Introduction

1 Measurement

A major difference between a "well-developed" science such as physics and some of the less "well-developed" sciences such as psychology or sociology is the degree to which things are measured. In this volume, we develop a theory of measurement that can act as a foundation for measurement in the social and behavioral sciences. Starting with such classical measurement concepts of the physical sciences as temperature and mass, we extend a theory of measurement to the social sciences. We discuss the measurement of preference, loudness, brightness, intelligence, and so on. We also apply measurement to such societal problems as air and noise pollution, weather forecasting, and public health, and comment on the development of pollution indices and consumer price indices.

Throughout, we apply the results to decisionmaking. The decisionmaking applications deal with transportation, consumer behavior, environmental problems, energy use, and medicine, as well as with laboratory situations involving human and animal subjects.

In this introduction, we try to give the reader a quick preview of the contents and organization of the book, and of the problems we shall address. The reader might prefer to read the introduction rather quickly the first time, and to return to it later.

The following questions are some of those we shall ask. A few of these are stated here in very general terms, and of course we shall try to be more specific in what follows.

1. What does it mean to measure preference, likes and dislikes, etc.?
2. When does it make sense to say goal a is twice as important as goal b? Or twice as worthwhile?

ENCYCLOPEDIA OF MATHEMATICS and Its Applications, Gian-Carlo Rota (ed.).
Vol. 7: Fred S. Roberts, Measurement Theory

3. Does it make sense to assert that the average IQ of one group of individuals is twice the average IQ of a second group?

4. Is it possible to measure air pollution with one index that takes account of many different pollutants? If so, does it make sense to assert that the pollution level today is 20% lower than it was yesterday?

5. Is it meaningful to assert that the consumer price index has increased by 20% in a given period?

6. How can we quantitatively relate subjective judgments of loudness of a sound to the physical intensity of the sound? And how can we use such quantitative relationships to develop indices of noise pollution?

7. How can we use expert judges to judge the relative importance, or significance, or merit of alternative candidates, and then combine the judgments of the experts into one measure of importance, significance, or merit?

8. How can we choose between two different treatments for a disease, given that we are not sure exactly what the outcomes of the treatments will be?

9. Can subjective judgments that one event is more likely to occur than another be quantified?

10. Is measurement still possible if judgments of preference, relative importance, loudness, etc., are inconsistent?

At an early stage of scientific development, measurement is usually performed at only the crudest level, that of classification. Some philosophers of science (e.g., Torgerson [1958]) do not even wish to call this measurement. What are the advantages of performing measurement that goes beyond simple classification? Hempel [1952] describes some of these. First, if we can measure things, we can begin to differentiate more than we can by simply classifying. For example, we can do more than simply distinguishing between warm objects and cold ones; we can assign degrees of warmth. Greater descriptive flexibility leads to greater flexibility in the formulation of general laws. (One should imagine trying to state the laws of physics using only classifications!) Measurement is usually performed by assigning numbers—though we shall argue below that this is not a prerequisite for measurement. Assignment of numbers makes possible the application of the concepts and theories of mathematics. General laws can now be stated in mathematical language—for example, as formal relations among quantities. Mathematical tools for analysis of numbers can help us to reason about objects and their properties, and to deduce general principles describing these properties. Thus, the existence of and experience with centuries of mathematical reasoning is a large part of the reason we find it useful to measure things.

We shall study two quite different types of measurement, fundamental measurement and derived measurement. (See Table 1.) Fundamental

Table 1. Measurement

A.	Types of Measurement
	1. Fundamental
	2. Derived
B.	Problems in Fundamental Measurement
	1. Representation
	2. Uniqueness
C.	Axioms in a Representation Theorem
	1. Prescriptive or normative
	2. Descriptive

measurement, as we shall describe it, takes place at an early stage of scientific development, when several fundamental concepts are measured for the first time. Mass, temperature, and volume are fundamental measures. Derived measurement takes place later, when some concepts have already been measured, and new measures are defined in terms of existing ones. Density can be thought of as a derived measure, defined as mass divided by volume. (Derived measurement is usually a relative matter. As we shall point out, the same scale—for example, density—can be introduced either as a derived scale or as a fundamental one.) Much of this work will be concerned with fundamental measurement. However, large parts of Chapters 2, 4, and 6 deal with derived measurement.

In the development of a scientific discipline, fundamental measurement is not usually performed in as formalistic a way as we describe in this volume. We usually do not attempt to describe an actual process that is undergone as techniques of measurement are developed. We are not interested in a measuring apparatus and in the interaction between the apparatus and the objects being measured. Rather, we attempt to describe how to put measurement on a firm, well-defined foundation. A number of the results, however, are of potential practical significance, and we shall attempt to mention practical techniques for actually computing measures or scales whenever possible. Once measurement has been put on a firm foundation, we develop a variety of tools for analyzing statements made in terms of scale values. These techniques have immediate application in a variety of practical problems.

Putting measurement on a firm foundation is not a terribly important activity in the modern-day physical sciences; many physicists would probably not consider it physics, but only "philosophy of physics." Practicing physicists usually take measurement for granted, and anyway measurements are usually based on powerful, well-established theories. In the social sciences, on the other hand, much present activity can be categorized as the search for appropriate scales of measurement to describe behavior, aid in decisions, etc. Putting measurement on a firm foundation can potentially play a very important role in the development of the social sciences.

Since much of this volume is devoted to putting measurement on a firm foundation, it is not surprising that much of the discussion is axiomatic in nature. We try to develop axioms or conditions under which measurement is possible. These are usually conditions about an individual's judgments, preferences, reactions, and so on. For example, we shall show that preferences can be measured in a consistent way provided that they satisfy two axioms:

(A) If you prefer a to b, you do not prefer b to a.
(B) If you do not prefer a to b, and do not prefer b to c, then you do not prefer a to c.

Such axioms can be looked at in two ways. The *prescriptive* or *normative* interpretation looks at the axioms as conditions of rationality. A truly rational man, given ideal conditions (unlimited computational ability, unlimited resources, etc.), should make judgments that satisfy the axioms, or else he is not acting rationally. Theories can be built based on the definition of a rational man, and procedures for a rational man to make decisions based on his judgments can be developed.* A second interpretation is that these axioms are *descriptive*. They give conditions on behavior which, if satisfied, allow measurement to take place. Whether or not measurement can take place then depends on whether or not an individual's judgments satisfy the axioms, and we hope that this is a testable question. Whenever possible in this volume, we try to discuss tests of the axioms presented. Although it is not fair to generalize, it is often true that the prescriptive axiomatic theories of measurement in the social sciences have been developed in the economic literature, and the descriptive theories in the psychological literature. The theories of physical measurement in general are simultaneously prescriptive and descriptive, and have largely philosophical significance. However, both the prescriptive and descriptive theories of measurement in the social sciences propose new theories and lead to new laws (the axioms), suggest experiments to distinguish between these laws, and give rise to practical techniques of measurement where none was possible before. (See Krantz [1968, Section 2.1] for a more detailed discussion of this point.)

The problem of finding axioms under which measurement is possible will be called in this volume the *representation problem*. Once measurement has been accomplished, we also consider the *uniqueness problem*: How

*Keeney and Raiffa [1976] distinguish between the normative and the prescriptive. The normative interpretation refers to an "idealized,...superrational being with an all-powering intellect," while the prescriptive refers to "normally intelligent people who want to think hard and systematically..." We shall not make this distinction.

unique is the resulting measure or scale? This problem will be very important in telling us what kinds of comparisons and what kinds of mathematical manipulations are possible with the measures obtained. For example, we shall ask whether it is meaningful to say that one group's average IQ is twice that of a second group's, or that the average air pollution level in one city is greater than that in a second city. We shall discuss the uniqueness problem in depth in Chapter 2, and return to it often.

2 The Measurement Literature

Our approach to measurement theory follows very closely that of Scott and Suppes [1958], Suppes and Zinnes [1963], Pfanzagl [1968], Krantz [1968], and Krantz *et al*. [1971]. There is a long-standing literature on the nature of measurement, and the reader might wish to consult some of this literature for other points of view. Much of this literature concerns itself with measurement in the physical sciences. Early influential books were written by Campbell [1920, 1928, 1938]. Campbell and such writers as Helmholtz [1887], Cohen and Nagel [1934], Guild [1938], Reese [1943], and Ellis [1966] did not accept measurement unless it involved some sort of concatenation or addition operation of the type we shall describe in Section 3.2. Although this approach to measurement was quite appropriate for physics, it was not really broad enough to encompass the measurement problems of the social sciences. However, as late as 1940, a committee of the British Association for the Advancement of Science questioned specifically whether psychologists such as S. S. Stevens, who were measuring human sensations such as loudness, were really performing measurement, since they used no concept of addition (Final Report of the British Association for the Advancement of Science [1940]). Our approach is broader than that of the classical writers, and considers a measurement theory that can act as a basic foundation for measurement in the social sciences. Some other important references on measurement theory, closer to our point of view, are Stevens [1946, 1951, 1959, 1968] and Adams [1965]. Stevens was the first to observe that uniqueness of a measurement assignment defined scale type, in the sense we describe in Section 2.3. Although chemistry and biology have used scales not much different from those of physics, the social sciences, in measurement of preference, intelligence, etc., have given rise to entirely new types of scales, with somewhat different character. We shall explore these in detail.

3 Decisionmaking

A major application of measurement theory is to problems of decisionmaking. Various authors (for example, Luce and Raiffa [1957]) have classified decisionmaking problems in several ways. (See Table 2.) The first

Table 2. Classification of Decisionmaking Problems

A.	Who Makes the Decision?
	1. Individual
	2. Group
B.	How Much Information about Consequences of Actions Is Known?
	1. Certainty
	2. Risk
	3. Uncertainty

distinction to make is whether the decision is being made by an *individual* or a *group*. Some groups act as individuals; the difference is whether members of the group are expressing their own opinions from among which the group must choose. In this volume, we shall be almost exclusively concerned with the individual decisionmaker. For good summaries of the group decisionmaking problem, see Luce and Raiffa [1957], Sen [1970], or Fishburn [1972]. A classic reference is Arrow [1951]. More specifically for the decisionmaking problem involving elections, see Black [1958], Farquharson [1969], and Riker and Ordeshook [1973].

A second distinction between problems of decisionmaking involves how much certainty there is about outcomes of various actions. If we are trying to choose among alternative acts, we say that we are in a situation of *certainty* if for each act there is exactly one consequence. We are in a situation of *risk* if for each act there is a set of possible consequences, none of which occurs with certainty, but each of which occurs with a known probability. Finally, we are in a situation of *uncertainty* if the probabilities that consequences will occur are unknown. In practice many decisionmaking problems involve a mixture of risk and uncertainty (or even of certainty, risk, and uncertainty).

In this volume, we shall concentrate primarily on decisions involving certainty. However, in Chapter 7 we shall discuss decisionmaking under risk and under uncertainty. In Chapter 8, we shall discuss how to measure the probability of various outcomes, given subjective judgments about which of two outcomes is more probable.

4 Utility

If decisions are being made in a situation of certainty, then we often choose that act whose certain consequence maximizes (or minimizes) some index—for example, a measure of value, worth, satisfaction, or utility. The notion of *utility* goes back at least to the eighteenth century. Much of the original interest in this concept goes back to Jeremy Bentham, who defines utility in the first chapter of *The Principles of Morals and Legislation* (1789), as follows:

> By utility is meant that property in any object, whereby it tends to produce benefit, advantage, pleasure, good, or happiness (all this in the present case comes to the same thing), or (what comes again to the same thing) to prevent the happening of mischief, pain, evil, or unhappiness to the party whose interest is considered: if that party be the community in general, then the happiness of the community; if a particular individual, then the happiness of that individual.

Bentham formulated procedures for measuring utility, for he thought societies should strive for "the greatest good for the greatest number"—that is, maximum utility. In this volume, we shall usually look at the measurement of utility as a representation problem and list various axiom systems sufficient for the existence of *utility functions*, measures of utility. In the early applications of utility theory in economics, there was an emphasis by economists such as Walras and Jevons on obtaining utility functions that were additive in the sense that the utility of the combination of two objects was the sum of the individual utilities of the two objects. Such a utility function is called *cardinal*. (More generally, a utility function will be called *cardinal* if it is unique up to a positive linear transformation.) We shall discuss axioms for cardinal utility in Chapters 3 and 5. In the late nineteenth century, Edgeworth [1881] questioned the assumption of additivity. In the early twentieth century, the economist Pareto [1906] showed that much of economic theory depends only on the assumption that the utility function is *ordinal*—that is, that the utilities can be used only to decide which of two objects has a higher value, and addition in particular does not necessarily make sense. We shall discuss ordinal utility in Section 3.1.

Luce and Suppes [1965] classify theories of utility in several ways. (See Table 3.) A similar classification applies to theories of measurement in general. First, these theories can invòlve decisionmaking under certainty, risk, or uncertainty. We have already discussed these distinctions. A second distinction is whether the theories are *algebraic* and *deterministic* or *probabilistic*. Most of the traditional utility theories are algebraic, and the

Table 3. Classification of Theories of Utility

A.	How Much Information about Consequences of Actions Is Known to the Decisionmakers?
	1. Certainty
	2. Risk
	3. Uncertainty
B.	What Kind of Representation Theorems?
	1. Algebraic and deterministic
	2. Probabilistic
C.	What Judgments Is Measurement of Utility Based on?
	1. Simple choices
	2. Rankings or choices from a set

representation theorems state axioms that are very algebraic in nature. For example, the axioms for cardinal utility functions stated in Section 3.2 say that preference and combination of objects define a certain kind of ordered semigroup. (Sometimes, if there is an interest in continuous utility functions, topological axioms must be added.) Our basic formulation of fundamental measurement is algebraic, starting with the idea of a relation and an operation (Chapter 1), and then basing measurement on a relational system consisting of certain empirical relations and operations (Section 1.8). Sometimes it is useful to modify algebraic theories. This is the case when the fundamental judgments that representation theorems describe are made inconsistently or made consistently, but according to some statistical regularity. When data is inconsistent or there is no discernible deterministic pattern available, the algebraic theories must be replaced by probabilistic theories, which are built around this more random data. Falmagne [1976] argues that we shall (almost) always have to replace algebraic theories by probabilistic or random ones, at least in applications to the behavioral sciences. He is concerned with developing probabilistic analogues of the traditional algebraic theories of fundamental measurement. We discuss some probabilistic theories in Section 6.2, but otherwise our basic approach is algebraic.

Both the algebraic and probabilistic theories of utility can take two forms. The first form assumes that a utility function can be derived simply on the basis of *simple choices* among pairs of alternatives or objects. The second asks for a *ranking* among elements in each set of alternatives, or *choice of a best element* from the set, before deriving a utility function. We shall discuss only simple choice theories in this volume. Luce and Suppes [1965, Section 6] and Krantz *et al.* [to appear] summarize some probabilistic ranking theories.

Some excellent surveys of utility theory and decisionmaking are Fishburn [1968, 1970] and Luce and Suppes [1965] and, for utility with multi-attributed alternatives, Farquhar [1977] and Keeney and Raiffa [1976]. Aoki, Chipman, and Fishburn [1971] list many references on preferences, utility, and demand. Other references on utility theory include Adams [1960], Arrow [1951], Chipman [1960], Debreu [1959], Edwards [1954, 1961], Fischer and Edwards [1973], Fishburn [1964], Luce and Raiffa [1957], Majumdar [1958], Savage [1954], Slovic and Lichtenstein [1971], Stigler [1950], and von Neumann and Morgenstern [1944]. For a more complete list of references, see Luce and Suppes [1965] or Fishburn [1970].

5 Mathematics

The study of measurement, decisionmaking, and utility has given rise to interesting new mathematical results and questions, most of which are not very well known among mathematicians. A major purpose of this book is to organize a class of these mathematical results and to introduce the mathematically trained reader to them.

Most of the mathematics in this book is distinctly algebraic in flavor. It deals with ordered algebraic systems and homomorphic mappings of one such system into another. Many of these systems are finite, so there is a discrete flavor to much of the mathematics. Fundamental properties of the real number system are used throughout, and many of the results are related to branches of logic such as set theory and foundations of geometry. A variety of results and tools of an analytic or topological nature are scattered throughout. For example, solutions of certain functional equations are studied in Chapter 4.

Finally, much of the mathematics described here has a flavor of its own. It is to be expected that, as more mathematicians become interested in problems of the social sciences, new forms of mathematics will have to be developed to solve these problems. Hopefully a work like this one will stimulate some mathematically trained individuals to become involved in such developments.

6 Organization of the Book

Since relational systems form a basis for much of the theory of measurement, I have chosen to include an introductory chapter on the theory of relations, which includes most of the relation-theoretic concepts needed later on. The reader familiar with this theory can skip much of Chapter 1, though he should read Theorems 1.2, 1.3, and 1.4 and Section 1.8.

Chapter 2 introduces fundamental and derived measurement and defines the two basic problems of fundamental measurement—the representation problem and the uniqueness problem. It introduces scale type and uses scale type to study what statements involving scales are meaningful. It applies the results to a variety of practical problems, such as the making of index numbers (consumer price, consumer confidence) and the measurement of air pollution. Chapter 3 illustrates fundamental measurement by giving three basic representation theorems, stating conditions under which measurement can be performed; one is for ordinal measurement, one for extensive measurement, and one for difference measurement. In the utility interpretation, the first theorem gives conditions for the existence of ordinal utility functions, and the second and third conditions for the existence of cardinal utility functions. Chapter 4 is an applications chapter. It presents the theory of psychophysical scaling, and applies the results about scale type and fundamental and derived measurement to this theory. It concentrates on the measurement of loudness, and shows how one might derive a measure of loudness from known physical scales of intensity of a sound.

Chapters 5 through 8 introduce various complications into the measurement picture. In Chapter 5, we study "complicated" or multidimensional alternatives. A new kind of fundamental measurement, called conjoint measurement, is introduced. The results have application to combinations of psychological factors such as drive and incentive, to binaural additivity of loudness, to mental testing problems, and to measuring discomfort due to different weather factors. They also have applications to problems of urban services, allocation of funds for education, treatment of medical problems, design of large public facilities such as airports, etc. In Chapter 6, we ask what to do when it is not even possible to perform ordinal measurement in the sense of Section 3.1—that is, when the necessary conditions for ordinal measurement are violated. We widen our scope, and introduce the idea of measurement without numbers or measurement when the basic data is inconsistent. We study Luce's fundamental idea of a semiorder and then give examples of probabilistic theories of measurement. A variety of applications to data from pair comparison experiments are presented.

Chapter 7 discusses the problem of decisionmaking under risk or uncertainty. It introduces the famous expected utility hypothesis, which goes back to Bernoulli [1738], and which says that a decisionmaker chooses that action which maximizes his expected utility. Accepting this as a *prescription*, we give applications to decisionmaking problems from such fields as transportation, medicine, and public health, and to calculation of utility. A number of *descriptive* utility theorists believe that although we do not consciously calculate expected utilities, we act *as if* we are maximizing expected utility. We present axiom systems which give conditions on choices among acts with risky or uncertain consequences sufficient to guarantee that these choices are made as if expected utility were being maximized.

In a decisionmaking situation under uncertainty, and in other situations as well, it is sometimes useful or necessary to be able to calculate probabilities that reflect our judgments that certain events are subjectively more probable than others. In Chapter 8, we discuss the measurement of subjective probability.

Chapters 1 and 2 and Section 3.1 form a groundwork for the rest of the book. Much of the rest of the book can be read in any order. For example, Sections 3.2 and 3.3 and Chapters 4, 5, 6, 7, and 8 are essentially independent, though Section 7.3 depends on Chapter 5 and parts of Chapter 8 depend on earlier material in the beginning of Chapter 7 and in Section 3.2. Even Sections 6.1 and 6.2 are essentially independent, with the exception that Section 6.2.4 depends on Section 6.1.

There are so many topics to be covered in the growing field of measurement theory that it is impossible to survey them all. However, references have been provided at the end of each chapter, and it is hoped they will lead the reader into the literature.