STEVENS' HANDBOOK OF EXPERIMENTAL PSYCHOLOGY

THIRD EDITION

Volume 4: Methodology in Experimental Psychology

Editor-in-Chief

HAL PASHLER

Volume Editor

JOHN WIXTED

John Wiley & Sons, Inc.

STEVENS' HANDBOOK OF EXPERIMENTAL PSYCHOLOGY

THIRD EDITION

Volume 4: Methodology in Experimental Psychology

STEVENS' HANDBOOK OF EXPERIMENTAL PSYCHOLOGY

THIRD EDITION

Volume 4: Methodology in Experimental Psychology

Editor-in-Chief

HAL PASHLER

Volume Editor

JOHN WIXTED

John Wiley & Sons, Inc.

This book is printed on acid-free paper. ♾

Published by John Wiley & Sons, Inc., New York.
Published simultaneously in Canada.

Library of Congress Cataloging-in-Publication Data

Stevens' handbook of experimental psychology / Hal Pashler, editor-in-chief — 3rd ed.
p. cm.
Includes bibliographical references and index.
Contents: v. 1. Sensation and perception — v. 2. Memory and cognitive processes — v. 3. Learning, motivation, and emotion — v. 4. Methodology in experimental psychology.
ISBN 0-471-44333-6 (set) — ISBN 0-471-37777-5 (v. 1 : cloth : alk. paper) — ISBN 0-471-38030-X (v. 2 : cloth : alk. paper) — ISBN 0-471-38047-4 (v. 3 : cloth : alk. paper) — ISBN 0-471-37888-7 (v. 4 : cloth : alk. paper) — ISBN 0-471-44333-6 (set)
1. Psychology, Experimental. I. Pashler, Harold E.

BF181 H336 2001
150—dc21

2001046809

Contributors

Kimmo Alho, PhD
University of Helsinki, Finland

Norman H. Anderson, PhD
University of California, San Diego

F. Gregory Ashby, PhD
University of California,
Santa Barbara

Rachel Barr, PhD
Rutgers University

Nathan Brody, PhD
Wesleyan University

Max Coltheart, DSc
Macquarie University, Australia

M. Robin DiMatteo, PhD
University of California, Riverside

Shawn W. Ell, MA
University of California,
Santa Barbara

George A. Gescheider, PhD
Hamilton College

Ronald K. Hambleton, PhD
University of Massachusetts

Luis Hernandez-García, PhD
University of Michigan

John Jonides, PhD
University of Michigan

Daniel S. Levine, PhD
University of Texas at Arlington

Geoffrey R. Loftus, PhD
University of Washington

Gordon D. Logan, PhD
Vanderbilt University

R. Duncan Luce, PhD
University of California, Irvine

Neil A. Macmillan, PhD
Brooklyn College, City University
of New York

Lawrence E. Marks, PhD
John B. Pierce Laboratory,
Yale University

In Jae Myung, PhD
Ohio State University

Risto Näätänen, PhD
University of Helsinki, Finland

Stephen A. Petrill, PhD
Pennsylvania State University

Mary J. Pitoniak, MA
University of Massachusetts

Mark A. Pitt, PhD
Ohio State University

Patrick Rabbitt, PhD
The Victoria University of Manchester

Robert Rosenthal, PhD
University of California, Riverside

Carolyn Rovee-Collier, PhD
Rutgers University

Erich Schröger, PhD
University Leipzig, Germany

Patrick Suppes, PhD
Stanford University

Trisha Van Zandt, PhD
Ohio State University

Tor Wager, MA
University of Michigan

Contents

Preface

The precise origins of experimental psychology can be debated, but by any count the field is more than a hundred years old. The past 10 years have been marked by tremendous progress: a honing of experimental strategies and clearer theoretical conceptualizations in many areas combined with a more vigorous cross-fertilization across neighboring fields.

Despite the undeniable progress, vigorous debate continues on many of the most fundamental questions. From the nature of learning to the psychophysical functions relating sensory stimuli to sensory experiences and from the underpinnings of emotion to the nature of attention, a good many of the questions posed in the late 19th century remain alive and in some cases highly controversial.

Although some have viewed this fact as discouraging, it should scarcely be surprising. As in the biological sciences generally, early hopes that a few simple laws and principles would explain everything that needed to be explained have gradually given way to a recognition of the vast complexity of human (and nonhuman) organisms in general, and of their mental faculties in particular. There is no contradiction between recognizing the magnitude of the progress that has been made and appreciating the gap between current understanding and the fuller understanding that we hope to achieve in the future.

Stanley Smith ("Smitty") Stevens' *Handbook of Experimental Psychology,* of which this is the third edition, has made notable contributions to the progress of the field. At the same time, from one edition to the next, the *Handbook* has changed in ways that reflect growing recognition of the complexity of its subject matter. The first edition was published in 1951 under the editorship of the great psychophysical pioneer himself. This single volume (described by some reviewers as the last successful single-volume handbook of psychology) contained a number of very influential contributions in the theory of learning, as well as important contributions to psychophysics for which Stevens was justly famous. The volume had a remarkably wide influence in the heyday of a period in which many researchers believed that principles of learning theory would provide the basic theoretical underpinning for psychology as a whole.

Published in 1988, the second edition was edited by a team comprised of Richard Atkinson, Richard J. Herrnstein, Gardner Lindzey, and Duncan Luce. The editors of the second edition adopted a narrower definition of the field, paring down material that overlapped with physics or physiology and reducing the role of applied psychology. The result was a set of two volumes, each of which was substantially smaller than the single volume in the first edition.

Discussions of a third edition of the Stevens' *Handbook* began in 1998. My fellow editors and I agreed that experimental psychology had broadened and deepened to such a point that two volumes could no longer reasonably encompass the major accomplishments that have occurred in the field since 1988. We also felt that a greatly enlarged treatment of methodology would make the *Handbook* particularly valuable to those seeking to undertake research in new areas, whether graduate students in training or researchers venturing into subfields that are new to them.

The past 10 years have seen a marked increase in efforts to link psychological phenomena to neurophysiological foundations. Rather than eschewing this approach, we have embraced it without whittling down the core content of traditional experimental psychology, which has been the primary focus of the *Handbook* since its inception.

The most notable change from the previous edition to this one is the addition of a new volume on methodology. This volume provides rigorous but comprehensible tutorials on the key methodological concepts of experimental psychology, and it should serve as a useful adjunct to graduate education in psychology.

I am most grateful to Wiley for its strong support of the project from the beginning. The development of the new *Handbook* was initially guided by Kelly Franklin, now Vice President and Director of Business Development at Wiley. Jennifer Simon, Associate Publisher, took over the project for Wiley in 1999. Jennifer combined a great measure of good sense, good humor, and the firmness essential for bringing the project to a timely completion. Although the project took somewhat longer than we initially envisioned, progress has been much faster than it was in the second edition, making for an up-to-date presentation of fast-moving fields. Both Isabel Pratt at Wiley and Noriko Coburn at University of California at San Diego made essential contributions to the smooth operation of the project. Finally, I am very grateful to the four distinguished volume editors, Randy Gallistel, Doug Medin, John Wixted, and Steve Yantis, for their enormous contributions to this project.

Hal Pashler

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, $\cdot$. 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 actually 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 concluded, 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 collections 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 which 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 numerical—is necessary for precise prediction or control of phenomena. Of course, such a representation is useful only to the extent of the precision 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 distance 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

$$\begin{aligned} f(A \cup B) &= |A \cup B| = |A| + |B| \\ &= f(A) + f(B). \end{aligned}$$

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 the 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 invariance 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 uncertainty 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 classical 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 case, 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 be 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 affected 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

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

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

These, together with an Archimedean 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

$$(a, x) \succsim (b, y) \Leftrightarrow \psi_A(a) + \psi_X(x) \geq \psi_A(b) + \psi_X(y).$$

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

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

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