is not reflexive because $(y,y) \notin R$, and R is not irreflexive because $(x,x) \in R$. Therefore, R is nonreflexive.

Definition 9.

(i) Let A be a set and $R \subset A \times A$, then R is symmetric if for every $a,b \in A$, $(a,b) \in R$ implies $(b,a) \in R$.

(ii) R is asymmetric if for every $a,b \in A$, $(a,b) \in R$ implies $(b,a) \notin R$.

(iii) R is antisymmetric if for every $a,b \in A$, whenever $(a,b) \in R$ and $(b,a) \in R$, then $a = b$.

Examples

1. Equality is symmetric, since if $a = b$, then clearly $b = a$.

2. However, "is less than," "<" is not symmetric, since for example, $5 < 6$ does not imply $6 < 5$. Actually, "<" is asymmetric.

3. An example of an antisymmetric relation is less than or equal to, "$\leq$." It is neither symmetric or asymmetric. Because $5 \leq 6$, but $6 \leq 5$, we see that "$\leq$" is not symmetric. Further, since $5 \leq 5$ implies $5 \leq 5$, "$\leq$" is not asymmetric. "$\leq$" is antisymmetric because the only way one number can be both greater than or equal to, and less than or equal to another number is if it equals that number.

4. "$\subset$" or "is included in" is another example of an antisymmetric relation. We made use of this assertion in several proofs in Chapter 1. In proving that two sets were equal, for instance, $A = B$, we proved that $A \subset B$ and that $B \subset A$, from which we concluded that $A = B$.

5. An example that each of us can identify with is the relation "loves." Sam loves Sally, but Sally does not love Sam. Poor Sam, loving is not symmetric. Actually loving is not symmetric or asymmetric or antisymmetric. It is not asymmetric, since fortunately for us all, there exist cases where, for example, Romeo loves Juliet, and Juliet loves Romeo. Loving is not antisymmetric, since this would imply that if one person loves a second person, and conversely, then the two people must be the same person. This would mean a world without any couples in love. Romanticism aside, the relation "loves" would be an instance of a nonsymmetric relation.

6. More concrete examples of a symmetric relation would be "is exactly as tall as," "is exactly as intelligent as," etc., while relations such as "is taller than" and "weighs more than" would be asymmetric.

7. The relation "is the next door neighbor of" is an example of a symmetric relation, since if Jones lives next door to Smith, Smith lives next door to Jones.

8. A psychological example of an asymmetric relation would be "is reinforced if he chooses." In a particular trial an individual is reinforced if he makes the correct choice, and is not reinforced if he makes the wrong choice. Suppose A is correct and B is incorrect, then Tom is reinforced if he chooses A over B, but Tom is not reinforced if he chooses B over A.

Definition 10. Let A be a set and $R \subset A \times A$, then

(i) R is transitive if for every $a,b,c \in A$, if (a,b) and $(b,c) \in R$, then $(a,c) \in R$;

(ii) R is <u>intransitive</u> if for every $a,b,c \in A$, if $(a,b) \in R$ and $(b,c) \in R$, then it is not the case that $(a,c) \in R$;

(iii) If R is neither transitive or intransitive, then R is <u>nontransitive</u>.

Examples

1. Equality is transitive. If $a = b$ and $b = c$, then we have $a = c$.

2. Another transitive relation is "is less than," "<," for is $a < b$ and $b < c$, then $a < c$.

3. Set inclusion, "$\subset$," is transitive. If $A \subset B$ and $B \subset C$, then $A \subset C$.

4. Returning to our discussion of love, if Sam loves Sally and Sally loves Jim, it is most unlikely that Sam loves Jim. However, under some conditions Sam may love Jim. Therefore, "loves" is a nontransitive relation.

5. If Ann is Mary's mother, and Mary is Betty's mother, this does not imply that Ann is Betty's mother. "Is the mother of" is an example of an intransitive relation.

6. The height of people designates many relations. For example, "is taller than" is transitive. If Tom is taller than Dick, and Dick is taller than Harry, then Tom is taller than Harry.

7. Suppose $R = \{(1,2), (2,3), (3,4), (2,4)\}$, then R is not transitive. because $(1,2) \in R$ and $(2,3) \in R$, but $(1,3) \notin R$. Also, R is not intransitive, since $(2,3) \in R$ and $(3,4) \in R$, but $(2,4) \in R$, contrary to the definition of intransitivity. Therefore, R is nontransitive.

8. Piaget describes four levels of operations in his theory; sensori-motor, pre-operational, concrete, and formal. The ages of transition to a higher level may vary, but the order is fixed. Therefore, the relation "is a prerequisite to" is an example of a transitive relation. If sensori-motor operations are prerequisite to concrete operations, and concrete operations are prerequisite to formal operations, then sensori-motor operations are prerequisite to formal operations.

Those properties of relations discussed in Definitions 8, 9, and 10 are the ones we are most interested in, but for completeness we will include several additional ones.

<u>Definition 11</u>. If A is a set and $R \subset A \times A$, then R is <u>connected</u> if for every $a,b \in A$, whenever $a \neq b$, then $(a,b) \in R$ or $(b,a) \in R$.

Examples

1. The relation "is less than," "<" is connected. If $a \neq b$, then $a < b$ or $b < a$.

2. Set inclusion is not connected. If $A \neq B$, it is not necessarily the case that $A \subset B$ or $B \subset A$. It is possible that $A \cap B = \phi$, or that we do not have inclusion, but rather partial overlap.

3. The relation "loves" is not connected, because it is conceivable that Alan does not love Ellen, and Ellen does not love Alan.

Definition 12. If A is a set and $R \subset A \times A$, then R is circular if $(a,b) \in R$ and $(b,c) \in R$ imply that $(c,a) \in R$.

Examples

1. Equality is a circular relation. If $a = b$ and $b = c$, then $c = a$.

2. The relation "is a sibling of" is another circular relation. If Fred is a sibling of Harvey and Harvey is a sibling of Morty, then Morty is a sibling of Fred.

3. Proper set inclusion is an example of a relation that is not circular. If $A \subsetneq B$ and $B \subsetneq C$, it is not true that $C \subsetneq A$.

For those readers who would like to see the newly introduced properties used in a more formal way, the following two problems are included.

Problem 1. Suppose that a relation R is transitive and symmetric. Give an example to show that R need not necessarily be reflexive.

One may try to argue as follows: For $a, b \in A$, by symmetry $(a,b) \in R$ implies that $(b,a) \in R$. But if $(a,b) \in R$ and $(b,a) \in R$, then by transitivity $(a,a) \in R$, from which it is tempting to conclude that R is reflexive. We must investigate why the above argument is fallacious. Let $A = \{a,b\}$ and $R = \{(b,b)\}$. In this example (b,b) is the only element in R. R is not reflexive because $(a,a) \notin R$, and for R to be reflexive both (a,a) and (b,b) must belong to R. However, R is trivially symmetric and transitive.

Problem 2. Show that a relation is reflexive and circular if and only if it is reflexive, symmetric, and transitive. (This is a problem in A Survey of Modern Algebra by Birkhoff and MacLane, 1964.)

Proof: (i) Suppose R is reflexive and circular. Therefore, for every $a,b,c \in A$, $(a,a) \in R$, and further $(a,b) \in R$ and $(b,c) \in R$ implies that $(c,a) \in R$ by circularity.

Show R is symmetric. Suppose $(a,b) \in R$; by the reflexivity of R, $(b,b) \in R$ as well. Now by circularity, $(a,b) \in R$ and $(b,b) \in R$ imply $(b,a) \in R$. Thus, $(a,b) \in R$ implies $(b,a) \in R$.

Show R is transitive. If $(a,b) \in R$ and $(b,c) \in R$, by circularity we have $(c,a) \in R$, but by the just proven symmetry, it follows that $(a,c) \in R$. Hence, if $(a,b) \in R$, and $(b,c) \in R$, then $(a,c) \in R$.

(ii) Suppose R is reflexive, symmetric, and transitive. We must show only that R is circular, since it is given that R is already reflexive. If $(a,b) \in R$ and $(b,c) \in R$, then $(a,c) \in R$ by the transitivity. Next by symmetry $(a,c) \in R$ implies that $(c,a) \in R$. Therefore, $(a,b) \in R$ and $(b,c) \in R$ imply that $(c,a) \in R$.

To help clarify the descriptive capabilities of the properties that have been discussed, Table 1 has been constructed to indicate the properties of ten relations. In Table 1 a set of elements for which a cited relation is to be operative is indicated for each relation. The relations in Table 1 tend to fall into groupings according to their properties. Some relations such as "equals" and "is exactly as kind as" are reflexive, symmetric, and transitive. Relations with those properties are termed equivalence relations. Other relations such as "is greater than," "weighs more than," and "is less intelligent than" are irreflexive, asymmetric, and transitive. With the twelve properties cited in Table 1 the logical qualities of any relation can be richly articulated.

Equivalence Relations

Definition 13. If A is a set and $R \subset A \times A$, then R is an *equivalence relation* if:

(i) For every $a \in A$, $(a,a) \in R$ (reflexivity);

(ii) For every $a,b \in A$, $(a,b) \in R$ implies $(b,a) \in R$ (symmetry);

(iii) For every $a,b,c \in A$, $(a,b) \in R$ and $(b,c) \in R$ imply $(a,c) \in R$ (transitivity).

Examples

1. Equality for numbers is an equivalence relation, because "=" is reflexive, symmetric, and transitive.

Table 1

A Classification of Some Relations by Their Order Properties

Set	Relation	reflexive	irreflexive	nonreflexive	symmetric	asymmetric	antisymmetric	nonsymmetric	transitive	intransitive	nontransitive	connected	circular
1. real numbers	=	yes			yes		yes		yes				yes
2. real numbers	>		yes			yes			yes				
3. real numbers	≥	yes					yes		yes			yes	
8. sets of people	includes (c)	yes					yes		yes				
9. women	is a mother of		yes			yes				yes			
0. people	is a sibling of		yes		yes				yes				yes
1. buildings	has a different height than		yes		yes						yes		
2. Americans	loves			yes				yes			yes		
3. Americans	weighs more than		yes			yes			yes				
4. people	is as kind as	yes			yes				yes				yes
5. people	is less intelligent than		yes			yes			yes				
4. [1,2,3,4]	[(1,2),(2,3),(1,3),(1,4),(2,4),(3,4)]		yes			yes			yes			yes	
5. [1,2,3,4]	[(1,1),(1,2),(2,3),(2,4),(3,4)]			yes		yes					yes		
6. natural numbers	[(1,1),(2,1),(1,2)]			yes	yes						yes		
7. natural numbers	[(1,1)]			yes	yes		yes		yes				yes

2. The relation "is less than" is not an equivalence relation because "<" is not symmetric.

3. Let the states of the United States form the set under consideration. Then we could define a relation R by $(x,y) \in R$, where x,y are states, if both states x and y have governors whose last names begin with the same letter. For example, if the letter was S, the states would include Massachusetts (Sargent), Pennsylvania (Shapp), Texas (Smith), etc. The relation would be reflexive, because for instance (Texas, Texas) $\in$ R. The relations would be symmetric, because if for instance (Texas, Pennsylvania) $\in$ R, then (Pennsylvania, Texas) $\in$ R. The relation is transitive. Consider, if we look at (Texas, Pennsylvania) $\in$ R and (Pennsylvania, Massachusetts) $\in$ R, then (Texas, Massachusetts) $\in$ R. Therefore, R as defined above would be an equivalence relation.

   Actually we would be able to divide the states up into mutually disjoint groupings because each state would fall into only one category, depending on the initial letter of the state's governor's name. Granted that this particular partitioning does not reflect any real division according to national importance of political ideology of the individual governors, but it is an example of how we can often divide a collection of items or people into disjoint subcollections with each subcollection representative of some unique property. The actual significance of such a representation depends on the importance or value of the defined relation. We will follow up this idea of partitioning in a more precise and mathematical presentation later in the book.

4. Let $\mathbb{Z}$ be the set of all integers, i.e., $\mathbb{Z} = \{\ldots,-3,-2,-1,0,1,2,3,\ldots\}$. Define for $m,n \in \mathbb{Z}$, $(m,n) \in R$ if $m - n$ is a multiple of 5, i.e., if $m - n = 5t$ for some integer t. R is an equivalence relation.

   (i) $(m,m) \in R$ for every $m \in \mathbb{Z}$, because $m - m = 0 = 5(0)$, where $0 \in \mathbb{Z}$. Therefore, R is reflexive.

   (ii) If $(m,n) \in R$, then there exists an integer t such that $m - n = 5t$. Therefore $n - m = -5t$, but $-5t = 5(-t)$ and $-t$ is an integer. Hence, $(n,m) \in R$ and R is symmetric.

   (iii) If $(m,n) \in R$ and $(n,p) \in R$, then for some integers k and j, we have $m - n = 5k$ and $n - p = 5j$. Therefore, $m - p = (m - n) + (n - p) = 5k + 5j = 5(k + j) = 5i$, where $i = k + j$ is some integer. Hence $(m,p) \in R$ and R is transitive.

5. The next example will at first appear to be quite difficult, but at a closer inspection, it may be observed that we are merely establishing the equivalence of fractions such as 2/5, 4/10, 10/25, etc., by stating that the product of the means equals the product

of the extremes. For example, $2/5 = 4/10$ because $2(10) = 5(4)$. Now to the example, let $a,b,c,d \in \mathbb{Z}$, and let M = the set of all ordered pairs of integers (a,b) with $b \neq 0$. Define R as $((a,b), (c,d)) \in R$ if and only if $ad = bc$. (Notice this is the same as saying $((2,5), (4,10)) \in R$ if and only if $2(10) = 4(5)$.

It must be shown that R is an equivalence relation.

(i) If $(a,b) \in M$, then $((a,b), (a,b))$ is an element of R, because $ab = ba$. Thus, we have proven that R is reflexive.

(ii) If $(a,b) \in M$ and $(c,d) \in M$, and suppose further that $((a,b), (c,d)) \in R$, then by definition we have $ad = bc$, which by rearrangement implies $cb = da$, and therefore, $((c,d), (a,b)) \in R$, and symmetry has been demonstrated.

(iii) Let (a,b), (c,d), and (e,f) be elements of M, and suppose that $((a,b), (c,d)) \in R$ and $((c,d), (e,f)) \in R$, then we have that $ad = bc$ and $cf = de$. Therefore, upon multiplying $ad = bc$ by f we obtain $adf = bcf$, and multiplying of $cf = de$ by b, we obtain $bcf = bde$. Hence, $adf = bcf$ and $bcf = bde$, and by the transitivity of the equality relation, we have $adf = bde$, which we may rewrite as $afd = bed$. By hypothesis $d \neq 0$, and therefore $d^{-1} = 1/d$ exists. Multiplying both sides of the equality $afd = bed$ by d^{-1} we obtain $af = de$, i.e., $((a,b), (e,f)) \in R$. Hence, R is transitive.

In the example about states having governors whose last names begin with the same letter, a brief description was included of how the states could be broken up into disjoint groupings. This is a very valuable procedure in considering sets, and will be now presented in a more thorough manner.

Definition 14. Let A be a set and $R \subset A \times A$, then the equivalence class of a A is the set, $\{x \in A \mid (a,x) \in R\}$, which we shall denote by either [a] or c l (a).

Examples

1. We have already shown that equality is an equivalence relation. If $a \in A$, then $[a] = \{a\}$, since $(a,x) \in R$ if and only if $a = x$.

2. Let $\mathbb{Z}$ be the set of all integers. Define for $m,n \in \mathbb{Z}$, $(m,n) \in R$ if $m - n$ is a multiple of 5. We demonstrated already that R is an equivalence relation. Then for $a \in \mathbb{Z}$,

$$[a] = \{x \in A \mid (a,x) \in R\}$$

$$= \{x \in A \mid a - x = 5i,\ i = 0, \pm 1, \pm 2, \cdots\}.$$

Therefore, we may divide the integers into five distinct equivalence classes, [0], [1], [2], [3], [4]. Note that, for example [2] = [7], since both are equal to $\{\cdots,-8,-3,2,7,12,\cdots\}$. Thus, we may divide the integers according to whether the remainder is 0,1,2,3, or 4 upon division by 5. Hence, $\mathbb{Z} = \{[0], [1], [2], [3], [4]\}$.

3. Joseph Scandura has created a new language to describe what (rule) is learned in a particular task. He calls this language set function language (SFL), and it utilizes the term equivalence class. His theory is as follows: A particular stimulus is observed, and it is then assigned to the appropriate class of stimuli, on the basis of its defining properties. A rule is then a mapping from a class of stimuli to a class of responses, from which the required response is selected from the class of functionally equivalent responses. An example would be the following. Let [1 + 3 + 5] consist of elements such as 1 apple + 3 apples + 5 apples, $1 + $3 + $5, 1 dot + 3 dots + 5 dots, etc. Let [9] consist of elements such as 9 apples, $9, 9 dots, etc. Then the rule is an operation between equivalence classes of number series and their sums. An example, in computing $1 + $3 + $5, it is first recognized as an instance of [1 + 3 + 5] which by a rule is mapped into [9], from which the appropriate response $9 is selected. A more detailed account of the theory may be found in Scandura (1970).

4. Equivalence classes are used as part of the underlying structure in a paper by David H. Krantz (1964), "Conjoint Measurement: The Luce-Tukey Axiomatization and Some Extensions."

In Example 2 it was shown that the integers could be divided up into five equivalence classes, [0], [1], [2], [3], and [4]. This process of dividing up a set is referred to as a partition.

Definition 15. A partition of a set A is a collection of nonempty subsets of A that are disjoint and whose union is A.

There is a theorem in mathematics that describes the relationship between the equivalence classes of an equivalence relation and a partition of a set. This theorem is most useful in mathematics, because of the way it allows a set to be divided into meaningful subsets. It can be equally valuable in psychology as a means of dividing up experimental data, stimuli, concepts, etc. into important, distinctive subcategories. The theorem will not be proven in this book, but may be found in any standard abstract algebra book, such as Herstein (1964) or Dean (1966).

Theorem (i) The distinct equivalence classes of an equivalence relation on A provide us with a partition of A; i.e., they provide us with a decomposition of A into mutually disjoint nonempty subsets whose union equals A.

(ii) Conversely, given a partition of A into mutually disjoint nonempty subsets, we can define an equivalence relation on A, for which these subsets are the distinct equivalence classes.

Examples

1. We have already discussed that the integers may be divided into [0], [1], [2], [3], and [4], if the relation is, for $m,n \in \mathbb{Z}$, $(m,n) \in R$ if $m - n$ is a multiple of 5. That is, we divided the integers according to whether the remainder was found to be 0, 1, 2, 3, or 4, upon division by 5.

2. If the relation had been, for $m,n \in \mathbb{Z}$, $(m,n) \in R$ if $m - n$ is a multiple of 7. Then the integers would have been partitioned into [0], [1], [2], [3], [4], [5], and [6].

3. In a used car lot, if the owner divides his cars into groupings, where all the cars in one grouping are one make, all the cars in the next grouping are another make, etc., then he is partitioning the cars into disjoint nonempty subsets. For example, there is a grouping of Fords, Pontiacs, Plymouths, etc. We could define an equivalence relation on the set consisting of Friendly Freddie's Forever Lasting Cars. If a,b are cars in Freddie's lot, then $(a,b) \in R$ if a and b are the same make. We now show that R is an equivalence relation:

   (i) $(a,a) \in R$, because clearly a car is the same make as itself. Therefore, R is reflexive.

   (ii) If $(a,b) \in R$, then $(b,a) \in R$, because if a and b are the same make, certainly b and a are the same make: $\therefore$ R is symmetric.

   (iii) If $(a,b) \in R$ and $(b,c) \in R$, then a and b are the same make, also b and c are the same make, and therefore a and c are both the same make, or $(a,c) \in R$, from which we conclude that R is transitive. In this example we have illustrated the converse of the theorem, i.e., according to the way the set of cars was divided up it was possible to define an equivalence relation on the set of cars.

4. An application to psychology would be in a conditioning experiment. The animal is conditioned to push one of two buttons. His responses may be divided into two disjoint sets whose union consists of all his responses. The animal either presses the correct button or the wrong button.

5. In a discrimination task, the individual may be asked to divide up the stimuli according to color. Therefore, the set of stimuli are divided into classes, with each class consisting of stimuli of a particular color.

6. In a rule-oriented subject matter such as mathematics, a person learns to do many problems on the basis of one rule. He must analyze a problem, decide which rules are relevant, and then apply the rules. Therefore, each individual problem is not treated as an isolated case.

The examples have hopefully given further illustration of how fundamental this theorem is and how relevant it is to questions in psychology. The theorem essentially describes a person's ability to organize and classify.

Types of Ordering

With the completion of our discussion of equivalence relations, we begin a discussion of various types of ordering. The names attached to these orders vary in the literature, and one must be careful to make note of the possible distinctions between texts. The definitions and names that we will use in this book seem to be the most common. We begin with a list of definitions, and then follow the definitions with relevant examples and references as to where in the psychological literature applications of ordering may be found.

Definition 16. A relation "$<$" is a partial ordering for a set A if $<$ is reflexive, antisymmetric, and transitive, i.e.,

(i) for every $a \in A$, $a < a$;

(ii) for every $a,b \in A$, $a < b$ and $b < a$ implies $a = b$; and

(iii) for every $a,b,c \in A$, $a < b$ and $b < c$ imply $a < c$.

Definition 17. A relation "$<$" is a strict partial ordering of A if $<$ is antisymmetric and transitive. Therefore, we could call a partial ordering a reflexive strict partial ordering.

Definition 18. A relation "$<$" is a linear ordering (also called simple or total) of A if $<$ is reflexive, antisymmetric, transitive, and connected. That is, if $<$ is a partial ordering and in addition for every $a,b \in A$, if $a \neq b$, then $a < b$ or $b < a$.

Definition 19. A relation "$<$" is a strict linear ordering if $<$ is antisymmetric, transitive, and connected.

We may use diagrams to indicate the different types of ordering. For example, if one can reach one element of a set from another element in the set in a continually ascending manner, then the elements are ordered. Let us consider a set A, where $A = \{a,b,c,d\}$. Suppose that the elements of A are related as indicated in Figure 10. We may observe that $a < b$, $a < c$, $a < d$, $b < d$, and $c < d$, but $b \nless c$ and $c \nless b$. Therefore, the order defined by Figure 10 would be a partial ordering or a strict partial ordering, depending on whether we allow reflexivity. However, this ordering is not linear, since neither $b < c$ or $c < b$. The diagram for a linear or simple ordering would have to be along a single vertical line such as in Figure 11, where $a < b$, $a < c$, $a < d$, $b < c$, $b < d$, and $c < d$. Therefore, the connectivity property is satisfied, unlike in the previous illustration, where there existed a pair where $b \nless c$ and $c \nless b$.

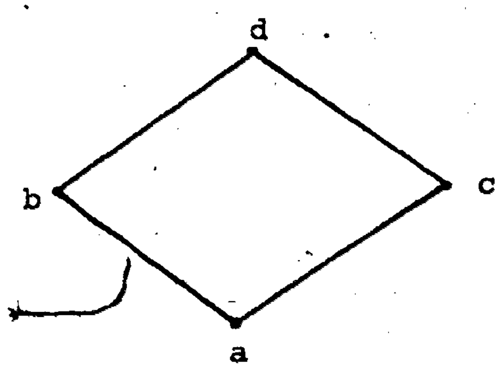

Figure 10

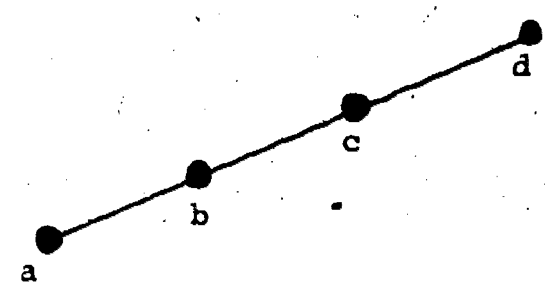

Figure 11

The figures were introduced as a visual aid in understanding the concepts of partial and linear ordering. We now give a series of examples to indicate the kinds of relations that are partially or linearly ordered. We will begin with a few relations that we have discussed in detail already.

Examples

1. Consider $\leq$ for integers. This is a linear ordering, because for any integers m, n, and p,

   (i) $m \leq m$ for all m, i.e., for any integer, it is less than or equal to itself;

   (ii) if $m \leq n$ and $n \leq m$, then the only possibility is that $m = n$;

(iii) if $m \leq n$ and $n \leq p$, this clearly implies that $m \leq p$ (for example, if $3 \leq 5$ and $5 \leq 11$, then $3 \leq 11$);

(iv) for any m,n where $m \neq n$, then either $m \leq n$ or $n \leq m$ (this means that if two numbers are not equal, then one of the two is the larger). Combining (i), (ii), (iii), and (iv) we have shown that $\leq$ is a linear ordering.

2. If we consider "<," we immediately notice that "less than" is not reflexive. The other properties hold. Therefore, < is a strict linear ordering.

3. Set inclusion, "$\subset$," is a partial ordering, but not a linear ordering.

   (i) For any set A, $A \subset A$. Every set is a subset of itself.

   (ii) For any sets A and B, if $A \subset B$, and $B \subset A$, then $A = B$.

   (iii) For any A, B, and C, if $A \subset B$ and $B \subset C$, then $A \subset C$. This is obviously true, but if there are any nonbelievers, the Venn diagram in Figure 12 gives an intuitive demonstration.

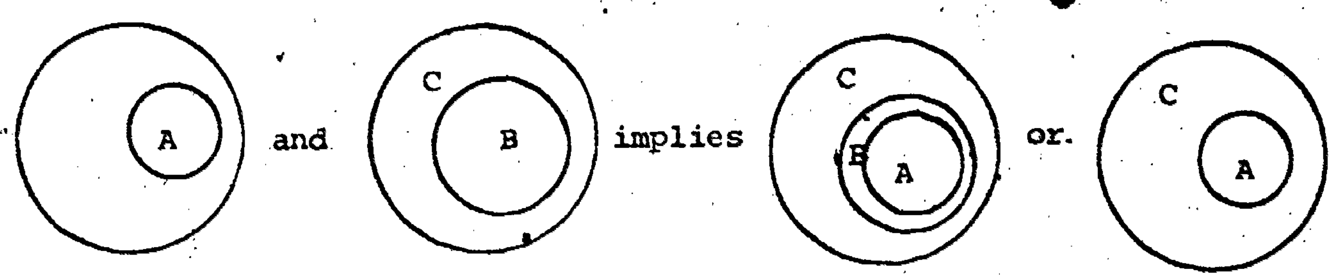

Figure 12

   (iv) For any sets A and B, where $A \neq B$, we need not have $A \subset B$ or $B \subset A$. In fact we could even have $A \cap B = \phi$. If $A = \{1,3,5,7\}$ and $B = \{2,4,6,8\}$; then $A \cap B = \phi$. Therefore, set inclusion is not connected, and the relation "$\subset$" is a partial ordering.

4. Proper set inclusion, "$\subsetneq$," is a strict partial ordering, since it is not reflexive. No set is a proper subset of itself.

5. Examples 1 through 4 served to illustrate the four new definitions. There are analogous real world parallels. For example, "is taller than" is a strict linear order. It is not reflexive, because no one is taller than himself.

6. Within many branches of psychology such as developmental psychology, there is discussion of hierarchies of events. For example, in the developmental psychological theory of Jean Piaget (Piaget & Inhelder, 1969) there is elaborated a linear hierarchy

of cognitive operations. Piaget contends that a cognitive preoperational schema such as grasping precedes the cognitive concrete operation of classification in development which in turn precedes the cognitive formal operation of hypothetico-deductive thinking in development. A sequence of behavioral forms of this type has the mathematical properties of a linear ordering with the relation being "is a prerequisite to" or "is a necessary condition for." To Piaget, classes of cognitive behaviors--preoperational, concrete, and formal--are reflexive, antisymmetric, transitive, and connected for the relation of "is a necessary condition for." To demonstrate that a class of behavioral phenomena comply to some ordering for some relation, empirical conditions must be formulated that will allow for the testing of the defining properties of the relation. For example, in the case of the Piagetian cognitive theory one may consider two types of cognitive operations and if one operation is not demonstrated to be a prerequisite to the other operation, then the connected property cannot be attributed to that relation and the relation is thus not a linear ordering. The terminology of relations and ordering can be used not only to describe qualitatively the structural properties of arrays of behavioral phenomena, but also aid in the formulation of the empirical conditions by which one can test the structures and hierarchies attributed to an array of behavioral phenomena.

7. Airasian and Bart (1971) have introduced ordering theory, formally referred to as tree theory, as an alternative measurement model. Ordering theory has as its primary purpose the testing of hypothesized hierarchies among items, or sometimes the determination of such hierarchies. Ordering theory is similar to other classical models in that it utilizes the item response matrix, but it differs in that it does not use summative scores. Also, the classical approaches assume that the trait measured is linearly ordered, which usually is never tested for. Order theory does not use summative scores as a starting point for statistical analysis, but rather is used to determine logical relationships between items represented in the item response matrix.

8. The next example is again a more mathematical one: It serves the purpose of illustrating the new terminology in a more abstract way. Let $\mathbb{Z}^+$ be the positive integers, i.e., $\{1,2,3,\cdots\}$. Define "$|$" to mean divides: Therefore, $a|b$ means $at = b$, for some $t \in \mathbb{Z}^+$. We will show that "$|$" is a partial ordering.

   (i) For any $m \in \mathbb{Z}^+$, $m|m$, since $m \cdot 1 = m$.

   (ii) For $m,n \in \mathbb{Z}^+$, if $m|n$ and $n|m$, then there exists t and s in $\mathbb{Z}+$, such that $mt = n$ and $ns = m$. Therefore, by substitution $(mt)s = m$, which we may rewrite as $m(ts) = m$. Hence, $ts = 1$, but both t and s being positive integers imply $t = s = 1$. Therefore, $mt = m = n$, and antisymmetry is proven.

(iii) For $m,n,p \in \mathbb{Z}^+$, if $m|n$ and $n|p$, then there exist t and s in $\mathbb{Z}^+$ such that $mt = n$ and $ns = p$. Which upon substitution yields $(mt)s = p$ or $m(ts) = p$, but ts equals an integer, say $q \in \mathbb{Z}^+$, and this implies $mq = p$. Thus, we may conclude that $m|p$ and "|" is transitive. We have now proven that "|" is a partial ordering. We may show that "|" is not a linear ordering.

(iv) For example, consider $3,7 \in \mathbb{Z}^+$, but $3 \nmid 7$ and $7 \nmid 3$. Therefore, divides is not connected.

9. The last example that we will consider is that of lexicographic ordering. Suppose that sets A and B are linearly ordered. Consider the Cartesian product of A and B, i.e., A X B. It may be proven that A X B may be linearly ordered by <, where we define $(a,b) < (a',b')$ if and only if $a <_1 a'$, or if $a = a'$, then if $b <_2 b'$. We are denoting the strict linear order for A by $<_1$ and the linear order for B by $<_2$. The proof that < is a linear ordering is not that difficult, but requires much cumbersome notation and the consideration of separate cases. Because of this fact, a proof will not be included. Instead several interesting applications will be discussed. Suppose that set A equals set B, and that the members or elements of the set are the letters of the alphabet, i.e., $A = B = \{a,b,c,\cdots,x,y,z\}$. The ordering of A (and B) will be the normal alphabetical ordering. Then lexicographic ordering is a precise and elegant way of describing how a dictionary is put together. If two words are compared, and if the first letters are different we order the two words on the basis of the alphabetical order of the first letters of the two words. If the first letters are the same, then we order the two words on the basis of the second letters, and so on.

   A second useful application is that lexicographic ordering offers a method of comparing points in the plane. The points could be compared by looking at the first coordinates, if they are the same, then we compare second coordinates. Therefore, one could say $(1,4) < (3,1)$, $(2,7) < (2,9)$, $(1,1000) < (2,2)$, etc.

If a set is linearly ordered by a relation $\prec$, we may consider an additional property that certain linearly ordered sets have.

Definition 20. Let A be a set and suppose $\prec$ is a linear ordering of A, then A is well ordered if and only if every nonempty subset of A has a least or smallest element, i.e., if for every nonempty subset $B \subset A$, there is an element $b_0 \in B$, such that $b_0 \prec b$ for every $b \in B$.

Examples

1. The set of all positive integers is well ordered by $\leq$, because every subset of the positive integers has a smallest element. This assertion is equivalent to Peano's axiom.

2. The set of all integers is not well ordered by $\leq$, because, for example, $\mathbb{Z}$ itself has no smallest element.

3. Clearly every finite set with a linear ordering defined on it is also well ordered, because there are only a finite number of elements to consider at a time, and the smallest one may always be picked out.

4. If the set under consideration consists of scores on an achievement test, then these scores are linearly ordered by "less than or equal to." Also the set is well ordered, because any subcollection of scores will always have a lowest score.

We have completed our discussion of relations and the special properties of relations. We have also examined equivalence relations and different types of orderings. The richness of these ideas should be evident from the ease with which they handle both abstract and real considerations. Psychologists have been utilizing these ideas in their justifications of various phenomena, so it would be reasonable to incorporate these terms into the language of psychology as a means of precise description.

CHAPTER 3

MAPPINGS

One of the most important ideas in all of mathematics is that of a function or mapping. This term is so fundamental that it is in common usage in most disciplines. Almost anyone will with great regularity refer to one thing as being a "function" of something else. In a very narrow mathematical sense, a function may be viewed as a formula that associates to a number another number. For example, according to a formula the number 5 may be associated with the number 7. This is a restricted definition of a function, and is highly limited in terms of its applicability. Therefore, as a first definition of a function, let us consider the following.

<u>Definition 21</u>. A <u>function</u> or <u>mapping</u> f, from one set U to another set V, is a rule that associates with each element x in a certain subset D_f of U, a uniquely determined element f(x) in V. The set of values in D_f is called the <u>domain</u>. The element $y = f(x)$ is called the <u>image</u> of f at x, where $x \in D_f$. The set of all image values of f is referred to as the <u>range</u> and will be denoted by R_f.

Even though this definition is more general than the previous one, in that the sets U and V do not have to be sets of numbers, there is still an ambiguity built into the definition.

In mathematics, as well as in psychology, when dealing with abstract ideas, it is important to be precise with one's language. In the definition of a function or mapping the key word is rule. A mapping from U to V is a rule, but what is a rule? The definition is highly intuitive and will be made use of in the book, but in order to be as rigorous as possible, another definition of a function will be given. The new definition, interestingly enough, will be in terms of the language introduced in the first two chapters.

<u>Definition 22</u>. Let U and V be nonempty sets, then a <u>mapping</u> or <u>function</u> from U into V is a set f of ordered pairs in the Cartesian product U X V, such that if (x,y) and (x,z) are elements of f, then $y = z$. In other words, a mapping f is a relation between sets U and V, such that for every admissible value x in U there is a unique y in V, such that $(x,y) \in f$. The collection of all first components, denoted by D_f, will be called the <u>domain</u>. Therefore, D_f is the set of all admissible values in U. The <u>range</u>, R_f, consists of all those values in V occurring as second components in the ordered pairs.

A function, then, is a special type of relation. It is a subset of the Cartesian product U X V, with the added condition that the second member of an ordered pair in f is uniquely determined by the first member. In order to take advantage of the intuitive nature of Definition 21, rather than writing $(x,y) \in f$, we will adhere to the more commonly recognized notation of $y = f(x)$, and will refer to $y = f(x)$ as the image of f at x.

Examples

1. Let $U = \{1,2,3,4,5\}$ and $V = \{3,5,9,16,17\}$, and define $f = \{(1,5), (2,3), (3,17), (4,16), (5,9)\}$, then $f \subset U \times V$, and further for every $x \in U$, there is a unique $y \in V$. Therefore, f is a function.

2. Let $U = \{1,2,3,4,5\}$ and $V = \{3,5,9,16,17\}$ and define $f = \{(1,5), (2,3), (3,5), (4,9), (5,9)\}$, then $f \subset U \times V$, and again for every $x \in U$, there is a unique $y \in V$. Both 1 and 3 are associated with 5, but this is not contrary to the definition of a mapping, since each $x \in U$ still has only one value in V associated with it. Notice also that in this example the range is $\{3,5,9\}$ and is not equal to all of V.

3. Let $U = \{1,2,3,4,5\}$ and $V = \{3,5,9,16,17\}$ and define $f = \{(1,3), (2,9), (2,5), (3,16)\}$. f is a subset of U X V, but f is not a function, since there are two different image values 5 and 9 assigned with 2.

4. Suppose Miss Nice is a second grade teacher in a small school and that she has ten students: Tom, Mary, Bill, Lola, Frankie, Jim, Paula, Farnsworth, Betty, and Tony. She gives them a spelling test of 20 words and makes a chart for the results like the one in Figure 13. This is an example of a function. Let

   U = {Tom, Mary, Bill, Lola, Frankie, Jim, Paula, Farnsworth, Betty, Tony}; and

   $V = \{0,1,2,\cdots,18,19,20\}$ = possible number of correct answers.

   Define f = {(Tom,12), (Mary,16), (Bill,17), (Lola,11), (Frankie,19), (Jim,14), (Paula,20), (Farnsworth,16), (Betty,15), (Tony,19)}.

   f is a subset of U X V and also for every element in the domain, there is a unique element in the range, namely for each child there is a test score associated. The range in the example is $\{11,12,14,15,16,17,19,20\}$.

42

Tom	12	Jim	14
Mary	16	Paula	20
Bill	17	Farnsworth	16
Lola	11	Betty	15
Frankie	19	Tony	19

Figure 13

5. A function may be thought of in terms of a machine. There is an input, an output, and a machine f performing the change. For input x, f(x) would represent the output. Put a quantity of heavy cream in a blender f and the result will be whipped cream. Put a coin in a bubble gum machine and out comes a piece of bubble gum. The parallel to a machine is indicated in Figure 14.

$$x \longrightarrow \boxed{f} \longrightarrow y = f(x)$$

f is the machine, f(x) is the output

Figure 14

6. The idea of a function as a collection of ordered pairs seems to indicate that it may be helpful to consider a function in terms of its graph. We will do this in a separate section at the end of the chapter.

7. The idea of a function may also be given a geometric interpretation. Consider the description of a mapping, f, in Figure 15.

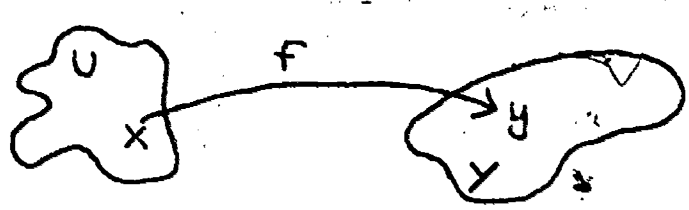

Figure 15

8. There are certain functions that are worthy of specific mention. One of them is the identity mapping. In effect an element is mapped into itself, that is the mapping does not change anything. We would write this as $f(x) = x$. For instance, $f(3) = 3$, $f(\cdot 271) = \cdot 271$, etc. The set of values that are left unchanged by a mapping are often said to be invariant with respect to the mapping. The idea of invariants is valuable in psychology. For example, if one understands what types of transformations leave an entity or concept unaltered, then one has a good understanding of what that entity or concept is.

9. The constant function is another very basic mapping. For this mapping, regardless of which element in the domain is selected, the function always assigns the same range value. Examples of a constant function would be $f(x) = 5$, where regardless of what the x value is, it is always assigned the value 5. Another example is in a store where every item costs the same amount, or in a conditioning experiment, where an animal is conditioned to always pick the element in the left position, regardless of whether the elements are balls, blocks, colors, etc.

10. In Scandura's (1970) SFL language mentioned before in chapter 2, the idea of a function is basic to the discussion. He distinguishes between a rule, a concept, and an association as follows. A rule he defines as a function whose domain is a set of stimuli and whose range is a set of responses. A rule is then a mapping between equivalence classes of stimuli and responses. A concept is a constant function, i.e., each stimulus in a class is paired with a common response. An association is a function whose domain consists of one stimulus and whose range consists of one response, i.e., an association is a single S-R pair.

11. Anyone who has debated whether it was necessary to put an additional stamp on a letter is familiar with the post office function. It is an example of a mapping where the domain is broken up into several parts as in Figure 16.

$$f(x) = \begin{cases} 8¢ \text{ if } 0 < x \leq 1 \text{ ounce} \\ 16¢ \text{ if } 1 < x \leq 2 \text{ ounces} \\ 24¢ \text{ if } 2 < x \leq 3 \text{ ounces} \\ \quad\quad \text{etc.} \end{cases}$$

Figure 16

12. Addition is another example of a mapping: Let $\mathbb{Z}$ be the set of integers, and define $U = \mathbb{Z} \times \mathbb{Z}$, to be the Cartesian product of the integers with themselves, i.e., U consists of all the ordered pairs of integers. Define f as a mapping from U into , and denote

it be $f: U \to \mathbb{Z}$, where $f((m,n)) = m + n$, with $m,n \in \mathbb{Z}$. Therefore, $f((10,4)) = 10 + 4 = 14$, $f((11,3)) = 11 + 3 = 14$, etc.

13. Another interesting function is called the characteristic function. Let U be any set, and suppose S is a subset of U, then define

$$f_S(x) = \begin{cases} 1 \text{ if } x \in S \\ 0 \text{ if } x \notin S. \end{cases}$$

This means that if $x \in S$, then the function value is 1, otherwise the function is 0. We could think of a discrimination problem in these terms. If the subject makes the correct discrimination he receives a reward, and if he does not, then he receives nothing. It is just necessary to think of 1 as reward and 0 as no reward.

14. Sequences are used with great frequency in psychology. An article may refer to the 1001 subjects as $S_0, S_1, S_2, \cdots, S_{1000}$, or in statistics one may be interested in the multiple correlation between variables $X_1, X_2, \cdots, X_k$. A sequence is a special case of a function. The domain of the mapping consists of $0,1,2, \cdots$, and the range consists of whatever is being described. Rather than write $S(0), S(1), S(2), \cdots$ we write $S_0, S_1, S_2, \cdots$, but nevertheless, a sequence is a special case of a function.

We have considered a rather extensive list of examples of functions. But, if one considers the frequency with which the word function occurs in daily life, in addition to its more technical uses in the sciences, it is clear why it is important that the definition and types of functions be discussed in this text. Keep in mind that a mapping is a relation, with the added condition that for each element in the domain there is associated a unique element in the range.

We have looked at examples where the range was the entire set V and others where $R_f \subsetneq V$. Those mappings that have $R_f = V$ are of special interest, and have been given a special name.

Definition 23. If f is a mapping from U into V, then f is said to map onto v if $R_f = V$, i.e., the range of f is all of V. This may also be stated as, f is a mapping from U onto V if for every $y \in V$, there exists an $x \in D_f$, the domain of f, such that $(x,y) \in f$, or equivalently, $y = f(x)$. An onto mapping is also called a surjective mapping or a surjection.

Examples

1. Consider the first two examples of functions. We were given that $U = \{1,2,3,4,5\}$ and $V = \{3,5,9,16,17\}$. In example 1, the range was equal to the set of elements 3,5,9,16, and 17. Therefore, the mapping is onto. But in example 2, the range was only 3,5, and 9. Therefore, $R_f \subsetneq V$, and this function is only a mapping from U into V, not onto V.

2. The example of the 2nd grade spelling test results is a case of another function that is not onto V. V equals the numbers $0,1,2,\cdots,20$, i.e., the potential number of correct answers, but the actual results only were $R_f = \{11,12,14,15,16,17,19,20\}$, and $R_f \subsetneq V$.

3. If we consider the identity function, and suppose the domain consists of all the real numbers, i.e., all the numbers along the number line. Also assume that V is equal to the real numbers. Then the identity mapping $f(x) = x$ is a mapping onto V since every real number is simply mapped into itself.

4. If we again consider the identity mapping, but suppose that $U = V = \mathbb{Z}$, $\mathbb{Z}$ recall is the set of integers. Then $f(x) = x$ is a mapping onto V because every integer is mapped onto itself. However, if the function were $f(x) = 2x$, i.e., each number is associated with twice itself, then the mapping would not be onto, because the range would consist of only the even integers, and not all of the integers. For example, $f(3) = 6$, $f(9) = 18$, etc. It is impossible to find an integer x, such that for instance $f(x) = 3$, since $2x = 3$ would imply that $x = 3/2$, which is not an integer.

5. If we again let $U = V = \mathbb{Z}$, we see that the constant function is not onto, since the range of the constant function is only one element.

6. The post office function is not onto because the price of mailing letters is always a multiple of 8¢. If the letter weighs too much, another 8¢ must be put on the letter.

7. The SFL theory of Scandura defines a concept in terms of function language. The domain is a set of stimuli, the set V is a set of responses, but the range of a learned concept consists of only one response, namely the correct one. Therefore, a concept is not onto.

8. On a true-false test, the domain consists of a set of questions and the answers are to be selected from the set $V = \{T,F\}$. If the answers to the set of questions consist of both true and false answers, then the mapping is onto; however, if all the answers are true or all the answers are false, then the mapping is into.

9. A matching test (like the one in Figure 17) would be an example of an onto mapping. The function consists of the following ordered pairs: (New York, Albany), (Minnesota, St. Paul), (New Jersey, Trenton), (California, Sacramento), (Pennsylvania, Harrisburg).

A.	New York	1.	Sacramento
B.	Minnesota	2.	Trenton
C.	New Jersey	3.	Albany
D.	California	4.	Harrisburg
E.	Pennsylvania	5.	St. Paul

Figure 17

There is another important type of function that is useful is establishing a correspondence between two sets. These mappings are called 1 - 1, or one to one.

Definition 24. A function f is 1 - 1, or one to one, if for any x_1 and x_2 in D_f, where $x_1 \neq x_2$, then $f(x_1) \neq f(x_2)$. Equivalently, if $f(x_1) = f(x_2)$, then x_1 must equal x_2. In other words, no element in the range of f, R_f, may occur more than once. A one to one mapping is also called an injective mapping or an injection.

Examples

1. If $U = \{1,2,3,4,5\}$ and $V = \{3,5,9,16,17\}$, define $f = \{(1,5), (2,3), (3,5), (4,9), (5,9)\}$. This function is not 1 - 1, because both 4 and 5 are mapped into 9, i.e., $4 \neq 5$, but $f(4) = f(5) = 9$.

2. However, if $f = \{(1,5), (2,3), (3,17), (4,16), (5,9)\}$, then P is one to one.

3. The example of a mapping corresponding to the results of a spelling test given before is not a 1 - 1 mapping, because both Mary and Farnsworth scored 16.

4. The identity mapping from one set to itself is an obvious example of a 1 - 1 function. Since this mapping is defined by $f(x) = x$, then trivially if $x_1 \neq x_2$, then $f(x_1) \neq f(x_2)$, because $f(x_1) = x_1$ and $f(x_2) = x_2$.

5. The constant function is 1 - 1 only if the domain consists of one element; otherwise there are many elements mapped into the same element. Therefore, a concept is generally not a 1 - 1 mapping.

6. A true-false test generally is not a one to one mapping, because more than one of the items is true and more than one of the items is false. For example, in a five question test it is impossible to have a 1 - 1 mapping.

7. A matching test is, however, 1 - 1, because each answer corresponds to only one question.

Some mappings are onto, but not one to one, others are 1 - 1, but not onto, and there are also mappings that are both 1 - 1 and onto.

Definition 25. A mapping f is a 1 - 1 correspondence between sets U and V if f is a 1 - 1 mapping onto V. A 1 - 1 correspondence is also called a bijective mapping or bijection. Thus, a mapping that is an injection and a subjection is a bijection.

Examples

1. The identity mapping is a 1 - 1 correspondence, since we have shown if U = V = real numbers, then $f(x) = x$ is both 1 - 1 and onto.

2. The mapping $f(x) = 2x$, where $U = V = \mathbb{Z}$, was shown to be into, not onto, but $f(x) = 2x$ is 1 - 1 since if $x_1 \neq x_2$, then $f(x_1) \neq f(x_2)$. This follows because $2x_1 \neq 2x_2$.

3. If $U = \{1,2,3,4,5\}$ and $V = \{3,6,9\}$, then for $f = \{(1,3), (2,9), (3,6), (4,3), (5,9)\}$, the function is onto, but not 1 - 1, since, for example, both 1 and 4 are mapped into 3.

4. Another example of a 1 - 1 correspondence would be a matching test. We have shown that this is both a 1 - 1 and onto mapping.

5. In any theory designed to describe the human mind such as automata theory, the psychologist hypothesizes a 1 - 1 correspondence between man and the simulated model.

Before we begin a discussion of different operations between mappings it is a good idea to define the equality of two functions.

Definition 26. If f and g are mappings of U into V, the f equals g, i.e., $f = g$, if $f(x) = g(x)$ for every $x \in U$.

We may define a sum, difference, production, and quotient of two functions f and g: In other words there exist methods of producing new functions.

48

Definition 27. Suppose f and g are mappings from U into V, with domains D_f and D_g respectively. Then we make the following definitions:

(i) $(f + g)(x) = f(x) + g(x)$;

(ii) $(f - g)(x) = f(x) - g(x)$; and

(iii) $(f \cdot g)(x) = f(x)g(x)$.

In (i), (ii), and (iii) the domain of the new function is $D_f \cap D_g$, i.e., those elements common to both domains.

(iv) $(f/g)(x) = \dfrac{f(x)}{g(x)}$, where $x \in D_f \cap D_g - \{D_g | g(x) = 0\}$, i.e.,

those elements in common to D_f and D_g with the exception of the elements in D_g, where $g(x) = 0$. This way the problem of division by zero is avoided, and the new function is defined everywhere on its domain.

Example

1. If $f(x) = x^2 + 1$ and $g(x) = x - 4$, and suppose the domain consists of the real numbers, i.e., all the numbers on the number line. Then,

$$(f + g)(x) = f(x) + g(x) = (x^2 + 1) + (x - 4) = x^2 + x - 3;$$
$$(f - g)(x) = f(x) - g(x) = (x^2 + 1) - (x - 4) = x^2 - x + 5;$$
$$(f \cdot g)(x) = f(x)g(x) = (x^2 + 1)(x - 4) = x^3 - 4x^2 + x - 4;$$
$$(f/g)(x) = \frac{f(x)}{g(x)} = \frac{x^2 + 1}{x - 4}, \text{ where } x \neq 4.$$

The operation that will have more psychological relevance than the others is probably the composition of functions.

Definition 28. Let f be a function with domain in U and range in V. Let g be a function with domain in V and range in W. Then the composition gof is the function from U into W, defined as

$$gof = \{(x,z) | \text{there exists a } y \in V \text{ such that } (x,y) \in f \text{ and } (y,z) \in g\}.$$

The domain of gof consists of all those x in U such that f(x) is in V, and the range consists of all those g(f(x)).

A few examples may help clarify this definition. Notice that a composition of functions is a means of going from one set of entities to another set, and then from this set, then going to a third set. An important warning to the reader is that in some textbooks and journals gof is taken to mean first applying g and then applying f. However, in this book gof will always be understood to mean that f is applied first, and then g is applied. As will be pointed out, gof need not equal fog, so it is important to determine which convention is being adhered to in the article you are reading.

Examples

1. Suppose f is the mapping that associates 1 yard with 3 feet, and that g is the rule that associates 1 foot with 12 inches, then gof is the mapping that associates 1 yard with 36 inches. The domain of gof is yards, and the range is inches. For example, (gof)(4 yards) = g(f(4 yards)) = g(12 feet) = 144 inches.

2. Suppose $f(x) = x^2 + 1$ and $g(x) = x - 4$, and suppose the domain of f and of g is the real numbers, then

   $$(gof)(x) = g(f(x)) = g(x^2 + 1) = (x^2 + 1) - 4 = x^2 - 3, \text{ but}$$
   $$(fog)(x) = f(g(x)) = f(x - 4) = (x - 4)^2 + 1 = x^2 - 8x + 17.$$

   This is an example of where gof $\neq$ fog, since $x^2 - 3 \neq x^2 - 8x + 17$ for all x, except when $x = 5/2$. Recall that for two functions to be equal they must be equal for all x.

3. Suppose a psychology class has an examination. Let f be the mapping that assigns a numerical score to each student. Let g be the grade-line mapping, i.e., certain scores receive an A, others a B, and so on. Then gof assigns each student a grade on the test.

4. Consider Harlow's oddity problem. Given three objects, with one of the objects different from the other two. The odd item should be selected. Let f be the function which represents the decision as to which element is the odd item. Let g be the function of selecting this item. Then gof is the successful performance of an oddity problem task.

An interesting theorem regarding the composition of functions will be stated without proof.

Theorem. Let f be a function with domain in U and range in V. Let g be a function with domain in V and range in W. Then,

(i) if f and g are each onto, then gof is also onto; and
(ii) if f and g are each 1 - 1, then gof is also one to one.

When we discussed 1 - 1 mappings, we pointed out that there were no elements in the range occurring more than once, i.e., if $f(x_1) = f(x_2)$, then $x_1 = x_2$. It may then be observed that if the ordered pairs constituting the function f have their first and second entries interchanged, then this new set of ordered pairs would also describe a function. Because of the 1 - 1 nature of f there is correspondence between a domain element and a range element, or conversely a matching of one element in the range with one element in the domain. The function obtained upon this interchange of components is called the inverse of f.

Definition 29. Let f be a 1 - 1 function from U into V: If f^{-1} is defined as $f^{-1} = \{(y,x) \mid (x,y) \in f\}$, then f^{-1} is a 1 - 1 function from V into U and is called the inverse of f.

Examples

1. If U = {Tom, Betty, Bill, Sally, Peter} and V = {18,17,20,15,16} represents their respective scores on a 20 question test, then f is a mapping from U onto V such that f = {(Tom,18), (Betty,17), (Bill,20), (Sally,15), (Peter,16)}. f is a 1 - 1 mapping, therefore the inverse function f^{-1} may be defined. f^{-1} = {(18,Tom), (17,Betty), (20,Bill), (15,Sally), (16,Peter)}. Here, each score is associated with a particular person, rather than assigning for each person a particular score.

2. Consider the matching test in Figure 18 which was introduced earlier in the chapter. We have already shown that this is a 1 - 1 mapping. Therefore, an inverse exists. If f = {(New York,Albany), (Minnesota, St. Paul), (New Jersey,Trenton), (California,Sacramento), (Pennsylvania,Harrisburg)}, then f^{-1} = {(Albany,New York), (St. Paul, Minnesota), (Trenton,New Jersey), (Sacramento,California), (Harrisburg,Pennsylvania)}.

A.	New York	1.	Sacramento
B.	Minnesota	2.	Trenton
C.	New Jersey	3.	Albany
D.	California	4.	Harrisburg
E.	Pennsylvania	5.	St. Paul

Figure 18

3. If $f(x) = 2x$, then f is a 1 - 1 mapping. We may show this easily; if $f(x_1) = f(x_2)$, i.e., $2x_1 = 2x_2$, then this implies $x_1 = x_2$, or f is 1 - 1. If f is defined as $y = 2x$, then $x = y/2$ would define the inverse function f^{-1}. For every y value, the x value is one half of this y value.

The ideas of the composition of functions, a 1 - 1 correspondence, and an inverse of a function may be connected by means of a useful theorem that will now be stated.

Theorem. The mapping f from U into V is a 1 - 1 correspondence, i.e., a 1 - 1, onto mapping if and only if there exists a mapping f^{-1} from V into U such that $f^{-1}of$ and fof^{-1} are the identity mappings on U and V respectively, i.e., $(f^{-1}of)(x) = f^{-1}(f(x)) = f^{-1}(y) = x$ and $(fof^{-1})(y) = f(f^{-1}(y)) = f(x) = y$.

Examples

1. In other words, if a function and its inverse are consecutively applied, one ends up where one started. If an individual travels from New York to Boston and then from Boston to New York, he ends up where he started. The person's trip may be described as

$$f(\text{New York}) = \text{Boston}$$
$$f^{-1}(\text{Boston}) = \text{New York},$$

then $(f^{-1}of)(\text{New York}) = f^{-1}(f(\text{New York})) = f^{-1}(\text{Boston}) = \text{New York}$, or $(fof^{-1})(\text{Boston}) = f(f^{-1}(\text{Boston})) = f(\text{New York}) = \text{Boston}$, which would describe the trip from Boston to New York and then a return to Boston.

2. Another example would be if we define $y = f(x) = 2x$. We have already proven that f is 1 - 1. The inverse function was shown to be $x = f^{-1}(y) = y/2$. Then, $(f^{-1}of)(x) = f^{-1}(f(x)) = f^{-1}(y) = x$ and specifically this is $(f^{-1}of)(x) = f^{-1}(f(x)) = f^{-1}(2x) = f^{-1}(y) = x$. Similarly, $(fof^{-1})(y) = y$.

We conclude this chapter with an elementary discussion of graphing techniques, and to illustrate these procedures we will graph some of the functions described in this chapter.

Our examination of graphing will be on a rectangular coordinate system, which has two axes, a horizontal one called the x axis and a vertical one called the y axis. Any point in the plane may be located in this system. The directed distance along the horizontal from the point of intersection of the axes called the origin is referred to as the x coordinate or the abscissa. The directed distance along the vertical is called the y coordinate or ordinate. The abscissa and ordinate of a point are indicated by an ordered pair called the coordinates of a point. The graphical representation of the following ordered pairs, (7,3), (-2,4), (5,1/2), (-1,-4), (2,-1), is illustrated in Figure 19.

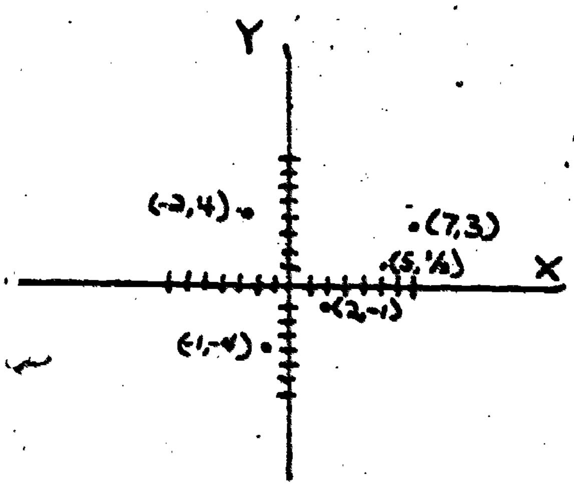

Figure 19

The connection between a function and its graph should be clear. The function consists of all those ordered pairs or points indicated in the graph. In other words, every point satisfying a function lies on the graph of the function, and conversely, every point on the graph and only those points are points that satisfy the function. That is, there is a 1 - 1 correspondence between those points satisfying a function, and the points of the graph of the function.

Examples

1. Let U = {1,2,3,4,5} and V = {3,5,9,16,17} and define f = {(1,5), (2,3), (3,17), (4,16), (5,9)}. This function is graphed in Figure 20.

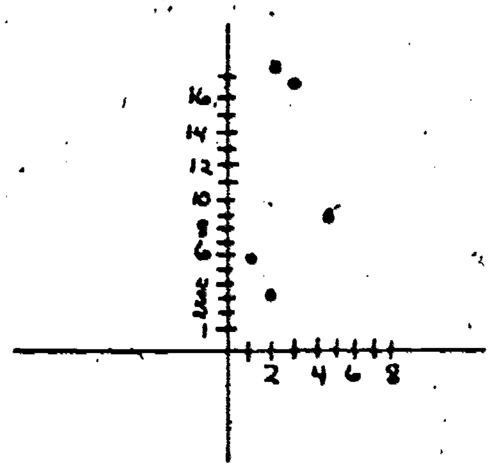

Figure 20

2. Let U = {1,2,3,4,5} and V = {3,5,9,16,17} and define f = {(1,5), (2,3), (3,5), (4,9), (5,9)}. Figure 21 illustrates the graph of f.

CHAPTER 4

GROUPS

A class of algebraic entities useful in psychology is groups. The presentation on groups will be made in two chapters. The first chapter includes a discussion of the definition of a group and the related terms of groupoid, semigroup, and monoid. Elementary examples from mathematics are included to illustrate the relevant terminology. The use of multiplication tables for finite groups will be explained, and then used in the verification of certain sets as groups. To gain familiarity with the new concepts a number of direct consequences will be proven. Other key terms such as subgroup, generators, and different types of mappings such as homomorphism, isomorphism, and automorphism will be introduced and the chapter will be concluded with an examination of several important examples, or types of, groups.

The second chapter will be concerned with the application of groups to psychology. Examples will be given from Piagetian theory, the theory of kinship relations, the studies of measurement, perception, language, automata theory, habit family hierarchies, cross-context matching, symmetric choice experiments, and the use of groups in the application of parallel tasks.

We now define a group. First, a group is more than a set of elements. It is a set for which there is defined an operation such that certain properties are satisfied.

Definition 30. A group is a nonempty set of elements G together with an operation * defined on ordered pairs of elements in G, such that the following four properties are satisfied.

(i) For every $a,b \in G$, the element $a*b \in G$, i.e., the product of any two elements a and b in the set G gives an element a*b that is also in the set G. This property is called closure.

(ii) For every $a,b,c \in G$, $(a*b)*c = a*(b*c)$, i.e., whether we first perform the operation (a*b) and then combine it with c, or if we first perform b*c, and then combine a with b*c, the final outcome is the same. This property is referred to as the associative property.

(iii) For every $a \in G$, there exists an element $e \in G$, such that $a*e = e*a = a$, i.e., there exists an element e, such that regardless of which element of G is considered, when e is combined with that element, the element is unchanged, or in other words, is identical to the way it was before the operation was performed. This element e is called the identity element.

(iv) For every $a \in G$, there exists an element $a^{-1} \in G$, such that $a*a^{-1} = a^{-1}*a = e$, i.e., for every element in G there exists an element a^{-1} such that when the two are combined, the resultant product is the identity element. This element a^{-1} is called the inverse element.

Recapitulating, a group is a nonempty set of elements G together with an operation *, such that G is closed, associative, has an identity element, and every element in G has an inverse. A group is an example of a mathematical system. Actually a group G should be written as (G, *) to indicate that it is a set of elements and a specific operation, but for simplicity of notation a group will be written as G. The reader should, however, also remember that the operation is implicitly understood. Certain sets when combined with particular operations will satisfy only some of the properties. We give names to specific subcollections of the four properties.

Definition 31. A groupoid is a nonempty set G together with an operation *, that has closure, i.e., for any $a,b \in G$, than $a*b$ is also an element of G.

Definition 32. A semigroup is a nonempty set G together with an operation *, that satisfies the closure and associative properties. In other words, a semigroup is an associative groupoid.

Definition 33. A monoid is a nonempty set G together with an operation *, that satisfies the closure and associative properties, and further has an identity element. That is, a monoid is a semigroup that has an identity element.

There is one more important property concerning groups, or for that matter groupoids, semigroups, and monoids. The commutative property is not a requirement of being a group, but it is very important in a discussion of groups. As will become evident in the examples of the following pages, it is not always possible to interchange the order of combining two elements and obtain the same element. We earlier saw that the composition of two functions f and g gave different results in considering fog and gof.

Definition 34. The operation * defined on the set G is said to be commutative or abelian if for every $a,b \in G$, $a*b = b*a$. Therefore, a group satisfying the added property that $a*b = b*a$ for every a,b in G, is called a commutative or abelian group. Similarly, an abelian groupoid, semigroup, or monoid could be defined.

We will encounter groups that have a finite number of elements and others that have an infinite number. Naturally, the question of how many elements are in a group is more interesting in the finite case.

Definition 35. The order of a group G, denoted o(G) or G is the number of elements in the group.

In the case of finite groups a multiplication table may be made to indicate all the possible products. Suppose the group G is defined as $G = \{x_1, x_2, \cdots, x_n\}$. List the elements $x_1, x_2, \cdots, x_n$ across the uppermost row and down the farthest left column, as in Figure 27. The element appearing in the i^{th} row and the j^{th} column would be the element $x_i * x_j$, which equals some x_k in G, since G is a group, and is, therefore, closed. We will make use of the multiplication table in some of the examples.

*	x_1	x_2	...	x_i	x_j	...	x_n
x_1							
x_2							
$\vdots$							
x_i					x_k		
x_j							
$\vdots$							
x_n							

Figure 27

Examples

1. Suppose that the set G equals the elements 1 and -1, and the operation is multiplication. A table of the products is shown in Figure 28.

$\bullet$	1	-1
1	1	-1
-1	-1	1

Figure 28

(i) G is closed, because every product is 1 or -1.

(ii) G is associative, because with multiplication it does not matter which way the elements are grouped.

(iii) G has an identity element, namely the element 1, because $1 \cdot 1 = 1$ and $(-1) \cdot 1 = -1$.

(iv) Each element has an inverse; in fact, each element is its own inverse; $1 \cdot 1 = 1$ and $(-1) \cdot (-1) = 1$.

Therefore, G is a group, and G is actually an abelian group since the order of multiplication does not matter.

2. Let $\mathbb{Z}$ be the integers, i.e., $\mathbb{Z} = \{\cdots,-2,-1,0,1,2.\cdots\}$ and let the operation be addition. $\mathbb{Z}$ is an abelian group. The sum of any two integers is another integer; therefore, $\mathbb{Z}$ is closed. $\mathbb{Z}$ is associative, because for $a,b,c \in \mathbb{Z}$, $(a+b) + c = a + (b+c)$. The identity element is 0, because any integer plus 0 is still the same integer. The inverse of an integer a is $-a$, since $a + (-a) = 0$. For example, the inverse of 3 is -3. Finally, G is abelian, since $a + b = b + a$.

3. If the set was changed to be the natural numbers or counting numbers, $N = \{1,2,3,\cdots\}$, then, if the operation is again addition, N is an abelian semigroup. It has no identity, because $0 \notin N$. Also, since the negative integers are not included, there are no inverses. If we consider $5 \in N$, the inverse would have to be -5, but $-5 \notin N$.

4. If we modify the set of natural numbers by adding the element 0, then the set under consideration is $G = \{0,1,2,\cdots\}$. This set is an example of a commutative monoid under addition, since 0 is the identity element.

5. In considering the set of natural numbers, but now with the operation of subtraction, it may be observed that the set is not even closed. If, for example, we consider the natural numbers 5 and 9, $5 - 9 = -4$, but -4 is not a natural number. The reader's immediate reaction may be to ask, suppose instead of the natural numbers we considered the integers with the operation of subtraction. We still would not be able to get a group, because the associative property does not hold. For instance, if we consider 15, 8, 12, notice that $(15 - 8) - 12 = 7 - 12 = -5$, but $15 - (8 - 12) = 15 - (-4) = 19$ and $-5 \neq 19$. Neither is there an identity element. It is true that, for example, $5 - 0 = 5$, but $0 - 5 = -5$ and -5 is not equal to 5. Recall that the identity property required that $a*e = a$. Therefore, we have an example of a groupoid.

6. Perhaps your curiosity is aroused as to what would happen if we looked at the integers together with the operation of multiplication. Closure, associativity, and the existence of an identity, namely $e = 1$, are all complied with, however, 1 and -1 are the only elements that have an inverse. If we consider 6 as an element of the integers, the inverse of 6 is 1/6 since $6 \cdot 1/6 = 1$, but 1/6 is not an integer. This set is then a monoid under multiplication.

7. If we would enlarge our set to the rational numbers and again consider the operation of multiplication, we then may observe that we have an abelian group. The rational numbers are the set consisting of all fractions. A whole number is a special case of a fraction, e.g., 3 = 3/1. Therefore, the integers are contained in the rational numbers. The only property in question would be the inverse, but with the inclusion of fractions in our set, the inverse of a fraction is just its reciprocal which again is a fraction. The inverse of 3 is 1/3, the inverse of 5/8 is 8/5, etc.

8. The next example is used to illustrate that for the same set G = {e,a,b,c} (see Figure 29) we may indicate multiplication tables of two distinct groups.

(i)

	e	a	b	c
e	e	a	b	c
a	a	e	c	b
b	b	c	a	e
c	c	b	e	a

Figure 29

It is a group. Clearly there is closure, the identity is e, and the inverse of e is e, of a is a, of b is c, and of c is b. The associativity requires verification, left to the reader. For example a*(b*c) = a*e = a and (a*b)*c = c*c = a; therefore, a*(b*c) = (a*b)*c. The other products of this type should be examined.

(ii)

	e	a	b	c
e	e	a	b	c
a	a	e	c	b
b	b	c	e	a
c	c	b	a	e

Figure 30

Figure 30 describes a group. In this example each element is its own inverse.

Example (i) is an example of a cyclic group and a more detailed discussion of cyclic groups will be given at the end of the chapter. Example (ii) is usually referred to as the "4-group." The discussion of Piaget's INRC group (Piaget & Inhelder, 1958) will be based on the "4-group," and will occur in the next chapter as an application of groups to psychology.

9. Consider a square, and observe that the center of the square is the point at the intersection of the diagonals of the square. Let the set G consist of the rotations of the square around its center through 90°, 180°, 270°, and 360° in the clockwise direction. Denote these rotations by $R_{90°}$, $R_{180°}$, $R_{270°}$, $R_{360°}$, respectively. Define A*B to be the rotation A followed by the rotation B. For example, $R_{180°} * R_{270°} = R_{90°}$, because $R_{450°} = R_{90°}$. The multiplication table for G is given in Figure 31. Notice that $R_{360°}$ is the identity rotation and the inverse of any particular rotation is that rotation needed to complete a 360° rotation. G is a group and if one compares Example 8(i) with this example with the correspondence of e with $R_{360°}$, b with $R_{90°}$, a with $R_{180°}$, and c with $R_{270°}$, one sees that they are essentially the same group. Notice further that 90°, 180°, 270°, and 360° were chosen for the square because these rotations leave the vertices or corners in the same positions. In the case of a triangle these invariant rotations would be 120°, 240°, and 360°.

	$R_{90°}$	$R_{180°}$	$R_{270°}$	$R_{360°}$
$R_{90°}$	$R_{180°}$	$R_{270°}$	$R_{360°}$	$R_{90°}$
$R_{180°}$	$R_{270°}$	$R_{360°}$	$R_{90°}$	$R_{180°}$
$R_{270°}$	$R_{360°}$	$R_{90°}$	$R_{180°}$	$R_{270°}$
$R_{360°}$	$R_{90°}$	$R_{180°}$	$R_{270°}$	$R_{360°}$

Figure 31

10. A related but more complicated example that has geometric and visual significance is that of the group of the symmetries of a square. Consider a square, and it may not be a bad idea to actually use a square piece of paper to aid in the verification. Impose a coordinate system on the piece of paper with the origin at the intersection of the diagonals of the square and the sides of the square parallel to the coordinate axes. A sketch of the situation is given in Figure 32. Let the set under consideration consist of eight motions of the square. These motions are all rigid, i.e., the square is not in any way distorted or folded or squashed.

Further, notice that each motion is such that the square always coincides with its initial position after any one of the motions. Let the first four motions be clockwise rotations of the square through 90°, 180°, 270°, and 360°. Denote these motions by $R_{90°}$, $R_{180°}$, $R_{270°}$, and $R_{360°}$, respectively.

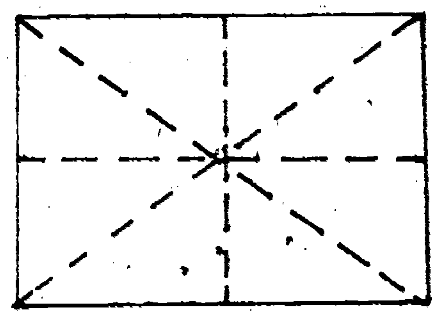

Figure 32

Let X represent the reflection of the square around the x axis, and let Y represent the reflection of the square around its y axis. Let D_1 represent the reflection of the square around the diagonal going from the upper left corner to the lower right corner. Finally, let D_2 be the reflection of the square around the diagonal going from the lower left corner to the upper right corner. Therefore, $G = \{R_{90°}, R_{180°}, R_{270°}, R_{360°}, X, Y, D_1, D_2\}$. Define $A*B$ to mean perform motion A and then motion B on the square. For example, $D_1*R_{180°}$ would mean reflect the square around the diagonal going from the upper left to lower right and then rotate through 180°. The result in this case would be D_2. The completion of the multiplication table may be greatly simplified by using a square piece of paper with the numbers 1,2,3, and 4 in the corners on both sides of the paper. Perform the indicated motions and determine what new motion is obtained. From Table 2 it may be verified that G is a group, but not an abelian group.

The identity element is $R_{360°}$, and also observe that $R_{90°}$ and $R_{270°}$ are each other's inverses. Otherwise the other six elements are self inverses, i.e., $R^2_{180°} = R^2_{360°} = X^2 = Y^2 = D_1^2 = D_2^2 = R_{360°}$ = identity element. In general, groups of the symmetries of regular (equal sided) n sided polygons are called dihedral groups.

11. Let G be the collection of all subsets, which is also often called the power set, of some set S. Define an operation * on G, where $A*B = (A - B) \cup (B - A)$, i.e., * is the symmetric difference operation discussed in the first chapter. Recall that we proved $A \Delta B = B \Delta A$ in Chapter 1, i.e., Δ, the symmetric difference, is commutative. The closure of * (or Δ) is obvious. The identity

Table 2

	$R_{90°}$	$R_{180°}$	$R_{270°}$	$R_{360°}$	X	Y	D_1	D_2
$R_{90°}$	$R_{180°}$	$R_{270°}$	$R_{360°}$	$R_{90°}$	D_2	D_1	X	Y
$R_{180°}$	$R_{270°}$	$R_{360°}$	$R_{90°}$	$R_{180°}$	Y	X	D_2	D_1
$R_{270°}$	$R_{360°}$	$R_{90°}$	$R_{180°}$	$R_{270°}$	D_1	D_2	Y	X
$R_{360°}$	$R_{90°}$	$R_{180°}$	$R_{270°}$	$R_{360°}$	X	Y	D_1	D_2
X	D_1	Y	D_2	X	$R_{360°}$	$R_{180°}$	$R_{90°}$	$R_{270°}$
Y	D_2	X	D_1	Y	$R_{180°}$	$R_{360°}$	$R_{270°}$	$R_{90°}$
D_1	Y	D_2	X	D_1	$R_{270°}$	$R_{90°}$	$R_{360°}$	$R_{180°}$
D_2	X	D_1	Y	D_2	$R_{90°}$	$R_{270°}$	$R_{180°}$	$R_{360°}$

element is the null or empty set, because, for every $A \subset S$, $A*\phi = (A - \phi) \cup (\phi - A) = A \cup \phi = A$. By the commutative property $\phi*A$ also equals A. The inverse of any set A is A itself, because $A*A = (A - A) \cup (A - A) = \phi \cup \phi = \phi$. The only property that remains to be demonstrated is the associative property, i.e., that for arbitrary sets A, B, and C in S, $(A*B)*C = A*(B*C)$. The verification gets quite messy, and requires more computational expertise than would be expected of the reader. Observe that $(A*B)*C = [(A*B) - C] \cup [C - (A*B)] = [((A - B) \cup (B - A)) - C] \cup [C - ((A - B) \cup (B - A))]$ and, $A*(B*C) = [A - (B*C)] \cup [(B*C) - A] = [A - ((B - C) \cup (C - B))] \cup [((B - C) \cup (C - B)) - A]$ and these two expressions must be proven to be the same. As a means of intuitive justification, but not an actual proof, the problem will be considered in terms of Venn diagrams in Figure 33. Therefore, we have an abelian group. This particular group will be used by Bart (1971) in his discussion of Piaget's model of formal operations, and how that model may be generalized, which follows in the next chapter.

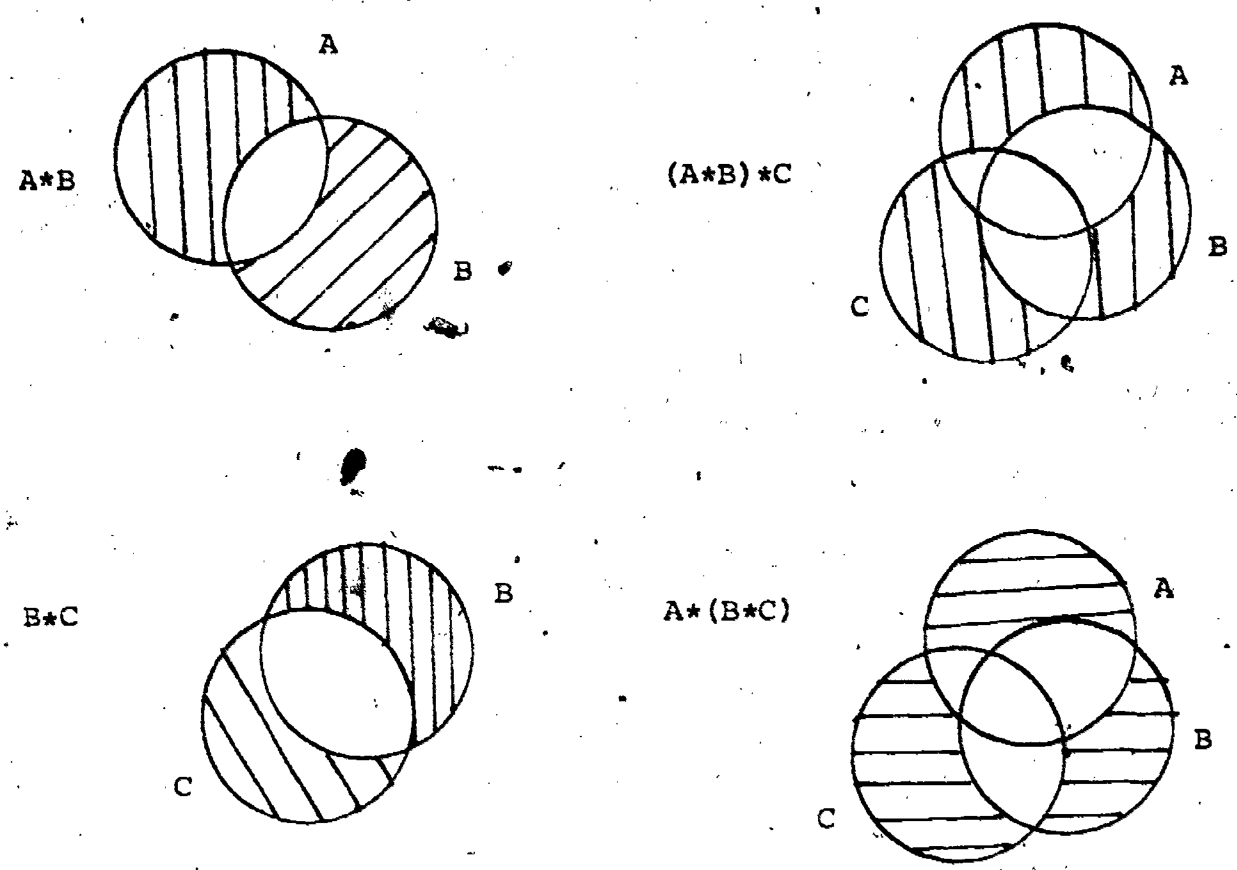

Figure 33

Before we begin to examine several useful consequences of the concept group, a small table is included reviewing the examples concerning the integers and rational numbers with the operations of addition, subtraction, and multiplication. Table 3 indicates that a particular set may be a group under one operation but not another, or that a particular operation imposes a group structure on some sets but not on all sets.

Consequences

In this section we include some direct consequences of the definition of a group.

Lemma. If G is a group, then the identity element is unique.

Proof: We must show that if there are two elements e and s such that $e*a = a*e = a$ and $a*s = s*a = a$ for every $a \in G$, then e and s are equal, i.e., there is only one identity element. If e is an identity element, then $e*a = a$ for any $a \in G$. But s is an element of G, therefore, $e*s = s$. If s is an identity element, then $a*s = a$ for any $a \in G$. In particular, since $e \in G$, $e*s = e$. Thus, we have shown that $e*s = e$ and $e*s = s$, from which we may conclude that $e = s$.

Lemma. If G is a group, then every element a in G has a unique inverse.

Proof: Suppose that there exist elements a^{-1} and b in G such that $a*a^{-1} = a^{-1}*a = e$ and $a*b = b*a = e$, we must prove that $a^{-1} = b$.

$a^{-1} = a^{-1}*e$, because any element combined with the identity is itself. Therefore, by substitution, $a^{-1} = a^{-1}*(a*b)$, since we have assumed $a*b = e$. By the associativity of $*$, $a^{-1} = (a^{-1}*a)*b = e*b = b$. Hence, $a^{-1} = b$ and the inverse of a is unique.

There are several other basic results that we state without proof. They may be in an introductory text in abstract algebra such as Herstein (1964), Dean (1966), or Burton (1965).

Lemma. (i) If G is a group, then for every $a \in G$, $a = (a^{-1})^{-1}$, i.e., the inverse of the inverse itself is the element you began with.

(ii) If G is a group, then for any $a,b \in G$, $(a*b)^{-1} = b^{-1}*a^{-1}$, and if G is abelian, then $(a*b)^{-1} = a^{-1}*b^{-1}$.

We conclude this section with a typical group theoretic exercise.

Table 3

Set	Operation	Closure	Associativity	Identity	Inverse	Groupoid	Semigroup	Monoid	Group	Abelian
Integers	+	YES	YES	YES	YES	YES	YES	YES	YES	YES
Natural numbers	+	YES	YES	NO	NO	YES	YES	NO	NO	YES
Natural numbers & {0}	+	YES	YES	YES	NO	YES	YES	YES	NO	YES
Natural numbers	−	NO	NO	NO	NO	NO	NO	NO	NO	NO
Integers	−	YES	NO	NO	NO	YES	NO	NO	NO	NO
Integers	·	YES	YES	YES	NO	YES	YES	YES	NO	YES
Rationals	·	YES	YES	YES	YES	YES	YES	YES	YES	YES
Groupoid	*	YES	NO	NO	NO	YES	NO	NO	NO	--
Semigroup	*	YES	YES	NO	NO	YES	YES	NO	NO	--
Monoid	*	YES	YES	YES	NO	YES	YES	YES	NO	--
Group	*	YES	YES	YES	YES	YES	YES	YES	YES	--

Theorem. If G is a group, satisfying the property that $(a*b)^2 = a^2*b^2$ for all a,b in G, then G is an abelian group.

Proof: We must show that for every $a,b \in G$, $a*b = b*a$, which would establish that G is commutative. By hypothesis, $a^2*b^2 = (a*b)^2$, where the operation is understood to be *. If $a^2b^2 = (ab)^2$, then since $(ab)^2 = (ab)(ab)$, we have $a^2b^2 = (ab)(ab)$. Upon multiplying both sides of the equality by a^{-1}, we have $a^{-1}a^2b^2 = a^{-1}(ab)(ab)$, or $a^{-1}aabb = (a^{-1}a)b(ab)$ by use of the associativity. Therefore, we obtain $eabb = eb(ab)$, or $abb = bab$. Next multiply on both sides by b^{-1}, to obtain $abbb^{-1} = babb^{-1}$, from which we conclude that $abe = bae$, or $ab = ba$, i.e., G is abelian.

Subgroups

After we introduced the idea of a set, we followed it up with an examination of subsets. We will analogously now introduce the idea of a subgroup.

Definition 36. A subset H of a group G, is said to be a subgroup of G, if H itself is a group under the same operation * that is defined on G.

Examples

1. Under addition we have shown that both the integers and rational numbers are groups. Therefore, the integers and rationals could be considered as H and G, respectively, in the above definition, and we may say that the integers are a subgroup of the rationals under addition. Notice that if the operation were multiplication, the integers would not form a group, and thus would not be a subgroup.

2. In our discussion of the square, we first considered the set, $\{R_{90°}, R_{180°}, R_{270°}, R_{360°}\}$ and proved it was a group. Next we examined $\{R_{90°}, R_{180°}, R_{270°}, R_{360°}, X, Y, D_1, D_2\}$ and proved that it too was a group. Hence, the set of rotations would be a subgroup of the set of motions.

On first inspection it would appear that in order to prove that a subset H of a group G is a subgroup, i.e., is actually a group itself, it appears that the set H must be tested for the four basic properties. Actually the situation is simpler than this. Since the associative property holds for the larger set G it certainly holds for H. Therefore, the associativity does not have to be verified. Two lemmas will be stated that indicate what must in actuality be tested.

Lemma. A subset H of a group G is a subgroup of G if and only if,

(i) $a,b \in H$ imply that $a*b \in H$; and

(ii) $a \in H$ implies that $a^{-1} \in H$.

By combining (i) and (ii) the existence of the identity element may be demonstrated. Suppose $a \in H$, then be (ii) $a^{-1} \in H$, but by (i) we have $a \in H$ and $a^{-1} \in H$ implies that $a * a^{-1} = e$ is also in H. In the case of a group of finite order, i.e., H has only finitely many members, the verification is even easier.

Lemma. If G is a finite group, and H is a subset of G, then H is a subgroup if H is closed under the operation of G, i.e., if $a,b \in H$, then $a*b \in H$.

Suppose we consider a group G, and G has subgroups H and K. The question may be posed, is $H \cap K$ a subgroup of G? The answer is yes but the question still remains, why?

Theorem. If G is a group and H and K are subgroups of G, then $H \cap K$ is also a subgroup of G.

Proof: $H \cap K$ is nonempty because $e \in H \cap K$, since $e \in H$ and $e \in K$. Now, suppose x and y are elements of $H \cap K$, we must show $x*y \in H \cap K$. The fact that $x \in H \cap K$ implies x is an element of H and of K, similarly $y \in H$ and $y \in K$. Because H and K are subgroups, $x \in H$ and $y \in H$ imply $x*y \in H$, and $x \in K$ and $y \in K$ imply $x*y \in K$, but $x*y \in H$ and $x*y \in K$ together imply $x*y \in H \cap K$. Secondly, if $x \in H \cap K$, we must show that $x^{-1} \in H \cap K$. $x \in H \cap K$ implies $x \in H$ and $x \in K$, but the fact that H and K are subgroups implies $x^{-1} \in H$ and $x^{-1} \in K$, from which deduce $x^{-1} \in H \cap K$. Therefore, $H \cap K$ is a subgroup by the stated lemma.

A useful result concerning subgroups is called Lagrange's Theorem for finite groups.

Lagrange's Theorem. If G is a finite group and H is a subgroup of G, then the order of the group $|G|$ is a multiple of the order of the subgroup $|H|$.

For example, if a group has eight elements, then there can be no subgroup of three elements. Be cautious in applying the theorem. Just because a group of eight elements has a particular subset of four elements, it does not imply that this set is a subgroup. What the theorem guarantees is that if H is a subgroup of G, then the number of elements in H must divide the number of elements in G. In other words, this theorem is a necessary, but not sufficient, condition for being a subgroup.

GENERATORS

A concept related to the ideas of groups and subgroups is that of generators. It would be most desirable if the group could be produced by considering a subset of the elements of the group in various combinations.

Definition 37. Let G be a group, and suppose $S = \{g_1,\cdots,g_N\}$ is a subset of G, such that all the elements in G may be produced as products involving only the elements in S, then we call the elements of S the generators of G.

Definition 38. Let G be a group and suppose that there is a single generator a, i.e., $G = \{a^i | i = 0,\pm1,\cdots\}$, or in other words, for every $x \in G$, there exists an integer n such that $x = a^n = \underbrace{a*a\cdots*a}_{n \text{ times}}$. G is written as G = (a), and G is called a cyclic group with generator a.

Examples

1. We have shown that $G = \{R_{90°},R_{180°},R_{270°},R_{360°}\}$, i.e., the rotations of the square leaving the vertices fixed is a group. This is a cyclic group with generator $R_{90°}$, because any other rotation may be obtained by repeated application of $R_{90°}$.

2. Consider the set of even integers, i.e., $G = \{\cdots,-4,-2,0,2,4,\cdots\}$ with the operation of addition. It may easily be shown that G is a group. The set of even integers is a generator group. $S = \{2,-2\}$, where we mean that any element in G is a multiple of 2 or -2.

Definition 39. If G is a group, and $a \in G$, then $(a) = \{a^i | i = 0,\pm1,\cdots\}$ and (a) is called a cyclic subgroup of G. (If there exists an element, a, such that G = (a), then G is a cyclic group.)

Definition 40. If G is a group and $a \in G$, then the smallest positive integer K, such that $a^K = e$ is called the period of a.

68

Example

1. In the case where $G = \{R_{90°}, R_{180°}, R_{270°}, R_{360°}\}$, $K = 4$, because $(R_{90°})^4 = R_{90°}*R_{90°}*R_{90°}*R_{90°} = R_{360°} = e$.

HOMOMORPHISMS AND ISOMORPHISMS

In this section we will relate two groups by means of mappings between them. These mappings will indicate the similarities of structure of the two groups.

Definition 41. A homomorphism ϕ is a mapping from one group G_1 into another group G_2, such that for all a,b in G_1, $\phi(a*b) = \phi(a) o \phi(b)$, where * is the operation for G_1 and o is the operation for G_2. If G_1 and G_2 are the same group then the operations * and o are the same.

Examples

1. Suppose ϕ is a mapping from G into G, defined by $\phi(x) = 2x$, and assume addition is the operation involved, then ϕ is a homomorphism. This is true because, for x and y in G, $\phi(x+y) = 2(x+y) = 2x + 2y = \phi(x) + \phi(y)$.

2. Suppose ϕ is a mapping from G_1 into G_2, and that G_1 is the real numbers together with the operation addition and G_2 is the real numbers together with the operation of multiplication. Define ϕ by $\phi(x) = 2^x$. Then, $\phi(x+y) = 2^{x+y} = 2^x \cdot 2^y = \phi(x) \cdot \phi(y)$. Therefore, ϕ is a homomorphism.

3. Suppose ϕ is a mapping from G into G_1 and G equals the integers, and the operation under consideration is addition. Define $\phi(x) = x+1$, then ϕ is not a homomorphism, because

$$\phi(x+y) = x + y + 1, \text{ but}$$

$$\phi(x) + \phi(y) = x + 1 + y + 1 = x + y + 2.$$

Definition 42. A mapping ϕ from G_1 into G_2, with G_1 and G_2 being groups, is an isomorphism if

(i) ϕ is a homomorphism, i.e., $\phi(a*b) = \phi(a) o \phi(b)$, where * and o are the operations of G_1 and G_2 respectively; and

(ii) ϕ is 1 - 1. That is, an isomorphism is a 1 - 1 homomorphism. An automorphism is an isomorphism of G onto itself.

Definition 43. Two groups G_1 and G_2 are isomorphic if there exists an isomorphism of G_1 onto G_2, i.e., there exists a mapping that is a 1 - 1 and onto mapping such that $\phi(a*b) = \phi(a)o\phi(b)$, where $*$ and o are the respective operations for G_1 and G_2.

It is important to realize what it means to say that two groups are isomorphic. It does not mean that the two groups are equal or identical. They may be, but they don't have to be. It does, however, indicate that the two groups are structurally alike or parallel. To establish an isomorphic relationship between a man and a computer does not say that the computer is the same as the man, but that there is a 1 - 1 correspondence between actions of the man and simulated actions of the machine.

In a poker game you are given a chip for every dollar you have; therefore, there is a 1 - 1 correspondence between the amount of chips you have and the amount of money you have, but a chip is not the same as a dollar. Try getting one chip's worth of gas at your local service station. The key idea of speaking of isomorphic sets or groups is to say that a structural parallelism exists between them.

Example

1. If we let $G_1 = \{R_{90°}, R_{180°}, R_{270°}, R_{360°}\}$, i.e., the rotations of a square, and for G_2 consider your watch. Set it at 12 o'clock. Define four elements: changing the watch to 3 o'clock, 6 o'clock, 9 o'clock, 12 o'clock, and denote these changes be A_3, A_6, A_9, A_{12} respectively. We may find a 1 - 1 onto mapping between G_1 and G_2: $\phi(R_{90°}) = A_3$, $\phi(R_{180°}) = A_6$, $\phi(R_{270°}) = A_9$, $\phi(R_{360°}) = A_{12}$. Also, $\phi(x*y) = \phi(x)o\phi(y)$, e.g., $\phi(R_{90°}*R_{180°}) = \phi(R_{270°}) = A_9 = A_3 o A_6 = \phi(R_{90°})o\phi(R_{180°})$. Therefore, G_1 and G_2 are isomorphic, but certainly a square piece of paper is not a watch, yet the imposed structures on G_1 and G_2 are the same.

We close this section with a few descriptive lemmas concerning homomorphisms.

Lemma. If ϕ is a homomorphism for G_1 into G_2, then ϕ maps the identity element of G_1 into the identity element of G_2, i.e., $\phi(e) = \bar{e}$.

Proof: Let x G, then $\phi(x)\bar{e} = \phi(x)$, since $\bar{e}$ is the identity element of G_2. But $\phi(x) = \phi(xe)$, since e is the identity element of G_1. Therefore, $\phi(x)\bar{e} = \phi(xe)$, but ϕ being a homomorphism implies $\phi(xe) = \phi(x)\phi(e)$. We thus have $\phi(x)\bar{e} = \phi(x)\phi(e)$, from which we deduce that $\phi(e) = \bar{e}$. We make use of what is called the cancellation law.

A valuable term related to a discussion of homomorphisms is that of the kernel of the homomorphism.

Definition 44. The kernel of a homomorphism ϕ, denoted kerϕ is defined for a homomorphism ϕ from G_1 into G_2, to be the set of elements in G_1 that are mapped into the identity element of G_2. kerϕ = $\{x \in G_1 | \phi(x) = \overline{e}\}$.

Recall that an isomorphism is a 1 - 1 homomorphism. An alternative to proving that ϕ is 1 - 1 is the next lemma.

Lemma. A homomorphism ϕ from G_1 into G_2 is an isomorphism if and only if the kernel of ϕ consists of the identity element of G_1 alone, i.e., kerϕ = $\{e\}$.

ADDITIONAL IMPORTANT GROUPS

We finish up this chapter with a discussion of a few very important groups that deserve special mention. In chapter 3 we examined mappings. In particular we investigated 1 - 1 mappings of a group, or actually at that time we just spoke of a set, onto itself. A result we stated without proof was that the composition of two 1 - 1 functions was a 1 - 1 mapping and similarly, the composition of two onto mappings was an onto mapping. Therefore, the composition of two 1 - 1 onto mappings would be also 1 - 1 onto, i.e., composition of mappings is a closed operation. It turns out that the composition of mappings is associative as well. There exists an identity mapping, namely f(x) = x, and this function we could denote it i would be the identity element for the set of 1 - 1 onto mappings. Finally, a 1 - 1 onto mapping has an inverse function that is also a 1 - 1 onto mapping. Therefore, the set of all 1 - 1 mappings of a set onto itself together with the operation of composition of functions is a group. It is not an abelian group, because if we return to the discussion of Chapter 3, it is clear that fog and gof generally are different.

A closely related example concerns the set of automorphisms. An automorphism was defined as an isomorphism of a group G onto itself. Therefore, an automorphism is a 1 - 1 mapping of G onto G, such that $\phi(a*b) = \phi(a)*\phi(b)$, where $*$ is the operation for G. It turns out that the set of automorphisms which are a subset of all 1 - 1 onto mappings are also a group.

The last example is tied in with the discussion of 1 - 1 onto mappings. We will briefly examine permutation groups.

Definition 45. Let S be a set, then a permutation, denoted by π, is a 1 - 1 mapping of S onto itself.

Therefore, a permutation is a mapping. The distinction between a permutation and an automorphism is that S does not have to be a group for permutations. We have just shown that the set of all 1 - 1 mappings of a set onto itself is a group, i.e., the set of all permutations of a set forms a group with the operation being composition. This group is referred to as the symmetric group.

Definition 46. The symmetric group is the group formed by the set of all 1 - 1 mappings of a set S mapped onto itself under the operation of composition.

Permutation groups are most valuable when the set under consideration is finite. If $S = \{a_1, \cdots, a_n\}$, then the permutation π is described by

$$\pi = \begin{pmatrix} a_1 & a_2 & \cdots & a_n \\ \pi(a_1) & \pi(a_2) & & \pi(a_n) \end{pmatrix},$$

i.e., the action of π on the element s in S is indicated in the second row. We will give a detailed analysis of the symmetric group S_3 on the three elements a_1, a_2, a_3, which for convenience we denote 1,2,3. For example if π is such that 1 goes to 3, 2 goes to 2, and 3 goes to 1, then

$$\pi' = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}.$$

For the three elements 1,2,3 there are six possible permutations, namely

$$\pi_1 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix} \quad \pi_2 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix} \quad \pi_3 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix} \quad \pi_4 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$$

$$\pi_5 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix} \quad \pi_6 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}.$$

We now will show that $\mathcal{S}_3 = \{\pi_1, \pi_2, \pi_3, \pi_4, \pi_5, \pi_6\}$ is a group. The operation will be composition and will be performed as follows. If we compute

$\pi_2 \circ \pi_4 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$, we start with the 1 in the left permutation.

Below the 1 is another 1, so we say 1 goes to 1, and then go the second permutation in the 1 spot. Here, 1 goes to 2. So we have $1 \to 1 \to 1 \to 2$, and therefore, $1 \to 2$. Next we start with the 2 in the left permutation, $2 \to 3$, so we go to the 3 in the right permutation, and see that 3 goes to 1. Therefore, $2 \to 3 \to 3 \to 1$, or 2 goes to 1. Finally, we start at 3 in the left permutation. $3 \to 2$ and so we go to 2 in the right permutation and $2 \to 3$. Therefore, $3 \to 2 \to 2 \to 3$ or $3 \to 3$. Combining our results we have

$$\pi_2 \circ \pi_4 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix} = \pi_3.$$

Another example would be

$$\pi_5 \circ \pi_3 = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix} = \pi_6,$$

where $1 \to 3 \to 3 \to 3$, $2 \to 1 \to 1 \to 2$, and $3 \to 2 \to 2 \to 1$.

A complete table would look like the one in Table. Notice that $\pi_1 = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}$ is the identity, because it maps each element into itself. From the table it may now be verified that $\mathcal{S}_3$ is a group. We have already proven that the set of all 1 - 1 onto mappings of a set onto itself is a group, but it would be interesting practice for the reader to try to verify some of the entries in Table 4.

The terms transitive and regular permutation group appear frequently in the literature.

Definition 47. A permutation group is said to be transitive if it has the property of containing a permutation which replaces any given letter, or a_i, by any other letter, i.e., each of the letters of the group may be replaced by each of the other letters of the group.

Our group $\mathcal{S}_3$ is an example of a transitive group.

Definition 48. A regular permutation group is a transitive group whose order, or number of mappings in the group is equal to its degree of elements or letters being transformed.

Table 4

	π_1	π_2	π_3	π_4	π_5	π_6
π_1	π_1	π_2	π_3	π_4	π_5	π_6
π_2	π_2	π_1	π_4	π_3	π_6	π_5
π_3	π_3	π_5	π_1	π_6	π_2	π_4
π_4	π_4	π_6	π_2	π_5	π_1	π_3
π_5	π_5	π_3	π_6	π_1	π_4	π_2
π_6	π_6	π_4	π_5	π_2	π_3	π_1

We now have completed a fairly rich description of elementary group theory. The examples were included to illustrate the new definitions. The precision and elegance of the theory hopefully impresses the reader. If there would be any way that psychology could draw on this theory, it would be most desirable. The next chapter includes an impressive list of examples of how group theory has already entered the domain of psychology.

CHAPTER 5

THE APPLICATION OF GROUPS TO PSYCHOLOGY

There will not be any new mathematical terminology introduced in this chapter. The chapter is devoted to the description of various applications of group theory in the behavioral sciences.

In order to understand Piaget's theory of formal operations (Piaget & Inhelder, 1958), the reader should be familiar with basic propositional logic which is an area outside of the discussion in this book, and the INRC group, which is now within the realm of our understanding. There are four elements in this group, namely,

(i) I, the identity operator, which when applied to any proposition leaves the proposition unaltered;

(ii) N, the negation or inverse operator, which means one can return to the starting point by cancelling an operation already performed;

(iii) R, the reciprocal operator, which means that one may return to the starting point by compensating a difference, i.e., the product of two reciprocal transformations is not the identity but an equivalence; and

(iv) C, the correlative operator which is the negation of the reciprocal operator.

The multiplication table in Table 5 is the same as that of the "4-group" discussed in the preceding chapter

Table 5

	I	N	R	C
I	I	N	R	C
N	N	I	C	R
R	R	C	I	N
C	C	R	N	I

To fully appreciate the role of the INRC transformation would require a discussion of propositional logic and Boolean algebra, but we can give an illustration of how the INRC group would be applied in the task of establishing equilibrium for a balance.

Suppose that a balance is in equilibrium, we may cause disequilibrium by changing one of the weights or altering the distance of one of the weights from the fulcrum, or performing some combination of a weight and distance change. Assume we replace a weight of five pounds with a new weight of ten pounds. Then the negation or inverse of this action would be to remove the ten-pound weight and replace it again with the original five-pound weight. An example of a reciprocal operation would be to replace the weight on the other arm of the balance with a weight of twice the original. This action compensates for the original action, but does return the balance to equilibrium in the exact same way as it originally was. The correlate would be the negation of the reciprocal transformation.

The most important changes for Piaget are the negation and reciprocal transformations. They are the two forms of reversibility, i.e., the original situation may be restored by either cancelling a performed operation or by compensating for the operation. An understanding of the role of reversibility in Piagetian theory cannot be whole without an appreciation of the underlying mathematical framework of his theory.

There are certain weaknesses and limitations in the Piagetian logical-mathematical model for the stage of formal operations. Bart (1971) points out that the INRC transformation group is inadequate in explaining how certain logical propositions that are operations can be transformed into other element operations. Therefore, Bart has formulated a generalization of this model. The generalization presupposes an understanding of the Boolean algebraic structure of combinatorial thinking and the regular Boolean permutation group structure of hypothetico-deductive thinking. The method of designating the formal transformations in the groups descriptive of formal thought is in terms of the symmetric difference operation that we have already examined in detail.

One weakness in Piaget's theory is that it does not distinguish the level of cognitive complexity of one level of combinatorial ability from another level. Suppose Γ_k represents one individual's level, and Γ_{k+1} another individual's level, then the second person would be at a higher level. A type of mapping or transformation ϕ defined on Γ_k will be a permutation, and will be called the symmetric difference transformation. These transformations form a group, in fact a regular permutation group. From this framework a method of positive intersection generators is employed to indicate the primitive formal transformations proper to a level of formal thought.

The generalization model can describe any situation that Piaget's INRC model can, and in addition those cases where the Piagetian approach is inadequate. Also the generalization has qualitatively distinct levels within the stage of formal operations.

Group theory may be used to study the kinship of different primitive societies. Boyd (1969) has written an article on this topic. He offers a justification for applying groups to model marriage class systems. For example, if one group G_1 evolves into a second group G_2, then G_1 and G_2 are related through homomorphic images. The actual kinship systems are generated by means of grammars and the kinship system may be clarified by componential analysis through the use of Cartesian products. Boyd points out that if the dimensions are generation and sex, then (+1, female) would be someone's mother. His goal is to use a mathematical model to bring seemingly different problems into a larger all-encompassing theory. The theories of kinship grammars and componential analysis are related by a regular permutation group.

Boyd gives a study of the Arunta tribe, an Australian tribe that has marriage classes. The Arunta make distinction between older and younger siblings, and the sex of the speaker influences which kinship term is required. The set of one word kinship terms are: a man's father; a man's mother; a woman's father; a woman's mother; elder brother; elder sister; younger brother; younger sister; a man's child; a man's son; a man's daughter; a woman's son; a woman's daughter; wife; and husband. Boyd calls this set A. Any other relatives may be formed by composing some of the above terms.

The Arunta tribe may be partitioned into eight marriage classes. All the fathers of children in a particular class, themselves came from the same class, and conversely all the children of men in a given class belong to the same class. This relation of fatherhood, F, describes a permutation, and similarly the relation of motherhood, M, describes a permutation. Other relations may be derived from M and F. The set of all possible compositions of the permutations F and M generate a permutation group. In fact, the group is a regular permutation group. From this group the other kinship terms may be incorporated into this network.

For Boyd, the meaningful way to apply groups to psychology is to study the permutation or transformation groups of a structure onto itself, because it is the study of actions or transformations that offer insight into problems.

Group theory has been applied to questions in perception. Hoffman (1966) demonstrated that perceptual constancies such as image location in the field of view, size constancy, shape constancy, and others may be described in terms of Lie groups of transformations. Our discussion of his articulation must of necessity be rather superficial, since a Lie group is more than a group. It is also a differential manifold, and Lie theory is on a much higher plane than our elementary examination of groups. The interested reader would have to consult mathematical textbooks on Lie theory. Hoffman offers an explanation of how a Lie theory of visual perception may be used to account for complementary after-images, i.e., the after-effect of seen movement, and the visual analog of relativistic length contraction.

Groups also have application in the theory of measurement. Luce and Tukey (1963) provided a theory for interval measurement based on the ordering of objects, so long as the contributions of at least two distinct factors are simultaneously considered. This theory is called conjoint measurement. Krantz (1964) considers an approach in which an equivalence relation may be defined in a Cartesian product in such a way that the resulting set of equivalence classes form a commutative group. Different group structures in the same product set will be isomorphic, i.e., there exists an isomorphism of one group onto the other. Krantz introduces an ordered group, which is defined as a group with a partial ordering on it, such that for $x, y \in G$ and $x \prec y$, then for any $z \in G$, $x*z \prec y*z$ and $z*x \prec z*y$. Further, if $\prec$ is a linear or simple ordering, then G is a simply ordered group. An Archimedean simply ordered group is defined to be a group where for $x \neq e$, e the identity element, and y any element in G, there then exists an integer n such that $x^n \succ y$.

He then establishes that an Archimedean simply ordered group is isomorphic to a subgroup of the real numbers under addition, which in turn then leads to interval scale measurement.

Cross-context matching is the situation where an observer states that certain stimuli in one context match other stimuli in another context. Krantz (1968) points out that the changing from one context S to another T, describes a function he denotes by $g_{S,T}$, where $g_{S,T}(A) = B$, if A is a stimulus in context A and B is a stimulus in T. In perceiving something it is not enough to ask about the particular stimulus; the spatial and temporal context must also be considered. If there exists a set of transformations of the stimulus elements such that these mappings form a semigroup, i.e., a closed associative set, and if the collection of mappings are context-invariant, then the $g_{S,T}$ are transformations of commutative groups, and knowledge of certain context effects may be utilized for predicting other context effects. What makes this article fairly involved is that the discussion is going on at three levels:

(i) transformations of stimuli;

(ii) isomorphisms of transformation groups; and

(iii) functions from pairs of contexts into the group of automorphisms of a transformation group. This third level is where the predictive power of context changes is richest.

In psychology it is crucial to be able to replicate a test or task, and for this reason Levine's (1970) article on transformations that produce parallel curves or sets should be of interest. In stimulus generalization studies, Thurstonian psychophysics, mental test theory, JND scales, and Fechnerian psychophysics and utility theory, Levine points out the value of comparison between two tests, two curves, etc. He sees the finding of all the functions that render a given set of functions parallel to be a major task. These functions are referred to as scales. Any scale that renders a set of scales parallel is called a solution for the set, and a set having a solution is called a uniform system. As an illustration, a set of two scales is a uniform system if and only if the set is uncrossed, i.e., if F and G are the scales, $F(x) < G(x)$ for all x, or $F(x) = G(x)$ for all x, or $F(x) > G(x)$ for all x. This is then generalized to any arbitrary number of scales. More precisely, each scale may be thought of as a 1 - 1 continuous mapping of the real numbers onto themselves. The operation involved is composition, and each set of scales is associated with a unique group under the operation of composition.

It turns out, for example, that if two sets of scales have the same associated group, they also have the same set of solutions rendering them parallel. By an associated group, Levine means that for each pair of scales F and G in a set of scales, $\mathcal{J}$, the associated group of that set of scales is the group generated by $F^{-1}G$.

By following Levine's procedures, the psychologist can determine whether his sets of curves may be rendered parallel or if he must modify his approach.

A relatively new area in psychology where mathematics is used is the study of language and communication. Chomsky (1963) has been considering the question of how is it that a person has the ability to comprehend sentences that he has never heard, and on other occasions, provide appropriate novel responses. Chomsky describes the flow of speech as a sequence of discrete atoms that are concatenated, i.e., right after each other.

He defines a system, with L being the set of all finite sequences that can be formed from the elements of some arbitrary finite set V. He defines an operation $\frown$, that represents the result of concatenating two sequences ϕ and $\chi \in L$. If $\phi \frown \chi = \psi$, where $\psi \in L$, i.e., ψ is a new finite sequence, then L is closed under $\frown$. The operation $\frown$ is also associative $(\phi \frown \chi) \frown \psi = \phi \frown (\chi \frown \psi)$, provided that one carefully formulates what he means by associativity. The empty or null sequence is the identity element, so L under the operation $\frown$ may be viewed as a monoid or semigroup with an identity element.

Chomsky gives an example of why associativity must be carefully defined. Notice that "they$\frown$(are$\frown$(flying$\frown$planes))" has a different meaning from "they$\frown$(are$\frown$flying)$\frown$planes)." This difficulty is avoided by assuming that a language has several distinct levels. Lower levels are specified by how they relate to higher levels. It is necessary then to have several concatenation systems. These systems are used in the attempt to characterize a grammar in such a way that an explicit enumeration of grammatical sentences is possible.

The process of coding is the mapping of one monoid into another. Chomsky illustrates this by considering one monoid to be all the strings that can be formed from the characters of a finite alphabet A, and the other monoid to be all the strings that can be formed by words in a finite vocabulary. A code would be an isomorphism of U into a subset of A. The theory is then extended to states, where a state of a coding system represents the memory at a given moment. The memory is augmented with time.

Arbib (1968) has edited a book on the algebraic theory of machines and languages in which the discussion is in terms of semigroups. In one particular chapter, Assmus and Florentin (ibid.) explain machine theory using semigroups as the fundamental connection between algebra and machines. The semigroup is used to form a standard version of any machine, methods of decomposing semigroups describe parallel decompositions of the machine into components, and also the definitions of irreducible component machines are in terms of the decompositions of semigroups, and then these irreducible component machines are used to build all other machines. If the state transition maps are permutations, then a machine with only permutations as mappings has a semigroup that is actually a group. The set of permutations are transitive, i.e., any state can be reached from any other state.

An examination of the book clearly reveals that the parallel study of machines and the theory of semigroups is necessary to have any real appreciation of the foundations of machine or automata theory.

Berlyne (1964) has a chapter on group structures and equilibrium in his book. He begins by talking about habit family structures, i.e., there exist parallel strands joined together at their beginnings and ends, which indicate that each has the same stimuli situation, and each led to the same response. He then describes how the habit family hierarchies in thinking must be more complex, and suggests that the study of transformation groups may be helpful. He draws on the work of people like Piaget and Poincaré.

For example, a group has an inverse, which may either be a compensation or a cancellation. The importance of reversibility in thinking and questions of equilibrium is of the utmost. The ability to consider an action, and then determine whether it is appropriate or not, without actually carrying it out, is fundamental to thinking. Any behavior system possessing a group structure also would have a habit family hierarchy, but Berlyne points out that the converse is not true. The system may for instance have a groupoid, semigroup, or monoid structure.

In situations where group structures are relevant, a transitive transformation group is the most desirable, because it always allows the possibility to get from any one element to any other element by means of one transformation. This offers great efficiency and economy of effort in assessing any situation. For this reason, the consideration of transitive groups should be applied to questions of equilibrium.

In a transitive group structure, no starting point is needed, because no matter what situation a person encounters, the person has the ability to compensate or modify it.

Natapoff (1970) illustrates how groups may be used in symmetric choice experiments. He defines a symmetric choice experiment as an experiment where the way the distribution of choices among alternatives that appear almost identical depends on those minor differences among the alternatives. The seeming equivalence reflects the symmetry of the problem, while the differences indicate the restrictions or limitations of the symmetry. Group theory is helpful in analyzing such experiments.

If $S_1, \cdots, S_N$ are N similar alternatives, he calls them states, of some fixed quantity that is to be symmetrically distributed, then $f(S_i)$ will represent the fractional share of the quantity that is given to the i^{th} choice. If two states are the extent of the choices, then $f(S_1) + f(S_2) = 1$, where 1 represents the entire quantity under consideration. In general $f(S_1) + \cdots + f(S_N) = 1$.

Suppose that all of the states are essentially the same; the choose one as a reference state and form a set G, $G = \{g_1, \cdots, g_N\}$, where the g_i are transformations mapping the reference state into each of the original N states. Therefore, one of the g_i will be the identity transformation.

The focus of the task is no longer on N states, but one reference state and a set of transformations. G reflects the symmetry of the set of states, and the set of transformations g form a group. Actually, which state is used as the reference state is immaterial. The set G will always produce the N states $S_1, \ldots, S_N$, only the order for $g_1 S_j$, $g_2 S_j, \ldots g_N S_j$ may be different. For example, $g_2 S_i$ may be S_5 and $g_2 S_j$ may be S_7.

From here Natapoff shows that every symmetric choice function may be reduced to a simpler type of function, from which greater amounts of information may be extracted than if the built-in symmetry of the experiment was not taken advantage of.

Hopefully, the ll examples of the application of groups to psychology have illustrated the broad range of uses of groups already in the psychological literature. Yet the value of mathematical analysis has not been fully appreciated. If this chapter has served as a motivation to begin a closer examination of the potential power of mathematical structures, then this book has fulfilled its purpose.

CHAPTER 6

RINGS AND FIELDS

This chapter will be relatively short, because presently there are very few applications of rings and fields to psychology. This does not mean that rings and fields will not be helpful in analyzing psychological questions, but rather that their applicability has not really been tested yet. In this chapter we will define the important terminology and illustrate these definitions through fairly elementary mathematical examples. A few basic properties of rings and fields will be proven to give the reader a greater feeling of how these new concepts may be used.

All the algebraic structures that will be introduced have the common quality of having two operations. Remember, the group concept has only one operation. The ring is the most fundamental of the two-operation structures.

Definition 49. A ring R is a nonempty set of elements with two operations defined on it; for convenience they are denoted by + and $\cdot$, such that

(i) For all $a,b \in R$, $a + b \in R$;

(ii) For all $a,b,c \in R$, $a + (b + c) = (a + b) + c$;

(iii) There exists an element 0 in R, such that $a + 0 = 0 + a = a$ for all $a \in R$;

(iv) For every a in R there exists an element $-a$ in R, such that $a + (-a) = (-a) + a = 0$;

(v) For every $a,b \in R$, $a + b = b + a$;

(vi) For every $a,b \in R$, $a \cdot b \in R$;

(vii) For all $a,b,c \in R$, $a \cdot (b \cdot c) = (a \cdot b) \cdot c$;

(viii) For all $a,b,c \in R$, $a \cdot (b + c) = a \cdot b + a \cdot c$ and $(b + c) \cdot a = b \cdot a + c \cdot a$. This law is called the distributive law.

In reading through these eight conditions that must be satisfied for a set to be a ring, perhaps the reader observed that this definition may be written more compactly.

Definition 50. A <u>ring</u> R is a nonempty set of elements with two operations, denoted by + and $\cdot$, such that

(i) R is an abelian group under +;

(ii) R is a semigroup under $\cdot$; and

(iii) R satisfies the distributive property, i.e., for all $a,b,c \in R$, $a\cdot(b + c) = a\cdot b + a\cdot c$ and $(b + c)\cdot a = b\cdot a + c\cdot a$.

The other algebraic structures that we will consider are built up from a ring by adding additional properties.

Definition 51. A <u>ring with an identity</u> R is a ring where the operation $\cdot$ has an identity element, i.e., there exists an element $1 \in R$ such that for every $a \in R$, $a\cdot 1 = 1\cdot a = a$. Therefore, R is a monoid under the operation $\cdot$.

Definition 52. A <u>commutative ring</u> R is a ring for which the operation $\cdot$ is commutative, i.e., for every $a,b \in R$, $a\cdot b = b\cdot a$.

Definition 53. A ring is called an <u>integral domain</u> if it is a commutative ring with an identity and satisfies the additional property, that if for $a,b \in R$ we have $a\cdot b = 0$, then either $a = 0$ or $b = 0$ or both a and b equal 0.

This added property has a name.

Definition 54. In a commutative ring, if for $a \neq 0$ there exists an element $b \neq 0$, such that $a\cdot b = 0$, then a is called a <u>zero divisor</u>.

Definition 55. A <u>division ring</u> R is a ring where its nonzero elements form a group under the operation $\cdot$.

The final related definition is that of a field.

Definition 56. A <u>field</u> F is a ring whose nonzero elements form a commutative group under the operation $\cdot$, or in other words, a field is a commutative division ring.

Figure 34 in a sense indicates an ordering among the related concepts and may aid in learning the new definitions. A similar diagram appears in Dean (1966). In his figure a line from one definition A to a definition B, higher on the figure, indicates that every system in A is also a system in B.

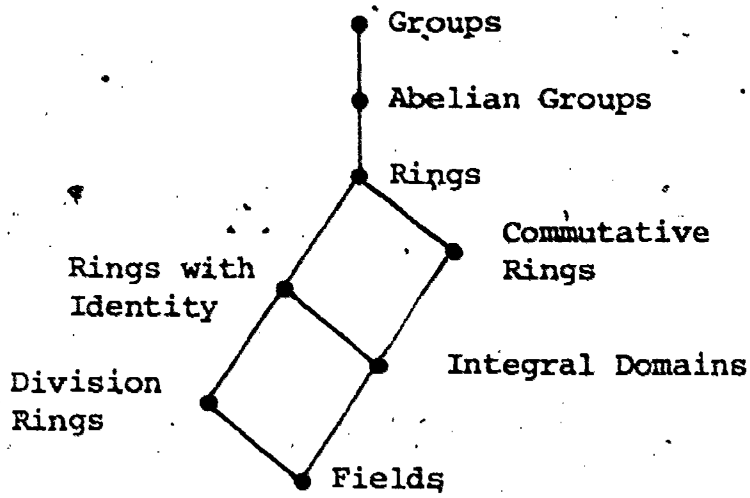

Figure 34

Before we begin to look at some examples, it should be pointed out that the operations + and · do not have to be normal arithmetic addition and multiplication. They may represent any pair of operations satisfying the list of conditions.

Examples

1. Consider the integers with the operations of arithmetic addition and multiplication. We have already proven that the integers form a group under addition, in fact in abelian group. The integers are closed under multiplication and are also associative and commutative under multiplication and the distributive property holds. There is an identity element, namely 1, since any integer times 1 is the same integer. However, the integers with the exception of 1 and -1 do not have their multiplicative inverses in the integers. For example, the inverse of 5 is 1/5. Therefore, the integers with + and · form a commutative ring with identity element. If we now observe that there are no zero divisors in the integers, i.e., the only way the product of two integers can be zero is if at least one of them is zero, then we may conclude that the integers are an integral domain.

2. The even integers with the operations of addition and multiplication would be a commutative ring. The even integers are equal to {····, -4,-2,0,2,4,····}, and therefore, there is no multiplicative identity.

3. An example of a field would be the rational numbers with the operations of addition and multiplication. The multiplicative identity is 1, and the rationals have the multiplicative inverse of any element. For example, the inverse of 9 would be 1/9, of 2/3 would be 3/2, etc. Therefore, the rationals form a commutative group under addition, a commutative group under multiplication, and clearly the distributive property holds.

4. If the set under consideration is the set of functions from the real numbers into the real numbers, and the operations are defined by

$$(f + g)(x) = f(x) + g(x)$$
$$(f \cdot g)(x) = f(x) \cdot g(x),$$

then

(i) Closure under + follows from the definition.

(ii) $((f + g) + h)(x) = (f + g)(x) + h(x) = f(x) + g(x) + h(x) = f(x) + (g + h)(x) = (f + (g + h))(x)$. Therefore, + is associative.

(iii) The identity element for + is the function that is identically 0, i.e., $(f + 0)(x) = f(x) + 0(x) = f(x)$.

(iv) The inverse of a function f will be $-f$ under addition, since $(f + (-f))(x) = f(x) - f(x) = 0$.

(v) The set of functions is abelian, since $(f + g)(x) = f(x) + g(x) = g(x) + f(x) = (g + f)(x)$.

(vi) Closure under $\cdot$ follows from the definition.

(vii) Similarly, the associativity of $\cdot$ follows.

(viii) The distributive laws hold. We prove one of them, and the other follows in the same manner.

$(f \cdot (g+h))(x) = f(x) \cdot (g+h)(x) = f(x) \cdot [g(x) + h(x)] = f(x) \cdot g(x) + f(x) \cdot h(x) = (f \cdot g + f \cdot h)(x)$.

Therefore, the set of functions from the real numbers into the real numbers is a ring.

There is an identity element, namely the function identical to 1, since $(f \cdot 1)(x) = f(x) \cdot 1(x) = f(x)$. The commutivity of $\cdot$ follows immediately from the definition. The set of function is not an integral domain, because there exists a function not equal to zero, whose product is the zero function. For example, if f is defined as

$f(x) = \begin{cases} 1 \text{ if } x \geq 0 \\ 0 \text{ if } x < 0 \end{cases}$ and g is defined by $g(x) = \begin{cases} 0 \text{ if } x \geq 0 \\ 1 \text{ if } x < 0 \end{cases}$, then the product function $(f \cdot g)(x) = f(x)g(x) = 0$ for all x.

5. In the chapter on relations we showed that the relation, the remainder upon division by 5, partitioned the integers up into five classes, namely, $[0] = \{\ldots,-10,-5,0,5,10,\ldots\}$, $[1] = \{\ldots,-9,-4,1,6,11,\ldots\}$, $[2] = \{\ldots,-8,-3,2,7,12,\ldots\}$, $[3] = \{\ldots,-7,-2,3,8,13,\ldots\}$, and $[4] = \{\ldots,-6,-1,4,9,14,\ldots\}$. Let $R = \{[0],[1],[2],[3],[4]\}$; we will show that if $[m] + [n]$ is defined to be the remainder of $m + n$ upon division by 5, and $[m]\cdot[n]$ is defined to be the remainder of $m \cdot n$ upon division by 5, then R is a commutative ring with a unit element. In fact, we will be able to show that R is a field. That R is a ring is easily verifiable from the definitions of the operations. For example, [0] would serve as the identity element in addition. The additive inverses of [0] would be [0], of [1] would be [4], of [2] would be [3], of [3] would be [2], and of [4] would be [1], since in each case the sum is equal to [0]. The distributive property may be verified rather easily. One illustration of the distributive law is $[2]\cdot([3] + [4]) = [2]\cdot[7] = [2]\cdot[2] = [4]$, and $[2]\cdot[3] + [2]\cdot[4] = [6] + [8] = [1] + [3] = [4]$. Therefore, $[2]\cdot([3] + [4]) = [2]\cdot[3] + [2]\cdot[4]$. If we now assume that R is a ring, we observe that [1] serves as the multiplicative identity. The commutativity of $\cdot$ is an immediate consequence of the commutativity of the integers since for two integers n and m, $n \cdot m = m \cdot n$. Each element has a multiplicative inverse; the inverse of [1] is [1], of [2] is [3], of [3] is [2], and of [4] is [4], because in each case the product equals [1]. Therefore, R is a field. As a means of reviewing the example, we include product tables for the two operations in Tables 6 and 7.

Table 6

+	[0]	[1]	[2]	[3]	[4]
[0]	[0]	[1]	[2]	[3]	[4]
[1]	[1]	[2]	[3]	[4]	[0]
[2]	[2]	[3]	[4]	[0]	[1]
[3]	[3]	[4]	[0]	[1]	[2]
[4]	[4]	[0]	[1]	[2]	[3]

Table 7

[·]	[1]	[2]	[3]	[4]
[1]	[1]	[2]	[3]	[4]
[2]	[2]	[4]	[1]	[3]
[3]	[3]	[1]	[4]	[2]
[4]	[4]	[3]	[2]	[1]

6. An interesting observation is that if we defined the relation to be the remainder upon division by 4, then there would have been four classes [0], [1], [2], [3]. However, in this example, the nonzero elements do not form a group under multiplication. There is a zero divisor, namely [2], because [2]·[2] = [0] and [2] certainly is not the zero element. The multiplication table in Table 8 shows that [2] does not have an inverse for the operation of multiplication. What are the differences between division by 4 and by 5 that cause such a drastic difference in the structures of the two systems? As an exercise, the reader should do a similar analysis for division by 6 and 7 and then on the basis of these results try to generalize when a system will be a field and when it will not.

Table 8

[·]	[1]	[2]	[3]
[1]	[1]	[2]	[3]
[2]	[2]	[0]	[2]
[3]	[3]	[2]	[1]

7. If we consider our set to consist of all the subsets of some given set, and let the two operations be the symmetric difference and intersection, then we have a commutative ring with identity (Burton, 1965). We have already proven in the chapter on groups that for the set of all subsets of some universal set, the symmetric difference yields a group structure. The intersection operation is closed and associative. Therefore, if the distributive law holds, then we have a ring. $A \cap (B \Delta C) = A \cap [(B - C) \cup (C - B)] = [A \cap (B - C)] \cup [A \cap (C - B)]$. By an argument analogous to those of the first chapter, $A \cap (B - C) = (A \cap B) - (A \cap C)$ and $A \cap (C - B) = (A \cap C) - (A \cap B)$. Therefore, $A \cap (B \Delta C) = [A \cap (B - C)] \cup [A \cap (C - B)] = [(A \cap B) - A \cap C] \cup [A \cap C - (A \cap B)] = (A \cap B) \Delta (A \cap C)$. Similarly, that $(B \Delta C) \cap A = (B \cap A) \Delta (C \cap A)$ may be demonstrated. Therefore,

our system is a ring. The ring is commutative because $A \cap B = B \cap A$, and the ring also has an identity, namely the universal set, since $A \cap U = A$, where U is the universal set.

An interesting problem that may be proven by an application of the distributive law is that any number times zero is zero. If someone asked you why $a \cdot 0 = 0$, you would probably say because anything times zero equals zero, and he would again say why, and suddenly you are in the midst of a vicious circle. Let us actually prove that $a \cdot 0 = 0$.

<u>Lemma</u>. Let R be a ring, then for any $a \in R$, $a \cdot 0 = 0$.

Proof: Let a be any element in R. If o is the identity element under addition, then in particular $0 = 0 + 0$. Therefore, $a \cdot 0 = a(0 + 0) = a \cdot 0 + a \cdot 0$. But, since R is a group under addition, each element has an inverse, and we may cancel out an $a \cdot 0$ from each side of the equation. Therefore, $0 = a \cdot 0$, or equivalently $a \cdot 0 = 0$.

A rather important result that we hinted at in our discussion of the various rings or fields formed on the basis of the relation defined by the remainder upon division by a particular number will be stated without proof.

<u>Theorem</u>. A finite integral domain, i.e., an integral domain with a finite number of elements, is a field.

In the example based on division by 5, we had a field structure, however, with division by 4; there were zero divisors; hence we did not have an integral domain, and consequently we did not have a field. Notice that this theorem only holds for finite sets.

A third interesting question is would it be possible in a ring to have the identity element under addition and under multiplication be the same element? The answer is no; they are distinct provided that the ring is not the ring consisting of 0 alone.

<u>Theorem</u>. Let R be a ring with an identity, and assume $R \neq \{0\}$, then the elements 0 and 1 are distinct.

Proof: Let a be a nonzero element of R. If 1 is the identity element, then $a \cdot 1 = a$. We also have just proven that for $a \in R$, $a \cdot 0 = 0$. Therefore, 0 is not possibly equal to 1, unless $a = 0$, but by assumption $a \neq 0$.

In the discussion of groups we spoke of subgroups, and it is reasonable that in our examination of rings we would like to have the corresponding idea of a subring.

Definition 57. Let R be a ring and suppose that S is a subset of R, such that under the same operation, + and ·, that are used in R, that S is itself a ring, then S is called a subring.

It is not necessary to check all the properties of a ring, because several of them are built into the ring structure. For example, if R is associative, clearly a subset of R, namely S, is associative. It turns out the crucial properties to check are essentially three in number.

Theorem. A nonempty subset S of a ring R is a subring if and only if

(i) For all $a,b \in S$, $a + b \in S$, where + is the additive operation of R;

(ii) For every $a \in S$, $-a$ is also an element of S, i.e., the additive inverse is in S for every element of S; and

(iii) For all $a,b \in S$, $a \cdot b \in S$, where $\cdot$ is the multiplicative operation.

It is not necessary to have a separate condition that 0 belong to S because if $a \in S$, then by (ii) $-a$ also belongs to S. Now applying (i), since a and $-a$ both belong to S, then $a + (-a) = 0$ also is in S.

Examples

1. The even integers are a subring of the integers under normal addition and multiplication. If we apply the previous theorem, we see that the set of even integers is closed under addition, has an additive inverse for every element, and is closed under multiplication.

2. The odd integers would not be a subring because they are not closed under addition. For example, 3 and 5 are both odd integers, but $3 + 5 = 8$, and 8 is not an odd integer.

3. Another example of a subring is the ring (or actually the field) of rational numbers which has the integers as a subring.

We introduced the concept of a homomorphism in the discussion on groups. We will now introduce a parallel idea for ring theory. The distinction being that the ring has two operations and the group just one, so that the definition of homomorphism must involve both operations.

<u>Definition 58</u>. Let R_1 and R_2 be two rings. A mapping ϕ from R_1 into R_2 is called a <u>homomorphism</u> if for all $a,b \in R$

(i) $\phi(a+b) = \phi(a) + \phi(b)$; and

(ii) $\phi(a \cdot b) = \phi(a) \cdot \phi(b)$.

It must be stressed that the + and $\cdot$ in R_1 and R_2 need not necessarily be the same operations.

Examples

1. The identity mapping $\phi(x) = x$ from the real numbers onto the reals numbers is a ring homomorphism:

   (i) $\phi(a+b) = a+b = \phi(a) + \phi(b)$; and

   (ii) $\phi(a \cdot b) = a \cdot b = \phi(a) \cdot \phi(b)$.

2. The mapping $\phi(x) = 5x$, however, is not a ring homomorphism. In fact, $\phi(x) = kx$, where k is any number other than 1 is not a ring homomorphism:

   (i) $\phi(a+b) = 5(a+b) = 5a + 5b = \phi(a) + \phi(b)$; however,

   (ii) $\phi(a \cdot b) = 5a \cdot b$ and $\phi(a) \cdot \phi(b) = (5a) \cdot (5b)$, and clearly $5a \cdot b = 25a \cdot b$, or in other words $\phi(a \cdot b) \neq \phi(a) \cdot \phi(b)$.

3. We have proven that the relation, the remainder upon division by 5, defined a field consisting of the elements [0], [1], [2], [3], and [4]. If we consider the mapping $\phi(x) = [x]$, then ϕ is a ring homomorphism:

   (i) $\phi(a+b) = [a+b] = [a] + [b] = \phi(a) + \phi(b)$; and

   (ii) $\phi(a \cdot b) = [a \cdot b] = [a] \cdot [b] = \phi(a) \cdot \phi(b)$.

   This example is an illustration of the difference between the operations in one ring and another. The + in R_1 is normal addition, while the + in R_2 is the addition of equivalence classes of numbers. For instance, 27 + 16 = 43, while [27] + [16] = [43] = [3] with respect to the relation the remainder upon division by 5.

There are several related definitions that we now introduce.

Definition 59. If ϕ is a homomorphism from ring R_1 into ring R_2, then the kernel of ϕ is defined to be the set of all elements in R_1 such that ϕ applied to any of these elements yields the additive identity of R_2, i.e., if a is an element of the kernel, then $\phi(a) = 0$.

Examples

1. In the case of $\phi(x) = x$, the kernel consists of only the element 0, since every other element is mapped onto a nonzero value.

2. In the example $\phi(x) = [x]$, where $[x]$ represents the class determined by the remainder upon division of x by 5, the kernel consists of all multiples of 5. This is true, because any multiple of 5 is mapped into the class [0], and [0] is the additive identity for the field consisting of [0], [1], [2], [3], and [4].

Definition 60. An isomorphism ϕ is a homomorphism of ring R_1 into ring R_2 such that ϕ satisfies the additional condition of being a 1 - 1 mapping.

If we carry the analogy of rings to groups one step further we may now define when two rings are isomorphic.

Definition 61. Rings R_1 and R_2 are isomorphic is there exists an isomorphism of R_1 onto R_2, i.e., there is a 1 - 1 mapping from R_1 onto R_2 that satisfies

(i) $\phi(a+b) = \phi(a)+\phi(b)$; and

(ii) $\phi(a \cdot b) = \phi(a) \cdot \phi(b)$.

The overall discussion of rings and fields was not as deep as that of groups, the reason being that the chapter on groups could be followed up by a rich collection of explanatory examples from the behavioral sciences. Unfortunately, little work has been done in psychology that uses rings and fields. Perhaps the difficulty is that rings and fields require two operations and in addition these operations are interrelated by the distributive properties. It, therefore, stands to reason that any behavioral system that may be described by a ring or field structure must be quite involved. Only after the full potential of group theory is realized in the behavioral sciences will we really be able to pass judgment as to the applicative value of rings and fields.

CHAPTER 7

VECTOR SPACES AND LINEAR TRANSFORMATIONS

In this chapter we introduce another algebraic system. A vector space will have structural similarities to the other systems that we have examined in preceding chapters, but it differs from the other systems in that it has an operation that is defined with respect to a field whose elements serve as operators on the vector space.

The value of particular vector spaces in statistical and measurement analyses of psychological questions has been widely recognized, as may be indicated by the fact that many graduate psychology departments required students to have training in statistics and measurement. In these classes the students learn techniques and methods that are based on vector space theory. The examination of vector spaces will be in two parts. The first chapter introduces the concept of a vector space, offers examples of vector spaces, and then includes a discussion of linear combinations, linear independence and dependence, and bases, that serve in a sense as the building blocks, structurally speaking, of a vector space. A detailed study of linear transformations follows, in which, among other things, it is shown that the set of linear transformations is itself a vector space.

The second chapter is directed at the concept of a matrix. The matrix is an excellent concept to conclude the book with, because it will be proved that the set of matrices may be used in defining a group, or a ring, or a vector space, or under certain special conditions, in defining a field. This will serve as a review of the key structures introduced in the book. Matrices also are valuable to discuss because they have a wide range of applications outside of mathematics.

We now begin the examination of vector spaces by giving a definition of a vector space.

Definition 62. A nonempty set V is called a vector space over field, F, if V under the operation + satisfies the following conditions:

(i) For every $v,w \in V$, $v+w$ is also an element of V, i.e., V is closed under +;

(ii) For every u,v,w in V, $(u+v) + w = u + (v+w)$, i.e., V is associative under +;

(iii) There exists an element 0 in V such that for every $v \in V$, $v+0 = v$, i.e., there exists an additive identity in V;

(iv) For every $v \in V$, there exists an element $-v$ in V such that $v+(-v) = 0$, i.e., each element in V has its additive inverse in V; and

(v) For every $v, w \in V$, $v+w = w+v$, i.e., V is commutative under $+$.

In addition to (i) through (v), there is defined for every $\lambda \in F$ and $v \in V$, an element λv belonging to V that satisfies the following four conditions:

(vi) For every $\lambda \in F$, $v \in V$, $w \in V$, $\lambda(v+w) = \lambda v + \lambda w$;

(vii) For every $\lambda \in F$, $\delta \in F$, $v \in V$, $(\lambda+\delta)v = \lambda v + \delta v$;

(viii) For every $\lambda \in F$, $\delta \in F$, $v \in V$, $\lambda(\delta v) = (\lambda\delta)v$; and

(ix) For the multiplicative identity of F, denote it by 1, and for any $v \in V$, $1v = v$.

A few instructive remarks about the definition of a vector space may prove helpful. Conditions (i) through (v) are equivalent to saying that V under the operation $+$ is an abelian group. Conditions (vi) through (ix) relate the vector space to a particular field, and to emphasize the connection between the set of elements V, referred to as a vector space, and the particular field, V is often called a vector space over a field, rather than just a vector space. The operation joining the elements of V and those of F is often referred to as the operation of scalar multiplication. A convention that will be adhered to in this book is to use Greek letters such as λ, β, δ, to represent elements in the field. This should reduce the possible confusion of whether a given element is to be considered an element of V or of F.

Examples

1. If we consider V to be the set of all ordered pairs of real numbers, i.e., all points in the plane, and take the field F to be the real numbers, then we may show that V is a vector space of F. We define the addition to be, for a,b,c,d real numbers, $(a,b) + (c,d) = (a+c, b+d)$, i.e., we are defining the operation of addition of ordered pairs in terms of the sums of the individual components. Notice, therefore, that the plus sign on the left and right hand side of the equality has a different meaning. Scalar multiplication is defined in the following manner. For a,b real numbers and λ a real number, $\lambda(a,b) = (\lambda a, \lambda b)$, or in other words, the scalar multiple of an ordered pair is the multiple of each coordinate. The verification that V is a vector space is a simple one.

   (i) $(a,b) + (c,d) = (a+c, b+d)$, which is another point in the plane. Therefore, we have closure.

(ii) $[(a,b) + (c,d)] + (e,f) = (a,b) + [(c,d) + (e,f)]$, because of the underlying associativity of the real numbers.

(iii) The identity element is the ordered pair $(0,0)$.

(iv) The additive inverse of (a,b) is $(-a,-b)$, because $(a,b) + (-a,-b) = (0,0)$.

(v) The commutative property is a consequence of the commutativity of the real numbers.

(vi) $\lambda[(a,b) + (c,d)] = \lambda(a+c, b+d) = (\lambda(a+c), \lambda(b+d)) = (\lambda a+\lambda c, \lambda b+\lambda d) = (\lambda a,\lambda b) + (\lambda c,\lambda d) = \lambda(a,b) + \lambda(c,d)$.

(vii) $(\lambda+\delta)(a,b) = ((\lambda+\delta)a,(\lambda+\delta)b) = (\lambda a + \delta a, \lambda b + \delta b) = (\lambda a,\lambda b) + (\delta a,\delta b) = \lambda(a,b) + \delta(a,b)$.

(viii) $(\lambda\delta)(a,b) = (\lambda\delta a,\lambda\delta b) = \lambda(\delta a, \delta b) = (\lambda)(\delta)(a,b)$.

(ix) $1(a,b) = (1a, 1b) = (a,b)$.

Therefore, V is a vector space.

2. For those readers familiar with vectors, (a,b) would correspond to the vector with x component a and y component b, emanating from the origin. Therefore, the addition of (a,b) and (c,d) is actually the operation of vector addition. Scalar multiplication is the same as multiplying a vector by a scalar. This is indicated graphically in Figure 35. Anyone who has taken courses in physics must realize the importance of vectors in physics.

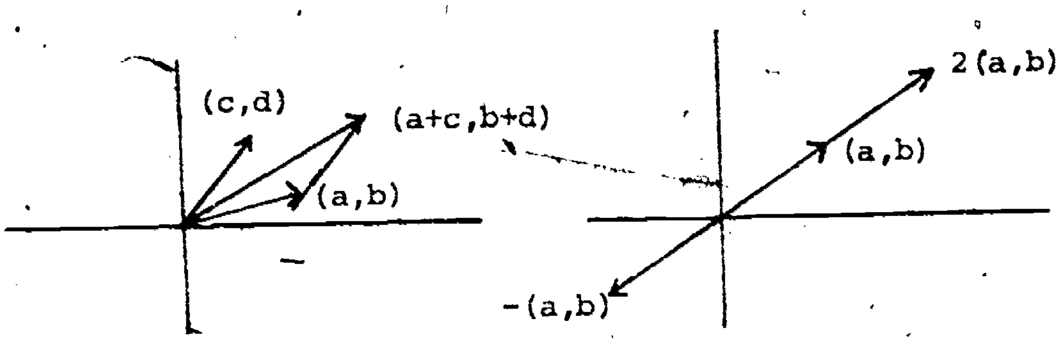

Figure 35

3. Another example of a vector space is the set of all ordered triples of real numbers, i.e., all points in 3-dimensional space, with the operations, $(a,b,c) + (d,e,f) = (a+d, b+e, c+f)$ and $\lambda(a,b,c) = (\lambda a, \lambda b, \lambda c)$. Three dimensional space is precisely the world we are a part of. The verification is identical to that in example 1.

4. If we consider the set of functions from the real numbers into the real numbers to be V and define addition by $(f+g)(x) = f(x) + g(x)$, for any real number x, then V is an abelian group under +. We have already shown this in an earlier example on groups. The operation of scalar multiplication is defined by $(\lambda f)(x) = \lambda(f(x))$, where λ is an element of the field of real numbers. That properties (vi) through (ix) of a vector space hold is simple enough to show.

5. An interesting way of defining a vector space is by considering two fields F_1 and F_2, where F_2 is a subfield of F_1. Then F_1 is a vector space over F_2. Clearly, F_1 is a group under addition if F_1 is a field. If scalar multiplication is taken to be multiplication in F_1, then the product of an element in F_2 and in F_1 is certainly in F_1 because F_2 is a subfield of F_1, and further multiplication in F_1 is closed. Property (vi) and (vii) correspond to the distributive laws in the field, (viii) to the associative property for multiplication, and (ix) to the existence of a multiplicative identity in a field.

After introducing concepts such as a group or a ring, we followed by defining a subgroup and subring. We have a corresponding term in the algebraic system called a vector space.

Definition 63. A subspace S of a vector space V over field F is a subset of V, that itself is a vector space under the operations of V.

In actuality it is only necessary to prove that S is closed under addition and that for $\lambda \in F$ and $v \in S$, $\lambda v \in S$. The other properties of a vector space are consequences of these. For example, (vi) through (viii) hold in S because they already hold in the larger set V. Similarly, (ix) holds because we are considering the same field F, and, thus, the same multiplicative identity. Further, if S is closed under addition, we need to only prove that the additive inverse also belongs to S, in order to prove that S is a subgroup of V under addition. But, if $v \in S$, then $-v = (-1)v$ is also an element of S by the scalar multiplication. Therefore, we have an alternative way of proving a set to be a subspace.

Theorem. S is a subspace of V a vector space if S is a subset of V and

(i) if for $v, w \in S$, $v+w \in S$; and
(ii) if $v \in S$ and $\lambda \in F$ imply $\lambda v \in S$.

Example

1. We proved that the set of all functions from the real numbers into the real numbers may be defined to be a vector space over the real numbers. If we take a subset, namely all the continuous functions from the real numbers into the real numbers, then we have a subspace. This follows because the sum of two continuous functions is a continuous function, which means additive closure. Scalar multiplication of a continuous function is again a continuous function.

We shift gears a bit now and rather than discussing the structure called a vector space, try to describe how this structure is built up.

In a vector space, a series of elements are often added, and by the closure property these sums yield new elements. Sums of this type have a particular name.

Definition 64. Let $v_1,\cdots,v_N$ be elements of a vector space V and suppose $\lambda_1,\cdots,\lambda_N$ belong to the field F, then an element $\lambda_1v_1+\lambda_2v_2+\cdots+\lambda_Nv_N$ is called a linear combination of $v_1,v_2,\cdots,v_N$.

If we form all the possible combinations of the elements $v_1,v_2,\cdots,v_N$, we form a new set.

Definition 65. If $v_1,v_2,\cdots,v_N$ are elements of a vector space V, then the linear span of $v_1,\cdots,v_N$ is the set of all possible linear combinations of $v_1,\cdots,v_N$. If $\{v_1,\cdots,v_N\}$ is such that its span exhausts all the vector space V, i.e., every element in V is expressible as a linear combination of $v_1,\cdots,v_N$, then $\{v_1,\cdots,v_N\}$ spans V.

If we can find a subset of V that spans V, then we are able to describe all of V by means of the information gained from a subset of V. This is certainly economical in terms of time and effort in studying the set V. But, we are not content at this even; we are greedy enough to ask if we can find an even smaller set that will give us as much information. Perhaps there is still some built-in redundancy of information. Keep in mind that the question we are asking is really the one we are posing in cognition. How does one utilize what he knows in learning something new? We will offer an analysis of cognition in a later section.

As a step in the direction of answering whether there is still redundancy in the information we learn from a spanning set, we introduce the important concepts of linear independence and dependence.

Definition 66. A set $v_1, v_2, \cdots, v_N$ in a sector space V is said to be linearly dependent if there exist elements $\lambda_1, \cdots, \lambda_N$ in the field F, some of which are not zero, such that $\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_N v_N = 0$.

Definition 67. A set $v_1, \cdots, v_N$ in a vector space V is said to be linearly independent if they are not linearly dependent, or equivalently if for $\lambda_1, \lambda_2, \cdots, \lambda_N$ in the field F, $\lambda_1 v_1 + \cdots + \lambda_N v_N = 0$ implies that $\lambda_1 = \lambda_2 = \cdots = \lambda_N = 0$.

We will state a number of theorems that show how independence and dependence of a set of vectors reveal information about the structure of a fector space, but first we include a few examples to clarify the abstract sounding definitions.

Examples

1. We have proved that the set of all ordered pairs may be made into a vector space. Suppose $v_1 = (1,0)$ and $v_2 = (0,1)$, we will show that v_1 and v_2 are linearly independent. Let λ_1 and λ_2 be real numbers, and suppose $\lambda_1 v_1 + \lambda_2 v_2 = 0$, i.e., $\lambda_1(1,0) + \lambda_2(0,1) = (\lambda_1, 0) + (0,\lambda_2) = (\lambda_1, \lambda_2) = (0,0)$, since $(0,0)$ is the zero element. Therefore, if $(\lambda_1, \lambda_2) = (0,0)$, we must have $\lambda_1 = \lambda_2 = 0$, which by definition means that v_1 and v_2 are linearly independent.

2. Suppose $v_1 = (2,5)$, $v_2 = (1,-2)$ and $v_3 = (4,4)$, and let λ_1, λ_2, and λ_3 be real numbers. If $\lambda_1 v_1 + \lambda_2 v_2 + \lambda_3 v_3 = 0$, or equivalently, $\lambda_1(2,5) + \lambda_2(1,-2) + \lambda_3(4,4) = (0,0)$, then $(2\lambda_1, 5\lambda_1) + (\lambda_2, -2\lambda_2) + (4\lambda_3, 4\lambda_3) = (0,0)$, and finally, $(2\lambda_1 + \lambda_2 + 4\lambda_3,\ 5\lambda_1 - 2\lambda_2 + 4\lambda_3) = (0,0)$. Notice, that if for example, $\lambda_1 = 4$, $\lambda_2 = 4$ and $\lambda_3 = -3$, we have that $(2\lambda_1 + \lambda_2 + 4\lambda_3,\ 5\lambda_1 - 2\lambda_2 + 4\lambda_3) = (8 + 4 - 12,\ 20 - 8 - 12) = (0,0)$, but this means that there exist $\lambda_1, \lambda_2, \lambda_3$ not all zero, such that $\lambda_1 v_1 + \lambda_2 v_2 + \lambda_3 v_3 = 0$. Therefore, v_1, v_2, and v_3 are linearly dependent.

We give the following theorems without proof, but we want some of these results to be at the reader's disposal.

Theorem. A set $v_1, v_2, \cdots, v_N$ in a vector space V is linearly dependent if any one of the following conditions is met:

(i) The set includes the zero vector;

(ii) The set contains a nonempty subset that is linearly dependent; or

(iii) There exists at least one element, say v_i, that is expressible as a linear combination of the remaining elements.

We do not include any proofs, but the reader is invited to convince himself that these statements are true. For instance, suppose one of the elements, say v_i is the zero element. Then it is possible to find a linear combination of $v_1, \cdots, v_N$ that equals zero, but has at least one λ not equal to zero. An obvious choice would be $\lambda_1 v_1 + \ldots + 1v_i + \ldots + \lambda_N v_N = 0$. Since $v_i = 0$, any nonzero λ_i may be selected, because whatever value is chosen for λ_i, $\lambda_i v_i = 0$. Thus, it is not necessarily the case that $\lambda_1 = \lambda_2 = \cdots = \lambda_N = 0$. Therefore, $v_1, \cdots, v_N$ are linearly dependent.

<u>Theorem</u>. If a set $v_1, \cdots, v_N$ of elements in a vector space V over F is linearly independent, then every linear combination of $v_1, \cdots, v_N$ has a unique representation of the form $\lambda_1 v_1 + \cdots + \lambda_N v_N$.

It may be proved that if the representation were assumed to be not unique, then the independence of $v_1, \ldots, v_N$ would be contradicted.

We have introduced two new concepts, the idea of a linear combination and spanning set, and then the idea of independence and dependence. A spanning set was capable of accounting for the entire vector space from some subset of the space. But the question remained as to whether an even smaller set could be found that still spanned all of the vector space V. The examination of linear independence offered a way to remove redundancy or duplication. If the set was dependent, then certain elements were expressible as linear combinations of the others, so in effect these elements offer no information that could not have been obtained by other means without them. It would be wonderful if a set could be found that spans all of V and at the same time is as small as possible in terms of the number of elements in it. Well, such a set exists, and is called a basis.

<u>Definition 68</u>. A vector space V is of <u>finite dimension</u> if it has a spanning set with a finite number of elements.

<u>Definition 69</u>. A subset B of a vector space V of finite dimension is called a <u>basis</u> for V if B spans all of V and B is a linearly independent set.

Examples

1. We earlier proved that $v_1 = (1,0)$ and $v_2 = (0,1)$ was a linearly independent set in the plane. The set consisting of v_1, v_2 also spans the plane, since for any point (x,y) in the plane $(x,y) = x(1,0) + y(0,1)$. Therefore, $(1,0)$ and $(0,1)$ form a basis for the plane.

2. As you might guess (1,0,0), (0,1,0), and (0,0,1) form a basis for three-dimensional space. The set consisting of (6,0,0), (0,-5,0), and (0,0,1/2) would be another basis for three-dimensional space. A particular vector space V may have more than one basis, but any two bases for V must have the same number of vectors.

Definition 70. The number of elements in a basis for a vector space V is called the dimension of V.

We have accomplished our original goal of finding from a spanning set for a vector space V a smaller set of minimum size that still spans V. This set is the basis, or perhaps better stated, a basis for V.

An interesting application of these terms is in the area of cognition. Suppose that there is a set of information from some subject matter that must be learned. At first the student has no idea of which are the relevant and irrelevant dimensions relating to the task. He has a certain body of knowledge that he draws from in various combinations to learn individual items. As he gains a greater understanding of the task he is considering, he begins to integrate the learning of individual items into a more cohesive and structured approach. By learning a specific rule, he may be able to master an entire class of items without mastering each item individually. The goal of learning a particular subject matter may then be described as the process of tending towards a cognitive basis capable of understanding any question in the given area, but at the same time free of any unnecessary overlap or redundancy.

One of the most important concepts that we have examined in previous chapters is that of a homomorphism. There is an analogous concept for vector spaces, but it is called a linear transformation.

Definition 71. Let V and W be vector spaces over a field F, then a mapping T from V into W is called a linear transformation if

(i) for $v_1, v_2 \in V$, $T(v_1+v_2) = T(v_1) + T(v_2)$; and

(ii) for $v \in V$ and $\lambda \in F$, $T(\lambda v) = \lambda T(v)$.

We will denote the set of all linear transformations from V into W by LT(V,W).

Examples

1. An obvious example is if V = W and T is defined by T(x) = 5x, then

(i) $T(v_1+v_2) = 5(v_1+v_2) = 5v_1 + 5v_2 = T(v_1) + T(v_2)$; and
(ii) $T(\lambda v) = 5(\lambda v) = \lambda(5v) = \lambda T(v)$.

2. A more involved example requires us to make a few assumptions. Let the field be the real numbers and let F[x] denote the set of all polynomials f, where $f = \lambda_0+\lambda_1x+\lambda_2x^2+\cdots+\lambda_kx^k$, and the λ's are real numbers. It may be shown that F[x] is a vector space. Define an operator D on f, such that $Df = \lambda_1+2\lambda_2x+\cdots+k\lambda_kx^{k-1}$. For those readers who have had an introductory calculus course, you might realize that D is the derivative. We will verify that D is a linear transformation. If $f = \lambda_0+\lambda_1x+\lambda_2x^2+\ldots+\lambda_kx^k$ and $g = \delta_0+\delta_1x+\delta_2x^2+\ldots+\delta_kx^k$, then $Df = \lambda_1+2\lambda_2x+\ldots+k\lambda_kx^{k-1}$ and $Dg = \delta_1+2\delta_2x+\ldots+k\delta_kx^{k-1}$, so $Df + Dg = (\lambda_1+\delta_1) + 2(\lambda_2+\delta_2)x + \cdots + k(\lambda_k+\delta_k)x^{k-1}$. On the other hand, $f+g = (\lambda_0+\delta_0) + (\lambda_1+\delta_1)x + \cdots + (\lambda_k+\delta_k)x^k$, which implies that $D(f+g) = (\lambda_1+\delta_1) + 2(\lambda_2+\delta_2)+\cdots k(\lambda_k+\delta_k)x^{k-1}$. Therefore, $d(f+g) = Df + Dg$, and similarly it may be demonstrated that $d(\delta f) = \delta(Df)$.

An interesting point about the set of all linear transformations from V into W, where V and W are vector spaces, is that LT(V,W) is itself a vector space.

<u>Theorem.</u> Let V and W be vector spaces over a field F, then LT(V,W) is a vector space over F, if the operations are defined by

(i) for $S,T \in LT(V,W)$, $(S+T)(v) = S(v) + T(v)$, where $v \in V$; and
(ii) for $S \in LT(V,W)$, $(\lambda S)(v) = \lambda(S(v))$, where $\lambda \in F$ and $v \in V$.

We will not give a detailed proof, but will sketch some of the important arguments. In order to prove that LT(V,W) is a vector space, we must show that if S and T belong to LT(V,W), then S+T also is an element. In other words, it is necessary to prove that $(S+T)(v_1+v_2) = (S+T)(v_1) + (S+T)(v_2)$ and that $(S+T)(\lambda v) = \lambda(S+T)(v)$. This would establish closure. The remaining properties with respect to the operation of addition are rather elementary. The only more complicated step remaining is to prove that if S belongs to LT(V,W), then λS is also in LT(V,W). In other words, $\lambda S(v_1+v_2) = \lambda S(v_1) + \lambda S(v_2)$ and $\lambda S(\delta v) = \delta(\lambda S)(v)$.

One of the most impressive qualities of algebraic systems is how they all are nicely interconnected. Each structure builds upon the others. We have defined vector spaces, and now have just demonstrated that the set of linear transformations from one vector space V into another W is itself a vector space. If the vector spaces V and W are the same, i.e., V = W, then a new operation between linear transformations may be introduced, namely the product of two linear transformations ST. The product transformation ST is another linear transformation. Therefore, LT(V,V) has both an addition and a multiplication operation. If you are thinking "Could LT(V,V) be made into a ring?", the answer is yes.

Theorem. Let V be a vector space over a field F, and LT(V,V) be the set of all linear transformations of V into itself, then LT(V,V) under the operations of addition and multiplication is a ring.

While we still have LT(V,V) under consideration it is a good idea to introduce a few more terms.

Definition 72. A linear transformation T in LT(V,V) is called regular or invertible if there exists another transformation, denote it be T^{-1}, such that $TT^{-1} = T^{-1}T = I$, where I is the identity transformation. If no such transformation exists, then T is called singular.

The linear transformation T maps V into itself. It may be important in some cases to know just how much of V is mapped into by T.

Definition 73. If $T \in LT(V,V)$, then the range of T is denoted by TV, and is the set of all elements in V that are mapped into by T.

One way of comparing V and the range of T is by examining the basis for the range to see if it has fewer elements.

Definition 74. For a finite dimensional vector space V, the rank of V is the number of elements in the basis of the range of V. That is, the rank is the dimension of the range.

The next chapter will begin where this one leaves off. A connection will be established between linear transformations and matrices. Most of the terminology of chapter 7 is needed in the development of the chapter on matrices. Once the connection is clear, an examination of matrix operations is included in order to better understand the techniques applied in the various psychological illustrations.

CHAPTER 8

MATRICES AND THEIR APPLICATIONS

The final chapter concerns itself with the study of matrices. The transition between linear transformations and matrices is a smooth one, because a matrix will be defined in terms of a linear transformation's action on a basis of a vector space V. The action of the transformation on any particular basis element will be expressible as a linear combination of the basis elements of a vector space W. Having defined a matrix it will be necessary to examine its structure, and in due course we will be able to form a group of matrices, a vector space of matrices, a ring of matrices, and with special consideration, a field of matrices. We will then shift our emphasis from theory to application, and study how matrices make many types of statistical analyses tolerable.

Let T be a linear transformation from a vector space V into a vector space W, i.e., $T \in LT(V,W)$. Suppose that V is of dimension n and that W is of dimension m, and that $v_1,v_2,\cdots,v_n$ and $w_1,w_2,\cdots,w_m$, are bases of V and W respectively. Further, assume that both V and W are defined over the same field of F. If we know how T acts on a basis of V, then in effect we know how T acts on any element in V, since every element in V is a linear combination of $v_1,v_2,\cdots,v_n$.

Let the action of T on $v_1,v_2,\cdots,v_n$ be described as follows:

$$\begin{aligned}
T(v_1) &= \alpha_{11}w_1 + \alpha_{12}w_2 + \dots + \alpha_{1m}w_m \\
T(v_2) &= \alpha_{21}w_1 + \alpha_{22}w_2 + \dots + \alpha_{2m}w_m \\
&\cdots\cdots\cdots\cdots\cdots\cdots \\
&\cdots\cdots\cdots\cdots\cdots\cdots \\
&\cdots\cdots\cdots\cdots\cdots\cdots \\
T(v_n) &= \alpha_{n1}w_1 + \alpha_{n2}w_2 + \dots + \alpha_{nm}w_m
\end{aligned}$$

That is, T maps an element of V into an element of W, and any element in W is expressible as a linear combination of $w_1,\dots,w_m$, the basis of W. The α_{ij} are elements of the field F. The double subscript is used to locate the particular entry. The first subscript indicates what row the element is in. For example, if it was a 4, this means we are in the 4th row from the top. The second subscript indicates the column under consideration. This means that the 7th column, for instance, would be the 7th column over from the left. So, α_{47} would be the element in the 4th row and the 7th column.

There are n rows and m columns in our system of equations. The array of α_{ij} elements completely describes the action of the linear transformation T. The rectangular array of the α_{ij} is called a matrix.

Definition 75. Let V and W be vector spaces over F of dimensions n and m, respectively, and assume that $v_1, v_2, \ldots, v_n$ is a basis for V and $w_1, w_2, \ldots, w_m$ is a basis for W. Let T be a linear transformation of V into W, the matrix of T with respect to the given bases is

$$\begin{pmatrix} \alpha_{11} & \alpha_{12} & \cdots & \cdots & \alpha_{1m} \\ \alpha_{21} & \alpha_{22} & \cdots & \cdots & \cdots \\ \alpha_{31} & \alpha_{32} & \alpha_{33} & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \alpha_{n1} & \alpha_{n2} & \alpha_{n3} & \cdots & \alpha_{nm} \end{pmatrix}$$

where $T(v_i) = \alpha_{i1}w_1 + \alpha_{i2}w_2 + \ldots + \alpha_{im}w_m$, for each i, $1 < i < n$. The matrix is an $n \times m$ matrix.

Before we go any further, a less abstract illustration of the definition of a matrix may be helpful. If we have,

$$\begin{aligned} T(v_1) &= 7w_1 - 3w_2 + w_3 - w_4 \\ T(v_2) &= w_1 \quad\quad + 5w_3 - 4w_4 \\ T(v_3) &= 2w_1 + w_2 - w_3 \end{aligned}$$

then the matrix of T would be

$$\begin{pmatrix} 7 & -3 & 1 & -1 \\ 1 & 0 & 5 & -4 \\ 2 & 1 & -1 & 0 \end{pmatrix}$$

We will be most interested in the structure of square matrices, i.e., matrices having the same number of rows and columns. In fact, the study of transformations T from a vector space V into itself will prove to be of the most theoretical value.

Suppose that T is a linear transformation of a vector space V into itself. Let V be an n dimensional vector space with basis $v_1,\dots,v_n$, and define the action of T on V by

$$T(v_1) = \alpha_{11}v_1 + \alpha_{12}v_2 + \dots + \alpha_{1n}v_n$$
$$T(v_2) = \alpha_{21}v_1 + \alpha_{22}v_2 + \dots + \alpha_{2n}v_n$$
$$\dots\dots\dots\dots\dots\dots\dots\dots\dots$$
$$T(v_n) = \alpha_{n1}v_1 + \alpha_{n2}v_2 + \dots + \alpha_{nn}v_n$$

Then the matrix associated with this sytem would be the following:

$$\begin{pmatrix} \alpha_{11} & \alpha_{12} & \dots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \dots & \alpha_{2n} \\ \dots & \dots & \dots & \dots \\ \alpha_{n1} & \alpha_{n2} & \dots & \alpha_{nn} \end{pmatrix}$$

which is an n x n matrix.

To ask what the matrix for a particular linear transformation looks like is an ambiguous question unless the basis of the vector space is specified. Suppose T is a mapping from a vector space V into V, and let U be the set of all ordered pairs of real numbers, i.e., the Cartesian plane. Also assume that the field is the field of real numbers. A basis for V would be $v_1 = (1,0)$ and $v_2 = (0,1)$. If the linear transformation T is defined by $T(v_1) = v_2$ and $T(v_2) = v_1$, then the matrix of T with respect to this basis is

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Another basis for V would be $w_1 = (1,1)$ and $w_2 = (1,-1)$. It would be good practice to verify that w_1 and w_2 are independent and that they span all of V. We may describe the action of T on this basis as well.

$T(w_1) = T(v_1+v_2)$, because $w_1 = v_1+v_2$, since $(1,1) = (1,0) + (0,1)$. But since T is a linear transformation, $T(w_1) = T(v_1+v_2) = T(v_1) + T(v_2) = v_2 + v_1 = v_1 + v_2 = w_1$. Similarly, $T(w_2) = T(v_1 - v_2)$, because $w_2 = v_1 - v_2$, and further, $T(w_1) = T(v_1 - v_2) = T(v_1) - T(v_2) =$

$v_2 - v_1 = -(v_1 - v_2) = -w_2$. Therefore, the matrix of T with respect to this basis is

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Clearly, the two matrices are different, even though both are for the same linear transformation. This is why it is so important to know what the basis for the vector space is.

Having established what a matrix is, and how it ties in with the theory of vector spaces, it would be fruitful to examine various operations on matrices. For example, what is the sum of two matrices? In order to be able to add two matrices, they must be of the same size, that is, a 3 x 3 matrix cannot be added to a 4 x 4 or a 3 x 5 matrix, but only with another 3 x 3 matrix. Suppose we have two n x m matrices

$$\begin{pmatrix} \alpha_{11} & \cdots & \alpha_{1m} \\ \alpha_{21} & \cdots & \alpha_{2m} \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \alpha_{n1} & \cdots & \alpha_{nm} \end{pmatrix} \text{ and } \begin{pmatrix} \gamma_{11} & \cdots & \gamma_{1m} \\ \gamma_{21} & \cdots & \gamma_{2m} \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \gamma_{n1} & \cdots & \gamma_{nm} \end{pmatrix}.$$

For simplicity we will denote them by $[\alpha_{ij}]$ and $[\gamma_{ij}]$ respectively.

<u>Definition 76</u>. If $[\alpha_{ij}]$ and $[\gamma_{ij}]$ are two n x m matrices, then the <u>sum</u> of $[\alpha_{ij}]$ and $[\gamma_{ij}]$ is the matrix obtained by adding their corresponding elements. Therefore, $[\alpha_{ij}] + [\gamma_{ij}] = [\alpha_{ij}+\gamma_{ij}]$, or

$$\begin{pmatrix} \alpha_{11}+\gamma_{11} & \alpha_{12}+\gamma_{12} & \cdots & \alpha_{1m}+\gamma_{1m} \\ \alpha_{21}+\gamma_{21} & \alpha_{22}+\gamma_{22} & \cdots & \alpha_{2m}+\gamma_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \alpha_{n1}+\gamma_{n1} & \alpha_{n2}+\gamma_{n2} & \cdots & \alpha_{nm}+\gamma_{nm} \end{pmatrix}$$

Examples

1. $$\begin{pmatrix} 2 & 5 & 0 & 1 \\ 3 & -2 & 7 & 2 \\ 4 & 1 & 5 & 3 \\ 2 & 0 & -4 & -1 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 4 & 6 \\ 3 & 2 & 1 & 5 \\ 0 & 4 & -2 & -3 \\ 3 & -1 & -5 & 6 \end{pmatrix} = \begin{pmatrix} 2+0 & 5+0 & 0+4 & 1+6 \\ 3+3 & -2+2 & 7+1 & 2+5 \\ 4+0 & 1+4 & 5-2 & 3-3 \\ 2+3 & 0-1 & -4-5 & -1+6 \end{pmatrix}$$

$$= \begin{pmatrix} 2 & 5 & 4 & 7 \\ 6 & 0 & 8 & 7 \\ 4 & 5 & 3 & 0 \\ 5 & -1 & -9 & 5 \end{pmatrix}.$$

2. $$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} + \begin{pmatrix} 2 & 2 & 2 \\ 4 & 4 & 4 \\ -3 & -3 & -3 \end{pmatrix} = \begin{pmatrix} 3 & 4 & 5 \\ 8 & 9 & 10 \\ 4 & 5 & 6 \end{pmatrix}.$$

If we consider the set of all n x m matrices together with the operation of matrix addition, this set $\mathcal{M}$ forms a group. This is what makes the study of mathematical systems so nice; they have a way of becoming interwoven.

Theorem. If $\mathcal{M}$ is the set of all n x m matrices whose matrices have real number entries, and if the operation is matrix addition, then $\mathcal{M}$ is an abelian group.

Proof:

(i) Closure follows directly from the definition, since the sum of two n x m matrices with real elements in another n x m matrix with real elements.

(ii) The associativity is a consequence of the definition of addition and the associativity of the real numbers. Each entry will have an equality between expressions of the type $(\alpha_{ij}+\gamma_{ij}) + \theta_{ij} = \alpha_{ij} + (\gamma_{ij}+\theta_{ij})$.

(iii) The identity element is the matrix, all of whose elements are 0, i.e.,

106

$$\begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}$$

$$\begin{pmatrix} \alpha_{11} & \cdots & \alpha_{1m} \\ \alpha_{21} & \cdots & \alpha_{2m} \\ \vdots & & \vdots \\ \alpha_{n1} & \cdots & \alpha_{nm} \end{pmatrix} + \begin{pmatrix} 0 & \cdots & 0 \\ 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 0 \end{pmatrix} = \begin{pmatrix} \alpha_{11} & \cdots & \alpha_{1m} \\ \alpha_{21} & \cdots & \alpha_{2m} \\ \vdots & & \vdots \\ \alpha_{n1} & \cdots & \alpha_{nm} \end{pmatrix}$$

(iv) The inverse of the matrix $[\alpha_{ij}]$ is the matrix $[-\alpha_{ij}]$. In other words, that matrix whose elements are the negative of the corresponding elements in $[\alpha_{ij}]$.

$$\begin{pmatrix} \alpha_{11} & \cdots & \alpha_{1m} \\ \alpha_{21} & \cdots & \alpha_{2m} \\ \vdots & & \vdots \\ \alpha_{n1} & \cdots & \alpha_{nm} \end{pmatrix} + \begin{pmatrix} -\alpha_{11} & \cdots & -\alpha_{1m} \\ -\alpha_{21} & \cdots & -\alpha_{2m} \\ \vdots & & \vdots \\ -\alpha_{n1} & \cdots & -\alpha_{nm} \end{pmatrix} =$$

$$\begin{pmatrix} \alpha_{11}-\alpha_{11} & \alpha_{12}-\alpha_{12} & \cdots & \alpha_{1m}-\alpha_{1m} \\ \alpha_{21}-\alpha_{21} & \alpha_{22}-\alpha_{22} & \cdots & \alpha_{2m}-\alpha_{2m} \\ \vdots & & & \vdots \\ \alpha_{n1}-\alpha_{n1} & \alpha_{n2}-\alpha_{n2} & \cdots & \alpha_{nm}-\alpha_{nm} \end{pmatrix} = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}$$

Therefore, $\mathcal{M}$ is a group.

(v) The commutative property holds because it holds for the real numbers, and, therefore, $\alpha_{ij} + \gamma_{ij} = \gamma_{ij} + \alpha_{ij}$ for each element in the matrices. Hence, $\mathcal{M}$ is an abelian group.

Another operation that we have examined in the last chapter is that of scalar multiplication.

Definition 77. If $[\alpha_{ij}]$ is an n x m matrix whose entries are real numbers and if λ is a real number, then the scalar product of λ and the matrix, denoted by $\lambda[\alpha_{ij}]$ is that matrix whose entries are obtained by multiplying each entry of $[\alpha_{ij}]$ by λ. In other words,

$$\lambda\begin{pmatrix} \alpha_{11} & \cdots & \alpha_{1m} \\ \alpha_{21} & \cdots & \alpha_{2m} \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \alpha_{n1} & \cdots & \alpha_{nm} \end{pmatrix} = \begin{pmatrix} \lambda\alpha_{11} & \cdots & \lambda\alpha_{1m} \\ \lambda\alpha_{21} & \cdots & \lambda\alpha_{2m} \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \lambda\alpha_{n1} & \cdots & \lambda\alpha_{nm} \end{pmatrix}$$

Examples

1. $$5\begin{pmatrix} 2 & 4 & 1 & 6 \\ 3 & -2 & 4 & 9 \\ 0 & 0 & -2 & -4 \end{pmatrix} = \begin{pmatrix} 10 & 20 & 5 & 30 \\ 15 & -10 & 20 & 45 \\ 0 & 0 & -10 & -20 \end{pmatrix}.$$

2. $$-1/3\begin{pmatrix} 6 & 2 & -1 \\ -5 & 4 & 12 \\ 3 & 0 & 8 \end{pmatrix} = \begin{pmatrix} -2 & -2/3 & 1/3 \\ 5/3 & -4/3 & -4 \\ -1 & 0 & -8/3 \end{pmatrix}.$$

By forming a system consisting of the set of all n x m matrices with real entries, $\mathcal{M}$, and the two operations of matrix addition and scalar multiplication, we have a vector space.

Theorem. If $\mathcal{M}$ consists of all the n x m matrices with real number entries, and there are two operations, matrix addition and scalar multiplication, defined on $\mathcal{M}$, then $\mathcal{M}$ is a vector space.

Proof: We have already shown that $\mathcal{M}$ under matrix addition is a commutative group. Also $\lambda[\alpha_{ij}]$ is a well defined operation that yields another element in $\mathcal{M}$. Therefore, only conditions (vi) - (ix) of a vector space must be substantiated.

To verify that property (vi) is valid, we must show that $\lambda([\alpha_{ij}] + [\gamma_{ij}]) = \lambda[\alpha_{ij}] + \lambda[\gamma_{ij}]$.

$$\lambda([\alpha_{ij}]+[\gamma_{ij}]) = \lambda \begin{pmatrix} \alpha_{11}+\gamma_{11} & \cdots & \alpha_{1m}+\gamma_{1m} \\ \alpha_{21}+\gamma_{21} & \cdots & \alpha_{2m}+\gamma_{2m} \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \alpha_{n1}+\gamma_{n1} & \cdots & \alpha_{nm}+\gamma_{nm} \end{pmatrix}$$

$$= \begin{pmatrix} \lambda(\alpha_{11}+\gamma_{11}) & \cdots & \lambda(\alpha_{1m}+\gamma_{1m}) \\ \lambda(\alpha_{21}+\gamma_{21}) & \cdots & \lambda(\alpha_{2m}+\gamma_{2m}) \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \lambda(\alpha_{n1}+\gamma_{n1}) & \cdots & \lambda(\alpha_{nm}+\gamma_{nm}) \end{pmatrix} = \begin{pmatrix} \lambda\alpha_{11}+\lambda\gamma_{11} & \cdots & \lambda\alpha_{1m}+\lambda\gamma_{1m} \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \lambda\alpha_{n1}+\lambda\gamma_{n1} & \cdots & \lambda\alpha_{nm}+\lambda\gamma_{nm} \end{pmatrix}$$

$$= \begin{pmatrix} \lambda\alpha_{11} & \cdots & \lambda\alpha_{1m} \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \lambda\alpha_{n1} & \cdots & \lambda\alpha_{nm} \end{pmatrix} + \begin{pmatrix} \lambda\gamma_{11} & \cdots & \lambda\gamma_{1m} \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \lambda\gamma_{n1} & \cdots & \lambda\gamma_{nm} \end{pmatrix}$$

$$= \lambda \begin{pmatrix} \alpha_{11} & \cdots & \alpha_{1m} \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \alpha_{n1} & \cdots & \alpha_{nm} \end{pmatrix} + \lambda \begin{pmatrix} \gamma_{11} & \cdots & \gamma_{1m} \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \gamma_{n1} & \cdots & \gamma_{nm} \end{pmatrix} = \lambda[\alpha_{ij}] + \lambda[\gamma_{ij}].$$

Properties (vii) and (viii) may be verified in a manner analogous to that above. Property (ix), $1[\alpha_{ij}] = [\alpha_{ij}]$, holds because

$$1 \begin{pmatrix} \alpha_{11} & \cdots & \alpha_{1m} \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \alpha_{n1} & \cdots & \alpha_{nm} \end{pmatrix} = \begin{pmatrix} 1\alpha_{11} & \cdots & 1\alpha_{1m} \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ 1\alpha_{n1} & \cdots & 1\alpha_{nm} \end{pmatrix} = \begin{pmatrix} \alpha_{11} & \cdots & \alpha_{1m} \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \alpha_{n1} & \cdots & \alpha_{nm} \end{pmatrix}$$

Therefore, $\mathcal{M}$ is a vector space.

Before going any further it would be advisable to formally define what may already be intuitively clear.

<u>Definition 78</u>. Two matrices $[\alpha_{ij}]$ and $[\gamma_{ij}]$, both n x m, are <u>equal</u> if and only if all their corresponding entries are equal.

Therefore, even though the following two matrices are very similar they are not equal.

$$\begin{pmatrix} 3 & 2 & 1 \\ 5 & -1 & 4 \\ 8 & 2 & -5 \end{pmatrix} \text{ and } \begin{pmatrix} 3 & 2 & 1 \\ 5 & 2 & 4 \\ 8 & 2 & -5 \end{pmatrix} \text{ differ only in the } \alpha_{22} \text{ position.}$$

Another important term in matrix theory is that of the transpose of a matrix.

<u>Definition 79</u>. Let $[\alpha_{ij}]$ be an n x m matrix, then the <u>transpose</u> of $[\alpha_{ij}]$, denoted by $[\alpha_{ij}]'$, is the m x n matrix obtained by interchanging the rows and columns of $[\alpha_{ij}]$. In other words, the rows of $[\alpha_{ij}]$ are the columns of $[\alpha_{ij}]'$ and the columns of $[\alpha_{ij}]$ are the rows of $[\alpha_{ij}]'$.

The operation of matrix multiplication is more complicated than matrix addition or scalar multiplication. It is interesting that the two matrices do not have to be of the same size. It is only necessary that the number of columns of the first matrix be the same as the number of rows of the second matrix. In other words, we may compute the matrix product of an n x m and a m x p matrix, but not the product of a m x p and n x m matrix.

<u>Definition 80</u>. Suppose $[\alpha_{ij}]$ is an n x m matrix and $[\gamma_{ij}]$ is an m x p matrix, then the <u>matrix product</u> $[\theta_{ij}]$ of $[\alpha_{ij}]$ and $[\gamma_{ij}]$ is an n x p matrix, whose elements are determined by the following rule: The entry in the ij position is obtained by multiplying the first entry in the ith row of $[\alpha_{ij}]$ by the first entry in the jth column of $[\gamma_{ij}]$ and then adding to it the product of the second entry in the ith row of $[\alpha_{ij}]$ and the second entry of the jth column of $[\gamma_{ij}]$, and so on, until the product of the mth element in the ith row of $[\alpha_{ij}]$ and the mth element of the jth column. A formula for this would be

$$\theta_{ij} = \alpha_{i1}\gamma_{1j} + \alpha_{i2}\gamma_{2j} + \ldots + \alpha_{im}\gamma_{mj} .$$

The double subscripts may cause some readers difficulty, so we include several concrete illustrations.

Examples

1. $$\begin{bmatrix} 5 & 3 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} -4 & 1 \\ 5 & 8 \end{bmatrix} = \begin{bmatrix} 5(-4)+3(5) & 5(1)+3(8) \\ (-1)(-4)+2(5) & (-1)(1)+2(8) \end{bmatrix}$$

$$= \begin{bmatrix} -5 & 29 \\ 14 & 14 \end{bmatrix}.$$ The entry in the 11 position is obtained by taking the first row of the first matrix times the first column of the second matrix. The entry in the 12 position is obtained by taking the first row of the first matrix times the second column of the second matrix, and so on. The resultant matrix is a 2 x 2 matrix since it is the product of a 2 x 2 matrix and a 2 x 2 matrix.

2. $$\begin{pmatrix} 1 & 4 & -2 \\ 5 & 0 & 3 \end{pmatrix} \begin{pmatrix} 3 & 0 \\ 4 & -3 \\ 2 & -2 \end{pmatrix} = \begin{pmatrix} 1(3)+4(4)-2(2) & 1(0)+4(-3)-2(-2) \\ 5(3)+0(4)+3(2) & 5(0)+0(-3)+3(-2) \end{pmatrix}$$

$$= \begin{pmatrix} 15 & -8 \\ 21 & -6 \end{pmatrix}.$$

The product of a 2 x 3 and a 3 x 2 matrix is a 2 x 2 matrix.

3. $$\begin{pmatrix} 3 & 0 \\ 4 & -3 \\ 2 & -2 \end{pmatrix} \begin{pmatrix} 1 & 4 & -2 \\ 5 & 0 & 3 \end{pmatrix} = \begin{pmatrix} 3(1)+0(5) & 3(4)+0(0) & 3(-2)+0(3) \\ 4(1)-3(5) & 4(4)-3(0) & 4(-2)-3(3) \\ 2(1)-2(5) & 2(4)-2(0) & 2(-2)-2(3) \end{pmatrix}$$

$$= \begin{bmatrix} 3 & 12 & -6 \\ -11 & 16 & -17 \\ -8 & 8 & -10 \end{bmatrix}.$$

The product of a 3 x 2 and a 2 x 3 matrix is a 3 x 3 matrix. It is important to notice that the matrices in examples 2 and 3 are the same, but the order of multiplication is reversed. The size of the matrices is not even the same, one is 2 x 2 and the other is 3 x 3. In general, the product of two matrices is not commutative, that is the order of multiplication makes a difference.

4. An important matrix is the identity matrix, which has 1's down the diagonal from upper left to lower right in a n x n matrix, and every other entry is a zero. A 3 x 3 identity matrix would be

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

It is called an identity matrix because, for example,

$$\begin{pmatrix} 2 & -3 & 1 \\ -5 & 0 & 4 \\ 1 & 1 & 3 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 2(1)-3(0)+1(0) & 2(0)-3(1)+1(0) & 2(0)-3(0)+1(1) \\ -5(1)+0(0)+4(0) & -5(0)+0(1)+4(0) & -5(0)+0(0)+4(1) \\ 1(1)+1(0)+3(0) & 1(0)+1(1)+3(0) & 1(0)+1(0)+3(1) \end{pmatrix}$$

$$= \begin{pmatrix} 2 & -3 & 1 \\ -5 & 0 & 4 \\ 1 & 1 & 3 \end{pmatrix}.$$

If we restrict our consideration to the set of all n x n matrices with real entries and define the operations of matrix addition and matrix multiplication on it, then the set is a ring.

Theorem. Let M be the set of all n x n matrices with real entries, and suppose that the operations of matrix addition and multiplication are defined on M, then M is a ring with a multiplicative identity.

Proof: We have already proved that $\mathcal{M}$ is an abelian group under matrix addition. The closure of matrices under matrix multiplication follows because the product of an n x n matrix with another n x n matrix is an n x n matrix. The associativity of matrix multiplication requires a great deal of paperwork, but does follow directly from the definition of matrix multiplication and the associativity of the real numbers. The distributive law requires the proof that $[\alpha_{ij}]([\gamma_{ij}]+[\theta_{ij}]) = [\alpha_{ij}][\gamma_{ij}] + [\alpha_{ij}][\theta_{ij}]$. This, too, is a rather lengthy calculation. On the left hand side, $[\gamma_{ij}]$ and $[\theta_{ij}]$ are added, and then we compute the product of $[\alpha_{ij}]$ and the matrix we obtained by addition. The right hand side of the equality requires the products of $[\alpha_{ij}]$ and $[\gamma_{ij}]$, and $[\alpha_{ij}]$ and $[\theta_{ij}]$, and then the two resultant matrices are added. The results of the left and right hand sides will be the same. The identity element is the matrix

$$\begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & & 1 \end{pmatrix}$$

The ring is not commutative.

A related remark concerns the existence of zero divisors. For example, $\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$, but neither $\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ nor $\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$ is the zero element.

The only question that remains is that of the multiplicative inverse of a matrix. To begin with, only square matrices possess an inverse, and not even all square matrices have an inverse. The usual approach to finding the inverse of a matrix involves the study of determinants. A rather formal approach to determinants is very messy because of the great amount of notation required. For this reason the topic of determinants will not be examined in this book. However, there exists an alternate approach to finding the inverse of a matrix. This procedure involves what are called <u>elementary row operations</u>. We state these operations without a thorough description. A matrix may have a particular row multiplied by a nonzero constant, two rows may be interchanged, and the multiple of one row may be added to another. If the n x n matrix $[\alpha_{ij}]$ is altered by performing a series of these elementary row operations and at the same time we are performing each one of these operations on the n x n identity matrix. Once our original n x n matrix has been altered until it is now the n x n identity matrix, whatever the matrix that was originally the identity matrix looks now is the inverse of $[\alpha_{ij}]$. If it is impossible to reduce $[\alpha_{ij}]$ to the identity matrix, then $[\alpha_{ij}]$ has no inverse.

This entire discussion may seem incredible, but keep in mind just how complicated it should be to find the inverse of not one number, but an entire array of numbers.

Example

1. Let $[\alpha_{ij}] = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}$. We start with $\begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}$ and $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. Now we add the second row to the first row and do the same for the identity matrix, and rewrite the first row as this sum $\begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}$ and $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. Now again, add the first and second rows, but this time rewrite the second row as the sum of $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ and $\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$. Then the inverse of $\begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}$ should be $\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$. Do you believe it? Let us check: $\begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 2(1)-1(1) & 2(1)-1(2) \\ (-1)(1)+1(1) & (-1)(1)+1(2) \end{pmatrix}$

$= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 1(2)+(1)(-1) & 1(-1)+1(1) \\ 1(2)+2(-1) & 1(-1)+2(1) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

It works!

An amazing result is that if we restrict ourselves to the consideration of all 2 x 2 matrices with real entries that are of the form $\begin{pmatrix} \alpha & \gamma \\ -\gamma & \alpha \end{pmatrix}$, and we take the operations of matrix addition and matrix multiplication, then this set under the given operations forms a field. The reader is urged to go through the verification that the set is closed under both operations, that it has an additive and multiplicative identity, additive and multiplicative inverse, and all the other required properties.

Before we make the transition from theory to the practical and applied use of matrices, we show how matrices are helpful in solving systems of equations. We will give an illustration for a 2 x 2 case, i.e., when we have two equations in two unknowns, but the method is technically the same for a twenty equation in twenty unknown systems. Consider,

$$2x - y = 3$$

$$-x + y = -1.$$

We could easily show that $x = 2$ and $y = 1$ by using the normal procedures of solving simultaneous equations. An alternative procedure uses matrices. We can rewrite the system of equations as

$$\begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 3 \\ -1 \end{pmatrix},$$

because this is ~~the~~ same as $\begin{pmatrix} 2x - y \\ -x + y \end{pmatrix} = \begin{pmatrix} 3 \\ -1 \end{pmatrix}$, and by the definition of equality of matrices, $2x - y = 3$ and $-x + y = -1$. In considering, $\begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 3 \\ -1 \end{pmatrix}$, if we could find the inverse of $\begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}$, we would have $\begin{pmatrix} x \\ y \end{pmatrix}$ all alone on the left hand side of ~~the~~ equality and we could read off the answer to the problem. We have already computed the inverse of $\begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}$. It is $\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$. Therefore, multiply both sides of the equation by $\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$:

$$\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} 3 \\ -1 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} 3 \\ -1 \end{pmatrix}$$

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} 3 \\ -1 \end{pmatrix}$$

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1(3)+1(-1) \\ 1(3)+2(-1) \end{pmatrix}$$

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

Therefore, we have that $x = 2$ and $y = 1$.

The applications of matrices are fairly well known. The value of matrices in any statistical analysis of experimental data will be illustrated with a series of examples. Other applications will also be cited.

Examples

1. The theory of <u>Markov chains</u> concerns itself with the study of an experimental situation where the outcome on any given trial depends only on the outcome of the immediately preceding trial. Therefore, an outcome E_j does not have a fixed probability, but rather a conditional probability, p_{ij}, which represents the following. Given that outcome E_i has occurred, the probability that outcome E_j will occur on the next trial is p_{ij}. For example, if we have outcomes E_1, E_5, and E_9 occurring on succession, then the probability of this event is $p_1 p_{15} p_{59}$, where p_1 is the probability that E_1 occurs on the first trial. The outcomes, E_i, are generally referred to as the <u>states</u> of the system, and the p_{ij} are called the <u>transition probabilities</u>. An array or matrix can be formed that includes all the transition probabilities in an experiment that has $E_1, \ldots, E_N$ as possible states.

$$\begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1N} \\ p_{21} & p_{22} & \cdots & p_{2N} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ p_{N1} & p_{N2} & \cdots & p_{NN} \end{pmatrix}$$ is called the matrix of <u>transition probabilities</u>. From this matrix we may determine the probability of going from any state to any other state on the next trial. A necessary condition concerning the rows of the matrix is that the sum of the transition probabilities across any row is equal to one. Markov chains have many applications in probability, physics, and genetics. Recently they have also been used in forming models for classical conditioning, paired associate learning, and recall learning.

116

2. Theios and Brelsford (1966) have written an article concerning the use of a Markov model to describe eye blink conditioning in rabbits. They developed a theory to describe the changes taking place in the trial by trial probability of eliciting a response by the rabbit. The experiment used a tone as the conditioned stimulus (CS), an air puff to the eye served as the unconditioned stimulus (UCS), and the desired response was an eye blink to the conditioned stimulus. Theios and Brelsford used the following Markov model to reflect the changes in the probabilities during the experiment. The matrices that we will consider are

$$
\begin{array}{cc}
 & \begin{array}{ccc} C & A & N \end{array} \\
\begin{array}{c} C \\ A \\ N \end{array} & \begin{pmatrix} 1 & 0 & 0 \\ c & 1-c & 0 \\ 0 & a & 1-a \end{pmatrix}
\end{array}
\qquad
\begin{array}{c} P_r \\ \begin{pmatrix} P_C \\ P_A \\ P_N \end{pmatrix} \end{array}
\qquad
\begin{array}{c} P_r \text{ (start)} \\ \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \end{array}
$$

The remaining discussion is intended to clarify Theios and Brelsford's reasoning and choice of notation. The rows of the 3 x 3 transition matrix are the states of responsiveness on any given trial, while the columns are the possible states on the next trial. The entries in the matrix are the probabilities of moving from the given state to another during N, the intertrial period. The first 1 x 3 matrix has entries representing the probability of a response during the observation interval for each of the three states of responsiveness. The second 1 x 3 matrix gives the probabilities of a rabbit beginning the experiment in a particular state.

The rabbit begins the experiment in the naive state, N, where the probability of a response to the conditioned stimulus is P_N. After each application of the unconditioned stimulus (UCS), there is a probability, a, that the rabbit will become aroused. We represent this by saying the rabbit moves to state A. Once it is activated, the rabbit may give a response to the CS. We denote the probability of this by P_A. After arousal, there is then a certain likelihood that the response will become conditioned to the CS. Let us call this probability, c, and this represents the transition into the third state, C. Once conditioning has occurred, there is a probability P_C that the rabbit will respond by blinking to the CS before the UCS occurs. Therefore, there are actually three distinct levels of performance, P_N, P_A, and P_C in a conditioning experiment.

3. There are other areas in psychology where Markov models are being used. These models are valuable in studies of paired associate learning, recall learning, and avoidance conditioning. In the list of references at the end of the chapter, a number of articles are included that contain discussions of these topics.

4. Hay (1966) poses the question as to how many different object displacements can produce the same optical motion at the eye. He uses a matrix model to aid in determining how many different object displacements are optically equivalent to the same transformation of the optical array. The main focus, no pun intended, is on the characteristics of these classes of optical stimuli. There is a mapping of a three-dimensional object into an optical array that essentially has a two-dimensional structure. Hay feels that optically equivalent object displacements have certain commonalities, and the optical motions that they produce give information about these common features. Object displacements are in some type of correspondence with optical transformations.

5. G. A. Miller (1968) uses matrices in an examination of the value of algebraic models in psycholinguistics. He includes incidence matrices for clustering that are associated with the hierarchical semantic system as part of his discussion of methods for investigating semantic relations.

6. The use of matrices in the study of linear and multiple regression is of paramount importance (Draper and Smith, 1966). If data is studied by the method of least squares, in order to draw conclusions about dependency relationships between variables, then this approach is called regression. In the case of linear regression we try to show that for a given value X, a corresponding value Y may be predicted that is an estimate of the actual observed value Y. For each trial we could describe the observed value Y_j as a linear function of X_j, $Y_j = \alpha_0 + \alpha_1 X_j + \varepsilon_j$, where ε_j is the error. If we were to do this for a large number of trials, N, we would have a system of equations for which we would like to find those estimates a_0 and a_1 for α_0 and α_1, that produce the smallest value of $\varepsilon_1^2 + \ldots + \varepsilon_a^2$. We could express our system for estimates a_0 and a_1 as

$$
\begin{aligned}
Y_1 &= a_0 + a_1X_1 \\
Y_2 &= a_0 + a_1X_2 \\
&\cdots\cdots\cdots \\
&\cdots\cdots\cdots \\
&\cdots\cdots\cdots \\
Y_N &= a_0 + a_1X_N
\end{aligned}
$$

In matrix notation we could write this as $\begin{pmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ \cdot \\ Y_N \end{pmatrix} = \begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & X_N \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \end{pmatrix}$.

Let us next multiply both sides of the equality by the transpose of

$$\begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & X_N \end{pmatrix}:$$

$$\begin{pmatrix} 1 & 1 & \cdots & 1 \\ X_1 & X_2 & \cdots & X_N \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ \cdot \\ Y_N \end{pmatrix} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ X_1 & X_2 & \cdots & X_N \end{pmatrix} \begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & X_N \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \end{pmatrix}$$

which simplifies to

$$\begin{pmatrix} Y_1 + Y_2 + \cdots + Y_N \\ X_1Y_1+X_2Y_2+ \cdots +X_NY_N \end{pmatrix} = \begin{pmatrix} N & X_1+X_2+\ldots+X_N \\ X_1+X_2+\ldots+X_N & X_1^2+X_2^2+\ldots+X_N^2 \end{pmatrix} \cdot \begin{pmatrix} a_0 \\ a_1 \end{pmatrix}$$

For convenience, let us denote the equality as $X'Y = (X'X)a$. As we can observe $X'X$ is an example of a square matrix. It is a 2 x 2 matrix. If we compute its inverse $(X'X)^{-1}$, then we can solve the system for a. Therefore, $(X'X)^{-1}X'Y = a$, from which we can find those values for $a = \begin{pmatrix} a_0 \\ a_1 \end{pmatrix}$ that give us the least square estimates.

Draper and Smith also use matrices in the analysis of variance and the variance and covariance of a_0 and a_1. The regression analysis may be shifted to an examination of correlations between variables. Correlations are desirable because their values range between -1 and 1. In general, we may form a matrix of correlations,

$$\begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1N} \\ r_{21} & r_{22} & \cdots & r_{2N} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ r_{N1} & r_{N2} & \cdots & r_{NN} \end{pmatrix}$$

from which we may analyze the interdependence of variables.

A real strength of the use of matrices over other techniques of solving systems of equations is that the approach is general and theoretically is the same for 30 equations in 30 unknowns as it is for 3 equations in 3 unknowns. This means that it is easier to write a computer program that can analyze the data.

7. A similar use of matrices is in the technique of factor analysis (Guilford, 1959). A correlation matrix has as many rows and columns as there are tests or variables; however, a factor matrix has as many rows as there are tests, but only as many columns as there are common factors. These two matrices are related by the equation $FF' = R$, where R is the correlation matrix, F is the factor matrix, and F' is the transpose of the factor matrix. A further result shows that the number of common factors is equal to the rank of the correlation matrix. In other words, it is equal to the number of linearly independent rows in the correlation matrix.

This concludes the chapter on matrices. Matrices were interesting to study because of the theoretical systems that sets of matrices may be formed into. For example, a group, ring, or vector space. The properties of matrices served as a nice transition from the theoretical to the applied, and the applications revealed the rich potential of matrices in questions of learning and in data analysis.

Before we end the book a few concluding words may be in order. Mathematics is a fascinating subject in itself. There are thousands of theoretical mathematicians who will attest to this. But it is also potentially a rich instrument in structuring and analyzing questions in psychology. The algebraic systems we have examined are most worthy of close scrutiny as to how and where they should be used. A mathematical model is like a fashion model, it looks good no matter what you put on it, but remember you are selling the clothes, not the model.

BIBLIOGRAPHY

Airasian, P. W., & Bart, W. M. Tree-Theory: A Theory-Generative Measurement Model. Paper read at the annual meeting of the American Educational Research Association, New York, February, 1971.

Arbib, M. (Ed.). Algebraic Theory of Machines, Languages, and Semigroups. New York: Academic Press, 1968.

Bart, W. M. A Generalization of Piaget's Logical-Mathematical Model for the Stage of Formal Operations. Journal of Mathematical Psychology, 1971, 8, 539-553.

Berlyne, D. E. Structure and Direction in Thinking. New York: Wiley, 1965.

Birkhoff, G., & MacLane, S. A Survey of Modern Algebra. New York: Macmillan, 1964.

Boyd, J. P. The Algebra of Group Kinship. Journal of Mathematical Psychology, 1969, 6, 139-167.

Burton, D. M. An Introduction to Abstract Mathematical Systems. Reading, Mass.: Addison-Wesley, 1965.

Chomsky, N. Introduction to the Formal Analysis of Natural Languages. In R. D. Luce, R. Bush, & E. Galanter (Eds.), Handbook of Mathematical Psychology. Vol. I. New York: Wiley, 1963.

Dean, R. A. Elements of Abstract Algebra. New York: Wiley, 1966.

Draper, N. R., & Smith, H. Applied Regression Analysis. New York: Wiley, 1966.

Guilford, J. P. Psychometric Methods. New York: McGraw-Hill, 1959.

Harlow, H. F. Learning Set and Error Factor Theory. In S. Koch (Ed.), Psychology, A Study of a Science. Vol. I. New York: McGraw-Hill, 1959.

Hay, J. C. Optical Motions and Space Perception: An Extension of Gibson's Analysis. Psychological Review, 1966, 73, 550-565.

Herstein, I. N. Topics in Algebra. New York: Blaisdell, 1964.

Hoffman, W. C. The Lie Algebra of Visual Perception. Journal of Mathematical Psychology, 1966, 3, 65-98.

Krantz, D. A Theory of Context Effects Based on Cross-Context Matching. Journal of Mathematical Psychology, 1968, 5, 1-48.

Krantz, D. Conjoint Measurement: The Luce-Tukey Axiomatization and Some Extensions. Journal of Mathematical Psychology, 1964, 1, 248-277.

Levine, M. V. Transformations that Render Curves Parallel. Journal of Mathematical Psychology, 1970, 7, 410-443.

Luce, R. D., & Tukey, J. W. Simultaneous Conjoint Measurement: A New Type of Fundamental Measurement. Journal of Mathematical Psychology, 1964, 1, 1-27.

Miller, G. A. Algebraic Models in Psycholinguistics. In A. J. Vlek (Ed.), Algebraic Models in Psychology. Proceedings of the NUFFIC international summer session in science at Hel Dude Hof, The Hague, August 1968, sponsored by NATO.

Natapoff, A. How Symmetry Restricts Symmetric Choice. Journal of Mathematical Psychology, 1970, 7, 445-465.

Piaget, J., & Inhelder, B. The Growth of Logical Thinking from Childhood to Adolescence. New York: Basic Books, 1958.

Piaget, J., & Inhelder, B. The Psychology of the Child. New York: Basic Books, 1969.

Scandura, J. M. Role of Rules in Behavior: Toward an Operational Definition of What Rule Is Learned. Psychological Review, 1970, 77, 516-533.

Theios, J., & Brelsford, J. A Markov Model for Classical Conditioning: Application to Eye-Blink Conditioning in Rabbits. Psychological Review, 1966, 73, 393-405.

Theios, J., & Brelsford, J. Theoretical Interpretation of a Markov Model for Avoidance Conditioning. Journal of Mathematical Psychology, 1966, 3, 140-162.

General References

The following are references which the authors found useful in the preparation of this book. The reader will find many of the topics that we have touched upon in this book in a more elaborated form in the following sources.

Birkhoff, G., & MacLane, S. A Survey of Modern Algebra. New York: Macmillan, 1964.

Burton, D. M. An Introduction to Abstract Mathematical Systems. Reading, Mass.: Addison-Wesley, 1965.

Dean, R. A. Elements of Abstract Algebra. New York: Wiley, 1966.

Herstein, J. N. Topics in Algebra. New York: Blaisdell, 1964.

Suppes, P. Introduction to Logic. Princeton, N.J.: D. van Nostrand, 1957.