J. Wixted & H. Pashler (Eds.), Stevens' Handbook of Experimental Psychology, 3rd Edition, Vol. 4: Methodology in Experimental Psychology. New York: Wiley, pp. 1-41.

CHAPTER 1

Representational Measurement Theory

R. DUNCAN LUCE AND PATRICK SUPPES

CONCEPT OF REPRESENTATIONAL MEASUREMENT

Representational measurement is, on the one hand, an attempt to understand the nature of empirical observations that can be usefully recoded, in some reasonably unique fashion, in terms of familiar mathematical structures. The most common of these representing structures are the ordinary real numbers ordered in the usual way and with the operations of addition, +, and/or multiplication, ·. Intuitively, such representations seems a possibility when dealing with variables for which people have a clear sense of "greater than." When data can be summarized numerically, our knowledge of how to calculate and to relate numbers can usefully come into play. However, as we will see, caution must be exerted not to go beyond the information aetually coded numerically. In addition, more complex mathematical structures such as geometries are often used, for example, in multidimensional scaling.

On the other hand, representational measurement goes well beyond the mere construction of numerical representations to a careful examination of how such representations relate to one another in substantive scientific theories, such as in physics, psychophysics, and utility theory. These may be thought of as applications of measurement concepts for representing various kinds of empirical relations among variables.

In the 75 or so years beginning in 1870, some psychologists (often physicists or physicians turned psychologists) attempted to import measurement ideas from physics, but gradually it became clear that doing this successfully was a good deal trickier than was initially thought. Indeed, by the 1940s a number of physicists and philosophers of physics concluded that psychologists really did not and could not have an adequate basis for measurement. They eonclnded, correctly, that the classical measurement models were for the most part unsuited to psychological phenomena. But they also concluded, incorrectly, that no scientifically sound psychological measurement is possible at all. In part, the theory of representational measurement was the response of some psychologists and other social scientists who were fairly well trained in the necessary physics and mathematics to understand how to modify in substantial ways the classical models of physical measurement to be better suited to psychological issues. The purpose of this chapter is to outline the high points of the 50-year effort from 1950 to the present to develop a deeper understanding of such measurement.

The authors thank János Aczél, Ehtibar Dzhafarov, Jean-Claude Falmagne, and A.A.J. Marley for helpful comments and criticisms of an earlier draft.

Empirical Structures

Performing any experiment, in particular a psychological one, is a complex activity that we never analyze or report completely. The part that we analyze systematically and report on is sometimes called *a model of the data* or, in terms that are useful in the theory of measurement, *an empirical structure.* Such an empirical structure of an experiment is a drastic reduction of the entire experimental activity. In the simplest, purely psychological cases, we represent the empirical model as a set of stimuli, a set of responses, and some relations observed to hold between the stimuli and responses. (Such an empirical restriction to stimuli and responses does not mean that the theoretical considerations are so restricted; unobservable concepts may well play a role in theory.) In many psychological measurement experiments such an empirical structure consists of a set of stimuli that vary along a single dimension, for example, a set of sounds varying only in intensity. We might then record the pairwise judgments of loudness by a binary relation on the set of stimuli, where the first member of a pair represents the subject's judgment of which of two sounds was louder.

The use of such empirical structures in psychology is widespread because they come close to the way data are organized for subsequent statistical analysis or for testing a theory or hypothesis.

An important cluster of objections to the concept of empirical structures or models of data exists. One is that the formal analysis of empirical structures includes only a small portion of the many problems of experimental design. Among these are issues such as the randomization of responses between left and right hands and symmetry conditions in the lighting of visual stimuli. For example, in most experiments that study aspects of vision, having considerably more intense light on the left side of the subject than on the right would be considered a mistake. Such considerations do not ordinarily enter into any formal description of the experiment. This is just the beginning. There are understood conditions that are assumed to hold but are not enumerated: Sudden loud noises did not interfere with the concentration of the subjects, and neither the experimenter talked to the subject nor the subject to the experimenter during the collection of the data—although exceptions to this rule can certainly be found, especially in linguistically oriented experiments.

The concept of empirical structures is just meant to isolate the part of the experimental activity and the form of the data relevant to the hypothesis or theory being tested or to the measurements being made.

Isomorphic Structures

The prehistory of mathematics, before Babylonian, Chinese, or Egyptian civilizations began, left no written record but nonetheless had as a major development the concept of number. In particular, counting of small colleetions of objects was present. Oral terms for some sort of counting seem to exist in every language. The next big step was the introduction, no doubt independently in several places, of a written notation for numbers. It was a feat of great abstraction to develop the general theory of the constructive operations of counting, adding, subtracting, multiplying, and dividing numbers. The first problem for a theory of measurement was to show how this arithmetic of numbers could be constructed and applied to a variety of empirical structures.

To investigate this problem, as we do in the next section, we need the general notion of isomorphism between two structures. The intuitive idea is straightforward: Two structures are isomorphic when they exhibit the same structure from the standpoint of their basic concepts. The point of the formal definition of isomorphism is to make this notion of *same structure* precise.

As an elementary example, consider a *binary relational structure* consisting of a nonempty set A and a binary relation R defined on this set. We will be considering pairs of such structures in which both may be empirical structures, both may be numerical structures, or one may be empirical and the other numerical. The definition of isomorphism is unaffected by which combination is being considered.

The way we make the concept of having the same structure precise is to require the existence of a function mapping the one structure onto the other that preserves the binary relation. Formally, a binary relation structure (A, R) is *isomorphic* to a binary relation structure (A', R') if and only if there is a function f such that

(i) the domain of f is A and the codomain of f is A', i.e., A' is the image of A under f,
(ii) f is a one-one function,[1] and
(iii) for a and b in A, aRb iff[2] $f(a)R'f(b)$.

To illustrate this definition of isomorphism, consider the question: Are any two finite binary relation structures with the same number of elements isomorphic? Intuitively, it seems clear that the answer should be negative, because in one of the structures all the objects could stand in the relation R to each other and not so in the other. This is indeed the case and shows at once, as intended, that isomorphism depends not just on a one-one function from one set to another, but also on the structure as represented in the binary relation.

[1]In recent years, conditions (i) and (ii) together have come to be called *bijective*.

[2]This is a standard abbreviation for "if and only if."

Ordered Relational Structures

Weak Order

An idea basic to measurement is that the objects being measured exhibit a qualitative attribute for which it makes sense to ask the question: Which of two objects exhibits more of the attribute, or do they exhibit it to the same degree? For example, the attribute of having greater mass is reflected by placing the two objects on the pans of an equal-arm pan balance and observing which deflects downward. The attribute of loudness is reflected by which of two sounds a subject deems as louder or equally loud. Thus, the focus of measurement is not just on the numerical representation of any relational structures, but of ordered ones, that is, ones for which one of the relations is a *weak order*, denoted $\succsim$, which has two defining properties for all elements a, b, c in the domain A:

(i) *Transitive:* if $a \succsim b$ and $b \succsim c$, then $a \succsim c$.
(ii) *Connected:* either $a \succsim b$ or $b \succsim a$ or both.

The intuitive idea is that $\succsim$ captures the ordering of the attribute that we are attempting to measure.

Two distinct relations can be defined in terms of $\succsim$:

$$a \succ b \text{ iff } a \succsim b \text{ and not } (b \succsim a);$$
$$a \sim b \text{ iff both } a \succsim b \text{ and } b \succsim a.$$

It is an easy exercise to show that $\succ$ is transitive and irreflexive (i.e., $a \succ a$ cannot hold), and that $\sim$ is an equivalence relation (i.e., transitive, symmetric in the sense that $a \sim b$ iff $b \sim a$, and reflexive in the sense that $a \sim a$). The latter means that $\sim$ partitions A into equivalence classes.

Homomorphism

For most measurement situations one really is working with weak orders—after all, two entities having the same weight are not in general identical. But often it is mathematically easier to work with isomorphisms to the ordered real numbers, in whieh case one must deal with the following concept of simple orders. We do this by inducing the preference order over the equivalence classes defined by $\sim$. When $\sim$ is $=$, each element is an equivalence class, and the weak order $\succeq$ is called a *simple order*. The mapping from the weakly ordered structure via the isomorphisms of the (mutually disjoint) equivalences classes to the ordered real numbers is called a *homomorphism*. Unlike an isomorphism, which is one to one, an homomorphism is many to one. In some cases, such as additive conjoint measurement, discussed later, it is somewhat difficult, although possible, to formulate the theory using the equivalence classes.

Two Fundamental Problems of Representational Measurement

Existence

The most fundamental problem for a theory of representational measurement is to construct the following representation: Given an empirical structure satisfying certain properties, to which numerical structures, if any, is it isomorphic? These numerical structures, thus, represent the empirical one. It is the existence of such isomorphisms that constitutes the representational claim that measurement of a fundamental kind has taken place.

Quantification or measurement, in the sense just characterized, is important in some way in all empirical sciences. The primary significance of this fact is that given the isomorphism of structures, we may pass from the particular empirical structure to the numerical one and then use all our familiar computational methods, as applied to the isomorphic arithmetical structure, to infer facts about the isomorphic empirical structure. Such passage from simple qualitative observations to quantitative ones—the isomorphism of structures passing from the empirical to the numerieal—is necessary for precise prediction or control of phenomena. Of course, such a representation is useful only to the extent of the preeision of the observations on which it is based. A variety of numerical representations for various empirical psychological phenomena is given in the sections that follow.

Uniqueness

The second fundamental problem of representational measurement is to discover the uniqueness of the representations. Solving the representation problem for a theory of measurement is not enough. There is usually a formal difference between the kind of assignment of numbers arising from different procedures of measurement, as may be seen in three intuitive examples:

1. The population of California is greater than that of New York.
2. Mary is 10 years older than John.
3. The temperature in New York City this afternoon will be 92 °F.

Here we may easily distinguish three kinds of measurements. The first is an example of counting, which is an absolute scale. The number of members of a given collection that is counted is determined uniquely in the ideal case, although that can be difficult in practice (witness the 2000 presidential election in Florida). In contrast, the second example, the measurement of difference in age, is a ratio scale. Empirical procedures for measuring age do not determine the unit of age—chosen in the example to be the year rather than, for example, the month or the week. Although the choice of the unit of a person's age is arbitrary—that is, not empirically prescribed—that of the zero, birth, is not. Thus, the ratio of the ages of any two people is independent of its choice, and the age of people is an example of a ratio scale. The measurement of distance is another example of such a ratio scale. The third example, that of temperature, is an example of an interval scale. The empirical procedure of measuring temperature by use of a standard thermometer or other device determines neither a unit nor an origin.

We may thus also describe the second fundamental problem for representational measurement as that of determining the scale type of the measurements resulting from a given procedure.

A BRIEF HISTORY OF MEASUREMENT

Pre-19th-Century Measurement

Already by the fifth century B.C., if not before, Greek geometers were investigating problems central to the nature of measurement. The Greek achievements in mathematics are all of relevance to measurement. First, the theory of number, meaning for them the theory of the positive integers, was closely connected with counting; second, the geometric theory of proportion was central to magnitudes that we now represent by rational numbers (= ratios of integers); and, finally, the theory of incommensurable geometric magnitudes for those magnitudes that could not be represented by ratios. The famous proof of the irrationality of the square root of two seems arithmetic in spirit to us, but almost certainly the Greek discovery of incommensurability was geometric in character, namely, that the length of the diagonal of a square, or the hypotenuse of an isosceles right-angled triangle, was not commensurable with the sides. The Greeks well understood that the various kinds of results just described applied in general to magnitudes and not in any sense only to numbers or even only to the length of line segments. The spirit of this may be seen in the first definition of *Book 10* of Euclid, the one dealing with incommensurables: "Those magnitudes are said to be commensurable which are measured by the same measure, and those incommensurable which cannot have any common measure" (trans. 1956, p. 10).

It does not take much investigation to determine that theories and practices relevant to measurement occur throughout the centuries in many different contexts. It is impossible to give details here, but we mention a few salient examples. The first is the discussion of the measurement of pleasure and pain in Plato's dialogue *Protagoras*. The second is the set of partial qualitative axioms, characterizing in our terms empirical structures, given by Archimedes for measuring on unequal balances (Suppes, 1980). Here the two qualitative concepts are the distanee from the focal point of the balance and the weights of the objects placed in the two pans of the balance. This is perhaps the first partial qualitative axiomatization of conjoint measurement, which is discussed in more detail later. The third example is the large medieval literature giving a variety of qualitative axioms for the measurement of weight (Moody and Claggett, 1952). (Psychologists concerned about the difficulty of clarifying the measurement of fundamental psychological quantities should be encouraged by reading O'Brien's 1981 detailed exposition of the confused theories of weight in the ancient world.) The fourth example is the detailed discussion of intensive quantities by Nicole Oresme in the 14th century A.D. The fifth is Galileo's successful geometrization in the 17th century of the motion of heavenly bodies, done in the context of stating essentially qualitative axioms for what, in the earlier tradition, would be called the quantity of motion. The final example is also perhaps the last great, magnificent, original treatise of natural science written wholly in the geometrical tradition—Newton's *Principia* of 1687. Even in his famous three laws of motion, concepts were formulated in a qualitative, geometrical way, characteristic of the later formulation of qualitative axioms of measurement.

19th- and Early 20th-Century Physical Measurement

The most important early 19th-century work on measurement was the abstract theory of extensive quantities published in 1844 by H. Grassmann, *Die Wissenschaft der Extensiven Grösse oder die Ausdehnungslehre*. This abstract and forbidding treatise, not properly appreciated by mathematicians at the time of its appearance, contained at this early date the important generalization of the concept of geometric extensive quantities to n-dimensional vector spaces and, thus, to the addition, for example, of n-dimensional vectors. Grassmann also developed for the first time a theory of barycentric coordinates in n dimensions. It is now recognized that this was the first general and abstract theory of extensive quantities to be treated in a comprehensive manner.

Extensive Measurement

Despite the precedent of the massive work of Grassmann, it is fair to say that the modern theory of one-dimensional, extensive measurement originated much later in the century with the fundamental work of Helmholtz (1887) and Hölder (1901). The two fundamental concepts of these first modern attempts, and later ones as well, is a binary operation $\circ$ of combination and an ordering relation $\succsim$, each of which has different interpretations in different empirical structures. For example, mass ordering $\succsim$ is determined by an equal-arm pan balance (in a vacuum) with $a \circ b$ denoting objects a and b both placed on one pan. Lengths of rods are ordered by placing them side-by-side, adjusting one end to agree, and determining which rod extends beyond the other at the opposite end, and $\circ$ means abutting two rods along a straight line. The ways in which the basic axioms can be stated to describe the intertwining of these two concepts has a long history of later development. In every case, however, the fundamental isomorphism condition is the following: For a, b in the empirical domain,

$$f(a) \geq f(b) \Leftrightarrow a \succsim b, \qquad (1)$$

$$f(a \circ b) = f(a) + f(b), \qquad (2)$$

where f is the mapping function from the empirical structure to the numerical structure of the additive, positive real numbers, that is, for all entities a, $f(a) > 0$.

Certain necessary empirical (testable) properties must be satisfied for such a representation to hold. Among them are for all entities a, b, and c,

Commutativity: $a \circ b \sim b \circ a$.
Associativity: $(a \circ b) \circ c \sim a \circ (b \circ c)$.
Monotonicity: $a \succsim b \Leftrightarrow a \circ c \succsim b \circ c$.
Positivity: $a \circ a \succ a$.

Let a be any element. Define a *standard sequence based on* a to be a sequence $a(n)$, where n is an integer, such that $a(1) = a$, and for $i > 1$, $a(i) \sim a(i-1) \circ a$. An example of such a standard sequence is the centimeter marks on a meter ruler. The idea is that the elements of a standard sequence are equally spaced. The following (not directly testable) condition ensures that the stimuli are commensurable:

Archimedean: For any entities a, b,
there is an integer n such that $a(n) \succ b$.

These, together with the following structural condition that ensures very small elements,

Solvability: if $a \succ b$,
then for some c, $a \succ b \circ c$,

were shown to imply the existence of the representation given by Equations (1) and (2). By formulating the Archimedean axiom differently, Roberts and Luce (1968) showed that the solvability axiom could be eliminated.

Such empirical structures are called *extensive*. The uniqueness of their representations is discussed shortly.

Probability and Partial Operations

It is well known that probability P is an additive measure in the sense that it maps events into [0, 1] such that, for events A and B that are disjoint,

$$P(A \cup B) = P(A) + P(B).$$

Thus, probability is close to extensive measurement—but not quite, because the operation is limited to only disjoint events. However, the theory of extensive measurement can be generalized to partial operations having the property that if a and b are such that $a \circ b$ is defined and if $a \succsim c$ and $b \succsim d$, then $c \circ d$ is also defined. With some adaptation, this can be applied to probability; the details can be found in Chapter 3 of Krantz, Luce, Suppes, and Tversky (1971). (This reference is subsequently cited as FM I for Volume I of *Foundations of Measurement*. The other volumes are Suppes, Krantz, Luce, & Tversky, 1990, cited as FM II, and Luce, Krantz, Suppes, & Tversky, 1990, cited as FM III.)

Finite Partial Extensive Structures

Continuing with the theme of partial operation, we describe a recent treatment of a finite extensive structure that also has ratio scale representation and that is fully in the spirit of the earlier work involving continuous models. Suppose X is a finite set of physical objects, any two of which balance on an equal-arm balance; that is, if $a_1, \ldots, a_n$ are the objects, for any i and j, $i \neq j$, then $a_i \sim a_j$. Thus, they weigh the same. Moreover, if A and B are two sets of these objects, then on the balance we have $A \sim B$ if and only if A and B have the same number of objects. We also have a concatenation operation, union of disjoint sets. If $A \cap B = \emptyset$, then $A \cup B \sim C$ if and only if the objects in C balance the objects in A together with the objects in B. The qualitative strict ordering $A \succ B$ has an obvious operational meaning, which is that the objects in A, taken together, weigh more on the balance than the objects in B, taken together.

This simple setup is adequate to establish, by fundamental measurement, a scheme for numerically weighing other objects not in X. First, our homomorphism f on X is really simple. Since for all a_i and a_j and X, $a_i \sim a_j$, we have

$$f(a_i) = f(a_j),$$

with the restriction that $f(a_i) > 0$. We extend f to A, a subset of X, by setting $f(A) = |A| =$ the cardinality of (number of objects in) A. The extensive structure is thus transparent: For A and B subsets of X, if $A \cap B = \emptyset$ then

$$f(A \cup B) = |A \cup B| = |A| + |B| \\ = f(A) + f(B).$$

If we multiply f by any $\alpha > 0$ the equation still holds, as does the ordering. Moreover, in simple finite cases of extensive measurement such as the present, it is easy to prove directly that no transformations other than ratio transformations are possible. Let f^* denote another representation. For some object a, set $\alpha = f(a)/f^*(a)$. Observe that if $|A| = n$, then by a finite induction

$$\frac{f(A)}{f^*(A)} = \frac{nf(a)}{nf^*(a)} = \alpha,$$

so the representation forms a ratio scale.

Finite Probability

The "objects" $a_1, \ldots, a_n$ are now interpreted as possible outcomes of a probabilistic measurement experiment, so the a_is are the possible atomic events whose qualitative probability is to be judged.

The ordering $A \succsim B$ is interpreted as meaning that event A is at least as probable as event B; $A \sim B$ as A and B are equally probable; $A \succ B$ as A is strictly more probable than B.

Then we would like to interpret $f(A)$ as the numerical probability of event A, but if f is unique up to only a ratio scale, this will not work since $f(A)$ could be 50.1, not exactly a probability.

By adding another concept, that of the probabilistic independence of two events, we can strengthen the uniqueness result to that of an absolute scale. This is written $A \perp B$. Given a probability measure, the definition of independence is familiar: $A \perp B$ if and only if $P(A \cap B) = P(A)P(B)$. Independence cannot be defined in terms of the qualitative concepts introduced for arbitrary finite qualitative probability structures, but can be defined by extending the structure to elementary random variables (Suppes and Alechina, 1994). However, a definition can be given for the special case in which all atoms are equiprobable; it again uses the cardinality of the sets: $A \perp B$ if and only if $|X| \cdot |A \cap B| = |A| \cdot |B|$. It immediately follows from this definition that $X \perp X$, whence in the interpretation of $\perp$ we must have

$$P(X) = P(X \cap X) = P(X)P(X),$$

but this equation is satisfied only if $P(X) = 0$, which is impossible since $P(\emptyset) = 0$ and $X \succ \emptyset$, or $P(X) = 1$, which means that the scale type is an absolute—not a ratio—scale, as it should be for probability.

Units and Dimensions

An important aspect of 19th century physics was the development, starting with Fourier's work (1822/1955), of an explicit theory of units and dimensions. This is so commonplace now in physics that it is hard to believe that it only really began at such a late date. In Fourier's famous work, devoted to the theory of heat, he announced that in order to measure physical quantities and express them numerically, five different kinds of units of measurement were needed, namely, those of length, time, mass, temperature, and heat. Of even greater importance is the specific table he gave, for perhaps the first time in the history of physics, of the dimensions of various physical quantities. A modern version of such a table appears at the end of FM I.

The importance of this tradition of units and dimensions in the 19th century is to be seen in Maxwell's famous treatise on electricity and magnetism (1873). As a preliminary, he began with 26 numbered paragraphs on the measurement of quantities because of the importance he attached to problems of measurement in electricity and magnetism, a topic that was virtually unknown before the 19th century. Maxwell emphasized the fundamental character of the three fundamental units of length, time, and mass. He then went on to derive units, and by this he meant quantities whose dimensions may be expressed in terms of fundamental units (e.g., kinetic energy, whose dimension in thc usual notation is ML^2T^{-2}). Dimensional analysis, first put in systematic form by Fourier, is very useful in analyzing the consistency of the use of quantities in equations and can also be used for wider purposes, which are discussed in some detail in FM I.

Derived Measurement

In the Fourier and Maxwell analyses, the question of how a derived quantity is actually to be measured does not enter into the discussion. What is important is its dimensions in terms of fundamental units. Early in the 20th century the physicist Norman Campbell (1920/1957) used the distinction between fundamental and derived measurement in a sense more intrinsic to the theory of measurement itself. The distinction is the following: Fundamental measurement starts with qualitative statements (axioms) about empirical structures, such as those given earlier for an extensive structure, and then proves the existence of a representational theorem in terms of numbers, whence the phrase "representational measurement."

In contrast, a derived quantity is measured in terms of other fundamental measurements. A classical example is density, measured as the ratio of separate measurements of mass and volume. It is to be emphasized, of course, that calling density a derived measure with respect to mass and volume does not make a fundamental scientific claim. For example, it does not allege that fundamental measurement of density is impossible. Nevertheless, in understanding the foundations of measurement, it is always important to distinguish whether fundamental or derived measurement, in Campbell's sense, is being analyzed or used.

Axiomatic Geometry

From the standpoint of representational measurement theory, another development of great importance in the 19th century was the perfection of the axiomatic method in geometry, which grew out of the intense scrutiny of the foundations of geometry at the beginning of that century. The driving force behind this effort was undoubtedly the discovery and development of non-Euclidean geometries at the beginning of the century by Bolyai, Lobachevski, and Gauss. An important and intuitive example, later in the century, was Pasch's (1882) discovery of the axiom named in his honor. He found a gap in Euclid that required a new axiom, namely, the assertion that if a line intersects one side of a triangle, it must intersect also a second side. More generally, it was the high level of rigor and abstraction of Pasch's 1882 book that was the most important step leading to the modern formal axiomatic conception of geometry, which has been so much a model for representational measurement theory in the 20th century. The most influential work in this line of development was Hilbert's *Grundlagen der Geometrie,* first edition in 1899, much of its prominence resulting from Hilbert's position as one of the outstanding mathematicians of this period.

It should be added that even in one-dimensional geometry numerical representations arise even though there is no order relation. Indeed, for dimensions ≥ 2, no standard geometry has a weak order. Moreover, in geometry the continuum is not important for the fundamental Galilean and Lorentz groups. An underlying denumerable field of algebraic numbers is quite adequate.

Invariance

Another important development at the end of the 19th century was the creation of the explicit theory of invariance for spatial properties. The intuitive idea is that the spatial properties in analytical representations are invariant under the transformations that carry one model of the axioms into another model of the axioms. Thus, for example, the ordinary Cartesian representation of Euclidean geometry is such that the geometrical properties of the Euclidean space are invariant under the Euclidean group of transformations of the Cartesian representation. These are the transformations that are composed from translations (in any direction), rotations, and reflections. These ideas were made particularly prominent by the mathematician Felix Klein in his Erlangen address of 1872 (see Klein, 1893). These important concepts of invariance had a major impact in the development of the theory of special relativity by Einstein at the beginning of the 20th century. Here the invariance is that under the Lorentz transformations, which are those that leave invariant geometrical and kinematic properties of the space-time of special relativity. Without giving the full details of the Lorentz transformations, it is still possible to give a clear physical sense of the change from classical Newtonian physics to that of special relativity.

In the case of classical Newtonian mechanics, the invariance that characterizes the Galilean transformations is just the invariance of the distance between any two simultaneous points together with the invariance of any temporal interval, under any permissible change of coordinates. Note that this characterization requires that the units of measurement for both spatial distance and time be held constant. In the case of special relativity, the single invariant is what is called the *proper time* τ_{12} between two space-time points (x_1, y_1, z_1, t_1) and (x_2, y_2, z_2, t_2), which is defined as

$$\tau_{12} = \sqrt{(t_1 - t_2)^2 - \frac{1}{c^2}\left[(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2\right]},$$

where c is the velocity of light in the given units of measurement. It is easy to see the conceptual nature of the change. In the case of classical mechanics, the invariance of spatial distance between simultaneous points is separate from the invariance of temporal intervals. In the case of special relativity, they are intertwined. Thus, we properly speak of space-time invariance in the case of special relativity. As will be seen in what follows, the concepts of invariancc developed so thoroughly in the 19th and early 20th century in geometry and physics have carried over and are an important part of the representational theory of measurement.

Quantum Theory and the Problem of Measurement

Still another important development in the first half of the 20th century, of special relevance to the topic of this chapter, was the creation of quantum mechanics and, in particular, the extended analysis of the problem of measurement in that theory. In contrast with the long tradition of measurement in classical physics, at least three new problems arose that generated what in the literature is termed the *problem of measurement* in quantum mechanics. The first difficulty arises in measuring microscopic objects, that is, objects as small as atoms or electrons or other particles of a similar nature. The very attempt to measure a property of these particles creates a disturbance in the state of the particle, a disturbance that is not small relative to the particle itself. Classical physics assumed that, in principle, such minor disturbances of a measured object as did occur could either be eliminated or taken into account in a relatively simple way.

The second aspect is the precise limitation on such measurement, which was formulated by Heisenberg's unccrtainty principle. For example, it is not possible to measure both position and momentum exactly. Indeed, it is not possible, in general, to have a joint probability distribution of the measurements of the two. This applies not just to position and momentum, but also to other pairs of properties of a particle. The best that can be hoped for is the Heisenberg uncertainty relation. It expresses an inequality that bounds away from zero the product of the variances of the two properties measured, for example, the product of the variance of the measurement of position and the variance of the measurement of velocity or momentum. This inequality appeared really for the first time in quantum mechanics and is one of the principles that separates quantum mechanics drastically from classical physics. An accessible and clear exposition of these ideas is Heisenberg (1930), a work that few others have excelled for the quality of its exposition.

The third aspect of measurement in quantum mechanics is the disparity between the object being measured and the relatively large, macroscopic object used for the measurement. Here, a long and elaborate story can be told, as it is, for example, in von Neumann's classieal book on the foundations of quantum mechanics, which includes a detailed treatment of the measurement problem (von Neumann, 1932/1955). The critical aspect of this problem is deciding when a measurement has taken place. Von Neumann was inclined to the view that a measurement had taken place only when a relevant event had occurred in the consciousness of some observer. More moderate subsequent views are satisfied with the position that an observation takes place when a suitable recording has been made by a calibrated instrument.

Although we shall not discuss further the problem of measurement in quantum mechanics, nor even the application of the ideas to measurement in psychology, it is apparent that there is some resonance between the difficulties mentioned and the difficulties of measuring many psychological properties.

19th- and Early 20th-Century Psychology

Fechner's Psychophysics

Psychology was not a separate discipline until the late 19th century. Its roots were largely in philosophy with significant additions by medical and physical scientists. The latter brought a background of successful physical measurement, which they sought to re-create in sensory psychology at the least. The most prominent of these were H. Helmholtz, whose work among other things set the stage for extensive measurement, and G. T. Fechner, whose *Elemente der Psychophysik* (*Elements of Psychophysics*; 1860/1966) set the stage for subsequent developments in psychological measurement. We outline the problem faced in trying to transplant physical measurement and the nature of the proposed solution.

Recall that the main measurement device used in 19th-century physics was concatenation: Given two entities that exhibit the attribute to be measured, it was essential to find a method of concatenating them to form a third entity also exhibiting the attribute. Then one showed empirically that the structure satisfies the axioms of extensive measurement, as discussed earlier. When no empirical concatenation operation can be found, as for example with density, one could not do fundamental measurement. Rather, one sought an invariant property stated in terms of fundamentally measured quantities called derived measurement. Density is an example.

When dealing with sensory intensity, physical concatenation is available but just recovers the physical measure, which does not at all well correspond with subjective judgments such as the half loudness of a tone. A new approach was required. Fechner continued to accept the idea of building up a measurement scale by adding small increments, as in the standard sequences of extensive measurement, and then counting the number of such increments needed to fill a sensory interval. The question was: What are the small equal increments to be added? His idea was that they correspond to "just noticeable differences" (JND); when one first encounters the idea of a JND it seems to suggest a fixed threshold, but it gradually was interpreted to be defined statistically. To be specific, suppose x_0 and $x_1 = x_0 + \xi(x_0, \lambda)$ are stimulus intensities such that the probability of identifying x_1 as larger than x_0 is a constant λ, that is, $\Pr(x_0, x_1) = \lambda$. His idea was to fix λ and to measure the distance from x to y, $x < y$, in terms of the number of successive JNDs between them. Defining $x_0 = x$ and assuming that x_i has been defined, then define x_{i+1} as

$$x_{i+1} = x_i + \xi(x_i, \lambda).$$

The sequence ends with $x_n \leq y < x_{n+1}$. Fechner postulated the number of JNDs from x to y as his definition of distance without, however, establishing any empirical properties of justify that definition. Put another way, he treated without proof that a sequence of JNDs forms a standard sequence.

His next step was to draw on an empirical result of E. H. Weber to the effect that

$$\xi(x, \lambda) = \delta(\lambda)x, \quad \delta(\lambda) > 0,$$

which is called *Weber's law*. This is sometimes approximately true (e.g., for loudness of white noise well above absolute threshold), but more often it is not (e.g., for pure tones).

His final step was to introduce, much as in extensive measurement, a limiting process as λ approaches $\frac{1}{2}$ and δ approaches 0. He called this an auxiliary mathematical principle, which amounts to supposing without proof that a limit below exists. If we denote by ψ the counting function, then his assumption that, for fixed λ, the JNDs are equally distant can be interpreted to mean that for some function η of λ

$$\begin{aligned}\eta(\lambda) &= \psi[x + \xi(x, \lambda)] - \psi(x) \\ &= \psi([\delta(\lambda) + 1]x) - \psi(x).\end{aligned}$$

Therefore, dividing by $\delta(\lambda)x$

$$\frac{\psi([\delta(\lambda) + 1]x) - \psi(x)}{\delta(\lambda)x} = \frac{\eta(\lambda)}{\delta(\lambda)x} = \frac{\alpha(\lambda)}{x},$$

$$\text{where } \alpha(\lambda) = \frac{\eta(\lambda)}{\delta(\lambda)}.$$

Assuming that the limit of $\alpha(\lambda)$ exists, one has the simple ordinary differential equation

$$\frac{d\psi(x)}{dx} = \frac{k}{x}, \quad k = \lim_{\lambda \to \frac{1}{2}} \alpha(\lambda),$$

whose solution is well known to be

$$\psi(x) = r \ln x + s, \quad r > 0.$$

This conclusion, known as *Fechner's law*, was soon questioned by J. A. F. Plateau (1872), among others, although the emprical evidence was not conclusive. Later, Wiener (1915, 1921) was highly critical, and much later Luce and Edwards (1958) pointed out that, in fact, Fechner's mathematical auxiliary principle, although leading to the correct solution of the functional equation $\eta(\lambda) = \psi[x + \xi(x, \lambda)] - \psi(x)$ when Weber's law holds, fails to discover the correct solution in any other case—which empirically really is the norm. The mathematics is simply more subtle than he assumed.

In any event, note that Fechner's approach is not an example of representational measurement, because no empirical justification was provided for the definition of standard sequence used.

Reinterpreting Fechner Geometrically

Because Fechner's JND approach using infinitesimals seemed to be flawed, little was done for nearly half a century to construct psychophysical functions based on JNDs—that is, until Dzhafarov and Colonius (1999, 2001) reexamined what Fechner might have meant. They did this from a viewpoint of distances in a possible representation called a Finsler geometry, of which ordinary Riemannian geometry is a special case. Thus, their theory concerns stimuli of any finite dimension, not just one. The stimuli are vectors, for which we use bold-faced notation. The key idea, in our notation, is that for each person there is a universal function Φ such that, for λ sufficiently close to $\frac{1}{2}$, $\Phi(\psi[\mathbf{x} + \xi(\mathbf{x}, \lambda)] - \psi(\mathbf{x}))$ is comeasurable[3] with $\mathbf{x}$. This assumption means that this transformed differential can be integrated along any sufficiently smooth path between any two points. The minimum path length is defined to be the Fechnerian distance between them. This theory, which is mathematically quite elaborate, is testable in principle. But doing so certainly will not be easy because its assumptions, which are about the behavior of infinitesimals, are inherently difficult to check with fallible data. It remains to be seen how far this can be taken.

Ability and Achievement Testing

The vast majority of what is commonly called "psychological measurement" consists of various elaborations of ability and achievement testing that are usually grouped under the label "psychometrics." We do not cover any of this material because it definitely is neither a branch of nor a precursor to the representational measurement of an attribute. To be sure, a form of counting is employed, namely, the

[3]For the precise definition, see the reference.

items on a test that are correctly answered, and this number is statistically normed over a particular age or other feature so that the count is transformed into a normal distribution. Again, no axioms were or are provided. Of the psychometric approaches, we speak only of a portion of Thurstone's work that is closely related to sensory measurement. Recently, Doignon and Falmagne (1999) have developed an approach to ability measurement, called knowledge spaces, that is influenced by representational measurement considerations.

Thurstone's Discriminal Dispersions

In a series of three 1927 papers, L. L. Thurstone began a reinterpretation of Fechner's approach in terms of the then newly developed statistical concept of a random variable (see also Thurstone, 1959). In particular, he assumed that there was an underlying psychological continuum on which signal presentations are represented, but with variability. Thus, he interpreted the representation of stimulus x as a random variable $\Psi(x)$ with some distribution that he cautiously assumed (see Thurstone, 1927b, p. 373) to be normal with mean ψ_x and standard deviation (which he called a "discriminal dispersion") σ_x and possibly covariances with other stimulus representations. Later work gave reasons to consider extreme value distributions rather than the normal. His basic model for the probability of stimulus y being judged larger than x was

$$P(x, y) = \Pr[\Psi(y) - \Psi(x) > 0], \quad x \leq y. \tag{3}$$

The relation to Fechner's ideas is really quite close in that the mean subjective differences are equal for fixed $\lambda = P(x, y)$.

Given that the representations are assumed to be normal, the difference is also normal with mean $\psi_y - \psi_x$ and standard deviation

$$\sigma_{x,y} = \left(\sigma_x^2 + \sigma_y^2 - 2\rho_{x,y}\sigma_x\sigma_y\right)^{1/2}$$

so if $z_{x,y}$ is the normal deviate corresponding to $P(x, y)$, Equation (3) can be expressed as

$$\psi_y - \psi_x = z_{x,y}\sigma_{x,y}.$$

Thurstone dubbed this "a law of comparative judgment." Many papers before circa 1975 considered various modifications of the assumptions or focused on solving this equation for various special cases. We do not go into this here in part because the power of modern computers reduces the need for specialization.

Thurstone's approach had a natural one-dimensional generalization to the absolute identification of one of $n > 2$ possible stimuli. The theory assumes that each stimulus has a distribution on the sensory continuum and that the subject establishes $n - 1$ cut points to define the intervals of the range of the random variable that are identified with the stimuli. The basic data are conditional probabilities $P(x_j | x_i, n)$ of responding x_j when x_i, i, $j = 1, 2, \ldots, n$, is presented. Perhaps the most striking feature of such data is the following: Suppose a series of signals are selected such that adjacent pairs are equally detectable. Using a sequence of n adjacent ones, absolute identification data are processed through a Thurstone model in which $\psi_{x,n}$ and $\sigma_{x,n}$ are both estimated. Accepting that $\psi_{x,n}$ are independent of n, then the $\sigma_{x,n}$ definitely are not independent of n. In fact, once n reaches about 7, the value is independent of size, but $\sigma_{x,7} \approx 3\sigma_{x,2}$. This is a challenging finding and certainly casts doubt on any simple invariant meaning of the random variable $\Psi(x)$—apparently its distribution depends not only on x but on what might have been presented as well. Various authors have proposed alternative solutions (for a summary, see Iverson & Luce, 1998).

A sophisticated treatment of Fechner, Thurstone, and the subsequent literature is provided by Falmagne (1985).

Theory of Signal Detectability

Perhaps the most important generalization of Thurstone's idea is that of the theory of signal detectability, in which the basic change is to assume that the experimental subject can establish a response criterion β, in general different from 0, so that

$$P(x, y) = \Pr[\Psi(y) - \Psi(x) > \beta], \quad x \leq y.$$

Engineers first developed this model. It was adoped and elaborated in various psychological sources, including Green and Swets (1974) and Macmillan and Creelman (1991), and it has been widely applied throughout psychology.

Mid-20th-Century Psychological Measurement

Campbell's Objection to Psychological Measurement

N. R. Campbell, a physicist turned philosopher of physics who was especially concerned with physical measurement, took the very strong position that psychologists, in particular, and social scientists, in general, had not come up with anything deserving the name of measurement and probably never could. He was supported by a number of other British physicists. His argument, though somewhat elaborate, actually boiled down to asserting the truth of three simple propositions:

(i) A prerequisite of measurement is some form of empirical quantification that can be accepted or rejected experimentally.

(ii) The only known form of such quantification arises from binary operations of concatenation that can be shown empirically to satisfy the axioms of extensive measurement.

(iii) And psychology has no such extensive operations of its own.

Some appropriate references are Campbell (1920/1957, 1928) and Ferguson et al. (1940).

Stevens's Response

In a prolonged debate conducted before a subcommittee of the British Association for the Advancement of Sciences, the physicists agreed on these propositions and the psychologists did not, at least not fully. They accepted (iii) but in some measure denied (i) and (ii), although, of course, they admitted that both held for physics. The psychophysicist S. S. Stevens became the primary spokesperson for the psychological community. He first formulated his views in 1946, but his 1951 chapter in the first version of the *Handbook of Experimental Psychology,* of which he was editor, made his views widely known to the psychological community. They were complex, and at the moment we focus only on the part relevant to the issue of whether measurement can be justified outside physics.

Stevens' contention was that Proposition (i) is too narrow a concept of measurement, so (ii) and therefore (iii) are irrelevant. Rather, he argued for the claim that "Measurement is the assignment of numbers to objects or events according to rule. . . . The rule of assignment can be any consistent rule" (Stevens, 1975, pp. 46–47). The issue was whether the rule was sufficient to lead to one of several *scale types* that he dubbed nominal, ordinal, interval, ratio, and absolute. These are sufficiently well known to psychologists that we need not describe them in much detail. They concern the uniqueness of numerical representations. In the *nominal* case, of which the assignment of numbers to football players was his example, any permutation is permitted. This is not generally treated as measurement because no ordering by an attribute is involved. An *ordinal* scale is an assignment that can be subjected to any strictly increasing transformation, which of course preserves the order and nothing else. It is a representation with infinite degrees of freedom. An *interval* scale is one in which there is an arbitrary zero and unit; but once picked, no degrees of freedom are left. Therefore, the admissible transformation is $\psi \longmapsto r\psi + s$, $(r > 0)$. As stated, such a representation has to be on all of the real numbers. If, as is often the easc, especially in physics, one wants to place the representation on the positive real numbers, then the transformation becomes $\psi_+ \longmapsto s'\psi_+^r$, $(r > 0, s' > 0)$. Stevens (1959, pp. 31–34) called a representation unique up to power transformations a *log-interval* scale but did not seem to recognize that it is merely a different way of writing an interval scale representation ψ in which $\psi = \ln \psi_+$ and $s = \ln s'$. Whichever one uses, it has two degrees of freedom. The *ratio* case is the interval one with $r = 1$. Again, this has two forms depending on the range of ψ. For the case of a representation on the reals, the admissible transformations are the translations $\psi \longmapsto \psi + s$. There is a different version of ratio measurement that is inherently on the reals in the sense that it cannot be placed on the positive reals. In this case, 0 is a true zero that divides the representation into inherently positive and negative portions, and the admissible transformations are $\psi \longmapsto r\psi, r > 0$.

Stevens took the stance that what was important in measurement was its uniqueness properties and that they *could* come about in ways different from that of physics. The remaining part of his career, which is summarized in Stevens (1975), entailed the development of new methods of measurement that can all be encompassed as a form of sensory matching. The basic instruction to subjects was to require the match of a stimulus in one modality to that in another so that the subjective ratio between a pair of stimuli in the one dimension is maintained in the subjective ratio of the matched signals. This is called *cross-modal matching*. When one of the modalities is the real numbers, it is one of two forms of magnitude matching—*magnitude estimation* when numbers are to bc matched to a sensory stimuli and *magnitude production* when numbers are the stimuli to be matched by some physical stimuli. Using geometric means over subjects, he found the data to be quite orderly—power functions of the usual physical measures of intensity. Much of this work is covered in Stevens (1975).

His argument that this constituted a form of ratio scale measurement can be viewed in two distinct ways. The least charitable is that of Michell (1999), who treats it as little more than a play on the word "ratio" in the scale type and in the instructions to the subjects. He feels that Stevens failed to understand the need for empirical conditions to justify numerical representations. Narens (1996) took the view that Stevens' idea is worth trying to formalize and in the process making it empirically testable. Work along these lines continues, as discussed later.

REPRESENTATIONAL APPROACH AFTER 1950

Aside from extensive measurement, the representational theory of measurement is largely a creation by behavioral scientists and mathematicians during the second half of the 20th century. The basic thrust of this school of thought can be summarized as accepting Campbell's conditions (i), quantification based on empirical properties, and (iii), the social sciences do not have concatenation operations (although even that was never strictly correct, as is shown later, because of probability based on a partial operation), and rejecting the claim (ii) that the only form of quantification is an empirical concatenation operation. This school disagreed with Stevens' broadening of (i) to any rule, holding with the physicists that the rules had to be established on firm empirical grounds.

To do this, one had to establish the existence of schemes of empirically based measurement that were different from extensive measurement. Examples are provided here. For greater detail, see FM I, II, III, Narens (1985), or for an earlier perspective Pfanzagl (1968).

Several Alternatives to Extensive Measurement

Utility Theory

The first evidence of something different from extensive measurement was the construction by von Neumann and Morgenstern (1947) of an axiomatization of *expected utility theory*. Here, the stimuli were gambles of the form $(x, p; y)$ where consequence x occurs with probability p and y with probability $1 - p$. The basic primitive of the system was a weak preference order $\succsim$ over the binary gambles. They stated properties that seemed to be at least rational, if not necessarily descriptive; from them one was able to show the existence of a numerical utility function U over the consequences and gambles such that for two binary gambles g, h

$$g \succsim h \Leftrightarrow U(g) \geq U(h),$$
$$U(g, p; h) = U(g)p + U(h)(1 - p).$$

Note that this is an averaging representation, called *expected utility*, which is quite distinct from the adding of extensive measurement (see the subsection on averaging).

Actually, their theory has to be viewed as a form of derived measurement in Campbell's sense because the construction of the U function was in terms of the numerical probabilities built into the stimuli themselves. That limitation was overcome by Savage (1954), who modeled decision making under uncertainty as acts that are treated as an assignment of consequences to chance states of nature.[4] Savage assumed that each act had a finite number of consequences, but subsequent generalizations permitted infinitely many. Without building any numbers into the domain and using assumptions defended by arguments of rationality, he showed that one can construct both a utility function U and a subjective probability function S such that acts are evaluated by calculating the expectation of U with respect to the measure S. This representation is called *subjective expected utility* (SEU). It is a case of fundamental measurement in Campbell's sense. Indirectly, it involved a partial concatenation operation of disjoint unions, which was used to construct a subjective probability function.

These developments led to a very active research program involving psychologists, economists, and statisticians. The basic thrust has been of psychologists devising experiments that cast doubt on either a representation or some of its axioms, and of theorists of all stripes modifying the theory of accommodate the data. Among the key summary references are Edwards (1992), Fishburn (1970, 1988), Luce (2000), Quiggin (1993), and Wakker (1989).

Difference Measurement

The simplest example of difference measurement is location along a line. Here, some point is arbitrarily set to be 0, and other points are defined in terms of distance (length) from it, with those on one side defined to be positive and those on the other side negative. It is clear in this case that location measurement forms an example of interval scale measurement

[4] Some aspects of Savage's approach were anticipated by Ramsey (1931), but that paper was not widely known to psychologists and economists. Almost simultaneously with the appearance of Savage's work, Davidson, McKinsey, and Suppes (1955) drew on Ramsey's approach, and Davidson, Suppes, and Segal (1957) tested it experimentally.

that is readily reduced to length measurement. Indeed, all forms of difference measurement are very closely related to extensive measurement, but with the stimuli being pairs of elements (x, y) that define "intervals." Axioms can be given for this form of measurement where the stimuli are pairs (x, y) with both x, y in the same set X. The goal is a numerical representation φ of the form

$$(x, y) \succsim (u, v) \Leftrightarrow \varphi(x) - \varphi(y) \geq \varphi(u) - \varphi(v).$$

One key axiom that makes clear how a concatenation operation arises is that if $(x, y) \succsim (x', y')$ and $(y, z) \succsim (y', z')$, then $(x, z) \succsim (x', z')$.

An important modification is called *absolute difference measurement,* in which the goal is changed to

$$(x, y) \succsim (u, v) \Leftrightarrow |\varphi(x) - \varphi(y)| \geq |\varphi(u) - \varphi(v)|.$$

This form of measurement is a precursor to various ideas of similarity measurement important in multidimensional scaling. Here the behavioral axioms become considerably more complex. Both systems can be found in FM I, Chap. 4.

An important generalization of absolute difference measurement is to stimuli with n factors; it underlies developments of geometric measurement based on stimulus proximity. This can be found in FM II, Chap. 14.

Additive Conjoint Measurement

Perhaps the single development that most persuaded psychologists that fundamental measurement really could be different from extensive measurement consisted of two versions of what is called additive conjoint measurement. The first, by Debreu (1960), was aimed at showing economists how indifference curves could be used to construct cardinal (interval scale) utility functions. It was, therefore, naturally cast in topological terms. The second (and independent) one by Luce and Tukey (1964) was cast in algebraic terms, which seems more natural to psychologists and has been shown to include the topological approach as a special case. Again, it was an explanation of the conditions under which equal-attribute curves can give rise to measurement. Michell (1990) provides a careful treatment aimed at psychologists.

The basic idea is this: Suppose that an attribute is affected by two independent stimulus variables. For example, preference for a reward is affeeted by its size and the delay in receiving it; mass of an object is affected by both its volume and the (homogeneous) material of which it is composed; loudness of pure tones is affected by intensity and frequency; and so on. Formally, one can think of the two factors as distinct sets A and X, so an entity is of the form (a, x) where $a \in A$ and $x \in X$. The ordering attribute is $\succsim$ over such entities, that is, over the Cartesian product $A \times X$. Thus, $(a, x) \succsim (b, y)$ means that (a, x) exhibits more of the attribute in question than does (b, y). Again, the ordering is assumed to be a weak order: transitive and connected. Monotonicity (called independence in this literature) is also assumed: For $a, b \in A, x, y \in X$

$$\begin{aligned} &(a, x) \succsim (b, x) \Leftrightarrow (a, y) \succsim (b, y). \\ &(a, x) \succsim (a, y) \Leftrightarrow (b, x) \succsim (b, y). \end{aligned} \tag{4}$$

This familiar property is often examined in psychological research in which a dependent variable is plotted against, say, a measure of the first component with the second component shown as a parameter of the curves. The property holds if and only if the curves do not cross.

It is easy to show that this condition is not sufficient to get an additive representation of the two factors. If it were, then any set of nonintersecting curves in the plane could be rendered parallel straight lines by suitable nonlinear transformations of the axes. More is required, namely, the Thomsen condition, which arose in a mathematically closely related area called the theory of webs. Letting $\sim$ denote the indifference relation of $\succsim$, the *Thomsen condition* states

$$\left.\begin{array}{l}(a, z) \sim (c, y)\\(c, x) \sim (b, z)\end{array}\right\} \Rightarrow (a, x) \sim (b, y).$$

Note that it is a form of cancellation—of c in the first factor and z in the second.

These, together with an Arehimedean property establishing commensurability and some form of density of the factors, are enough to establish the following additive representation: There exist numerical functions ψ_A on A and ψ_X on X such that

$$\begin{aligned}&(a, x) \succsim (b, y)\\&\quad\Leftrightarrow \psi_A(a) + \psi_X(x) \geq \psi_A(b) + \psi_X(y).\end{aligned}$$

This representation is on all of the real numbers. A multiplicative version on the positive real numbers exists by setting $\xi_i = \exp \psi_i$. The additive representation forms an interval scale in the sense that ψ'_A, ψ'_X forms another equally good representation if and only if there are constants $r > 0, s_A, s_X$ such that

$$\begin{gathered}\psi'_A = r\psi_A + s_A,\\ \psi'_X = r\psi_X + s_X \Leftrightarrow \xi'_A = s'_A\xi^r_A, \quad \xi'_X = s'_X\xi^r_X,\\ s'_i = \exp s_i > 0.\end{gathered}$$

Additive conjoint measurement can be generalized to finitely many factors, and it is simpler in the sense that if monotonicity is generalized suitably and if there are at least three factors, then the Thomsen condition can be derived rather than assumed.

Although no concatenation operation is in sight, a family of them can be defined in terms of $\sim$, and they can be shown to satisfy the axioms of extensive measurement. This is the nature of the mathematical proof of the representation usually given.

Averaging

Some structures with a concatenation operation do not have an additive representation, but rather a weighted averaging representation of the form

$$\varphi(x \circ y) = \varphi(x)w + \varphi(y)(1 - w), \quad (5)$$

where the weight w is fixed. We have already encountered this form in the utility system if we think of the gamble $(x, p; y)$ as defining operations $\circ_p$ with $x \circ_p y \equiv (x, p; y)$, in which case $w = w(p)$. A general theory of such operations was first given by Pfanzagl (1959). It is much like extensive measurement but with associativity replaced by *bisymmetry:* For all stimuli x, y, u, v,

$$(x \circ y) \circ (u \circ v) \sim (x \circ u) \circ (y \circ v). \quad (6)$$

It is easy to verify that the weighted-average representation of Equation (5) implies bisymmetry, Equation (6), and $x \circ x \sim x$. The easiest way to show the converse is to show that defining $\succsim'$ over $X \times X$ by

$$(a, x) \succsim' (b, y) \Leftrightarrow a \circ x \succsim b \circ y$$

yields an additive conjoint structure, from which the result follows rather easily.

Nonadditive Representations

A natural question is: When does a concatenation operation have a numerical representation that is inherently nonadditive? By this, one means a representation for which no strictly increasing transformation renders it additive. Before exploring that, we cite an example of nonadditive representations that can in fact be transformed into additive ones. This is helpful in understanding the subtlety of the question.

One example that has arisen in utility theory is the representation

$$U(x \oplus y) = U(x) + U(y) - \delta U(x)U(y), \quad (7)$$

where δ is a real constant and U is the SEU or rank-dependent utility generalization (see

Luce, 2000, Chap. 4) with an intrinsic zero—no change from the status quo. Because Equation (7) can be rewritten

$$1 - \delta U(x \oplus y) = [1 - \delta U(x)][1 - \delta U(y)],$$

the transformation $V = -\kappa \ln(1 - \delta U)$, $\delta\kappa > 0$, is additive, that is, $V(x \oplus y) = V(x) + V(y)$, and order-preserving. The measure V is called a *value function*. The form in Equation (7) is called *p-additive* because it is the only polynomial with a fixed zero that can be put in additive form. The source of this representation is examined in the next major section. It is easy to verify that both the additive and the nonadditive representations are ratio scales in Stevens' sense. We know from extensive measurement that the change of unit in the additive representation is somehow reflecting something important about the underlying structure. Is that also true of the changes of units in the nonadditive representation? We will return to this point, which can be a source of confusion.

It should be noted that in probability theory for independent events, the p-additive form with $\delta = 1$ arises since

$$P(A \cup B) = P(A) + P(B) - P(A)P(B).$$

An earlier, similar example concerning velocity concatenation arose in Einstein's theory of special relativity. Like the psychological one, it entails a representation in the standard measure of velocity that forms a ratio scale and a nonlinear transformation to an additive one that also forms a ratio scale. We do not detail it here.

Nonadditive Concatenation

What characterizes an inherently nonadditive structure is the failure of the empirical property of associativity; that is, for some elements x, y, z in the domain,

$$x \circ (y \circ z) \not\sim (x \circ y) \circ z.$$

Cohen and Narens (1979) made the then-unexpected discovery that if one simply drops associativity from any standard axiomatization of extensive measurement, not only can one still continue to construct numerical representations that are onto the positive reals but, quite surprisingly, they continue to form a ratio scale as well; that is, the representation is unique up to similarity transformations. They called this important class of nonadditive representations *unit structures*. For a full discussion, see Chaps. 19 and 20 of FM III.

A Fundamental Account of Some Derived Measurement

Distribution Laws

The development of additive conjoint measurement allows one to give a systematic and fundamental account of what to that point had been treated as derived measurement. For classical physics, a typical situation in which derived measurement arises takes the form $\langle A \times X, \succsim, \circ_A\rangle$. For example, let A denote a set of volumes and X a set of homogeneous substances; the ordering is that of mass as established by an equal-arm pan balance in a vacuum. The operation $\circ_A$ is the simple union of volumes. For this case we know that $m = V\rho$, where m is the usual mass measure, V is the usual volume measure, and ρ is an inferred measure of density.

Observe that $\langle A \times X, \succsim\rangle$ forms an additive conjoint structure. By the monotonicity assumption of conjoint measurement, Equation (4), $\succsim$ induces the weak order $\succsim_A$ on A. It is assumed that $\langle A, \succsim_A, \circ_A\rangle$ forms an extensive structure. Thus we have the extensive representation φ_A of $\langle A, \succsim_A, \circ_A\rangle$ onto the positive real numbers and a multiplicative conjoint one $\xi_A\xi_X$ of $\langle A \times X, \succsim\rangle$ onto the positive real numbers.

The question is how φ_A and ξ_A relate. Because both preserve the order $\succsim_A$, there must be a strictly increasing function F such that $\xi_A = F(\varphi_A)$. Beyond that, we can say nothing without some assumption describing how the two structures interlock. One that holds for many physical cases, including the motivating mass example, is a qualitative *distribution law* of the form: For all a, b, c, d in A and x, y in X,

$$\left.\begin{array}{l}(a,x)\sim(c,y)\\(b,x)\sim(d,y)\end{array}\right\}\Rightarrow(a\circ_A b, x)\sim(c\circ_A d, y).$$

Using this, one is able to prove that, for some $r > 0, s > 0, F(z) = rz^s$. Because the conjoint representation is unique up to power transformations, we may select $s = 1$, that is, choose $\xi_A = \varphi_A$.

Note that distribution is a substantive, empirical property that in each specific case requircs verification. In fact, it holds for many of the classical physical attributes. From that fact one is able to construct the basic structure of (classical) physical quantities that underlies the technique called *dimensional analysis,* which is widely used in physical applications in engineering. It also accounts for the fact that physical units are all expressed as products of powers of a relatively small set of units. This is discussed in some detail in Chap. 10 of FM I and in a more satisfactory way in Section 22.7 of FM III.

Segregation Law

Within the behavioral sciences we have a situation that is somewhat similar to distribution. Suppose we return to the gambling structure, where some chance "experiment" is performed, such as drawing a ball from an urn with 100 red and yellow balls of which the respondent knows that the number of red is between 50 and 80. A typical binary gamble is of the form $(x, C; y)$, where C denotes a chance event such as drawing a red ball, and the consequence x is received if C occurs and y otherwise, that is, x if a red ball and y if a yellow ball. A weak preference order $\succsim$ over gambles is postulated. Let us distinguish gains from losses by supposing that there is a special consequence, denoted e, that means no change from the status quo. Things preferred to e are called gains, and those not preferred to it are called losses. Assume that for gains (and separately for losses) the axioms leading to a subjective expected utility representation are satisfied. Thus, there is a utility function U over gains and subjective probability function S such that

$$U(x, C; y) = U(x)S(C) + U(y)[1 - S(C)] \quad (8)$$

$$U(e) = 0. \quad (9)$$

Let $\oplus$ denote the operation of rceiving two things, called *joint receipt.* Therefore, $g \oplus h$ denotes receiving both of the gambles g and h. Assume that $\oplus$ is a commutative[5] and monotonic operation with e the identity element; that is, for all gambles g perceived as a gain, $g \oplus e \sim g$. Again, some law must link $\oplus$ to the gambles. The one proposed by Luce and Fishburn (1991) is *segregation:* For all gains x, y,

$$(x, C; e) \oplus y \sim (x \oplus y, C; y). \quad (10)$$

Observe that this is highly rational in the sense that both sides yield $x \oplus y$ when C occurs and y otherwise, so they should be seen as representing the same gamble. Moreover, there is some empirical evidence in support of it (Luce, 2000, Chap. 4). Despite its apparent innocence, it is powerful enough to show that $U(x \oplus y)$ is given by Equation (7). Thus, in fact, the operation $\oplus$ forms an extensive structure with additive representation $V = -\kappa \ln(1 - \delta U)$, $\delta\kappa > 0$. Clearly, the sign of δ greatly affects the relation between U and V: it is a negative exponential for $\delta > 0$, proportional for $\delta = 0$, and an exponential for $\delta < 0$.

[5]Later we examine what happens when we drop this assumption.

Applications of these ideas are given in Luce (2000). Perhaps the most interesting occurs when dealing with $x \oplus y$ where x is a gain and y a loss. If we assume that V is additive throughout the entire domain, then with $x \succsim e \succsim y$, $U(x \oplus y)$ is not additive. This carries through to mixed gambles that no longer have the simple bilinear form of binary SEU, Equation (8).

Invariance and Meaningfulness

Meaningful Statistics

Stevens (1951) raised the following issues in connection with the use of statistics on measurements. Some statistical assertions do not seem to make sense in some measurement schemes. Consider a case of ordinal measurement in which one set of three observations has ordinal measures 1, 4, and 5, with a mean of 10/3, and another set has measures 2, 3, and 6, with a mean of 11/3. One would say the former set is, on average, smaller than the second one. But since these are ordinal data, an equally satisfactory representation is 1, 5, and 6 for the first set and 2, 2.1, and 6.1 for the latter, with means respectively 12/3 and 10.2/3, reversing the conclusion. Thus, there is no invariant conclusion about means. Put another way, comparing means is meaningless in this context. By contrast, the median is invariant under monotonic transformations. It is easy to verify that the mean exhibits suitable invariance in the case of ratio scales.

These observations were immediately challenged and led to what can best be described as a tortured discussion that lasted many years. It was only clarified when the problem was recognized to be a special case of invariance principles that were well developed in both geometry and dimensional analysis.

The main reason why the discussion was confused is that it was conducted at the level of numerical representations, where two kinds of transformations are readily confused, rather than in terms of the underlying structure itself. Consider a cubical volume that is 4 yards on a side. An appropriate change of units is from yards to feet, so it is also 12 feet on a side. This is obviously different from the transformation that enlarges each side by a factor of 3, producing a cube that is 12 yards on a side. At the level of numerical representations, however, these two factor-of-3 changes are all too easily confused. This fact was not recognized when Stevens wrote, but it clearly makes very uncertain just what is meant by saying that a structure has a ratio or other type of representation and that certain invariances should hold.

Automorphisms

These observations lead one to take a deeper look into questions of uniqueness and invariance. Mapping empirical structures onto numerical ones is not the most general or fundamental way to approach invariance. The key to avoiding confusion is to understand what it is about a structure that corresponds to correct admissible transformations of the representation. This turns out to be isomorphisms that map an empirical structure onto itself. Such isomorphisms are called *automorphisms* by mathematicians and *symmetries* by physicists. Their importance is easily seen, as follows. Suppose α is an automorphism and f is a homomorphism of the structure into a numerical one, then it is not difficult to show that $f * \alpha$, where $*$ denotes function composition, is another equally good homomorphism into the same numerical structure. In the case of a ratio scale, this means that there is a positive numerical constant r_α such that $f * \alpha = r_\alpha f$. The automorphism captures something about the structure itself, and that is just what is needed.

Consider the utility example, Equation (7), where there are two nonlinearly related representations, both of which are ratio scales in Stevens' sense. Thus, calculations of the mean utility are invariant in any one representation, but they certainly are not across representations. Which should be used, if either? It turns out on careful examination that the one set of transformations corresponds to the automorphisms of the underlying extensive structure. The second set of transformations corresponds to the automorphisms of the SEU structure, not $\oplus$. Both changes are important, but different. Which one should be used depends on the question being asked.

Invariance

An important use of automorphisms, first emphasized for geometry by Klein (1872/1893) and heavily used by physicists and engineers in the method of dimensional analysis, is the idea that meaningful statements should be invariant under automorphisms. Consider a structure with various primitive relations. It is clear that these are invariant under the automorphisms of the structure, and it is natural to suppose that anything that can be meaningfully defined in terms of these primitives should also be invariant. Therefore, in particular, given the structure of physical attributes, any physical law is defined in terms of the attributes and thus must be invariant. This definitely does not mean that something that is invariant is necessarily a physical law. In the case of statistical analyses of measurements, we want the result to exhibit invariance appropriate to the structure underlying the measurements.

To answer Stevens' original question about statistics then entails asking whether the hypothesis being tested is meaningful (invariant) when translated back into assertions about the underlying structure. Doing this correctly is sometimes subtle, as is discussed in Chap. 22 of FM III and much more fully by Narens (2001).

Trivial Automorphisms and Invariance

Sometimes structures have but one automorphism, namely the function that maps each element of the structure into itself—the identity function. For example, in the additive structure of the natural numbers with the standard ordering, the only automorphism is the one that simply matches each number to itself: 0 to 0, 1 to 1, and so on.

Within the weak ordering $\succsim$ of a structure, there are trivial automorphisms beyond the identity mapping, namely, those that just map an element a to an equivalent element b; that is, the relation $a \sim b$ holds.

Consider invariance in such structures. We quickly see that the approach cannot yield any significant results because everything is invariant. This remark applies to all finite structures that are provided with a weak ordering. Thus, the only possibility is to examine the invariant properties of the structure of the set of numerical representations.

Let a finite empirical structure be given with a homomorphism f mapping the structure into a numerical structure. We have already introduced the concept of an admissible numerical transformation φ of f, namely, a one-one transformation of the range of f onto a possibly different set of real numbers, such that $\varphi * f$ is a homomorphism of the empirical structure. In order to fix the scale type and thus the nature of the invariance of the empirical structure, we investigate the set of all such homomorphisms for a given empirical structure. In the case of weight, any two homomorphisms f_1 and f_2 are related by a positive similarity transformation; that is, there is a positive real number $r > 0$ such that $rf_1 = f_2$. In the qualitative probability case with independence, $r = 1$, so the set of all homomorphisms has only one element. With $r \neq 1$ in the general similarity case, invariance is then characterized with respect to the multiplicative group of positive real numbers, each number in the group constituting a change of unit. A numerical statement about a set of numerical quantities is then *invariant* if and only if its truth value is constant under any changes of unit of any of the quantities. This definition is easily generalized to other groups of numerical transformations such as linear transformations for interval scales.

In contrast, consider a finite difference structure with a numerical representation as characterized earlier. In general, the set of all homomorphisms from the given finite structure to numerieal representations has no natural and simple mathematical characterization. For this reason, much of the general theory of representational measurement is concemed with empirical structures that map onto the full domain of real numbers. It remains true, however, that special finite empirical structures remain important in practice in setting up standard measurement procedures using well-defined units.

Covariants

In practice, physicists hold on to invariance by introducing and using the concept of covariants. Typical examples of such covariants are velocity and acceleration, neither of which is invariant from one coordinate frame to another under either Galilean or Lorentzian transformations, because, among other things, the direction of the velocity or acceleration vector of a particle will in general change from one frame to another. (The scalar magnitude of acceleration is invariant.)

The laws of physics are written in terms of such covariants. The fundamental idea is conveyed by the following. Let $Q_1, \ldots, Q_n$ be quantities that are functions of the space-time coordinates, with some Q_is possibly being derivatives of others, for example. Then, in general, as we go from one coordinate system to another (note that $'$ does not mean derivative) $Q'_1, \ldots, Q'_n$ will be covariant, rather than invariant, so their mathematical form is different in the new coordinate system. But any physical law involving them, say,

$$F(Q_1, \ldots, Q_n) = 0, \qquad (11)$$

must have the same form

$$F(Q'_1, \ldots, Q'_n) = 0$$

in the new coordinate frame. This same form is the important invariant requirement.

A simple example from classical mechanics is the conservation of momentum of two particles before and after a collision. Let v_i denote the velocity before and w_i the velocity after the collision, and m_i the mass, $i = 1, 2$, of each particle. Then the law, in the form of Equation (11), looks like this:

$$v_1 m_1 + v_2 m_2 - (w_1 m_1 + w_2 m_2) = 0,$$

and its transformed form will be, of course,

$$v'_1 m_1 + v'_2 m_2 - (w'_1 m_1 + w'_2 m_2) = 0,$$

but the forms of v_i and w_i will be, in general, covariant rather than invariant.

An Account of Stevens' Scale-Type Classification

Narens (1981a, 1981b) raised and partially answered the question of why the Stevens' classification into ratio, interval, and ordinal seales makes as much sense as it seems to. His result was generalized by Alper (1987), as described later. The question may be cast as follows: These familiar scale types have, respectively, one, two, and infinitely many degrees of freedom in the representation; are there not any others, such as ones having three or 10 degrees of freedom? To a first approximation, the answer is "no," but the precise answer is somewhat more complex than that.

To arrive at a suitable formulation, a special case may be suggestive. Consider a structure that has representations onto the reals—*continuous representations*—that form an interval scale. Then the representation has the following two properties. First, given numbers $x < y$ and $u < v$, there is a positive affine transformation that takes the pair (x, y) into

(u, v). It is found by setting $u = rx + s$, $v = ry + s$, whence $r = \frac{v-u}{y-x}$ and $s = \frac{yu-xv}{y-x}$. Thus, in terms of automorphisms we have the property that there exists one that maps any ordered pair of the structure into any other ordered pair. This is called *two-point homogeneity*. Equally well, if two affine transformations map a pair into the same pair, then they are identical. This follows from the fact that two equations uniquely determine r and s. In terms of automorphisms, this is called 2-*point uniqueness*. The latter can be recast by saying that any automorphism having two fixed points must be the identity automorphism.

In like manner, the ratio scale case is 1-point homogeneous and 1-point unique. The generalizations of these concepts to M—point homogeneity and N—point uniqueness are obvious. Moreover, in the continuous case it is easy to show that $M \leq N$. The question addressed by Narens was: Given that the structure is at least 1-point homogeneous and N—point unique for some finite N, what are the possibilities for (M, N)? Assuming $M = N$ and a continuous structure, he showed that the only possibilities are (1, 1) and (2, 2), that is, the ratio and interval scales. Alper (1987) dropped the condition that $M = N$ and showed that $(1, 2)$ can also occur, but that is the only added possibility. In terms of numerical representations on all of the real numbers, the (1, 2) transformations are of the form $x \longmapsto rx + s$ where s is any real and r is in some proper, nontrivial subgroup of the multiplicative, positive real group. One example is when r is of the form k^n, where $k > 0$ is fixed and n ranges over the positive and negative integers.

This result makes clear two things. First, we see that there can be no continuous scales between interval and ordinal, which of course is not finitely unique. Second, there are scales between ratio and interval. None of these has yet played a role in actual scientific measurement. Thus, for continuous structures Stevens' classification was almost complete, but not quite.

The result also raises some questions. First, how critical is the continuum assumption? The answer is "very": Cameron (1989) showed that nothing remotely like the Alper-Narens result holds for representations on the rational numbers. Second, what can be said about nonhomogeneous structures? Alper (1987) classified the $M = 0$ case, but the results are quite complex and apparently not terribly useful. Luce (1992) explored empirically important cases in which homogeneity fails very selectively. It does whenever there are *singular points*, which are defined to be points of the structure that remain fixed under all automorphisms. Familiar examples are 0 in the nonnegative, multiplicative real numbers and infinity if, as in relativistic velocity, it is adjoined to the system. For a broad class of systems, he showed that if a system has finitely many singular points and is homogeneous between adjacent ones, then there are at most three singular points—a minimum, an interior, and a maximum one. The detailed nature of these fixed points is somewhat complicated and is not discussed here. One specific utility structure with an interior singular point—an inherent zero—is explored in depth in Luce (2000).

Models of Stevens' Magnitude Methods

Stevens' (1975) empirical findings, which were known in the 1950s, were a challenge to measurement theorists. What underlies the network of (approximate) power function relations among subjective measures? Luce (1959) attempted to argue in terms of representations that if, for example, two attributes are each continuous ratio scales,[6] with typical physical representations φ_1 and φ_2, then

[6]Scale types other than ratio were also studied by Luce and subsequent authors.

a matching relation M between them should exhibit an invariance involving an admissible ratio-scale change of the one attribute corresponding under the match to a ratio-scale change of the other attribute, that is, $M[r\varphi_1(x)] = \alpha(r)\varphi_2(x)$. From this it is not difficult to prove that M is a power function of its argument. A major problem with this argument is its failure to distinguish two types of ratio scale transformations—changes of unit, such as centimeters to meters—and changes of scale, such as increasing the linear dimensions of a volume by a factor of three. Rozeboom (1962) was very critical of this failure. Luce (1990) reexamined the issue from the perspective of automorphisms. Suppose M is an empirical matching relation between two measurement structures, and suppose that for each translation (i.e., an automorphism with no fixed point) τ of the first structure there corresponds to a translation σ_τ of the second structure such that for any stimulus x of the first structure and any s of the second, then xMs holds if and only if for each automorphism τ of the first structure $\tau(x)M\sigma_\tau(s)$ also holds. This assumption, called *translation consistency,* is an empirically testable property, not a mere change of units. Assuming that the two structures have ratio scale representations, this property is equivalent to a power function relation between the representations.

Based on some ideas of R. N. Shepard, circulated privately and later modified and published in 1981, Krantz (1972) developed a theory that is based on three primitives: magnitude estimates, ratio estimates, and cross-modal matches. Various fairly simple, testable axioms were assumed that one would expect to hold if the psychophysical functions were power functions of the corresponding physical intensity and the ratios of the instructions were treated as mathematical ratios. These postulates were shown to yield the expected power function representations except for an arbitrary increasing function. This unknown function was eliminated by assuming, without a strong rationale, that the judgments for one continuum, such as length judgments, are veridical, thereby forcing the function to be a simple multiplicative factor. This model is summarized in Falmagne (1985, pp. 309–313). A somewhat related approaeh was offered by Falmagne and Narens (1983), also summarized in Falmagne (1985, pp. 329–339). It is based not on behavioral axioms, but on two invariance principles that they call meaningfulness and dimensional invariance. Like the Krantz theory, it too leads to the form $G(\varphi_i^{r_i}\varphi_j^{r_j})$, where G is unspecified beyond being strictly increasing.

Perhaps the deepest published analysis of the problem so far is Narens (1996). Unlike Stevens, he carefully distinguished numbers from numerals, noting that the experimental structure involved numerals whereas the scientists' representations of the phenomena involved numbers. He took seriously the idea that internally people are carrying out the ratio-preservation calculations embodied in Stevens' instructions. The upshot of Narens' axioms, which he carefully partitioned into those that are physical, those that are behavioral, and those that link the physical and the behavioral, was to derive two empirical predictions from the theory. Let $(x, \mathbf{p}, y)$ mean that the experimenter presents stimulus x and the numeral $\mathbf{p}$ to which the subject produces stimulus y as holding the p relation to x. So if $\mathbf{2}$ is given, then y is whatever the subject feels is twice x. The results are, first, a commutativity property: Suppose that the subject yields $(x, \mathbf{p}, y)$ and $(y, \mathbf{q}, z)$ when done in that order and $(x, \mathbf{q}, u)$ and $(u, \mathbf{p}, v)$ when the numerals are given in the opposite order. The prediction is $z = v$. A second result is a multiplicative one: Suppose $(x, \mathbf{pq}, w)$, then the prediction is $w = z$. It is clear that the latter property implies the former, but not conversely. Empirical data reported by Ellemeirer and Faulhammer

(2000) sustain the former prediction and unambiguously reject the latter.

Luce (2001) provides a variant axiomatic theory, based on a modification of some mathematical results summarized in Luce (2000) for utility theory. The axioms are formulated in terms of threc primitives: a sensory ordering $\succsim$ over physical stimuli varying in intensity, the joint presentation $x \oplus y$ of signals x and y (e.g., the presentation of pure tones of the same frequency and phase to the two ears), and for signals $x > y$ and positive number p denote by $z = (x, p, y)$ the signal that the subject judges makes interval $[y, z]$ stand in proportion p to interval $[y, x]$. The axioms, such as segregation, Equation (10), are behavioral and structural, and they are sufficient to ensure the existence of a continuous psychophysical mcasure ψ from stimuli to the positive real numbers and a continuous function W from the positive reals onto the positive reals and a constant $\delta > 0$ such that for $\oplus$ commutative

$$x \succsim y \Leftrightarrow \psi(x) \geq \psi(y), \qquad (12)$$

$$\psi(x \oplus y) = \psi(x) + \psi(y) + \delta\psi(x)\psi(y) \quad (\delta > 0), \quad (13)$$

$$\psi(x, p, y) - \psi(y) = W(p)[\psi(x) - \psi(y)]. \qquad (14)$$

We have written Equation (14) in this fashion rather than in a form comparable to the SEU equation for two reasons: It corresponds to the instructions given thc respondents, and $W(p)$ is not restricted to $[0, 1]$. Recent, currently unpublished, psychophysical data of R. Steingrimsson showed an important ease of $\oplus$ (two-ear loudness summation) that is rarely, if ever, commutative. This finding motivatcd Aczél, Lucc, and Ng (2001) to explore the noncommutative, nonassociative cases on the assumption $\oplus$ has a unit representation (mentioned earlier) and assuming Equations (12) and (14) and that certain unknown functions are differentiable. To everyone's surprise, the only new representations replacing (13) are either

$$\psi(x \oplus y) = \alpha\psi(x) + \psi(y), \quad (\alpha > 1)$$

when $x \oplus 0 \succ 0 \oplus x$, or

$$\psi(x \oplus y) = \psi(x) + \alpha'\psi(y), \quad (\alpha' > 1)$$

when $x \oplus 0 \prec 0 \oplus x$. These are called *left-* and *right-weighted additive* forms, respectively. These representations imply that some fixed dB correction can compensate the noncommutativity. Empirical studies evaluating this are underway.

One invariance condition limits the form of ψ to the exponential of a power function of deviations from absolute threshold, and another one limits the form of W to two parameters for $p \geq 1$ and two more for $p < 1$.

The theory not only is able to accommodate thc Ellemeirer and Faulhammer data but also predicts that the psychophysical function is a power function when $\oplus$ is not commutative and only approximately a power function for $\oplus$ commutative. Over eight or morc orders of magnitude, it is extremely close to a power function except near threshold and for very intense signals. Despite its not being a pure power function, the prcdictions for cross-modal matches are pure power functions.

Errors and Thresholds

To describe the general sources of errors and why they are inevitable in scientific work, we can do no better than quote the opening passage in Gauss's famous work on the theory of least squares, which is from the first part presented to the Royal Society of Göttingen in 1821:

> However much care is taken with observations of the magnitude of physical quantities, they are necessarily subject to more or less considerable errors. These errors, in the majority of cases, are not simple, but arise simultaneously from several distinct sources which it is convenient to distinguish into two classes.

Certain causes of errors depend, for each observation, on circumstances which are variable and independent of the result which one obtains: the errors arising from such sources are called *irregular* or *random*, and like the circumstances which produce them, their value is not susceptible of calculation. Such are the errors which arise from the imperfection of our senses and all those which are due to irregular exterior causes, such as, for example, the vibrations of the air which make vision less clear; some of the errors due to the inevitable imperfection of the best instruments belong to the same category. We may mention, for example, the roughness of the inside of a level, the lack of absolute rigidity, etc.

On the other hand, there exist causes which in all observations of the same nature produce an identical error, or depend on circumstances essentially connected with the result of the observation. We shall eall the errors of this category *constant* or *regular*.

It is evident that this distinction is relative up to a certain point and depends on how broad a sense one wishes to attach to the idea of observations of the same nature. For instance, if one repeats indefiniiely the measurement of a single angle, the errors arising from an imperfect division of the circular scale will belong to the class of constant errors. If, on the other hand, one measures successively several different angles, the errors due to the imperfection of the division will be regarded as random as long as one has not formed the table of errors pertaining to each division. (Gauss, 1821/1957, pp. 1–2)

Although Gauss had in mind problems of errors in physical measurement, it is quite obvious that his conceptual remarks apply as well to psychological measurement and, in fact, in the second paragraph refer directly to the "imperfection of our senses." It was really only in the 19th century that, even in physics, systematic and sustained attention was paid to quantitative problems of errors. For a historieal overview of the work preceding Gauss, see Todhunter (1865/1949). As can be seen from the references in the section on 19th- and early 20th-century psychology, quantitative attention to errors in psychological measurement began at least with Fechner in the second half of the 19th century. Also, as already noted, the analysis of thresholds in probabilistic terms really began in psychology with the cited work of Thurstone. However, the quantitative and mathematical theory of thresholds was discussed earlier by Norbert Wiener (1915, 1921). Wiener's treatment was, however, purely algebraic, whereas in terms of providing relatively direct methods of application, Thurstone's approach was entirely probabilistic in character. Already, Wiener (1915) stated very clearly and explicitly how to deal with the fact that with thresholds in perception, the relation of indistinguishability—whether we are talking about brightness of light, loudness of sound, or something similar—is not transitive.

The detailed theory was then given in the 1921 paper for construeting a measure up to an interval scale for such sensation-intensities. This is, without doubt, the first time that these important psychological matters were dealt with in rigorous detail from the standpoint of passing from qualitative judgments to a measurement representation. Here is the passage with which Wiener ends the 1921 paper:

In conclusion, let us consider what bearing all this work of ours can have on experimental psychology. One of the great defects under which the latter science at present labours is its propensity to try to answer questions without first trying to find out just what they ask. The experimental investigation of Weber's law[7] is a case in point: what most experimenters do take for granted before they begin their experiments is infinitely more important and interesting than any results to which their experiments lead. One of these unconscious assumptions is that sensations or sensation-intervals can be measured,

[7] Wiener means what is now called Fechner's logarithmic law.

> and that this process of measurement can be carried out in one way only. As a result, each new experimenter would seem to have devoted his whole energies to the invention of a method of procedure logically irrelevant to everything that had gone before: one man asks his subject to state when two intervals between sensations of a given kind appear different; another bases his whole work on an experiment where the observer's only problem is to divide a given colour-interval into two equal parts, and so on indefinitely, while even where the experiments are exactly alike, no two people choose quite the same method for working up their results. Now, if we make a large number of comparisons of sensation-intervals of a given sort with reference merely to whether one seems larger than another, the methods of measurement given in this paper indicate perfectly unambiguous ways of working up the results so as to obtain some quantitative law such as that of Weber without introducing such bits of mathematical stupidity as treating a "just noticeable difference" as an "infinitesimal," and have the further merit of always indicating *some* tangible mathematical conclusion, no matter what the outcome of the comparisons may be. (pp. 204–205)

The later and much more empirical work of Thurstone, already rcfcrred to, did not, however, give a representational theory of measurement as Wiener, in fact, in his own way did.

The work over the last few decades on errors and thresholds from the standpoint of representation theory of measurement naturally falls into two parts. The first part is the algebraic theory, and the second is the probabilistic theory. We first survey the algebraic results.

Algebraic Theory of Thresholds

The work following Wiener on algebraic thresholds was only revived in the 1950s and may be found in Goodman (1951), Halphen (1955), Luce (1956), and Scott and Suppes (1958). The subsequent literature is revicwed in some detail in FM II, Chap. 16. We follow the exposition of the algebraic ordinal theory there. We restrict ourselves here to finite semiorders, the concept first introduced axiomatically by Luce and in a modified axiomatization by Scott and Suppes.

Let A be a nonempty set, and let $\succ$ be a binary irreflexive relation on A. Then, $(A\ \succ)$ is a *semiorder* if for every a, b, c, and d in A

(i) If $a \succ c$ and $b \succ d$, then either $a \succ d$ or $b \succ c$.

(ii) If $a \succ b$ and $b \succ c$, then either $a \succ d$ or $d \succ c$.

For finite semiorders $(A, \succ)$ we can prove the following numcrical representational theorem with constant threshold, which in thc present case we will fix at 1, so the theorem asserts that thcre is a mapping f of A into the positive real numbers such that for any a and b in A,

$$a \succ b \quad \text{iff} \quad f(a) > f(b) + 1.$$

A wealth of more detailed and more delicate results on semiorders is to be found in Section 2 of Chap. 16 of FM II, and research continues on scmiorders and various gencralizations of them, such as interval orders.

Axioms extending the ordinal theory of semiorders to the kind of thing analyzed by Wiener (1921) are in Gerlach (1957); unfortunately, to obtain a full interval-scale reprcsentation with thresholds involves very complicated axioms. This is true to a lesser extent of the axioms for semiordered qualitativc probability structurcs given in Section 16.6.3 of FM II. The axioms are complicated when stated strictly in terms of the relation $\succ$ of semiorders.

Probabilistic Theory of Thresholds

For applications in experimental work, it is certainly the case that the probabilistic theory of thresholds is more natural and easicr to apply. From various directions, there are extensive developments in this area, many but not all of which are presented in FM II and III. We discuss here results that are simple to formulate and relevant to various kinds of experimental work. We begin with the ordinal theory.

A real-valued function P on $A \times A$ is called a *binary probability function* if it satisfies both

$$P(a, b) \geq 0,$$
$$P(a, b) + P(b, a) = 1.$$

The intended interpretation of $P(a, b)$ is as the probability of a being chosen over b. We use the probability measure P to define two natural binary relations.

$$aWb \quad \text{iff} \quad P(a, b) \geq \frac{1}{2},$$
$$aSb \quad \text{iff} \quad P(a, c) \geq P(b, c), \text{ for all } c.$$

In the spirit of semiorders we now define how the relations W and S are related to various versions of what is called *stochastic transitivity,* where stochastic means that the individual instances may not be transitive, but the probabilities are in some sense transitive. Here are the definitions. Let P be a binary probability function on $A \times A$. We define the following for all a, b, c, d in A:

Weak stochastic transitivity: If $P(a, b) \geq \frac{1}{2}$ and $P(b, c) \geq \frac{1}{2}$, then $P(a, c) \geq \frac{1}{2}$.

Weak independence: If $P(a, c) > P(b, c)$, then $P(a, d) \geq P(b, d)$.

Strong stochastic transitivity: If $P(a, b) \geq \frac{1}{2}$ and $P(b, c) \geq \frac{1}{2}$, then $P(a, c) \geq \max[P(a, b), P(b, c)]$.

The basic results for these concepts are taken from Block and Marschak (1960) and Fishburn (1973). Let P be a binary probability function on $A \times A$, and let W and S be defined as in the previous equations. Then

1. Weak stochastic transitivity holds if W is transitive.
2. Weak independence holds if S is connected.
3. Strong stochastic transitivity holds if $W = S$. Therefore strong stochastic transitivity implies weak independence; the two are equivalent if $P(a, b) \neq \frac{1}{2}$ for $a \neq b$.

Random Variable Representations

We turn next to random variable representations for measurement. In the first type, an essentially deterministic theory of measurement (e.g., additive conjoint measurement) is assumed in the background. But it is recognized that, for various reasons, variability in response occurs even in what are apparently constant circumstances. We describe here the approach developed and used by Falmagne (1976, 1985). Consider the conjoint indifference $(a, p) \sim (b, q)$ with a, p, and q given and b to be determined so that the indifference holds. Suppose that, in fact, b is a random variable which we may denote $\mathbf{B}(a, p; q)$. We suppose that such random variables are independently distributed. Since realizations of the random variables occur in repeated trials of a given experiment, we can define the equivalents we started with as holding when the value b is the Pth percentile of the distribution of the random variable $\mathbf{B}(a, p; q)$. Falmagne's proposal was to use the median, $P = \frac{1}{2}$, and he proceeded as follows. Let ϕ_1 and ϕ_2 be two numerical representations for the conjoint measurement in the usual deterministic sense. If we suppose that such an additive representation is approximately correct but has an additive error, then we have the following representation:

$$\varphi_1[\mathbf{B}(a, p; q)] = \varphi_1(a) + \varphi_2(q) - \varphi_2(p) + \epsilon(a, p; q),$$

where the ϵs are random variables. It is obvious enough how this equation provides a natural approximation of standard conjoint measurement. If we strengthen the assumptions a bit, we get an even more natural theory by assuming that the random variable

$\epsilon(a, p; q)$ has its median equal to zero. Using this stronger assumption about all the errors being distributed with a median of zero, Falmagne summarizes assumptions that must be made to have a measurement structure.

Let A and P be two intervals of real numbers, and let $\mathcal{U} = \{\mathbf{U}_{pq}(a) \mid p, q \in P, a \in A\}$ be a collection of random variables, each with a uniquely defined median. Then $\mathcal{U}$ is a structure for *random additive conjoint measurement* if for all p, q, r in P and a in A, the medians $m_{pq}(a)$ satisfy the following axioms:

(i) They are continuous in all variables p, q, and a.

(ii) They are strictly increasing in a and p, and strictly decreasing in q.

(iii) They map A into A.

(iv) They satisfy the cancellation rule with respect to function composition $*$, i.e.,

$$(m_{pq} * m_{qr})(a) = m_{pr}(a),$$

whenever both sides are defined.

For such random additive conjoint measurement structures, Falmagne (1985, p. 273) proved that there exist real-valued continuous strictly increasing functions ϕ_1 and ϕ_2, defined on A and P respectively, such that for any $\mathbf{U}_{pq}(a)$ in $\mathcal{U}$,

$$\varphi_1[\mathbf{U}_{pq}(a)] = \varphi_2(p) + \varphi_2(q) - \varphi_1(a) + \epsilon_{pq}(a),$$

where $\epsilon_{pq}(a)$ is a random variable with a unique median equal to zero. Moreover, if φ_1' and φ_2' are two other such functions, then

$$\varphi_1'(a) = \alpha\varphi_1(a) + \beta$$

and

$$\varphi_2'(p) = \alpha\varphi_2(p) + \gamma,$$

where $\alpha > 0$.

Statistical tests of these ideas are not a simple matter but have been studied in order to make the applications practical. Major references are Falmagne (1978); Falmagne and Iverson (1979); Falmagne, Iverson, and Marcovici (1979); and Iverson and Falmagne (1985). Recent important work on probability models includes Doignon and Regenwetter (1997); Falmagne and Regenwetter (1996); Falmagne, Regenwetter, and Grofman (1997); Marley (1993); Niederée and Heyer (1997); Regenwetter (1997); and Regenwetter and Marley (in press).

Qualitative Moments

Another approach to measuring, in a representational sense, the distribution of a random variable for given psychological phenomena is to assume that we have a qualitative method for measuring the moments of the distribution of the random variable. The experimental procedures for measuring such raw moments will vary drastically from one domain of experimentation to another. Theoretically, we need only to assume that we can judge qualitative relations of one moment relative to another and that we have a standard weak ordering of these qualitatively measured moments. The full formal discussion of these matters is rather intricate. The details can be found in Section 16.8 of FM II.

Qualitative Density Functions

As is familiar in all sorts of elementary probability examples, when a distribution has a given form, it is often much easier to characterize it by a density distribution of a random variable than by a probability measure on events or by the method of moments as just mentioned. In the discrete case, the situation is formally quite simple. Each atom (i.e., each atomic event) in the discrete density has a qualitative probability, and we need judge only relations between these qualitative probabilities. We require of a representing discrete density function p on $\{a_1, \ldots, a_n\}$ the following three properties:

(i) $p(a_i) \geq 0$.

(ii) $\sum_{i=1}^{n} p(a_i) = 1$.

(iii) $p(a_i) \geq p(a_j)$ iff $a_i \succsim a_j$.

Note that the a_i are *not* objects or stimuli in an experiment, but qualitative atomie events, exhaustive and mutually exclusive. Also note that in this discrete case it follows that $p(a_i) \leq 1$, whereas in the continuous case this is not true of densities.

We also need conditional discrete densities. For this purpose we assume that the underlying probability space X is finite or denumerable, with probability measures P on the given family $\mathcal{F}$ of events. The relation of the density p to the measure P is, for a_i an atom of X,

$$p(a_i) = P(\{a_i\})$$

Then if A is any event such that $P(A) > 0$,

$$p(a_i \mid A) = P(\{a_i\} \mid A),$$

and, of course, $p(a_i \mid A)$ is now a discrete density itself, satisfying (i) through (iii).

Here are two simple, but useful, examples of this approach. Let X be a finite set. Then the uniform density on X is characterized by all atoms being equivalent in the qualitative ordering $\succsim$, that is,

$$a_i \sim a_j.$$

We may then easily show that the unique density satisfying the equivalence and (i), (ii), and (iii) is

$$p(a_i) = \frac{1}{n},$$

where n is the number of atoms in X.

Among the many possible discrete distributions, we consider just one further example, whieh has application in experiments in which the model being tested assumes a probability of change of state independent of the time spent in the current state. In the case of discrete trials, such a mcmoryless process has a geometric distribution that can be tested or derived from some simple properties of the discrete but denumerable set of atomic events $\{a_1, \ldots, a_n, \ldots\}$, on each of which is a positive qualitative probability of the occurrence of the changc of state. The numbering of the atoms intuitively corresponds to the trials of an experiment. The atoms are ordered in qualitative probability by the relation $\succsim$. We also introduce a restricted conditional probability. If $i > j$ then $a_i \mid A_j$ is the conditional event that the change of state will occur on trial i given that it has *not* occurred on or before trial j. (Note that here A_j means no change of state from trial 1 through j.) The qualitative probability ordering relation is extended to include these special conditional events as well.

The two postulated properties, in addition to (i), (ii), and (iii) given above, are these:

(iv) Order property: $a_i \succsim a_j$ iff $j \geq i$;

(v) Memoryless property: $a_{i+1} \mid A_i \sim a_1$.

It is easy to prove that (iv) implies a weak ordering of $\succsim$. We can then prove that $p(a_n)$ has the form

$$p(a_n) = c(1-c)^{n-1} \quad (0 < c < 1).$$

Properties (i) through (v) are satisfied, but they are also satisfied by any other c', $0 < c' < 1$. For experiments testing only the memoryless property, no estimation of c is required. If it is desired to estimate c, the standard estimate is the sample mean m of the trial numhers on which the change of state was observed, since the mean μ of the density $p(a_n) = c(1-c)^{n-1}$ satisfies the following equation:

$$\mu = \frac{1-c}{c}.$$

For a formal characterization of the full qualitative probability for the algebra of events—not just atomic events—in the case of the geometric distribution, see Suppes (1987). For the closely related but mathematically more complicated continuous analogue (i.e., the exponential distribution), see Suck (1998).

GENERAL FEATURES OF THE AXIOMATIC APPROACH

Background

History

The story of the axiomatic method begins with the ancient Greeks, probably in the fifth century B.C. The evidence seems pretty convincing that it developed in response to the early crisis in the foundations of geometry mentioned earlier, namely, the problem of incommensurable magnitudes. It is surprising and important that the axiomatic method as we think of it was largely crystallized in Euclid's *Elements,* whose author flourished and taught in Alexandria around 300 B.C. From a modern standpoint, Euclid's schematic approach is flawed, but compared to any other standard to be found anywhere else for over two millennia, it is a remarkable achievement. The next great phase of axiomatic development occurred, as already mentioned, in the 19th century in connection with the crisis generated in the foundations of geometry itself. The third phase was the formalization within logic of the entire language used and the realization that results that could not be proved otherwise can be achieved by such complete logical formalization. In view of the historical review presented earlier in this article, we will concentrate on only this third phase in this section.

What Comes before the Axioms

Three main ingredients need to be fixed in an axiomatization before the axioms are formulated. First, there must be agreement on the general framework used. Is it going to be an informal, set-theoretical framework or one formalized within logic? These two alternatives are analyzed in more detail later. The second ingredient is to fix the primitive concepts of the theory being axiomatized. For example, in almost all theories of choice we need an ordering relation as a primitive concept, which means, formally, a binary relation. We also often need, as mentioned earlier, a binary operation as, for example, in the cases of extensive measurement and averaging. In any case, whatever the primitives may be, they should be stated at the beginning. The third ingredient, at least as important, is clarity and explicitness about what other theories are being assumed. It is a characteristic feature of empirical axiomatizations that some additional mathematics is usually assumed, often without explicit notice. This is not true, however, of many qualitative axiomatizations of representational measurement and often is not true in the foundations of geometry. In contrast, many varieties of probabilistic modeling in psychology do assume some prior mathematics in formulating the axioms. A simple example of this was Falmagne's axioms for random additive conjoint measurement, presented earlier. There, such statistical notions as the median and such elementary mathematical notions as that of continuity were assumed without further explanation or definition.

Another ingredient, less important from a formal standpoint but of considerable importance in practice, are the questions of whether notions defined in terms of the primitive concepts should be introduced when formulating the axioms and whether auxiliary mathematical notions are assumed in stating the axioms. The contrasting alternative is to state the axioms strictly in terms of the primitive notions. From the standpoint of logical purity, the latter course seems desirable, but in actual fact it is often awkward and intuitively unappealing to state all of the axioms in terms of the primitive concepts only. A completely elementary but good example of this is the introduction of a strict ordering and an equivalence relation defined in terms of a weak ordering, a move that is often used as a way of simplifying and making more perspicuous the formulation of axioms in choice or preference theory within psychology.

Theories with Standard Logical Formalization

Explicit and formally precise axiomatic versions of theories are those that are formalized within first-order logic. Such a logic can be easily characterized in an informal way. This logic assumes

(i) one kind of variable;
(ii) logical constants, mainly the sentential connectives such as *and;*
(iii) a notation for the universal and existential quantifiers; and
(iv) the identity symbol $=$.

A theory formulated within such a logical framework is called a *theory with standard formalization*. Ordinarily, three kinds of nonlogical constants occur in axiomatizing a theory within such a framework: the relation symbols (also called predicates), the operation symbols, and the individual constants.

The grammatical expressions of the theory are divided into terms and formulas, and recursive definitions of each are given. The simplest terms are variables or individual constants. New terms are built up by combining simpler terms with operation symbols in the manner spelled out recursively in the formulation of the language of the theory. Atomic formulas consist of a single predicate and the appropriate number of terms. Compound formulas are built up from atomic formulas by means of sentential connectives and quantifiers.

Theories with standard formalization are not often used in any of the empirical sciences, including psychology. On the other hand, they can play a useful conceptual role in answering some empirically important questions, as we illustrate later.

There are practical difficulties in casting ordinary scientific theories into the framework of first-order logic. The main source of the difficulty, which has already been mentioned, is that almost all systematic scientific theories assume a certain amount of mathematics a priori. Inclusion of such mathematics is not possible in any elegant and reasonable way in a theory beginning only with logic and with no other mathematical assumptions or apparatus. Moreover, a theory that requires for its formulation an Archimedean-type axiom, much needed in representational theories of measurement when the domain of objects considered is infinite, cannot even in principle be formulated within first-order logic. We say more about this well-known result later. For these and other reasons, standard axiomatic formulations of most mathematical theories, as well as scientific theories, follows the methodology to which we now turn.

Theories Defined as Set-Theoretical Predicates

A widely used alternative approach to formulating representational theories of measurement and other scientific theories is to axiomatize them within a set-theoretical framework. Moreover, this is close to the practice of much mathematics. In such an approach, axiomatizing a theory simply amounts to defining a certain set-theoretical predicate. The axioms, as we ordinarily think of them, are a part of the definition—its most important part from a scientific standpoint. Such definitions were (partially) presented earlier in a more or less formal way (e.g., weak orderings, extensive structures, and other examples of qualitative characterizations of empirical measurement structures). Note that the concept of isomorphism, or the closely related notion of homomorphism, is defined for structures satisfying some set-theoretical predicate. The language of set-theoretical predicates is not ordinarily used except in foundational talk; it is just a way of clarifying the status of the axioms. It means that the axioms are given within a framework that assumes set theory as the general framework for all, or almost all, mathematical concepts. It provides a seamless way of linking systematic scientific theories that use various kinds of mathematics with mathematics itself. An elementary but explicit discussion of the set-theoretical approach to axiomatization is found in Suppes (1957/1999, chap. 12).

Formal Results about Axiomatization

We sketch here some of the results that we think are of significance for quantitative work in experimental psychology. A detailed treatment is given in FM III, Chap. 21. We should emphasize that all the systematic results we state here hold only for theories formalized in first-order logic.

Elementary Languages

First, we need to introduce, informally, some general notions to be used in stating the results. We say that a language $\mathcal{L}$ of a theory is *elementary* if it is formulated in first-order logic. This means that, in addition to the apparatus of first-order logic, the theory only contains nonlogical relation symbols, operation symbols, and individual constants. Intuitively, a *model* of such a language $\mathcal{L}$ is simply an empirical structure, in the sense already discussed; in particular, it has a nonempty domain, a relation corresponding to each primitive relation symbol, an operation corresponding to each primitive operation symbol, and individuals in the domain corresponding to each individual constant.

Using such logical concepts, one major result is that there are infinite weak orders that cannot be represented by numerical order. A specific example is the lexicographic order of points in the plane, that is $(x, y) \succsim (x', y')$ if and only if either $x > x'$ or $x = x'$ and $y \geq y'$.

In examining the kinds of axioms given earlier (e.g., those for extensive measurement), it is clear that some form of an Archimedean axiom is needed to get a numerical representation, and such an axiom cannot be formulated in an elementary language $\mathcal{L}$, a point to which we return a little later.

A second, but positive, result arises when the domains of the measurement structures are finite. A class of such structures closed under isomorphism is called a *finitary* class of measurement structures. To that end, we need the concept of a language being recursively axiomatizable; namely, there is an algorithm for deciding whether a formula of $\mathcal{L}$ is an axiom of the given theory. It can be shown that any finitary class of measurement structures with respect to an elementary language $\mathcal{L}$ is axiomatizable but not necessarily recursively axiomatizable in $\mathcal{L}$.

The importance of this result is in showing that the expressive power of elementary languages is adequate for finitary classes but not necessarily for the stating of a set of recursive axioms. We come now to another positive result guaranteeing that recursive axioms are possible for a theory. When the relations, operations, and constants of an empirical structure are definable in elementary form when interpreted as numerical relations, functions, and constants, then the theory is recursively axiomatizable.

Nonaxiomatizability Results

Now we turn to a class of results of direct psychological interest. As early as the work of Wiener (1921), the nontransitive equivalence relation generated by semiorders was defined (see the earlier quotation); namely, if we think of a semiorder, then the indistinguishability or indifference relation that complements it will have the following numerical representation.

For two elements a and b that are indistinguishable or indifferent with respect to the semiorder, the following equivalence holds:

$$|f(a) - f(b)| \leq 1 \quad \text{iff} \quad a \sim b.$$

Now we have already seen that finite semiorders have a very simple axiomatization. Given how close the indistinguishability relation is to the semiorder itself, it seems plausible that this relation, too, should have a simple axiomatization. Surprisingly, Roberts (1968, 1969) proved that this is not the case. More precisely, let $\mathcal{L}$ be the elementary language whose only nonlogical symbol is the binary relational symbol $\sim$. Then the finitary class $\mathcal{J}$ of measurement structures for the binary relation of indistinguishability is not axiomatizable in $\mathcal{L}$ by a universal sentence.

Note that there is a restriction in the result. It states that $\sim$ is not axiomatizable by a universal sentence. This means that existential statements are excluded. The simple axiomatization of semiorders, given earlier, is such a universal axiomatization because no quantifiers were required. But that is not true of indistinguishability. A little later, we discuss the more general question of axioms with existential quantifiers for elementary languages.

This result about $\sim$ is typical of a group of theorems concerning familiar representations for which it is impossible to axiomatize the class of finite structures by adjoining a universal sentence to an elementary language $\mathcal{L}$. Scott and Suppes (1958) first proved this to be true for a quaternary relation symbol corresponding to a difference representation. Titiev (1972) obtained the result for additive conjoint measurement; he also showed that it holds for the n-dimensional metric structure using the Euclidean metric; and in 1980 he showed that it is true for the city-block metric when the number of dimensions $n \leq 3$. It is worth mentioning that the proof for $n = 3$ given by Titiev required computer assistance to examine 21,780 cases, each of which involved 10 equations and 12 unknowns in a related set of inequalities. To our knowledge, nothing is known about $n > 3$.

This last remark is worth emphasizing to bring out a certain point about the results mentioned here. For any particular case (e.g., an experiment using a set of 10 stimuli), a constructive approach, rather than the negative results given here, can be found for each particular case. One can simply write down the set of elementary linear inequalities that must be satisfied and ask a computer program to decide whether this finite set of inequalities in a fixed number of variables has a solution. If the answer is positive, then a numerical representation can be found, and the very restricted class of measurement structures built up around this fixed number of variables and fixed set of inequalities is indeed a measurement structure. What the theorems show is that the general elementary theory of such inequalities cannot be given in any reasonable axiomatic form. We cannot state for the various kinds of cases that are considered an elementary set of axioms that will guarantee a numerical solution for any finite model (i.e., a model with a finite domain) satisfying the axioms.

Finally, in this line of development, we mention a theorem requiring more sophisticated logical apparatus that was proved by Per Lindstrom (stated as Theorem 17, p. 243, FM III), namely, that even if existential quantifiers are permitted, the usual class of finite measurement structures for algebraic difference cannot be characterized by a finite set of elementary axioms.

Archimedean and Least-Upper-Bound Axioms

We have mentioned more than once that Archimedean axioms play a special role in formulating representational theories of measurement when the domain of the empirical structures is infinite. Recall that the Archimedean axiom for extensive measurement of weight or mass asserts that for any objects a and b, there exists some integer n such that n replicas of object a, written as $a(n)$, exceeds b, that is, $a(n) \succsim b$. This axiom, as well as other versions of it, cannot be directly formulated in an elementary language because of the existential quantification in terms of the natural numbers. In that sense, the fact that an elementary theory cannot include an Archimedean axiom has an immediate proof. Fortunately, however, a good deal more can be proved: For such elementary theories, of the kind we have considered in this chapter, there can be no elementary formulas of the elementary language $\mathcal{L}$ that are equivalent to an Archimedean axiom. After all, we might hope that one could simply replace the Archimedean axiom by a conjunction of elementary formulas, but this is not the case. For a proof, and references to the literature, see FM III, Section 21.7.

It might still be thought that by avoiding the explicit introduction of the natural numbers, we might be able to give an elementary formulation using one of the other axioms invoked in real analysis. Among these are Dedekind's (1872/1902) axiom of completeness, Cantor's (1895) formulation of completeness in terms of Cauchy sequences, and the more standard modern approach of assuming that a bounded nonempty set has a least-upper-bound in terms of the given ordering. We consider only the last example because its elementary form allows us to see easily what the problem is. To invoke this concept, we need to be able to talk in our elementary language not only about individuals in the given domain of an empirical structure, but also about sets of these individuals. But the move from individuals to sets of individuals is a mathematically powerful one, and it is not permitted in standard formulations of elementary languages. As in the case of the Archimedean axiom, then, we have an immediate argument for rejecting such an axiom. Moreover, as in the case of the Archimedean axiom, we can prove that no set of elementary formulas of an elementary language $\mathcal{L}$ is equivalent to the least-upper-bound axiom. The proof of this follows naturally from the Archimedean axiom, since in a general setting the least-upper-bound axiom implies an Archimedean axiom.

Proofs of Independence of Axioms and Primitive Concepts

All the theorems just discussed can be formulated only within the framework of elementary languages. Fortunately, important questions that often arise in discussions of axioms in various scientific domains can be answered within the purely set-theoretic framework and do not require logical formalization. The first of these is proving that the axioms are independent in the sense that none can be deduced from the others. The standard method for doing this is as follows. For each axiom, a model is given in which the remaining axioms are satisfied and the one in question is not satisfied. Doing this establishes that the axiom is independent of the others. The argument is simple. If the axiom in question could be derived from the remaining axioms, we would then have a violation of the intuitive concept of logical consequence. An example of lack of independence among axioms given for extensive measurement is the commutativity axiom, $a \circ b \sim b \circ c$. It follows from the other axioms with the Archimedean axiom playing a very important role.

The case of the independence of primitive symbols requires a method that is a little more subtle. What we want is an argument that will prove that it is not possible to define one of the primitive symbols in terms of the others. Padoa (1902) formulated a principle that can be applied to such situations. To prove that a given primitive concept is independent of the other primitive concepts of a theory, find two models of the axioms of the theory such that the primitive concept in question is essentially different in the two models and the remaining primitive symbols are the same in the two models.

As a very informal description of a trivial example, consider the theory of preference based on two primitive relations, one a strict preference and the other an indifference relation. Assume both are transitive. We want to show what is obvious—that strict preferencc cannot be defined in terms of indifference. We need only take a domain of two objects, for example, the numbers 1 and 2. Then for the indifference relation we just take identity: $1 = 1$ and $2 = 2$. But in one model the strict preference relation has 1 preferred to 2, and in the second preference model the preference relation has 2 preferred to 1. This shows that strict preference cannot be defined in terms of indifference because indifference is the samc in both models whereas preference is different.

CONCLUSIONS

The second half of the 20th century saw a number of devclopmcnts in our understanding of numerical measurement. Among these are the following: (a) examples of fundamental measurement different from extensive structures; (b) an increased understanding of how measurement structures interlock to yield substantive theories; (c) a classification of scale types for continuous measurement in terms of properties of automorphism groups; (d) an analysis of invariance principles in limiting the mathcmatical forms of various measures; (e) a logical analysis of what sorts of theories can and cannot be formulated using purely first-order logic without existential or Archimedean statements; and (f) a number of psychological applications especially in psychophysics and utility theory.

A major incompleteness remains in the socially important area of ability and achievement testing. Except for the work of Doignon and Falmange (1999), no rcpresentational results of significance exist for understanding how individuals differ in their grasp of certain concepts. This is not to deny the extensive development of statistical models, but only to remark that fundamental axiomatizations are rarely found. This is changing gradually, but as yet it is a small part of representational measurement theory.

REFERENCES

Aczél, J., Luce, R. D., & Ng, C. T. (2001). Functional equations arising in a theory of rank dependence and homogeneous joint receipts. Submitted.

Alper, T. M. (1987). A classification of all order-preserving homeomorphism groups of the real that satisfy finite uniqueness. *Journal of Mathematical Psychology, 31,* 135–154.

Block, H. D., & Marschak, J. (1960). Random orderings and stochastic theories of responses. In I. Olkin, S. Ghurye, W. Hoeffding, W. Madow, & H. Mann (Eds.), *Contributions to probability and statistics* (pp. 97–132). Stanford, CA: Stanford University Press.

Cameron, P. J. (1989). Groups of order-automorphisms of the rationals with prescribed scale type. *Journal of Mathematical Psychology, 33,* 163–171.

Campbell, N. R. (1920/1957). *Physics: The elements.* Cambridge: Cambridge University Press. [Reprinted under the title *Foundations of science: The philosophy of theory and experiment,* New York: Dover.] (Original work published 1920)

Campbell, N. R. (1928). *An account of the principles of theory and experiment.* London: Longmans, Green.

Cantor, G. (1895). Beiträge zur Begründung der transfiniten Mengenlehre. *Mathematische Annalen, 46,* 481–512.

Cohen, M. A., & Narens, L. (1979). Fundamental unit structures: A theory of ratio scalability. *Journal of Mathematical Psychology, 20,* 192–232.

Davidson, D., McKinsey, J. C. C., & Suppes, P. (1955). Outlines of a formal theory of value: I. *Philosophy of Science, 22,* 140–160.

Davidson, D., Suppes, P., & Segal, S. (1957). *Decision-making: An experimental approach.* Stanford, CA: Stanford University Press. [Reprinted as Midway Reprint, 1977, Chicago: University of Chicago Press.]

Debreu, G. (1960). Topological methods in cardinal utility theory. In K. J. Arrow, S. Karlin, & P. Suppes (Eds.), *Mathematical methods in the social sciences, 1959.* Stanford, CA: Stanford University Press (pp. 16–26).

Dedekind, R. (1872/1902). Stetigkeit und Irrationale Zahlen. Brunswick, 1872. [*Essays on the theory of numbers.* New York: Dover, 1963 (reprint of the English translation).]

Doignon, J.-P., & Falmagne, J.-C. (1999). *Knowledge spaces.* Berlin: Springer.

Doignon, J.-P., & Regenwetter, M. (1997). An approval-voting polytope for linear orders. *Journal of Mathematical Psychology, 41,* 171–188.

Dzhafarov, E. N., & Colonius, H. (1999). Fechnerian metrics in unidimensional and multidimensional spaces. *Psychonomic Bulletin and Review, 6,* 239–268.

Dzhafarov, E. N., & Colonius, H. (2001). Multidimensional Fechnerian scaling: Basics. *Journal of Mathematical Psychology, 45,* 670–719.

Edwards, W. (Ed.). (1992). *Utility theories: Measurements and applications.* Boston: Kluwer.

Ellemeirer, W., & Faulhammer, G. (2000). Empirical evaluation of axioms fundamental to Stevens' ratio-scaling approach: I. Loudness production. *Perception & Psychophysics, 62,* 1505–1511.

Euclid. (1956). *Elements, Book 10* (Vol. 3, T. L. Heath, Trans.) New York: Dover.

Falmagne, J.-C. (1976). Random conjoint measurement and loudness summation. *Psychological Review, 83,* 65–79.

Falmagne, J.-C. (1978). A representation theorem for finite random scale systems. *Journal of Mathematical Psychology, 18,* 52–72.

Falmagne, J.-C. (1985). *Elements of psychophysical theory.* New York: Oxford University Press.

Falmagne, J.-C., & Iverson, G. (1979). Conjoint Weber laws and additivity. *Journal of Mathematical Psychology, 20,* 164–183.

Falmagne, J.-C., Iverson, G., & Marcovici, S. (1979). Binaural "loudness" summation: Probabilistic theory and data. *Psychological Review, 86,* 25–43.

Falmagne, J.-C., & Narens, L. (1983). Scales and meaningfulness of quantitative laws. *Synthese, 55,* 287–325.

Falmagne, J.-C., & Regenwetter, M. (1996). A random utility model for approval voting. *Journal of Mathematical Psychology, 40,* 152–159.

Falmagne, J.-C., Regenwetter, M., Grofman, B. (1997). A stochastic model for the evolution of preferences. In A. A. J. Marley (Ed.), *Choice, decision, and measurement: Essays in honor of R. Duncan Luce* (pp. 111–129). Mahwah, NJ: Erlbaum.

Fechner, G. T. (1966). *Elements of psychophysics* (H. E. Adler, Trans.). New York: Holt, Rinehart & Winston. (Original work published 1860)

Ferguson, A., Meyers, C. S. (Vice Chairman), Bartlett, R. J. (Secretary), Banister, H., Bartlett, F. C., Brown, W., Campbell, N. R., Craik, K. J. W., Drever, J., Guild, J., Houstoun, R. A., Irwin, J. O., Kaye, G. W. C., Philpott, S. J. F., Richardson, L. F., Shaxby, J. H., Smith, T., Thouless, R. H., & Tucker, W. S. (1940). Quantitative estimates of sensory events. The advancement of science. *Report of the British Association for the Advancement of Science, 2,* 331–349.

Fishburn, P. C. (1970). *Utility theory for decision making.* New York: Wiley.

Fishburn, P. C. (1973). Binary choice probabilities: On the varieties of stochastic transitivity. *Journal of Mathematical Psychology, 10,* 327–352.

Fishburn, P. C. (1988). *Nonlinear preference and utility theory.* Baltimore, MD: Johns Hopkins Press.

Fourier, J. (1955). *The analytical theory of heat* (A. Freeman, Trans.). New York: Stechert. (Original work published 1822)

Gauss, K. F. (1821/1957). *Theory of least squares* (H. F. Trotter, Trans.). Princeton, NJ: Princeton University Press.

Gerlach, M. W. (1957). *Interval measurement of subjective magnitudes with subliminal differences* (Tech. Rep. No. 7). Stanford, CA: Stanford University, Behavioral Sciences Division, Applied Mathematics and Statistics Laboratory.

Goodman, N. (1951). *Structure of appearance*. Cambridge: Harvard University Press.

Grassmann, H. G. (1895). *Die Wissenschaft der Extensiven Grösse oder die Ausdehnungslehre*, republished in *Gesammelte Mathematische und Physikalische Werke*. Leipzig. (Original work published 1844)

Green, D. M., & Swets, J. A. (1974). *Signal detection theory and psychophysics*. Huntington, NY: Krieger. (Original work published 1966)

Halphen, E. (1955). La notion de vraisemblance. *Publication de l'Institut de Statistique de l'Université de Paris, 4*, 41–92.

Heisenberg, W. (1930). *Quantum theory*. Chicago: University of Chicago Press.

Helmholtz, H. V. (1887). Zählen und Messen erkenntnis-theoretisch betraehet. *Philosophische Aufsätze Eduard Zeller gewidmet*, Leipzig. Reprinted in *Gesammelte Abhandl.*, Vol. 3, 1895, 356–391. [English translation by C. L. Bryan, *Counting and Measuring*. Princecton, New Jersey: van Nostrand, 1930.]

Hilbert, D. (1899). *Grundlagen der Geometrie* (8th edition, with revisions and supplements by P. Bernays, 1956). Stuttgart: Teubner.

Hölder, O. (1901). Die Axiome der Quantität und die Lehre vom Mass. *Berichte über die Verhandlungen der Königlich Sächsischen Gesellschaft der Wissenschaften zu Leipzig, Mathematische-Physische Klasse, 53*, 1–64.

Iverson, G., & Falmagne, J.-C. (1985). Statistical issues in measurement. *Mathematical Social Sciences, 10*, 131–153.

Iverson, G., & Luce, R. D. (1998). The representational measurement approach to psychophysical and judgment problems. In M. H. Birnbaum (Ed.), *Measurement, judgment, and decision making* (pp. 1–79). San Diego: Academic Press.

Klein, F. (1872/1893). A comparative review of recent researches in geometry. *Bulletin of the New York Mathematical Society, 2*, 215–249.

Krantz, D. H. (1972). A theory of magnitude estimation and cross-modality matching. *Journal of Mathematical Psychology, 9*, 168–199.

Krantz, D. H., Luee, R. D., Suppes, P., & Tversky, A. (1971). *Foundations of measurement, Vol. I.* New York: Academic Press.

Luce, R. D. (1956). Semiorders and a theory of utility discrimination. *Econometrica, 24*, 178–191.

Luce, R. D. (1959). On the possible psychophysical laws. *Psychological Review, 66*, 81–95.

Luce, R. D. (1990). "On the possible psychophysical laws" revisited: Remarks on cross-modal matching. *Psychological Review, 97*, 66–77.

Luce, R. D. (1992). Singular points in generalized concatenation structures that otherwise are homogeneons. *Mathematical Social Sciences, 24*, 79–103.

Luce, R. D. (2000). *Utility of gains and losses: Measurement-theoretical and experimental approaches*. Mahwah, NJ: Erlbaum. [Errata: see Luee's Web page at http://socsei.uci.edu.]

Luce, R. D. (2001). A psychophysical theory of intensity proportions, joint presentations, and matches. *Psychological Review*, in press.

Luce, R. D., & Edwards, W. (1958). The derivation of subjective scales from just-noticeable differences. *Psychological Review, 65*, 227–237.

Luce, R. D., & Fishburn, P. C. (1991). Rank- and sign-dependent linear utility models for finite first-order gambles. *Journal of Risk and Uncertainty, 4*, 29–59.

Luce, R. D., Krantz, D. H., Suppes, P., & Tversky, A. (1990). *Foundations of measurement, Vol. III.* San Diego: Academic Press.

Luce, R. D., & Tukey, J. (1964). Simultaneous conjoint measurement: A new type of fundamental measurement. *Journal of Mathematical Psychology, 1,* 1–27.

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide.* Cambridge: Cambridge University Press.

Marley, A. A. J. (1993). Aggregation theorems and the combination of probabilistic rank orders. In D. E. Critchlow, M. A. Fligner, & J. S. Verducci (Eds.), *Probability models and statistical analyses for ranking data* (pp. 216–240). New York: Springer.

Maxwell, J. C. (1873). *A Treatise on electricity and magnetism.* London: Oxford University Press.

Michell, J. (1990). *An introduction to the logic of psychological measurement.* Hillsdale, NJ: Erlbaum.

Michell, J. (1999). *Measurement in psychology: A critical history of a methodological concept.* Cambridge: Cambridge University Press.

Moody, E., and Claggett, M. (Eds.). (1952). *The medieval science of weights (scientia de ponderibus): Treatises ascribed to Euclid, Archimedes, Thabit ibn Qurra, Jordanus de Nemore, and Blasius of Parma.* Madison: University of Wisconsin Press.

Narcns, L. (1981a). A general theory of ratio scalablity with remarks about the measurement-theoretic conccpt of meaningfulness. *Theory and Decision, 13,* 1–70.

Narens, L. (1981b). On the scalcs of measurement. *Journal of Mathematical Psychology, 24,* 249–275.

Narens, L. (1985). *Abstract measurement theory.* Cambridge: MIT Press.

Narens, L. (1996). A theory of magnitude estimation. *Journal of Mathematical Psychology, 40,* 109–129.

Narens, L. (2001). *Theories of meaningfulness.* Mahwah, NJ: Erlbaum, in press.

Niederée, R., & Heyer, D. (1997). Gencralized random utility models and the representational theory of measurement: A conceptual link, In A. A. J. Marley (Ed.), *Choice, decision, and measurement: Essays in honor of R. Duncan Luce* (pp. 153–187). Mahwah, NJ: Erlbaum.

O'Brien, D. (1981). *Theories of weight in the ancient world.* Paris: Les Belles Lettres.

Padoa, A. (1902). Un nouveau système irreductible de postulats pour l'algèbre. *Compte Rendu du deuxième Congres International de Mathématiciens,* 249–256.

Pasch, M. (1882). *Vorlesungen über Neuere Geometrie.* Leipzig: Springer.

Pfanzagl, J. (1959). A general theory of measurement-applications to utility. *Naval Research Logistics Quarterly, 6,* 283–294.

Pfanzagl, J. (1968). *Theory of measurement.* New York: Wiley.

Plateau, J. A. F. (1872). Sur la measure des sensations physiques, et sur loi qui lie l'intensité de ces sensation à l'intensité de la cause excitante. *Bulletin de l'Academie Royale de Belgique, 33,* 376–388.

Quiggin, J. (1993). *Generalized expected utility theory: The rank-dependent model.* Boston: Kluwer.

Ramsey, F. P. (1931). Truth and probability. In F. P. Ramsey (Ed.), *The foundations of mathematics and other logical essays* (pp. 156–198). New York: Harcourt, Brace.

Regenwetter, M. (1997). Random utility representations of finite n-ary relations. *Journal of Mathematical Psychology, 40,* 219–234.

Regenwetter, M., & Marley, A. A. J. (in press). Random relations, random utilities, and random functions. *Journal of Mathematical Psychology.*

Roberts, F. S. (1968). *Representations of indifference relations.* Unpublished doctoral dissertation, Stanford University, Stanford, CA.

Roberts, F. S. (1969). Indifference graphs. In F. Harary (Ed.), *Proof techniques in graph theory* (pp. 139–146). New York: Academic Press.

Roberts, F. S., & Luce, R. D. (1968). Axiomatic thermodynamics and extensive measurement. *Synthese, 18,* 311–326.

Rozeboom, W. E. (1962). The untenability of Luce's principle. *Psychological Review, 69,* 542–547. [With a reply by Luce, 548–551, and an additional comment by Rozeboom, 552].

Savage, L. J. (1954). *The foundations of statistics.* New York: Wiley.

Scott, D., & Suppes, P. (1958). Foundational aspects of theories of measurement. *Journal of Symbolic Logic, 23,* 113–128.

Shepard, R. N. (1981). Psychophysical relations and psychophysical scales: On the status of "direct" psychophysical measurement. *Journal of Mathematical Psychology, 24,* 21–57.

Stevens, S. S. (1951). Mathematics, measurement and psychophysies. In S. S. Stevens (Ed.), *Handbook of experimental psychology* (pp. 1–49). New York: Wiley.

Stevens, S. S. (1959). Measurement, psychophysics and utility. In C. W. Churchman & P. Ratoosh (Eds.), *Measurement: Definitions and theories* (pp. 18–64). New York: Wiley.

Stevens, S. S. (1975). *Psychophysics: Introduction to its perceptual, neural, and social prospects.* New York: Wiley.

Suck, R. (1998). A qualitative charaeterization of the exponential distribution, *Journal of Mathematical Psychology, 42,* 418–431.

Suppes, P. (1999). *Introduction to logic.* New York: Dover. (Original work published 1957)

Suppes, P. (1980). Limitations of the axiomatic method in ancient Greek mathematieal sciences. In P. Suppes (Ed.), *Models and methods in the philosophy of science: Selected essays* (pp. 25–40). Dordrecht, Netherlands: Kluwer.

Suppes, P. (1987), Propensity representations of probability. *Erkenntnis, 26,* 335–358.

Suppes, P., & Alechina, N. (1994). The definability of the qualitative independence of events in terms of extended indicator functions. *Journal of Mathematical Psychology, 38,* 366–376.

Suppes, P., Krantz, D. H., Luce, R. D., & Tversky, A. (1990). *Foundations of measurement, Vol. II.* San Diego: Aeademic Press.

Thurstone, L. L. (1927a). A law of comparative judgement. *Psychological Review, 34,* 273–287.

Thurstone, L. L. (1927b). Psychophysical analysis. *American Journal of Psychology, 38,* 368–389.

Thurstone, L. L. (1927c). Three psychophysical laws. *Psychological Review, 34,* 424–432.

Thurstone, L. L. (1959). *The measurement of values.* Chicago: University of Chicago Press.

Titiev, R. J. (1972). Measurement structures in classes that are not universally axiomatizable. *Journal of Mathematical Psychology, 9,* 200–205.

Titiev, R. J. (1980). Computer-assisted results about universal axiomatizability and the three-dimensional city-block metric. *Journal of Mathematical Psychology, 22,* 209–217.

Todhunter, I. (1949). *A history of the mathematical theory of probability.* New York: Chelsea. (Original work published 1865)

von Neumann, J. (1932/1955). *Mathematical foundations of quantum mechanics* (R. T. Byer, Trans.). Princeton: Princeton University Press.

von Neumann, J., & Morgenstern, O. (1947). *The theory of games and economic behavior.* Princeton, NJ: Princeton University Press.

Wakker, P. P. (1989). *Additive representations of preferences: A new foundation of decision analysis.* Dordrecht, Netherlands: Kluwer Academie Publishers.

Wiener, N. (1915). Studies in synthetic logic. *Procedings of the Cambridge Philosophical Society, 18,* 14–28.

Wiener, N. (1921). A new theory of measurement: A study in the logic of mathematics. *Proceedings of the London Mathematical Society, 17,* 181–205.