Birnbaum, M. H. (1978). Theories of psychophysical measurement. In N.J. Castellan & F. Restle, (Eds.), *Cognitive Theory, Vol. 3.* Hillsdale, N.J.: Lawrence Erlbaum Associates, pp. 1-5.

Part I

THEORIES OF PSYCHOPHYSICAL JUDGMENT

It seems fitting that the history of experimental psychology begins with Fechner's study of psychophysics, because measurement of subjective values is deemed prerequisite to a quantitative science of psychology. Although Fechner's original aim was to specify the relation between physical and psychological energy, it soon became apparent that the construction of psychological measurements could proceed even in cases where the stimuli had no obvious corresponding physical dimensions. Theories of psychophysics expanded to become theories of cognition and judgment. Thus experiments on problems in stimulus comparison and combination, for example, had implications not only for sensory psychologists who hoped to discover properties of neural transducers but also for a larger audience of experimental psychologists concerned with the understanding of information processing and judgment.

The study of judgment should concern students of all areas of psychology for at least three reasons. First, principles of judgment discovered in psychophysical research have been shown to be applicable in a wide variety of experimental situations in which the subject uses a numerical judgment scale to express a psychological value, attitude, opinion, belief, or feeling. Second, theories of measurement become relevant whenever a continuum or structure of psychological value is postulated, as in the recent developments on linear orderings, described elsewhere in this volume. Third, experimental and analytical techniques developed for the study of psychophysical theories can be applied to the study of algebraic models that arise in other areas of psychology.

This section brings together four authors who, though from diverse research traditions, share a common fundamental approach to the study of psychophysical judgment. Each chapter is concerned with criteria for evaluating theories, for ruling out theories that are wrong, and for retaining theories that deserve better treatment. Each is focused on the study of algebraic models; each introduces constraints, both experimental and theoretical, to help resolve theoretical issues of psychophysics.

In Chapter 1 is a charming dialogue, purportedly of recent discovery, between two Athenian scholars who presage important distinctions in the discussion of psychophysics. The dialogue has been translated and annotated by Marks, who compares it with the modern history of psychophysics, including the "direct" scaling approach. The dialogue makes clear the differences between physical measures of intensity and psychological magnitude, and it presents the concept of additivity as a fundamental device for constructing a scale of subjective value.

Marks reviews his recent work on loudness additivity in multicomponent tones. Two simultaneous tones of two different frequencies, in which the intensities of the tones are independently varied, are played to subjects who judge the loudness of the complex tone. Marks uses the principle of additivity together with factorial designs to derive scales of loudness. To account for the differences between scales derived from additive models of multicomponent tones and subtractive models of difference judgment, Marks proposes a stage theory of loudness in which different transformations of intensity occur depending on the subject's processing task.

Chapter 2, by Birnbaum, attacks a long-standing puzzle of psychophysics: the fact that scales derived from "ratio" techniques, such as magnitude estimation, do not agree with scales derived from "interval" techniques, such as category rating. This contradiction poses serious problems for attempts to measure sensation by operational definition, because two "direct" measures do not agree.

Birnbaum notes that "direct" numerical judgments of stimulus "ratios," for example, can be represented as the composition of three functions: a *psychophysical* function relating subjective magnitude to physical magnitude, a *comparison* function that computes the relationship between two stimuli, and a *judgment* function that relates the numerical judgment to subjective impression. Any theory of psychophysics can be considered as a set of premises about these processes, from which predictions for experimental results can be deduced. Birnbaum points out that a severe difficulty in psychophysics has been that too many theories, sets of premises, can account equally well for the data.

The chapter begins with a discussion of "direct" scaling of single stimuli, reviews problems with that approach, and describes the advantages and limitations of the factorial design approach. Key ideas of scale convergence and scale-free tests of algebraic models are introduced to provide diagnostic experiments among alternative theories of ratio and difference judgment. Scale convergence is the additional premise that subjective scale values are independent of the judgment task; that is, that scales of sensation derived from different models applied to different situations should agree. This requirement can be contrasted with Marks' stage theory, which permits different scales.

A series of experiments using the additional constraints are reviewed, showing that a coherent set of premises can account for the data and provide a simple solution to the long-standing disputes over "ratio" and "difference" judgments.

Instructions to judge "ratios" and "differences" of two stimuli lead to the same ordering of stimulus pairs, consistent with the idea that the same comparison process underlies both tasks. Birnbaum's theory contends that for a variety of continua, such as heaviness and loudness, there is no subjective zero point, and under these conditions a subtractive operation underlies judgments of "differences" and "ratios" of two stimuli. A ratio operation can be used to represent judgments of "ratios of differences," possibly because a difference has a well-defined zero point even when the stimuli are inherently only an interval scale. Of equal or greater importance to Birnbaum's conclusions are the methods of experimental and theoretical analysis used to reach those conclusions.

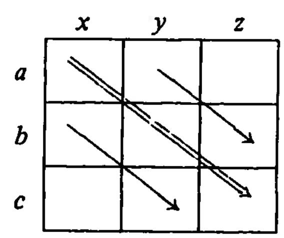
Chapter 3, by Restle, discusses theories that account for the effect of the background on the judgment of a stimulus. Restle devotes most of his attention to the Baldwin illusion, in which the apparent length of a line segment depends on the size of squares drawn at the ends of the lines. The relativity of judgment has been studied in the adaptation level approach (Helson, 1964; Restle & Greeno, 1970), which assumes that all effects of background stimulation can be summarized by one internal psychological state, the adaptation level. Restle notes that recent data for visual illusions appear to contradict a simple version of adaptation level theory, which assumes that the adaptation-level is a constant-weighted average of the stimuli in the field. The recent data show that the judgment of a stimulus is a nonmonotonic function of the background, apparently in contradiction to the theory.

Restle then discusses a modified form of adaptation level theory in which the weights used in the average depend on the similarity of the test and background stimuli. This additional premise, which allows the parameters of the context theory to depend on stimulus relationships, can account for the nonmonotonic effect of background size. An important point made by Restle is that the very strength of mathematical theories, their supreme testability, can be their downfall. There is a danger that theories could be prematurely ruled out by experiments that test auxiliary assumptions rather than the core of the theory. Chapter 3 illustrates how it may be possible to modify theories to account for what may seem at first to be condemning evidence.

In Chapter 4, Falmagne brings together concepts from the Fechner-Thurstone tradition with concepts of fundamental measurement in order to develop the beginnings of a statistical theory of psychological measurement. Falmagne presents a general theoretical treatment of polynomial measurement models, using examples from additive conjoint measurement, extensive measurement, and bisection measurement to illustrate his ideas. Because the paper is highly technical, it will be helpful to discuss some prerequisite concepts that provide a setting for his work.

One of the difficulties encountered by the conjoint measurement approach, Falmagne (1976) notes, is the question of its applicability to experimental data. *Foundations of Measurement* (Krantz, Luce, Suppes, & Tversky, 1971) attempts to specify a set of primitive ordinal assumptions from which one could deduce general premises of addition or subtraction, for example, considered in the chapters by Marks and Birnbaum. The difficulty is that the ordinal axioms of additive conjoint measurement do not have simple applications to data.

To illustrate this difficulty, it is helpful to consider Falmagne's example of loudness additivity. Suppose that the subject listens to dichotic tones consisting of two levels of intensity, one in each ear. Suppose the tone pairs are constructed from a factorial design in which the tones in the left ear are either a, b, or c, combined with tones in the right ear of x, y, or z. Let h and g represent the psychophysical functions for loudness for the left and right ears, respectively. The theory of additive conjoint measurement specifies the conditions under which it is possible to find a representation, h and g, such that (a, x) is louder than (b, y) if and only if $h(a) + g(x) > h(b) + g(y)$. One requirement of additivity is double cancellation: If (a, y) is louder than (b, z) and if (b, x) is louder than (c, y), then (a, x) should be louder than (c, z).



Double cancellation.
Single arrows represent premises;
double arrow represents conclusion.

This property follows from an additive representation, because if $h(a) + g(y) > h(b) + g(z)$ and if $h(b) + g(x) > h(c) + g(y)$, then $h(a) + h(b) + g(y) + g(x) > h(b) + h(c) + g(z) + g(y)$; cancelling $h(b)$ and $g(y)$ from both sides yields $h(a) + g(x) > h(c) + g(z)$, which implies that (a, x) is louder than (c,z).

This implication looks fairly straightforward to test. At first it may seem that all one need do is carry out the experiment and test whether the prediction is confirmed or refuted by the data. But the actual experiment is not that easy. First, what is the operationalization of "louder than"? Three popular definitions are among the posibilities: (1) (a, x) is louder than (b, y) if and only if $CJ(a, x) > CJ(b, y)$, where CJ is the category judgment of the loudness of the pair of tones; (2) magnitude estimation could be used to define the order, replacing category judgment in (1), as in Chapter 1; and (3) subjects could be asked to listen to two pairs and to report which of the pairs was the louder (or even judge the difference between two pairs, as in Chapter 2. A possible definition would be (a, x) is louder than (b, y) if $P(R_{ax;by}) > .5$ where $R_{ax;by}$ is the event that the subject judges the pair (a, x) to be greater than the pair (b, y). These three definitions would hopefully yield the same orderings.

Second, even with a definition, the experimental test remains unclear. Suppose one subject once judges (a, y) louder than (b, z), (b, x) louder than (c, y), and (c, z) louder than (a, x)? It could be attributed to a momentary fluctuation in loudness, to an error of memory, or to statistical fluctuations of proportions used to estimate probabilities. How many violations should it take for the investigator to seriously question the theory?

The statistical issue can be approached by representing the comparison process by an extension of the Thurstone approach. The psychophysical variability is attributed to variability in the psychological magnitudes of the stimuli.

In accord with Thurstone's case V, let $P(R_{12}) = F(\psi_1 - \psi_2)$ where $P(R_{12})$ is the probability that stimulus 1 is judged greater than stimulus 2, ψ_1 and ψ_2 are the subjective values of the stimuli, and F is the cumulative standard normal density function. Now suppose that $P(R_{ay;\,bz}) = .84$. Because .84 corresponds to a standard normal deviate of 1, it follows that $\psi_{ay} - \psi_{bz} = 1$, which implies that $h(a) + g(y) - h(b) - g(z) = 1$. Suppose also that $P(R_{bx;\,cy}) = .84$ then $h(b) + g(x) - h(c) - g(y) = 1$. It follows that $H(a) + h(b) + g(x) + g(y) - h(b) - h(c) - g(z) - g(y) = 2$ or $h(a) + g(x) - h(c) - g(z) = 2$, which implies the specific prediction that $P(R_{ax;\,cz}) = .98$, because F (2) = .98. Thus predictions can be checked against the obtained proportions by standard statistical techniques.

Chapter 4 shows how the method of maximum likelihood can be used to estimate parameters under ascending sequences of constraints to provide likelihood ratio tests to fit. Rather than fitting a model $P(R_{ax;\,by}) = F[g(a) + h(x) - g(b) - h(y)]$, Falmagne argues that it may be advantageous to fit the model $P(R_{ax;\,by}) = F(\theta_{ax;\,by})$ for the parameters $\theta_{ax;\,by}$ under constraints that specify the theory. The key idea is to use functional equations to define the polynomial rather than the polynomial itself. By combining the constraints of a theory of choice to those induced by the structure of the theory of combination, it should be possible to learn more about stimulus comparison and combination processes. Falmagne notes, however, that the development of axiomatic probability measurement theories still awaits a procedure that does not require an assumed form for F.

Any complete theory of psychophysics must deal with the issues of stimulus comparison, representation, combination, and contexual effects. The following chapters offer important contributions that advance the development of coherent psychological theory.

REFERENCES

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

Helson, H. *Adaptation-level theory*. New York: Harper & Row, 1964.

Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. *Foundations of measurement*. New York: Academic Press, 1971.

Restle, F., & Greeno, J. G. *Introduction to mathematical psychology*. Reading, Mass.: Addison-Wesley, 1970.