CHAPTER 1

PSYCHOPHYSICAL MEASUREMENT AND THEORY

J. C. FALMAGNE
New York University, New York, New York

CONTENTS

I am grateful to H. Levitt, M. Gizzi, and especially G. Iverson and M. Pavel for their comments on previous drafts of this material.
Based on *Elements of Psychophysical Theory*, by Jean-Claude Falmagne, copyright 1985, Oxford University Press. Used by arrangement with Oxford University Press.

OVERVIEW

LLOYD KAUFMAN
New York University, New York, New York

The first section of this *Handbook* focuses on conceptual and methodological issues that pertain broadly to the chapters of other sections. Many of the chapters of other sections also contain such material but in a form largely pertinent to the chapters themselves. The three chapters here are concerned with very general methods and techniques and, more important, with the theoretical underpinnings of the methods employed in other chapters. Thus this section examines the often unstated reasons for the methods used in fields as diverse as sensory psychophysics, cognition, and information processing.

The concepts described in the chapters by Falmagne and by Sperling and Dosher apply across virtually all the "boundaries" dividing the sections of this *Handbook*. Even though there are real differences between a study of the sensitivity of a sensory system and of, say, divided attention, scientists in both these areas often employ the same basic assumptions. Similarly, chapters in the section on information processing and those in the section on the perception of pattern and form are closely related to each other because both areas are strongly affected by a common set of conceptual tools.

Many of these conceptual tools are related to the problem of how one is to measure physical stimuli and patterns and relate those measures to the ways in which the organism transforms and responds selectively to attributes of the stimuli and patterns. This problem has a long and honorable history in psychology, and it is dealt with at length in Falmagne's chapter. In its more precise form it is often referred to as *Fechner's problem*, which can be described as the problem of finding a way to transform a scale of physical magnitudes so that they are proportionally related to psychological magnitudes. To understand this problem fully, one must grasp the notion of the *psychometric function*, dealt with at length in this first chapter of the *Handbook*. Falmagne bases his discussion on the theory of measurement and how it may be applied to measuring psychological phenomena. This introduction to measurement theory equips the reader to understand better the psychophysical methods and relationships discussed in later chapters.

Choices about method usually carry with them assumptions about underlying models and processes. For example, a simple forced-choice experiment implies a complex set of processes. The multiplicity and complexity of the models and assumptions entailed defy succinct summary. However, Falmagne's chapter will carry the reader a long way toward a fuller understanding of the implications of each choice about method.

The concepts of *threshold* and of *sensitivity* are given rigorous definitions, as are their relations to the now pervasive theory of signal detection. The latter theory comes to grips with the fact that all psychophysical tasks have cognitive components, and the assumptions we make about the guessing strategy adopted by the subject affect our interpretation of data. Falmagne provides the reader with a basic understanding of the theory of signal detection and the assumptions it entails. This portion of Falmagne's chapter serves as an introduction to the chapter by Sperling and Dosher that follows.

Sperling and Dosher provide a very general treatment of methods and theories designed to deal with the strategies employed by humans performing perceptual and cognitive tasks. The choice and sequencing of mental operations by subjects in the performance of many kinds of tasks may not be directly observable, but strong inferences can often be drawn using the methods and models discussed in the chapter. The authors demonstrate that these methods and theories are closely related to signal detection theory.

The authors choose to examine the application of general principles of decision optimization and resource allocation to mental processes occurring within relatively short periods of time. Signal detection theory is one application of the concepts of optimization to sensory-perceptual tasks. The *receiver operating characteristic* of signal detection theory is shown to be closely related to the *attention operating characteristic* and other *performance operating characteristics*. The authors also discuss how operating characteristics can be used to decide if independent resources are being tapped in a complicated task or if a single resource is being depleted by the several aspects of such a task. The discussion of concurrent and compound tasks in this chapter provides information that is essential to full understanding of the chapters by Welch and Warren, Gopher and Donchin, Wickens, and Moray (among others) in this *Handbook*. Despite the diversity of their substance, these chapters reflect a common need for basic methods, such as those dealt with here.

The chapter by Sperling and Dosher could have been placed in the section on human information processing, and this in fact was the original intention. However, it soon became clear that their message supplemented that of Falmagne and what they had to say was equally important to chapters dealing with topics as diverse as human performance, space perception, and pattern vision. Therefore, it was placed in this first section.

The chapter by Freeman is quite distinct from the other two chapters in this section. It does not address specific psychological or perceptual problems. We chose to include this chapter here because of the central role of the digital computer as a tool for the perception researcher. We recognize that the computer has changed the conduct of psychological inquiry, and it must be given the same type of treatment as is given to the physics of light in books on visual perception. Nearly all workers in perception and cognition employ computers and their graphics capabilities to produce the stimuli of their experiments. This chapter is designed to inform the reader about how computers are used to generate stimuli for perception research. It discusses methods of display and the ways in which lines and curves are generated, and it introduces the complexities of transformations and projections of images, all matters of vital concern to the perceptionist. It is also relevant to more practical matters, for example, the ability of today's computers to simulate scenes, such as those used in flight simulators.

The section of Freeman's chapter on stereoscopic displays is not as detailed as some of the other sections. The reason for this is that the details of stereoscopic display techniques are covered in the chapter by Arditi, and the reader wishing more information is referred to that source.

In conclusion, it should be noted that there was no single editor for this section. D. MacLeod, J. Thomas, M. Posner, K. Boff, H. Sedgwick, and L. Kaufman all contributed to the editorial process.

1. PRELIMINARIES

1.1. Outline of This Chapter

G. T. Fechner, the founder of psychophysics, was originally a professional physicist. At the age of 39, he turned to psychology and set out to apply the methods of experimental physics to the measurement of sensory events. To fully understand the details of Fechner's idea, as well as some of its difficulties, an excursion into physical measurement is necessary.

Section 2 contains a detailed description of the procedures that are the basis for the measurement of fundamental physical quantities, such as length or mass. The reader interested only in the applications of psychophysical models will be tempted to skip this section and may do so without much harm. However, we urge anyone striving for a solid understanding of the foundations of psychophysical measurement to study Section 2 carefully. A comparison between physical and psychophysical measurement is of interest for two major reasons.

First, measurement procedures proposed by Fechner for psychophysical phenomena result from a straightforward transposition of those applicable in physics. In both cases, the procedures are justified by a testable theory, and a detailed comparison of the two theories is instructive. Second, on the background of physical measurement, it is much easier to disentangle substantive from philosophical issues, the confusion of which has been an enduring plague in this field. A number of authors have contributed to a clarification of the foundations of psychophysical measurement, and key references are given in due place. The seminal role of R. D. Luce is recognized here, however.

Section 3 contains a description of Fechner's approach to psychophysics. The basic notion is that of a probability $P_{a,b}$ that stimulus a is perceived as exceeding stimulus b from the viewpoint of some sensory attribute. The theory justifying Fechner's procedure is that this probability only depends on the difference $u(a) - u(b)$, in which u is some unknown sensory scale, a candidate for a measure of "sensation." In symbols, this gives rise to the equation

$$P_{a,b} = F[u(a) - u(b)] , \qquad (1)$$

with F a strictly increasing continuous function. This equation occupies a central place in psychophysical theory. In fact, it would be only a mild exaggeration to say that a substantial part of psychophysical theory consists in comments on Eq. (1).

Section 4 reviews various discrimination models, many of which turn out to be special cases of Eq. (1). This section contains a discussion of the so-called law of comparative judgment of Thurstone, a particular instance of which (case V) is obtained when the function F in Eq. (1) is the normal integral.

The treatment of psychometric functions given in Section 5 may surprise the knowledgeable reader. It covers in detail a number of important questions often left to the intuition of the psychophysicist. Examples of such questions are, What does it mean to say that two or more psychometric functions are "parallel" or that they can be rendered so by a transformation of the physical scale? What is the relationship between "parallelism" and Eq. (1)? In a two-alternative forced-choice design, the probability $P_{a,b}$ is typically estimated by averaging the frequencies in the two alternatives. What is the theoretical impact, if any, of this standard practice, in particular with respect to Eq. (1)?

Sections 6 and 7 deal with the Weber functions Δ_π, the methods currently in use to estimate those functions experimentally (stochastic approximation, up–down), and a number of models proposed to explain typical data. Various generalizations of Weber's law are considered.

Section 8 is devoted to signal detection theory, which is presented so as to play down the notion that the subject is behaving as a statistician applying some optimal decision procedure. The prominent place usually given to this notion is misleading, in our opinion. However interesting, it is only a case of a general theory justifying a particular analysis of the so-called receiver operating characteristic (ROC) curves.

As suggested by the title of Section 9, the material there is rather mixed, discussing a variety of topics, among which are "probability summation," models for conjoint measurement (deterministic and probabilistic), bisection, and so on.

Finally, Section 10 is devoted to the many issues related to psychophysical scaling.

Although quite extensive, this chapter does not cover all the topics that its title could evoke. Two omissions among others are color theory, which is discussed by Pokorny and Smith, Chapter 8, and by Wyszecki, Chapter 9, and multidimensional scaling, which is also discussed by Wyszecki, Chapter 9.

In writing, we had in mind a reader with a minimal background in mathematics, corresponding to a few calculus courses and a *good* course in probability or statistics. Notions such as random variables, distributions, and expectations are assumed to be solid items of the reader's statistical equipment. A one-semester course in algebra may be helpful at some point but is by no means essential.

It is our firm belief that there has been in psychophysical theory a great deal more controversy than there was reasonable ground for. The attentive reader will notice a deliberate attempt to minimize the disputes and to give a unified presentation.

1.2. Key References

A useful complement to this chapter is a monograph by Gescheider (1976). As described in the preface, it is addressed to advanced undergraduate students with some background in statistics. This treatment differs from ours in that it covers a greater variety of empirical issues but pays less attention to the details of the mathematical aspects of the theories. Moreover, no attempt is made to cast psychophysical measurement in the general framework of measurement theory.

2. CONSTRUCTION OF A PHYSICAL SCALE FOR LENGTH (AN EXAMPLE OF EXTENSIVE MEASUREMENT)

2.1. Outline

By extensive measurement, we mean the measurement of fundamental physical quantities, such as mass or length, using qualitative devices. We shall give a concrete example. Consider a collection of thin rods. The problem at hand is the measurement of their length, but no rulers or other devices are available. A natural way of measuring the length of a given rod would involve the following steps:

1. Pick a particular, fixed rod as a "unit."

2. Count the maximum number of exact copies of this unit which can be placed along the rod to be measured without overlap.

The number so obtained is a measure of the length of the rod. If exact measurement is required, some refinements must be introduced. For the essentials, however, this algorithm is the usual one. The intuition supporting it is so compelling that it is at first difficult to realize (1) that quite a number of assumptions about physical reality are implicitly made and (2) that it involves a considerable amount of arbitrariness.

2.2. Notation

Our discussion of these issues will be facilitated by the adoption of a precise notation and terminology. The algorithm previously outlined can be analyzed into two distinct experimental procedures.

For any two rods a, b a *comparison procedure* is used to decide which of a, b has the greater length. The rods are placed alongside each other, in such a way that they coincide at one end. If they also coincide at the other end, we shall write

$$a \sim b \ .$$

Figure 1.1 illustrates the notions introduced in this subsection. The case where b covers a, but is not covered by it, will be denoted

$$a < b \ .$$

A more compact notation is also useful. Whenever either $a < b$, or $a \sim b$, we write $a \preccurlyeq b$. Thus

$$a \preccurlyeq b$$

simply means that b covers a (whether or not a covers b).

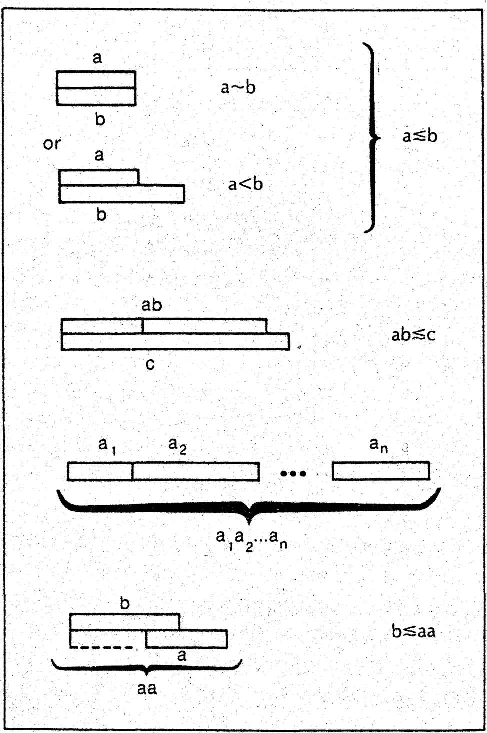


Figure 1.1. Measurement of length. Comparison and concatenation procedures.

The *concatenation procedure* for two rods a, b involves placing a and b end to end along a straight line, forming a new object, which we denote ab. Using the comparison procedure, this new object can be placed along some rod c to yield, for example, $ab \preccurlyeq c$.

When the rods of a sequence $a_1, a_2, \ldots, a_n$ are successively concatenated in the order: a_1 with a_2, then a_1a_2 with a_3, and so on, the result is denoted $a_1a_2 \ldots a_n$. A convenient abbreviation will be used to denote the successive concatenations of a with (exact copies of) itself. We shall write

$$n * a = \underbrace{aa \ldots a}_{n \text{ times}} \ .$$

Thus, in particular, $1 * a = a$, $2 * a = aa$, $b(3 * a) = baaa$, and so on. By convention, we shall admit that $n * a \sim n * a$, for $n = 1, 2, \ldots$. It will be convenient by extension to also refer to objects such as ab, aab, ... as rods. In the sequel, the letters x, y, z, ... will refer either to rods in the original sense or to objects resulting from some concatenation. Two methods for the measurement of the length of rods are described next.

2.3. First Method

Going back to the measurement algorithm proposed in Section 2.1, consider the task of measuring the length of some rod x. We pick arbitrarily some (small) rod y as a unit, and we form the successive concatenation of y with itself, until the following situation obtains

$$n * y \preccurlyeq x < (n + 1) * y \ . \qquad (2)$$

(In words: x covers $n * y$ but does not cover $(n + 1) * y$.)

We assign then the number n to x as its value on a scale measuring length, and we proceed similarly with the other rods in the collection. This method seems reasonable but encounters, in fact, a number of difficulties worth serious consideration, since they also occur in psychophysical application of the algorithm.

2.4. Difficulties

First, it is possible that $x < y$. In this case, we could assign the number 0 to x, but this would not be very satisfying. For example, there might be another rod x', such that $x \sim x' < y$, but $y < xx' < 2 * y$. In other words, both x and x' would have a scale value equal to 0, but xx' would have a scale value equal to 1. A very counterintuitive result! This shocking situation results from a general defect of the method; it is not very precise. When measuring the length of a rod by this method, we may commit an error, the size of which is smaller than the length of our unit, which by convention is equal to 1. The reason for this is the following. For any rod x, let us denote its true length by $l(x)$. We assume that, for any two rods z, w we have

$$z \preccurlyeq w \quad \text{iff} \quad l(z) \leqslant l(w) \ .$$

(We write *iff* for *if and only if*.) In the situation symbolized by Eq. (2), we obtain

$$l(n * y) \leqslant l(x) < l[(n + 1) * y] \ . \qquad (3)$$

The natural interpretation of the concatenation procedure for rods leads to the requirement that the length of a composite rod zw must be the sum of the length of z and w, that is,

$$l(zw) = l(z) + l(w) .$$

In general, for any sequence of rods $z_1, z_2, ..., z_n$ we must have

$$l(z_1 z_2 ... z_n) = \sum_{i=1}^{n} l(z_i) .$$

In particular

$$l(n * z) = n\, l(z) . \quad (4)$$

Going back to Eq. (3), this gives

$$n\, l(y) \leqslant l(x) < (n + 1) l(y) ,$$

which implies, since $l(y) = 1$ (y is the unit),

$$n \leqslant l(x) < n + 1 ,$$

that is,

$$l(x) = n + \gamma$$

with $0 \leqslant \gamma < 1$. Consequently, when we assign the number n to x as a scale value, we are making an error γ, the size of which is smaller than 1. Methods minimizing such an error—making it as small as one wishes—are not hard to design. One such method is considered in Section 2.5.

A second difficulty is that we have a priori no certainty that the method will work. Even if $y < x$, how can we be sure that by successively concatenating y with itself, we shall finally obtain Eq. (2) for some integer n? In the particular case analyzed here, considering the empirical interpretation of such expressions as $y < x$, $n * y$, and so on, it seems intuitively obvious that this will be the case. But where does this intuition come from? The answer is that we have learned from experience that the "physical" world around us satisfies a number of constraints, or "laws." An instance of such a law of immediate relevance to our discussion is the following:

Archimedean Axiom. For any rods x,z, we have either $x < z$, or there exists a positive integer n such that

$$n * z \leqslant x < (n + 1) * z .$$

This axiom is called *Archimedean*, since it evokes the so-called Archimedean property of real numbers: for any real numbers s,t with $t > 0$, there exists a positive integer n such that $s \leqslant nt$.

This method is based on the assumption that this law and various others are empirically true. In the sequel, such laws will be referred to as *axioms*. Another example of an axiom, intuitively consistent with the interpretation of the relation $\leqslant$ as meaning "is covered by," is the following:

Monotonicity Axiom. For any rods x,x',z,z', whenever $x \leqslant z$ and $x' \leqslant z'$, then $xx' \leqslant zz'$.

We stress the importance of this last axiom which, in some form or other, is the centerpiece of many axiom systems for extensive measurement. It would be tedious, and probably not very enlightening, to justify each step of the algorithm by the axiom or axioms on which it is based. It is sufficient for our purpose to remember that the algorithm described in Section 2.3, or its refinement, to which we turn in a moment, relies on axioms such as those exemplified here. In general, after a short reflection on their content, these axioms appear to be consistent with the reader's experience of the physical world (so that no experimental verification is required). We emphasize that this obvious character of the axioms will not extend into psychophysical applications of the algorithm.

2.5. Second Method

The precision of the algorithm described in Section 2.3 can be improved, provided that "fractions" of the "unit" can be used for the purpose of measurement. Since the unit is arbitrary, this means that any rod, no matter how small, can be divided more or less at will. A weak form of this notion is embodied in the following axiom.

Solvability Axiom. For any rod z, there exists a rod w such that $ww \leqslant z$.

In other terms, for any rod z, the formula $ww \leqslant z$ can always be solved for some rod w. The method based on this axiom requires more work than the preceding one and is based on a rather subtle idea, the details of which are worth careful study. (As before, our discussion will be heuristic; not all the axioms will be mentioned explicitly.) As a first step, we construct a "distinguished" sequence of shorter and shorter rods, as follows. We choose w_1 arbitrarily. Next, we pick w_2 such that $w_2 w_2 \leqslant w_1$, and so forth. In general, we shall have $w_n w_n \leqslant w_{n-1}$. Thus when n becomes large, w_n becomes shorter and shorter. In particular, if n is large enough, we can achieve $w_n \leqslant x$ and $w_n \leqslant y$, where x and y are, respectively, the rod to be measured and our "unit." Using the Archimedean axiom, we also know that

$$p_n(x) * w_n \leqslant x < [p_n(x) + 1] * w_n ,$$

$$p_n'(x) * w_n \leqslant y < [p_n'(x) + 1] * w_n$$

for some positive integers $p_n(x)$, $p'_n(x)$. (The index n in $p_n(x)$, $p'_n(x)$ is a reminder that these integers depend on the term w_n in the sequence; notice that $p'_n(x)$ does not depend on y, which is fixed.) Considering the (true) length of the rods involved in these expressions, we obtain

$$l(p_n(x) * w_n) \leqslant l(x) < l[(p_n(x) + 1) * w_n]$$

$$l(p_n'(x) * w_n) \leqslant l(y) < l[(p_n'(x) + 1) * w_n] ,$$

which implies, with $l(y) = 1$,

$$p_n(x)\, l(w_n) \leqslant l(x) < [p_n(x) + 1] l(w_n) ,$$

$$p_n'(x)\, l(w_n) \leqslant 1 < [p_n'(x) + 1] l(w_n) .$$

The basic idea is to use these inequalities to approximate the unknown quantity $l(x)$. Given w_n, the integers $p_n(x)$, $p'_n(x)$ are empirically determined (we can "compute" them). A little algebra involving some of the these inequalities permits the elimination of the bothersome quantities $l(w_n)$; we obtain

$$p'_n(x)\ l(x) < p_n(x) + 1 \ . \qquad (5)$$

A similar computation of the remaining inequalities yields

$$p_n(x) < l(x)[p'_n(x) + 1] \ . \qquad (6)$$

Combining Eqs. (5) and (6) and rearranging terms finally gives

$$p_n(x)/[p'_n(x) + 1] < l(x) < [p_n(x) + 1]/p'_n(x) \ , \qquad (7)$$

providing two bounds for $l(x)$. Let us investigate the situation when n becomes large; as indicated earlier, this means that w_n gets shorter and shorter. In turn, $p_n(x)$, $p'_n(x)$ must increase (a greater number of concatenations of w_n with itself are required to exceed x or y). In fact, when $n \to \infty$, we have both $p_n(x) \to \infty$ and $p'_n(x) \to \infty$. At this stage the consequences of this result on the two bounds in Eq. (7) are unclear. Fortunately, it can be shown that under the assumptions (i.e., axioms) underlying our discussion, the ratio $p_n(x)/p'_n(x)$ converges to some limit $\eta(x)$. This means, of course, that the two bounds in Eq. (7) converge to the same limit. That is,

$$p_n(x)/[p'_n(x) + 1] \to \eta(x) \ , \quad [p_n(x) + 1]/p'_n(x) \to \eta(x) \ ,$$

implying $l(x) = \eta(x)$. The outcome is that we can take either of the two bounds or $p_n(x)/p'_n(x)$ itself as a scale value of approximately $l(x)$, the approximation becoming increasingly accurate as n gets large. For example, we have

$$l(x) = [p_n(x)/p'_n(x)] + \gamma_{n'} \qquad (8)$$

with

$$-[p_n(x)/p'_n(x)] \times [p'_n(x) + 1]^{-1} < \gamma_n < 1/p'_n(x) \ .$$

(We leave it to the reader to check the algebra.)

Taking $p_n(x)/p'_n(x)$ as a scale value for x involves thus an error γ_n, the absolute value of which can be as small as required by practical or scientific applications.

At this point, the reader probably feels somewhat uneasy about the foundations of these methods. A proof of the key results, such as the convergence of $p_n(x)/p'_n(x)$, requires a more precise apparatus than was given here. In particular, a precise statement of all the axioms would be required. Such technical treatment of our subject is beyond the scope of this chapter, however.

Our aims in discussing these algorithms in such minute detail were as follows. We wanted to illustrate, with a minimum of formalism, the process by which qualitative observations, which are a typical outcome of an experiment, are progressively transformed into numerical statements regarding extensive measurement. This type of measurement is not only the most important example so far provided by science but also the cornerstone of various other types of measurement of interest to the psychologist. In particular, Fechner's enterprise must be regarded as an attempt to apply, in the context of psychophysical experiments, such algorithms to the measurement of sensory phenomena.

2.6. Representation Problem for Extensive Measurement

Reflecting on the position adopted so far in this section, it should be recognized that it is not devoid of obscurities. In particular, we discussed in detail two methods for the construction of a scale for length, without ever making exactly clear which problems such a scale was supposed to solve. At each step of these constructions it was somehow natural or obvious that this was the right course to follow. This approach leaves too many questions unanswered to be satisfying. Examples of puzzling questions are, What justifies the agreement existing in the scientific, as well as in the social, community that the scales obtained by such methods are appropriate? Is the agreement based on practical reasons, theoretical reasons, or both? Could a different scale have been used and, if so, under which conditions?

Here we shall take a more critical viewpoint regarding the methods of scale construction as previously discussed. Our manipulations involve two empirical procedures: the comparison procedure (symbolized by the relation $\preccurlyeq$), and the concatenation procedure (symbolized by writing xy for the two rods x, y). A scale is essentially a device by which the rods are represented by numbers. This suggests asking: *By which notions (operations, relations) of the real number system are we representing the two procedures?* Some hints were given earlier. In Section 2.4 it was argued that a function l defined on the set of rods and representing their true length, should be such that

$$x \preccurlyeq y \ \text{ iff } \ l(x) \leq l(y) \ . \qquad (9)$$

In the same context, it was also maintained that a natural interpretation of the concatenation procedure for rods would require that the length of xy should be equal to the length of x added to the length of y. Thus

$$l(xy) = l(x) + l(y) \ . \qquad (10)$$

In other words, the comparison procedure is represented by the ordering relation of the real numbers ($\leq$), and the concatenation is represented by the addition of the real numbers (+). Turning the question around leads to the following:

2.6.1. Representation Problem. Under which conditions does there exist a function l, defined on the set of rods and taking its values in the positive reals, such that Eqs. (9) and (10) are satisfied for all rods x,y?

A typical answer to this problem is a list, call it Λ, of conditions or axioms constraining the possible experimental results obtained from applying the two procedures. An example of such a list Λ would contain the monotonicity, Archimedean, and solvability axioms, plus some other conditions. The solution to the representation problem would then be given in the form of a representation theorem:

2.6.2. Representation Theorem. If all the axioms in the list Λ are satisfied, then there exists a function l mapping the set of rods in the positive real numbers, such that Eqs. (9) and (10) are satisfied for all rods x,y.

One proof of such a theorem is based on the following idea. We prove the existence of the function l by constructing it piecewise, so to speak. That is, we define $l(x)$ for every rod x, using essentially the second method described in Section 2.5. (Intuitively, the axioms are shown to imply that in Eq. (8), $p_n(x)/p'_n(x)$ converges and that $\gamma_n \to 0$. We define $l(x) = \lim_{n\to\infty} p_n(x)/$

$p'_n(x)$; thus, in particular, $l(y) = 1$.) Next, we show that the function l, as defined, satisfies Eqs. (9) and (10).

Note that, if *some* function l has been found to satisfy Eqs. (9) and (10), then any function l^* obtained by multiplying l by some constant $\alpha > 0$—that is,

$$l^*(x) = \alpha\, l(x)$$

for all rods x—also satisfies these formulas.

It is natural to ask whether *all* functions satisfying Eqs. (9) and (10) can be generated by this device. This question is of interest, since it corresponds to the situation commonly encountered; all usual scales for length are related by a multiplication, for example (approximately),

$$l_{\text{cm}}(x) = 2.54\, l^*_{\text{inch}}(x) .$$

When the axioms of a list Λ are sufficiently constraining, this situation obtains. The result is then formalized as follows.

2.6.3. Uniqueness Theorem. Suppose that all the axioms in a list Λ are satisfied. Let l, l^* be two functions satisfying Eqs. (9) and (10). Then, necessarily, $l(x) = \alpha\, l^*(x)$ for some constant $\alpha > 0$.

2.7. Summary and Remarks

The adoption of a measurement scale by the scientific community is a complex process, the various aspects of which have to be distinguished sharply. The formalism introduced in this section, with its central piece the representation problem, is standard in measurement theory. Its advantage is to make clear those aspects of the process which are susceptible to empirical verification.

Let us summarize. In the case of the measurement of the length of the rods in a collection, the scale was obtained by a succession of steps.

1. Two empirical procedures, comparison and concatenation, were chosen, more or less arbitrarily (no theoretical justifications were given).
2. A representation of each of these procedures by an entity (relation) of the real number system was adopted. The comparison procedure $\preccurlyeq$ was represented by the inequality ($\leqslant$) of the real numbers, and the concatenation was represented by the addition (+) of the reals.
3. The representation problem was formulated, involving the search for a positive-valued function l, satisfying for all rods x,y

$$l(x) \leqslant l(y) \quad \text{iff} \quad x \preccurlyeq y$$

and

$$l(xy) = l(x) + l(y)$$

4. A theory, that is, a list Λ of axioms, was proposed, implying the existence of the required scale l. This theory can be verified empirically. In particular, the validity of the monotonicity axiom:

whenever $x \preccurlyeq y$ and $x' \preccurlyeq y'$, then $xx' \preccurlyeq yy'$,

can in principle be checked.

5. A proof of the existence of the scale l was sketched, based on an algorithm (Section 2.5) permitting the construction of the scale within an arbitrarily small error.

At this stage, a scale for the measurement of length is available which, obviously, is the one commonly used. However, it must be realized that, in principle, the fact that steps 1–5 have been taken successfully does not guarantee that the resulting scale will be adopted for scientific or other practices. Other procedures could have been used, leading to a different scale. (An example will follow.) A consensus of the scientific community regarding a scale certainly requires the existence of a sound theoretical foundation, but it is also influenced by other considerations, such as, Is the scale convenient to construct and to use? Does it have the property of rendering the equations of models reasonably simple and intuitive?

Few people realize the extent to which the basic physical scales are arbitrary. As mentioned earlier, length, for example, could be measured by procedures essentially different from those discussed so far in this section, with no other consequences than that of rendering the writing of some physical laws more cumbersome and the actual application of the procedures more painful. We are not suggesting that mathematical or practical convenience is to be taken lightly. Clearly, however, neither of these has a bearing on "physical reality." It is of some importance for a psychophysicist to have a clear understanding of such facts. Ultimately, the measurement of sensation will rely on an agreement in the psychophysical community, based essentially on considerations of convenience. The alternative procedure for measuring length to be discussed, illustrates these remarks. (This example is due to Ellis, 1966.)

The comparison procedure for noncomposite rods is the same as before, but the concatenation differs. We write

$$ab \sim x$$

if the rods a, b, and x can be used to form a right triangle, with x as its hypotenuse, and a, b as the two other sides. Thus to check whether

$$ab \preccurlyeq cde ,$$

one forms, successively (see Figure 1.2),

$$ab \sim x ,$$

$$cd \sim y ,$$

$$ey \sim z .$$

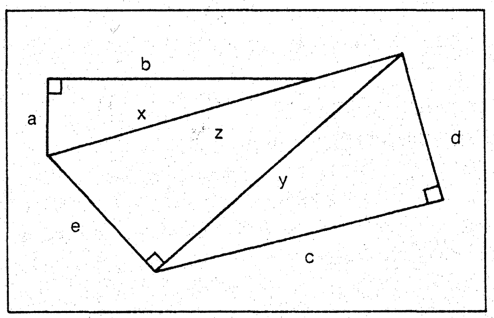


Figure 1.2. Alternative procedures for the measurement of the length of a collection of rods. We have $ab \preccurlyeq cde$, since $x \preccurlyeq z$ with $ab \sim x$, $cd \sim y$ and $ye \sim z$. The rods x, y, and z are hypotenuses of right triangles.

We define then

$$ab \preccurlyeq cde \quad \text{iff} \quad x \preccurlyeq z \ .$$

It is clear that these procedures cannot give rise to the same scale as the usual one. Nevertheless, as it turns out, all the axioms that would be satisfied in the usual case would also be satisfied here. (The reader may be tempted to check that the Archimedean, monotonicity, and solvability axioms are verified, and this may be moderately convincing. The key observation, however, is that for positive real numbers, addition is isomorphic to the operation $(s,t) \to (s^2 + t^2)^{1/2}$.) We may thus apply the representation theorem, and claim that there is some function f such that for all rods x,y,

$$x \preccurlyeq y \quad \text{iff} \quad f(x) \leqslant f(y)$$

and

$$f(xy) = f(x) + f(y) \ .$$

We emphasize that this concatenation is different from the usual one. We do *not* have $l(xy) = l(x) + l(y)$, where l is the usual scale. (A different notation could have been used to stress the fact that the two concatenations under consideration involve distinct empirical operations. We could have written, for example, $f(x\hat{\ }y) = f(x) + f(y)$.) From the viewpoint of physical reality, f is as defensible as l as a possible scale for length. The choice between l and f, or between the two concatenations, is in no way based on empirical data. It is clear that l is preferable because it is easier to construct empirically, and it renders the writing of physical laws somewhat easier. For a more detailed discussion of this practically minded, or positivistic, attitude toward measurement, the reader is referred to Ellis (1966), where several other empirical examples of extensive measurement will also be found. This raises the question of the relation between l and f. The answer is simple enough: $f(x) = l(x)^2$.

With minor adaptation (the details of which we shall not enter into here), the analysis of the measurement of length given in this section in terms of two empirical procedures (comparison and concatenation) is also applicable to the measurement of mass, using a two-pan, equal-arm balance. In this case, the experimenter has a collection of objects a,b ... and writes ab to signify that the objects a,b have been placed in the same pan of the balance. The experimenter also writes $cde \preccurlyeq ab$ if the pan containing c,d, and e does not stabilize itself at a lower level than the pan containing a,b. An examination of this situation indicates that essentially the same axioms will apply.

The examples of extensive measurement given so far in this section illustrate a type of measurement which is relatively well understood, one in which a measurement theory (a list Λ of axioms) is available, guaranteeing the existence of a representation of the empirical structure into the real number system, with a specified correspondence between the empirical procedures and some numerical relations. In this case, the scientist may feel relatively confident of the interpretation of the role played by a number assigned to an object by a scale, since the interpretation is based on an explicit theory.

Obviously, there are methods that are characteristically different for generating a measurement scale. For the measurement of mass, an example is provided by the spring balance, the readings of which could, in principle, be accepted a priori, without any theoretical justification. In fact, however, the numbers assigned by the spring balance are known, as a consequence of Hooke's law, to be proportional to those obtained through the two-pan, equal-arm balance, so that, indirectly, a measurement theory is available also in the case of the spring balance.

Whether a measurement theory can be dispensed with altogether is unclear. Some, such as S. S. Stevens's followers, would probably argue that this is the case. What cannot be disputed is that a measurement theory is a highly desirable rationale for any measurement scale designed to play an important role in the scientific formulation of the data.

2.8. Key References

Since Helmholtz (1887), various axiom systems for extensive measurement have been proposed. The discussion given here, even though it is only one of many possibilities, is representative of the mainstream of these theories. Generally, the axiom systems differ in the emphasis placed on side conditions or in the details of the representation. In most cases, these axiom systems deal with a deterministic situation. A probabilistic theory for extensive measurement has been presented by Falmagne (1980). A basic reference for this topic is Krantz, Luce, Suppes, and Tversky (1971) or, more recently, Roberts (1979).

3. FECHNER'S APPROACH TO PSYCHOPHYSICS

3.1. Construction of a Fechnerian Scale

Fechner's fundamental idea is that a sensory scale can be constructed by adapting, to a particular kind of sensory data, the standard measurement procedure for the measurement of length in physics. (We assume that, at a minimum, the material in Section 2.1 is familiar to the reader.) This is by no means obvious, and we shall proceed carefully.

Suppose that a, b, c ... are numbers representing, in conventional units, values of some physical magnitude, such as mass (or sound pressure, luminance, etc.). For simplicity, we shall refer to a, b, c ... as stimuli. Let $P_{a,b}$ be the probability that a subject, presented with the pair (a,b) of stimuli in some experimental paradigm, judges a at least as heavy as b. For the time being, consider only the data obtained for pairs (a,b) such that $a \geqslant b$. Let us assume that there exists a psychophysical scale, the properties of which govern important aspects of performance in this paradigm. Thus each stimulus a is mapped to a point $u(a)$ in this scale. We also assume that this mapping is order preserving (that is, $a < b$ iff $u(a) < u(b)$) and that $P_{a,b}$ is strictly increasing with the distance $u(a) - u(b)$ between the points representing a and b.

The following illustrative device is helpful: identify a and b in the pair (a,b) as names given to the two endpoints of a "rod"; a is the right endpoint, b the left endpoint. In fact, to stress the analogy with extensive measurement, the pairs (a,b) themselves, in this section, will be referred to as *rods*. Thus $P_{a,b}$ increases strictly monotonically with the length of the rod (a,b). We write

$$(a,b) \prec (c,d) \quad \text{iff} \quad P_{a,b} < P_{c,d} \ ,$$

$$(a,b) \sim (c,d) \quad \text{iff} \quad P_{a,b} = P_{c,d} \ ,$$

$$(a,b) \preccurlyeq (c,d) \quad \text{iff} \quad P_{a,b} \leqslant P_{c,d} \ .$$

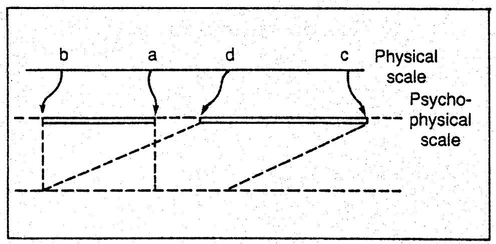


Figure 1.3. In a discrimination experiment, the two pairs of stumuli (a,b), (c,d) considered as rods; $(a,b) < (c,d)$.

Figure 1.3 summarizes the situation. We have thus a collection of rods and a comparison procedure represented by the relation $\leqslant$. This analogy with the extensive measurement of the length of rods discussed in Section 2 also suggests a concatenation procedure. For example, (a,b) concatenated with (b,c) should have a length equal to that of (a,c). We symbolize this fact by the formula

$$(a,b)(b,c) \sim (a,c) .$$

Thus a comparison of $(a,b)(b,c)$ with (d,e) is made possible by a comparison of (a,c) with (d,e). For example, if $(a,c) < (d,e)$, one concludes that

$$(a,b)(b,c) < (d,e) .$$

This, of course, is a special case. Generally, two rods to be concatenated need not have a common endpoint. A discussion of the more general situation, although quite straightforward, involves technical details and will be omitted here.

Let us proceed to construct a scale measuring the length of the rods (keeping in mind that the "length" of (a,b) is the distance between the stimuli a,b on some psychophysical scale). Suppose that we decide to use the method of Section 2.3 to measure the length of some rod (a,a'). We pick some rod (b_0, b_1) as a unit, together with any number of exact copies of that rod;

$$(b_0,b_1) \sim (b_1,b_2) \sim \ldots \sim (b_i,b_{i+1}) \sim \ldots .$$

We have thus by definition of the concatenation operation

$$(b_0,b_1)(b_1,b_2) \ldots (b_{i-1},b_i) \sim (b_0,b_i), \quad i = 1,2, \ldots .$$

The length n will be assigned to (a,a') (n being some positive integer) if

$$(b_0,b_n) \leqslant (a,a') < (b_0,b_{n+1}) ,$$

that is, if correspondingly, the probabilities satisfy the inequalities

$$P_{b_n,b_0} \leqslant P_{a,a'} < P_{b_{n+1},b_0} . \tag{11}$$

The number n is thus a measure of the length of the rod (a,a'), that is, of the distance between a and a' on the psychophysical scale. Let us apply this idea in an example. Suppose that the probabilities for all the pairs of stimuli in a set $\{a_0, a_1, a_2, a_3, a_4\}$ are given by the matrix

$$\begin{array}{c} \\ a_0 \\ a_1 \\ a_2 \\ a_3 \\ a_4 \end{array}\begin{array}{c} \begin{array}{ccccc} a_0 & a_1 & a_2 & a_3 & a_4 \end{array} \\ \begin{bmatrix} .5 & .75 & .80 & .90 & .95 \\ & .5 & .75 & .80 & .90 \\ & & .5 & .75 & .80 \\ & & & .5 & .75 \\ & & & & .5 \end{bmatrix} \end{array} \tag{12}$$

If we take (a_0,a_1) as a unit, this method leads us to assign the values in the matrix below, as measuring the distance between the points:

$$\begin{array}{c} \\ a_0 \\ a_1 \\ a_2 \\ a_3 \\ a_4 \end{array}\begin{array}{c} \begin{array}{ccccc} a_0 & a_1 & a_2 & a_3 & a_4 \end{array} \\ \begin{bmatrix} 0 & 1 & 2 & 3 & 4 \\ & 0 & 1 & 2 & 3 \\ & & 0 & 1 & 2 \\ & & & 0 & 1 \\ & & & & 0 \end{bmatrix} \end{array} . \tag{13}$$

We leave it to the reader to verify this in detail. (The values 0–3 follow from a straightforward application of the criterion represented by Eq. (11). The value 4 would require a refinement of that criterion.) The five stimuli can thus be represented as points on a straight line, say, with a_0 at the point 0, and the distance between a_i and a_{i+1} being constant, equal to 1.

Except for unessential details, this is Fechner's fundamental idea.

3.2. Remarks

In our deliberate emphasis of the relation between Fechner's scaling method and extensive measurement, we were led, for simplicity's sake, to make a somewhat unrealistic assumption. We supposed that any rod (a,b) could be "squeezed" between two rods (b_0,b_n) and (b_0,b_{n+1}), in the sense of Eq. (11). In practice, however, if a,b are far enough apart, the probability $P_{a,b}$ will be equal to 1 (or 0, if $b < a$) and will be unaffected by small changes of the values of a and b. This means of course that Eq. (11) cannot hold. A minor modification of the algorithm takes care of this difficulty, without altering the spirit of the method. We assign the number n to (a,b) if there exists a sequence a_0, a_1, ..., a_{n+1} such that

(i) $a = a_0 < a_1 < \ldots < a_n \leqslant b < a_{n+1}$.
(ii) $P_{a_{i+1},a_i} = .75$ for $0 \leqslant i \leqslant n$.

The distance $a_{i+1} - a_i$ is often referred to as a *just-noticeable difference* (jnd). Unfortunately, this term is used for a variety of closely related, but different, indices. To eliminate confusion, we shall reserve the term for one particular such index, which is defined in Section 7.4.

The Fechnerian method of scale construction described in Section 3.1 is an adaptation of the algorithm for the measurement of the length of rods outlined earlier (see the first method in Section 2.3). We have seen that such an algorithm lacks precision. In fact, it can be shown that a full psychophysical scale, one that would assign a scale value to each stimulus, could not be constructed using this method. A more sophisticated algorithm must be used, similar to the second method described in Section 2.5. This point was made by Luce and Edwards (1958).

Even assuming that an appropriate refinement of the algorithm is used, there is no guarantee that the method will holds for some strictly increasing, continuous functions u, F. Indeed, suppose that

$$P_{a,b} = \Phi\left[\frac{a - b}{(a + b)^{1/2}}\right] \tag{17}$$

for all positive real numbers a,b. (As customary, we denote by Φ the distribution function of a standard, normal random variable.) It is easy to check that, as defined by Eq. (17), the probabilities $P_{a,b}$ satisfy all the conditions of a balanced pychophysical discrimination system. (This verification is left to the reader.)

This model, however, is incompatible with Eq. (15). The reason for this is that the (functional) equation

$$P_{a,b} = \Phi\left[\frac{a - b}{(a + b)^{1/2}}\right] = F[u(a) - u(b)]$$

has no solution for the functions u, F. (That is, there are no functions u, F "solving" this equation; cf. Iverson, 1979). Thus additional conditions on the choice probabilities are needed, if Eq. (15) is to hold. Two such conditions are introduced in the next definition.

3.5.2. Definition. A psychophysical discrimination system (I,C,P) is called *Fechnerian* iff the equation $P_{a,b} = F[u(a) - u(b)]$ holds for some strictly increasing continuous functions u, F.

We say that (I,C,P) satisfies the *bicancellation condition* iff whenever $P_{a,b} \leq P_{a',b'}$, $P_{b,c} \leq P_{b',c'}$ and $(a,c),(a',c') \in C$, then $P_{a,c} \leq P_{a',c'}$.

A psychophysical discrimination system (I,C,P) satisfies the *quadruple condition* iff

$$P_{a,b} \leq P_{a',b'} \text{ iff } P_{a,a'} \leq P_{b,b'}$$

whenever all four probabilities are defined. The importance of the bicancellation condition has been emphasized earlier, in connection with a similar condition in extensive measurement (cf. Section 3.2). The relation between the three concepts in this definition is made clear in the theorem in Section 3.5.3.

3.5.3. Representation and Uniqueness Theorem. Let Ψ be a balanced psychophysical discrimination system. Then the following three conditions are equivalent:

(i) Ψ is Fechnerian
(ii) Ψ satisfies bicancellation
(iii) Ψ satisfies the quadruple condition.

Moreover, if any of these conditions is satisfied and (u, F), (u^*, F^*) are two pairs of functions satisfying Eq. (15), then $u^*(a) = \alpha u(a) + \beta$ and $F^*(s) = F(s/\alpha)$ for some constants $\alpha > 0$ and β.

The relation between the functions u and u^* in this theorem is sometimes expressed by stating that "u is an interval scale" (see Section 10 in this connection).

A full proof of this theorem would take many pages and is beyond the scope of this chapter. Some parts of this result are easy to obtain however, for example, the two implications (i) $\Rightarrow$ (ii) and (i) $\Rightarrow$ (iii). Assume, for example, that (i) holds, and let u, F be the two functions satisfying Eq. (15). Successively,

$$\begin{aligned} P_{a,b} \leq P_{a',b'} &\text{ iff } F[u(a) - u(b)] \leq F[u(a') - u(b')] \\ &\text{ iff } u(a) - u(b) \leq u(a') - u(b') \\ &\text{ iff } u(a) - u(a') \leq u(b) - u(b') \\ &\text{ iff } F[u(a) - u(a')] \leq F[u(b) - u(b')] \\ &\text{ iff } P_{a,a'} \leq P_{b,b'}, \end{aligned}$$

establishing the quadruple condition. The second implication, (i) $\Rightarrow$ (ii), is obtained by a similar method.

Placed on the background of the conditions defining a balanced psychophysical discrimination system, the bicancellation and the quadruple condition each constitutes a complete solution to the representation problem in Section 3.3.2. To put it another way: any model for choice probabilities $P_{a,b}$ satisfying either bicancellation or the quadruple condition can be put in the form of Eq. (15). In principle, these conditions can be tested experimentally. In practice, however, rather delicate statistical issues arise (cf. Iverson & Falmagne, in press).

The importance given in this chapter to Eq. (15) may surprise the reader. Actually, this representation has an impact beyond Fechner's scaling method. Many current models for choice probabilities are Fechnerian (in the sense of the definition in Section 3.5.2). As we shall see, these models differ in the specific assumptions made regarding the mechanisms of choice, which in turn determine the form of the function F in Eq. (15).

The critical issue remains of the status of the scale u, once it has been constructed. Does it make sense, as proposed by Fechner, to consider that such a scale measures the magnitude of the "sensation" evoked by the stimulus? We shall postpone this discussion for the moment (see Section 10).

3.6. Key References

The discussion of Fechner's scaling methods given here, even though perfectly compatible with Fechner's own presentation, was strongly influenced by the developments of measurement theory, as given, for example, in Krantz and colleagues (1971) or Roberts (1979). In this context, Fechner's problem is a case of difference measurement. The notions of a representation problem, representation theorem, and uniqueness theorem are standard in measurement theory.

This modern viewpoint regarding Fechner's enterprise is due to Luce and his collaborators (Luce, 1959a; Luce & Edwards, 1958; Luce & Galanter, 1963). The solution to the representation problem given here is mostly due to Doignon and Falmagne (1974; see also Falmagne, 1971, 1974). Related references are Levine (1971, 1972) and Krantz (1971). Eq. (15) also appears in the general context of choice theory, where it is dubbed the *strong utility model* (Luce & Suppes, 1965). The quadruple condition has been investigated by Marschak (1960) and Debreu (1960).

As indicated, statistical issues regarding the empirical testing of axioms such as bicancellation or the quadruple condition are discussed in Iverson and Falmagne (in press).

4. MODELS OF DISCRIMINATION

In Section 3 we considered a forced-choice paradigm, in which a subject is presented with pairs (a,b) of stimuli (a,b are real numbers, representing the stimulus values on some physical scale). The task is to select one of the two stimuli as exceeding the other, in terms of some subjective attribute, such as loudness or perceived weight, depending on the nature of the stimuli.

The basic theoretical notion was a probability $P_{a,b}$ that the subject chooses a over b. It was assumed that P is strictly increasing in a and strictly decreasing in b. A detailed theoretical analysis was made of the representation

$$P_{a,b} = F[u(a) - u(b)] \tag{18}$$

for these choice probabilities. In this equation, u and F are assumed to be real-valued, continuous, and strictly increasing functions, but are otherwise unspecified. Such a model says little regarding the details of the mechanism of choice. Certainly, the choice of a stimulus is the final stage of a complex process, involving physiological and psychological components. All these aspects are somehow captured by the functions u and F. This rather abstract viewpoint is open to criticisms, in particular regarding the interpretation of the functions u and F. Suppose, for example, that the subject is under time pressure. Say the choice response must be made within t sec after the presentation of the stimuli, with t varying across conditions (e.g., $t = 1, 3, 10$). Assuming that Eq. (18) holds in each condition, will the value of t affect u, F, both of these functions? Without a more explicit model, it is difficult to venture a guess. One could obviously *assume*, for instance, that only F will vary across conditions. However, some may feel uneasy about the (absence of) rationale for such a position. To take another example, suppose that the stimuli a,b ... are pure tones, presented on a background n of noise (say, n is the average sound pressure of a Gaussian noise). The values of n, if their range is chosen appropriately, will certainly affect the choice probabilities. Again, however, the impact of n on u or F is difficult to predict. In turn, one may argue, this uncertainty regarding the role of u and F in these experiments casts some doubt on the interpretation of u as a "sensation scale" (cf. Section 3).

This section is devoted to a discussion of a number of models consistent with Eq. (18). This means that a given model is either a special case of Eq. (18) (its assumptions imply a particular functional form for the function F) or has a special case that takes the form of Eq. (18), with F specified.

4.1. Random Utility Models

Let us assume that to each presented stimulus a, corresponds a random variable $\mathbf{U}_a$ symbolizing the effect of the stimulus on the subject's sensory apparatus. We also assume that a appears at least as intense as some other offered stimulus b if the sampled value of $\mathbf{U}_b$ does not exceed that of $\mathbf{U}_a$; formally

$$P_{a,b} = \text{Prob}\{\mathbf{U}_a \geqslant \mathbf{U}_b\} .$$

The distributions of the random variables $\mathbf{U}_a$ are unspecified.

In the literature of choice theory, this model is often referred to as the *random utility model* (Block & Marschak, 1960; Luce & Suppes, 1965; Marschak, 1960). Since no assumptions are made regarding the joint distribution of the random variables $\mathbf{U}_a$, one may ask whether this model sets any constraint on the data. Actually, it may be shown that if *some* collection of random variables $\mathbf{U}_a$ exists satisfying this model, then (in the case of a balanced system, cf. Section 3.5) we must have

$$1 \leqslant P_{a,b} + P_{b,c} + P_{c,a} \leqslant 2$$

for all stimuli a, b, and c (Block & Marschak, 1960). This is a rather weak condition, but one which can conceivably be rejected for some data.

This general model is consistent with the Fechnerian Eq. (18). In other words, under specific assumptions on the joint distribution of the random variables $\mathbf{U}_a$ and $\mathbf{U}_b$, Eq. (18) will be obtained. An example is given in Section 4.2.

4.2. Thurstone Law of Comparative Judgments

More specifically, suppose that $\mathbf{U}_a$ and $\mathbf{U}_b$ are independent and normally distributed, with respective means and variances $\mu(a)$, $\mu(b)$, $\sigma(a)^2$, $\sigma(b)^2$. Thus $\mathbf{U}_a - \mathbf{U}_b$ is normally distributed, with mean $\mu(a) - \mu(b)$ and variance $\sigma(a)^2 + \sigma(b)^2$. We obtain

$$P_{a,b} = \text{Prob}\{\mathbf{U}_a - \mathbf{U}_b \geqslant 0\} \tag{19}$$

$$= \Phi\{[\mu(a) - \mu(b)]/[\sigma(a)^2 + \sigma(b)^2]^{1/2}\} \tag{20}$$

where Φ is the distribution function of a unit normal random variable (i.e., a normal random variable with a mean equal to 0 and a variance equal to 1). Suppose, moreover, that the random variables have equal variances, say, $\sigma^2(c) = \alpha^2/2$ for all stimuli c. Then dividing by α in both the numerator and the denominator of Eq. (20), and writing $u(c) = \mu(c)/\alpha$, yields

$$P_{a,b} = \Phi[u(a) - u(b)] , \tag{21}$$

a special case of Eq. (18), with $F = \Phi$. The models embodied in Eqs. (20) and (21) are usually referred to as *cases III* and *V*, respectively, of Thurstone's *law of comparative judgment* (Thurstone, 1927a, 1927b; a very complete discussion of Thurstone's theory can be found in Bock & Jones, 1968). Thurstone case V has been given a special interpretation in a psychoacoustic context and has been applied to an impressive body of data by Durlach, Braida, and their coworkers (Braida & Durlach, 1972; Durlach & Braida, 1969; Jesteadt & Bilger, 1974; Jesteadt & Sims, 1975; Lim, Rabinowitz, Braida, & Durlach, 1977; Pynn, Braida, & Durlach, 1972).

4.3. Dropping the Normality Assumption

Notice that the normality assumption is not critical in the above discussion. Suppose that in Eq. (19) the random variables $\mathbf{U}_a$, $\mathbf{U}_b$ are independent and identically distributed except for a "shift" parameter. That is, suppose that for any stimulus c, $\mathbf{U}_c$ has the same distribution as $u(c) + \xi$, where u is a real-valued function and ξ is a fixed random variable. From Eq. (19), we have with ξ, ξ' independent and identically distributed

$$P_{a,b} = \text{Prob}\{u(a) + \xi - [u(b) + \xi'] \geqslant 0\}$$

$$= \text{Prob}\{\xi' - \xi \leqslant u(a) - u(b)\}$$

$$= G[u(a) - u(b)]$$

where G is the distribution function of $\xi' - \xi$. This is a special case of Eq. (18), generalizing case V of the law of comparative judgment.

4.4. Dropping the Constant Variance Assumption

The constant variance assumption used in the two preceding examples is not essential. Suppose that in Eq. (20) μ varies linearly with σ:

$$\mu(c) = \alpha\sigma(c) + \beta\ , \quad (22)$$

for some constants $\alpha > 0$ and β. Successively, from Eqs. (20) and (22)

$$P_{a,b} = \Phi\{\alpha[\sigma(a) - \sigma(b)]/[\sigma(a)^2 + \sigma(b)^2]^{1/2}\}$$

$$= \Phi\{\alpha[(\sigma(a)/\sigma(b)) - 1]/[(\sigma(a)/\sigma(b))^2 + 1]^{1/2}\} \quad (23)$$

Thus $P_{a,b}$ only depends on the ratio $\sigma(a)/\sigma(b)$. Defining

$$u(a) = \ln\ \sigma(a)\ ,$$

we rewrite the ratios $\sigma(a)/\sigma(b)$ in Eq. (23) as differences $u(a) - u(b)$, obtaining

$$P_{a,b} = F[u(a) - u(b)]\ , \quad (24)$$

where

$$F(s) = \Phi[\alpha(e^s - 1)/(e^{2s} + 1)^{1/2}]\ . \quad (25)$$

It is easy to check that F is strictly increasing. This model is sometimes referred to as *case VI* of Thurstone's law of comparative judgments (Bock & Jones, 1968; S. S. Stevens, 1959, 1966b). Again, the normality assumption is not essential in the above derivation.

4.5. A Timing Model

The linearity assumption, Eq. (22), linking mean and standard deviation of a random variable $\mathbf{U}_c$ may seem arbitrary. Actually, the above model arises quite naturally in psychoacoustics. Let a and b denote the sound pressure levels of two pure tones of the same frequency, say, 1000 Hz, presented successively and monaurally. Fairly detailed hypotheses will be made regarding the neural coding of physical sound intensity. We assume that a tone of level c applied in the auditory channel gives rise to a homogeneous Poisson process $L_t(c)$ of neural point events, with mean $\lambda(c)$. The interarrival times of these events (the interspike intervals) are thus independent and distributed exponentially, with expectation $\lambda(c)^{-1}$. Along lines explored by Luce and Green (1972, 1974a), suppose that a sample average $\mathbf{S}_{n,c}$ of these interarrival times is used as the basis for loudness discrimination (where n denotes the size of the sample). Stimulus a will be judged at least as loud as stimulus b if $\mathbf{S}_{n,a} \leq \mathbf{S}_{n,b}$; that is,

$$P_{a,b} = \text{Prob}\{\mathbf{S}_{n,a} \leq \mathbf{S}_{n,b}\}\ .$$

Since n can be assumed to be large ($n > 100$), $\mathbf{S}_{n(c)}$ is distributed very nearly normally, with expectation $\lambda(c)^{-1}$ and variance $\lambda(c)^{-2}/n$. The standard deviation is thus a linear function of the expectation, as in Eq. (22). We obtain

$$P_{a,b} = \Phi\{n^{1/2}[\lambda(b)^{-1} - \lambda(a)^{-1}]/[(\lambda(b)^{-2} + \lambda(a)^{-2}]^{1/2}\}$$

$$= \Phi\{n^{1/2}[(\lambda(a)/\lambda(b)) - 1]/[(\lambda(a)/\lambda(b))^2 + 1]^{1/2}\}\ ,$$

a special case of Eq. (23). In particular, Eqs. (24) and (25) follow with $u(a) = \ln \lambda(a)$ and $\alpha = n^{1/2}$.

4.6. An Extreme Value Model and the Strict Utility, or Logistic, Model

In the psychoacoustic paradigm used earlier, we suppose with Thompson and Singh (1967) that the neural coding of sound pressure is based on the combined effect of the stimulus on many independent, parallel channels. The sensory effect of a stimulus of level c in channel j, $(1 \leq j \leq n)$, is represented by a random variable $\mathbf{X}_{c,j}$.

We assume that $n(c)$ channels are triggered by stimulus c, the combined effect of which is represented by a random variable

$$\mathbf{U}_c = \max\{\mathbf{X}_{c,1}, \mathbf{X}_{c,2}, \ldots, \mathbf{X}_{c,n(c)}\}\ .$$

In words, the neural code of a stimulus c is the maximum of the excitation levels in $n(c)$ channels. As a basic equation specifying the choice probabilities, we have

$$P_{a,b} = \text{Prob}\{\mathbf{U}_a \geq \mathbf{U}_b\}\ .$$

It can be shown that if the random variables $\mathbf{X}_{c,j}$ are, with respect to c and j, independent and identically distributed (thus only the number of channels $n(a)$, $n(b)$ distinguishes the distribution of $\mathbf{U}_a$ from that of $\mathbf{U}_b$) and moreover satisfies some stability property, then we have approximately for large $n(a)$, $n(b)$,

$$P_{a,b} = \frac{n(a)}{n(a) + n(b)}$$

$$= \{1 + e^{-[\ln n(a) - \ln n(b)]}\}^{-1} \quad (26)$$

$$= F[u(a) - u(b)]\ ,$$

with $u(c) = \ln n(c)$ and

$$F(s) = (1 + e^{-s})^{-1}\ . \quad (27)$$

Taken by itself, Eq. (26) defines the *strict utility model* (Luce & Suppes, 1965), also called the *BTL* (Bradley-Terry-Luce) *system*, extensively investigated by Bradley (1954a, 1954b, 1955), Bradley and Terry (1952), and Luce (1959a) (cf. Suppes & Zinnes, 1963). Eq. (27) is the defining equation of the distribution function of a standard logistic random variable (Johnson & Kotz, 1970b, Chapter 22). This result, leading to Eq. (26), is due to Thompson and Singh (1967), based on extensive earlier work on the so-called extreme value distributions (Fisher & Tippett, 1928; Frechet, 1927; Gnedenko, 1943; Gumbel, 1958; von Mises, 1939). For some recent applications of these notions in choice theory and psychophysics, the reader is referred to Yellott (1977), and Wandell and Luce (1978), respectively.

4.7. Remarks

The diversity of these examples, which all lead to Eq. (18), justifies the central place given here to this equation. This diversity also carries an important lesson. In each of these examples, a key role is played for each stimulus c, by a basic random variable $\mathbf{U}_c$, formalizing the neural coding of the stimulus. The discrimination probabilities are symbolized by the equation

$$P_{a,b} = \text{Prob}\{\mathbf{U}_a \geqslant \mathbf{U}_b\} . \qquad (28)$$

Assuming that such a theoretical device is warranted and that the particular form of (the distribution function of) these random variables is taken seriously, it may seem sensible to assign a fundamental role to a central location index of these random variables. This would suggest adopting $E(\mathbf{U}_c)$—the expectation of the random variable $\mathbf{U}_c$—as a measure of the magnitude of the sensation evoked by the stimulus c. Notice, however, that $E(\mathbf{U}_c)$ does not necessarily coincide with $u(c)$ in Eq. (18). Such coincidence is obtained in Eqs. (20) and (21) but not in (22) and (23) (where we have $u(c) = \ln E(\mathbf{U}_c)$) and not, as we shall see, in our next model.

Thus even though Eq. (18) may play a fundamental role, the theoretical status of the scale u entering in this equation is not necessarily clear.

It is natural to ask, Are there reasonable models incompatible with Eq. (18)? The example in Section 4.8 provides an answer.

There is another lesson to be derived from these examples. Comparing the extreme value model Eq. (24) with the law of comparative judgment, case V, Eq. (20), it must be concluded that the mechanisms postulated are very different. Nevertheless, these models are extremely difficult to distinguish from an empirical viewpoint. The extreme value model predicts that the choice probabilities will satisfy the equation

$$P_{a,b} = F[u(a) - u(b)]$$

where F is the distribution function of a standard logistic random variable, while in Thurstone case V, the same equation is obtained, except that F is replaced by Φ, the distribution function of a standard normal random variable. It turns out that F and Φ are close approximations to each other (see Johnson & Kotz, 1970b, for details on this matter), so close, in fact, that choosing one model over the other by some empirical test is practically hopeless.

The reason for this paradox—drastically different assumptions but indistinguishable predictions—is that these models consist of very elaborate constructions concerning unobservable choice mechanisms for a relatively scarce data base. There are simply not enough data to support the edifice. This is especially true for the extreme value model.

It is certainly tempting to model the unobservable details of the choice mechanisms, and it may even be useful to do so, since this may provide insightful interpretations of the data and suggest useful experiments. The lesson is, however, that such detailed assumptions should probably not be taken too seriously, except in cases in which the data base is much richer, relative to the theoretical construction, than was assumed here.

4.8. A Neural Poisson Counting Model

Now let us consider a neural Poisson counting model incompatible with the equation $P_{a,b} = F[u(a) - u(b)]$. As in the psychoacoustic example in Section 4.6, suppose that a tone of level c generates a homogeneous Poisson process of spike events $L_t(c)$, of mean $\lambda(c)$. Suppose now, however, that intensity discrimination, rather than being based on the average spike intervals as in Section 4.5, relies on a count of the number of spikes during a fixed interval τ. Let $\mathbf{N}_a$, $\mathbf{N}_b$ be two random variables representing the number of spikes counted for each of the two stimuli a, b. Thus $\mathbf{N}_a$, $\mathbf{N}_b$ are two independent Poisson random variables, with expectations $\mu(a) = \lambda(a)\tau$, $\mu(b) = \lambda(b)\tau$, respectively. (We recall the variance of a Poisson random variable is equal to its expectation.) Assume further that

$$P_{a,b} = \text{Prob}\{\mathbf{N}_a \geqslant \mathbf{N}_b\} .$$

For large $\lambda(a)\tau$, $\lambda(b)\tau$, the random variables $\mathbf{N}_a$, $\mathbf{N}_b$ are nearly normal (Cramer, 1963, p. 250), yielding approximately

$$P_{a,b} = \Phi\{[\mu(a) - \mu(b)]/[\mu(a) + \mu(b)]^{1/2}\} .$$

This model, which, as far as we know, was proposed originally by Strackee and van der Gon (1962; see also Luce & Green, 1972, 1974a; McGill & Goldberg, 1968), is incompatible with Eq. (18); there are no (continuous, monotonic) functions μ, u, and F satisfying the equation

$$\Phi\{[\mu(a) - \mu(b)]/[\mu(a) + \mu(b)]^{1/2}\} = F[u(a) - u(b)] . \qquad (29)$$

The proof of this fact, based on a result due to Iverson (1979), will not be given here.

4.9. Remark on Statistical Testing

The models discussed in this section can be tested empirically by standard statistical techniques. A likelihood ratio method is sketched below for the logistic model, the principle of which is easily extended to other cases.

According to the logistic model defined by Eqs. (26) and (27), the choice probabilities must satisfy the equation

$$P_{a,b} = (1 + e^{-\theta_{ab}})^{-1} , \qquad (30)$$

with

$$\theta_{ab} = u(a) - u(b) . \qquad (31)$$

Notice that Eq. (31) implies—in fact, is equivalent to—the condition

$$\theta_{ab} + \theta_{bc} + \theta_{ca} = 0 , \qquad (32)$$

for all stimuli a, b, and c. In particular,

$$\theta_{aa} = 0 ,$$

$$\theta_{ba} = -\theta_{ab} .$$

There is a good reason for this reparameterization of the model. The new parameters θ_{ab} have to be estimated from the data, subject to the linear constraint, Eq. (32). This is a standard situation in statistics, which leads naturally to a likelihood ratio procedure. Let n_{ab} be the number of choices of stimulus a observed in the course of $n_{ab} + n_{ba}$ trials. Let Θ be the vector of all the parameters θ_{ab}. Under the usual conditions concerning the independence of trials, the likelihood of the data is the product

$$l(\Theta) = \prod_{(a,b)} (1 + e^{-\theta_{ab}})^{-n_{ab}}(1 + e^{-\theta_{ba}})^{-n_{ba}} . \qquad (33)$$

The unconstrained maximum likelihood estimates of the parameters θ_{ab} are given by

$$\hat{\theta}_{ab} = \ln(n_{ab}/n_{ba}) \ . \qquad (34)$$

(This corresponds to estimating the probabilities $P_{a,b}$ by their relative frequencies.) Let l_1 be the value of the likelihood function l in Eq. (33), when the parameters θ_{ab} are replaced by their unconstrained maximum likelihood estimates. Let l_2 be the value of the likelihood function l, when the parameters θ_{ab} are replaced by their maximum likelihood estimates, obtained under the linear constraint, Eq. (32). A classical result is that the ratio

$$-2 \ln(l_1/l_2)$$

is asymptotically (i.e., for a large number of trials) distributed as a chi-square random variable with a degree of freedom equal to the number of independent parameters remaining in l_2 (cf. Wilks, 1962, or any standard statistical text).

This procedure can be applied in principle to any model for binary choices, consistent with the Fechnerian equation

$$P_{a,b} = F[u(a) - u(b)] \ , \qquad (35)$$

in which the function F is specified exactly. This function being strictly increasing, its inverse F^{-1} exists, and Eq. (35) gives immediately

$$F^{-1}(P_{a,b}) + F^{-1}(P_{b,c}) + F^{-1}(P_{c,a}) = 0 \ , \qquad (36)$$

generalizing Eq. (32).

4.10. Key References

Some papers of general interest are Luce and Suppes (1965) and Luce (1977a, 1977b). Even though centered on applications in economics, the review paper by McFadden (1976) is a useful reference, in which special attention is paid to statistical matters. The book by Bock and Jones (1968) contains a very thorough discussion of Thurstone's discrimination models. Gumbel (1958) and Galambos (1978) are introductory texts on extreme value distributions. Other useful titles are listed below, organized by topics.

General Random Utility Models. Marschak (1960); Block and Marschak (1960); McFadden and Richter (1970, 1971), Manski (1977); Falmagne (1978).

Thurstone Law of Comparative Judgment. Thurstone (1927a, 1927b); Braida and Durlach (1972); Durlach and Braida (1969); Jesteadt and Bilger (1974); Jesteadt and Sims (1975); Lim, Rabinowitz, Braida, and Durlach (1977); Pynn, Braida, and Durlach (1972).

Extreme Value Model. Fisher and Tippett (1928); Frechet (1927); Gnedenko (1943); Thompson and Singh (1967); von Mises (1939); Wandell and Luce (1978).

Logistic Model; BTL Systems. Bradley (1954a, 1954b, 1955); Bradley and Terry (1952); Luce (1959a); Suppes and Zinnes (1963); Yellott (1977). (As indicated by its title, this paper of Yellott could also be placed in either of the two above categories.)

Timing and Counting Models. Luce and Green (1972, 1974a); Strackee and van der Gon (1962); McGill and Goldberg (1968).

The complete literature on probabilistic choice theory is huge, and the above list should not be taken as exhaustive. Only references of general interest, or having a potential relevance to psychophysics, were included.

Finally, a generally useful source for facts regarding the distribution function of commonly encountered random variables is Johnson and Kotz (1969, 1970a, 1970b).

5. PSYCHOMETRIC FUNCTIONS

Consider, for a fixed stimulus b, the probability $p_b(a)$ that stimulus a is judged as exceeding b. (Both b and a are in some real interval representing the physical scale.) A somewhat idealized graph of a function p_b, which is consistent in its main features with many data, is displayed in Figure 1.5.

Clearly, regarded as a function of two variables, $(a,b) \mapsto p_b(a)$ is (except for a change of notations) exactly the choice probability function $(a,b) \mapsto P_{a,b}$ analyzed in Sections 3 and 4. As we shall see, however, the change of notation is indicative of a change of viewpoint, which in turn leads to new theoretical insights.

Such a function p_b is traditionally referred to as a *psychometric function.* This term is also used in a different situation, when $p_b(a)$ denotes the probability of detecting a stimulus a embedded in some "noisy" background b. In other words, a and b may be different kinds of physical variables. Occasionally, we encounter the term in an even broader context, when the empirical measure under investigation is not a probability of discrimination or detection but of some other variable, such as a reaction time or a count of a neural spike firing. Our discussion will cover all these cases.

A central topic of this section will be whether the data support the assumption that two or more psychometric functions are "parallel," that is, can be made to coincide by rigid shifts along the horizontal axis. The rationale for this question is that parallelism is a criterion for an important class of model represented by the equation

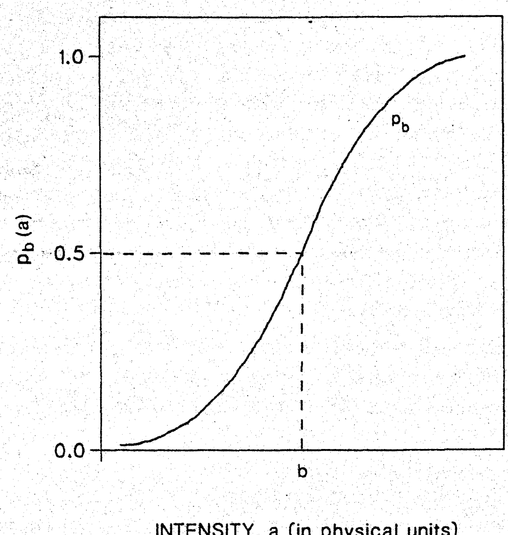


Figure 1.5. Idealized graph of a psychometric function.

$$p_b(a) = F[a - g(b)] , \quad (37)$$

in which the functions F and g depend on the particular model considered. In other words, any model satisfying this equation must predict parallel psychometric functions. The exact correspondence between Eq. (37) and parallelism will be described. A more general situation will also be investigated, corresponding to the equation

$$p_b(a) = F[u(a) - g(b)] . \quad (38)$$

In this case, the psychometric functions are not (necessarily) parallel but may be rendered so by some appropriate transformation u of the physical scale. Obviously, Eq. (38) generalizes the Fechnerian equation

$$P_{a,b} = F[u(a) - u(b)] \quad (39)$$

discussed at length in Sections 3 and 4. The importance of the issue of parallelism in psychophysical theory must be understood. Parallel psychometric functions indicate that the discrimination (or detection) acuity is uniform on the entire stimulus scale, a fact which may lead to adopting this scale as a measure of sensation magnitude.

Other topics are touched upon in this section. For instance, in the so-called two-alternative forced-choice (2AFC) design, the probability $P_{a,b}$ is often estimated by averaging the frequencies of the responses in the two alternatives. The theoretical consequences of this practice will be analyzed. It will be shown, for example, that it may have the unfortunate consequence of forcing nonparallelism.

We begin by considering a few empirical examples, leading to a basic definition.

5.1. Empirical Examples

5.1.1. Example. In an experiment reported by Engen (Kling & Riggs, 1971, p. 24), a subject was required to compare, by inspection, the length of two lines projected successively on a screen. In the course of the experiment, five lines of lengths 61, 62, 63, 64, and 65 mm were to be compared to a fixed line of length 63 mm. Thus in the above notations, b = 63 mm and a takes on five values. The pairs (a,b) of stimuli were presented randomly, with 100 trials per pair. On half of the trials, b was presented first. The subject was asked whether the perceived length of the first line exceeded that of the second. No feedback was given. Denote by $f_b(a)$ the relative frequency of the judgment that the perceived length of a exceeds that of b. The values of $f_b(a)$ are displayed in Figure 1.6. Such data are consistent with Figure 1.5 and suggest that p_b is a smooth function, strictly increasing on an interval bracketing b and such that $p_b(b)$ = .5. The method employed in this experiment is usually referred to as the method of *constant stimuli*, and the fixed stimulus b is called the standard stimulus.

5.1.2. Example. In an unpublished experiment of Graham and Hartline (1933; reported in Sirovich & Abramov, 1977), the frequency of spike firing of a single fiber in the lateral eye of the horseshoe crab, *Limulus*, was recorded as a function of the intensity of a visual stimulus for various monochromatic lights. The data (frequency of spike firing in the initial portion of the response immediately following the stimulus) are plotted in Figure 1.7, which is reproduced from Sirovich and Abramov

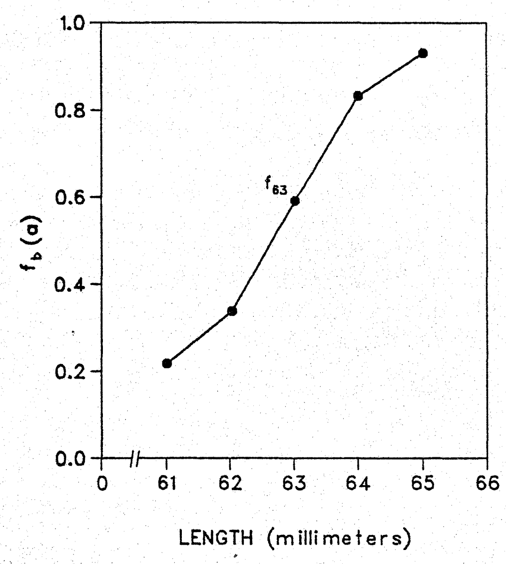


Figure 1.6. Proportion of "longer" judgment as a function of line length obtained with the method of constant stimuli. (From T. Engen, Psychophysics: Discrimination and detection, in J. W. Kling & L. A. Riggs (Eds.), *Experimental psychology* (3rd ed.). Copyright 1938, 1954, 1971 by Holt, Rinehart & Winston, Inc., CBS College Publishing. Reprinted with permission.)

(1977). The wavelength of the monochromatic light (in nm) is the parameter. It is clear that the five curves underlying the data in Figure 1.7 contain essentially the same information as traditional psychometric functions. Notice a difference, however, which concerns the ranges of the frequency of firing functions. As suggested by the data, these are real intervals bounded by, say, 0 and 80. This is easily taken care of. Any of a number of transformations would yield ranges bounded by 0 and 1. For example, with $\Psi_b(a)$ denoting the frequency of spike firing, for

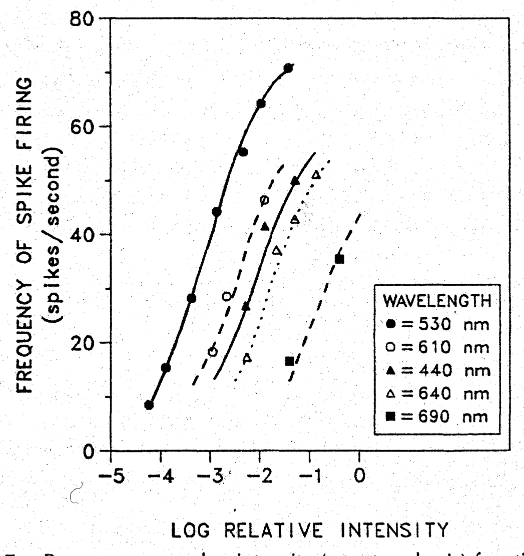


Figure 1.7. Response versus log intensity (quantum basis) functions from a single fiber in the lateral eye of the horseshoe crab, *Limulus*. Stimuli were monochromatic lights; the wavelength is indicated next to each curve. The response measure is frequency of spike firing in the initial portion of the response immediately following light onset. The five fitted curves are identical except for a shift along the abscissa. (From L. Sirovich & I. Abramov, Photopigments and pseudo-pigments, *Vision Research*, *17*. Copyright 1977 by Pergamon. Reprinted with permission.)

a stimulus b of intensity a, either of the two transformations below would be adequate:

$$\Psi_b(a) \mapsto p_b(a) = \Psi_b(a)/80$$

or

$$\Psi_b(a) \mapsto p_b(a) = \frac{\Psi_b(a)}{\Psi_b(a) + k}, \qquad (40)$$

where k is a positive constant. Such transformations would not affect an important property suggested by the data of Figure 1.7: the frequency of firing functions appear to be parallel, when plotted as functions of the logarithm of intensity. In fact, this parallelism would not be altered by any transformation

$$\Psi_b(a) \mapsto g[\Psi_b(a)],$$

where g is any continuous, strictly increasing function mapping the ranges of the functions Ψ_b into (0,1). For good reasons, much is made of this parallelism by Sirovich and Abramov, who point out that it supports (actually, is essentially equivalent to) the representation

$$\Psi_b(a) = R[a\mu(b)] \qquad (41)$$

where μ, R are real-valued functions, with R strictly increasing. The product $a\mu(b)$ is regarded as measuring the number of light quanta absorbed by the photoreceptor (cf. Naka & Rushton, 1966a, 1966b, 1966c). Notice that, with $p_b(a)$ as in Eq. (40) and $F(s) = R(e^s)/[R(e^s) + k$, Eq. (41) can be rewritten as

$$P_b(a) = F\{\ln a - \ln[1/\mu(b)]\},$$

a special case of Eq. (38).

In this example, a complete description of the stimulus involves a pair (b,a), where b denotes the wavelength and a the intensity. Thus the role of the standard in the example in Section 5.1.1 is played here by one coordinate of the stimulus. In the sequel, however, *background* will often be used as a generic term denoting the index of a psychometric function. For the sake of consistency, we shall also occasionally speak about the *masking effect* of the background even though such language refers only to particular applications.

5.1.3. Example. In the experiment described, Graham and Hartline also recorded the latency from light onset to first spike (see Figure 1.8). Most of the comments made concerning the example in Section 5.1.2 remain applicable here. To force the (average) latency $l_b(a)$ (where a,b are as in Section 5.1.2) into our theoretical framework, we can, to take an example among many, adopt the transformation

$$l_b(a) \mapsto p_b(a) = \frac{k}{k + l_b(a)}$$

where $k > 0$ is an appropriately chosen constant.

These examples pave the way to a general definition of a family of psychometric functions, in which the background (or index) is assumed to vary in some abstract set I, which may or may not be a real interval.

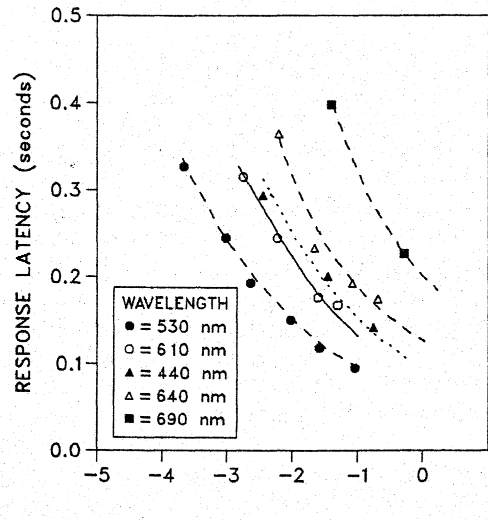


Figure 1.8. Response versus log intensity (quantum basis) functions from a single fiber in the lateral eye of the horseshoe crab, *Limulus*. Stimuli were monochromatic lights; the wavelength is indicated next to each curve. This set of records is identical to that of Figure 1.7, except that the response measure is the latency of the response from light onset to first spike. The five fitted curves are identical except for a shift along the abscissa. (From L. Sirovich & I. Abramov, Photopigments and pseudo-pigments, *Vision Research*, *17*. Copyright 1977 by Pergamon. Reprinted with permission.)

5.2. Psychometric Families—Definition

Unless one is interested in modeling their exact mathematical shape, psychometric functions are of little interest considered in isolation. Typically, the psychophysicist wishes to investigate how the shape of a psychometric function is affected by variation of the standard or the background. Accordingly, the definition in Section 5.2.1 is concerned with a *family* of psychometric functions. Notice the switch in notation, from $p_b(a)$ to $p_b(x)$, to emphasize that b, x may belong to different physical domains.

5.2.1. Definition. Let I be a set of backgrounds. For each background b in I, let C_b be a subset of the reals, and let p_b be a real-valued function defined on C_b. Suppose that, for some $b \in I$, the following axioms are satisfied:

1. C_b is an open interval.
2. $0 < p_b < 1$.
3. The function $x \mapsto p_b(x)$ is strictly increasing and continuous in the variable x.

Then p_b is called a *psychometric function*. The index b of a psychometric function p_b will be referred to as the *standard* or the *background*. A set $\{p_b | b \in I\}$ of psychometric functions is called *well linked* iff

4. For all a, $b \in I$ there exists a finite sequence $a_1 = a, a_2, \ldots, a_n = b$, such that

$$C_{a_i} \cap C_{a_{i+1}} \neq \emptyset, \quad \text{for } 1 \leqslant i \leqslant n.$$

A well-linked set of psychometric functions is called a psychometric family. Some comments on these conditions can be found in Section 5.3.

5.3. Remarks

Notice that, as defined in Section 5.2, a psychometric function resembles a distribution function (in the sense of statistics), but does not necessarily satisfy all the properties of this concept. Specifically, we do not require in general that a psychometric function take all the values between 0 and 1. Such property is not essential in most of our developments. More important, it would be a source of difficulty with various kinds of data.

The conditions defining a psychometric family should appear quite acceptable in many empirical situations. Axioms 1 and 2 are straightforward. Axiom 3 states that a psychometric function is strictly increasing and continuous. (This presupposes that the possibly constant upper and lower portions have been deleted.) This seems reasonable. (See, however, Falmagne, 1982.) The role of axiom 4 should be appreciated. This axiom states that any two psychometric functions can be linked by a finite sequence of psychometric functions, such that any two successive psychometric functions in the sequence have overlapping domains. This requirement is very natural from an empirical and especially a theoretical standpoint. A particular psychometric function provides precise but highly local information regarding the detectability (or discriminability) of the stimulus in a neighborhood of the stimulus scale. Axiom 4 ensures that these local informations can be pieced together to provide an overall picture of the subject sensitivity, for example, in the form of a psychophysical scale.

Examples of psychometric families are not difficult to manufacture, for example, by generalizing the models of discrimination discussed in Section 4.

5.4. Parallel Psychometric Families

Two empirical examples of "parallel" psychometric families were provided in Sections 5.1.2 and 5.1.3. Intuitively, a psychometric family is parallel if any two psychometric functions can be made to coincide by a horizontal "rigid" shift of one toward the other. This suggests that given one psychometric function, say, p_a, any other psychometric function p_b is completely characterized by the value of one parameter depending on b, which we denote by $g(b)$, expressing the length and direction of the rigid shift ($g(b)$ may be negative). This intuition is basically sound, but slightly misleading in its details. For instance, one or both of the psychometric functions p_a, p_b may be "truncated," and if both are, their truncation may be of a different kind, so that the coincidence after shift may not be complete (see Figure 1.9). The definition below takes care of this situation and is in fact consistent with a case in which for two particular psychometric functions p_a, p_b, no shift would achieve coincidence because the ranges of p_a, p_b do not overlap.

The concept of parallelism is of importance since it offers an easily testable criterion of the fact that the effects of the stimulus and the background combine "subtractively" (or "additively" as the case may be).

5.4.1. Definition. A psychometric family Ξ is called *parallel* iff for any two psychometric functions $p_a, p_b \in \Xi$,

$$p_a[p_a^{-1}(\pi) + \delta] = p_b[p_b^{-1}(\pi) + \delta] \tag{42}$$

for all $\pi \in (0,1)$ and $\delta \in$ Re such that both members of the equation are defined. (We recall that we write f^{-1} for the inverse function of a one-to-one function f; Re denotes the set of real numbers.)

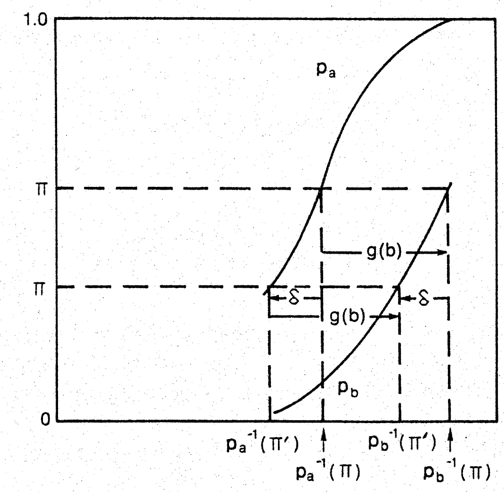


Figure 1.9. Two psychometric functions in a parallel psychometric family. The figure illustrates the notion of truncation, and the concepts of the definition in Section 5.4.1 and the theorem in Section 5.4.2. Notice that $g(b)$ is positive, and δ negative.

The simple result in Section 5.4.2 will help the reader to see the correspondence between this definition and Figure 1.9.

5.4.2. Theorem. A psychometric family Ξ is parallel iff for all $p_a, p_b \in \Xi$,

$$p_a^{-1}(\pi) - p_a^{-1}(\pi') = p_b^{-1}(\pi) - p_b^{-1}(\pi') \tag{43}$$

whenever all four terms are defined.

This means in particular that if p_a, p_b are distribution functions, they must have the same interquartile range:

$$p_a^{-1}(.75) - p_a^{-1}(.25) = p_b^{-1}(.75) - p_b^{-1}(.25) .$$

We include the proof of this result, which is very simple.

Proof. Suppose that Ξ is parallel, with

$$p_a^{-1}(\pi) - p_a^{-1}(\pi') = \delta, p_b^{-1}(\pi) = p_b^{-1}(\pi') = \delta' .$$

This implies

$$\pi = p_a[p_a^{-1}(\pi) + \delta] = p_b[p_b^{-1}(\pi') + \delta']$$

which, since Ξ is parallel, leads easily to $\delta = \delta'$.

Conversely, suppose that Eq. (43) holds whenever its terms are defined, but

$$\pi' = p_a[p_a^{-1}(\pi) + \delta] < p_b[p_b^{-1}(\pi) + \delta] = \pi'' .$$

Suppose also that $\delta \geq 0$. This implies that $\pi =< \pi' < \pi''$. Since the range of p_b is an interval, $p_b^{-1}(\pi')$ is defined, yielding successively

$$\delta = p_a^{-1}(\pi') - p_a^{-1}(\pi)$$

$$= p_b^{-1}(\pi') - p_b^{-1}(\pi)$$

$$< p_b^{-1}(\pi'') - p_b^{-1}(\pi)$$

$$= \delta ,$$

a contradiction.

The argument is similar in the case $\delta < 0$. ■

As mentioned earlier, the definition of a parallel psychometric family does not preclude the possibility that the ranges of some psychometric functions would not overlap. In a special case where such a situation does not arise, a useful representation of a psychometric family is available: the psychometric functions satisfy the equation

$$p_a(x) = F[x - g(a)] ,$$

for some functions F, g, where F is strictly increasing and continuous. This case is analyzed in the definition and theorem in Sections 5.4.3 and 5.4.4, respectively.

5.4.3. Definition. A psychometric family $\Xi = \{p_a | a \in I\}$ is called *anchored* iff there exists a number $\xi \in (0,1)$ such that:

(i) For all $a \in I$ there is an x satisfying $p_a(x) = \xi$.
(ii) For all $x \in \cup_{a \in I} C_a$, there is an $a \in I$ such that $p_a(x) = \xi$.

(We recall that C_a denotes the domain of the psychometric function p_a.) In words, conditions (i) and (ii) mean that for every background a there is a stimulus x and for every stimulus x there is a background a, such that $p_a(x) = \xi$. A number $\xi \in (0,1)$ satisfying these conditions will be called an *anchor* of Ξ.

These conditions are not very demanding. Suppose, for example, that the psychometric functions are defined from the choice probabilities $P_{a,b}$ of a balanced discrimination system (see Section 3.5.1) by the equation $p_a(b) = P_{a,b}$. It follows easily then that .5 is an anchor. Indeed, $p_a^{-1}(.5) = a$ is the identity function on I.

5.4.4. Theorem. An anchored psychometric family $\Xi = \{p_a | a \in I\}$ is parallel iff it has a representation

$$p_a(x) = F[x - g(a)] ,$$

where F is a continuous, strictly increasing function.

For a proof of this result, see Falmagne (1982). It must be realized that the property of parallelism of a psychometric family depends critically on the scale used to measure the stimulus and would not be preserved under nonlinear transformation of that scale. Consider, for example, an anchored, parallel psychometric family $\Xi = \{p_a | a \in I\}$ admitting a representation

$$p_a(x) = F[x - g(a)]$$

in the sense of the theorem in Section 5.4.4. Let v be a real-valued, strictly increasing, and continuous function defined on the interval of variation of x. Notice that, with $t = v(x)$, the equation $p_a^*(t) = p_a(x)$ defines a new anchored, psychometric family $\Xi^* = \{p_a^* | a \in I\}$. But Ξ^* need not be parallel. In fact, it is easy to show that Ξ^* is parallel if and only if v is a function of the form $v(x) = \mu x + \theta$, where $\mu > 0$ and θ is a constant. In general—that is, when v is not necessarily linear—the transformation of the stimulus scale generates a new psychometric family Ξ^* satisfying a subtractive representation

$$p_a^*(t) = F[u(t) - g(a)] . \qquad (44)$$

(Thus $u = v^{-1}$.) This suggests reversing the process. In Section 5.9.1 we ask, Under which conditions on a psychometric family does there exist a transformation of the stimulus scale which renders the psychometric functions parallel? Or in other terms, When does a psychometric family $\Xi^* = \{p_a^*\}$ have a subtractive representation of the form Eq. (44)?

5.5. Subtractive Families

5.5.1. Definition. A psychometric family $\Xi = \{p_a | a \in I\}$ is *subtractive* or a *subtractive family* iff there are three real-valued functions g, u, and F, the latter two being continuous and strictly increasing, such that

$$p_a(x) = F[u(x) - g(a)] \qquad (45)$$

for all $a \in I$ and $x \in C_a$. In such a case, we shall occasionally say that (g,u,F) is a *subtractive representation of* Ξ.

A special case of this representation has of course been encountered before, in the framework of a Fechnerian psychophysical discrimination system (definition in Section 3.5.1). It makes sense to adopt here a terminology consistent with the earlier one. Suppose, thus, that the psychometric family Ξ has in fact been obtained from a psychophysical discrimination system (I,C,P), through the equation

$$p_a(b) = P_{b,a} .$$

In this situation, Ξ will be referred to as a *discrimination family*, which will be called *balanced* iff (I,C,P) is balanced, that is, iff

$$p_a(b) + p_b(a) = 1 .$$

Thus when Ξ is a discrimination family, the functions g and u in Eq. (45) have the same domain. In the special case where $g = u$, Ξ will be called Fechnerian, or a Fechner family, and (u,F) will be labeled a Fechnerian representation of Ξ.

5.6. Remarks

A discrimination family $\Xi = \{p_a | a \in I\}$ can be subtractive without being Fechnerian. (Say, Eq. (45) is satisfied but u is not linearly related to g. An example is provided in Section 5.8). If, however, Ξ is a balanced discrimination family, then it is subtractive only if it is Fechnerian. Indeed, for all $a \in I$,

$$p_a(a) = F[u(a) - g(a)] = .5 ,$$

yielding, with $\alpha = F^{-1}(.5)$,

$$u(a) = g(a) + \alpha .$$

Defining $G(s) = F(s + \alpha)$, we obtain

$$p_b(a) = F[u(a) - gb)]$$

$$= G\{u(a) - [g(b) + \alpha]\}$$

$$= G[u(a) - u(b)] .$$

This indicates that our usage of the term Fechnerian is consistent with that in Section 3. (Notice that the above argument only uses the fact that $p_a(a) = .5$.)

5.7. A Remark on the Balancing Condition

Notice that if a discrimination family $\Xi = \{p_b\}$ is unbalanced, it can always be rendered balanced by a normalization such as

$$p_b^{**}(a) = p_b(a)/[p_b(a) + p_a(b)] .$$

More generally, any real-valued continuous function Ψ of two real variables, strictly increasing in the first variable and strictly decreasing in the second, satisfying

$$0 < \Psi < 1 , \quad (46)$$

$$\Psi(s,t) + \Psi(t,s) = 1 \quad (47)$$

achieves a similar normalization. The reader can check that the family $\Xi^{**} = \{p_b^{**}\}$ defined from the family Ξ by the equation

$$p_b^{**}(a) = \Psi[p_b(a),p_a(b)] \quad (48)$$

is indeed a balanced discrimination family. However, as demonstrated by the model in Section 5.8, it is not generally the case that if Ξ is subtractive, then the normalized family Ξ^{**} is subtractive. What is true, and easy to show, is that if Ξ is Fechnerian, then Ξ^{**} is also Fechnerian.

In some experimental situations, the order of presentation of the stimuli has an effect on the (probability of the) response. Such an effect is often of little interest, and the "careful experimenter" sometimes adopts a normalization procedure that suffers from the drawback just mentioned; namely, it does not necessarily preserve the subtractive character of a psychometric family. Let us demonstrate this. Denote by $n[a,(a,b)]$ the number of times stimulus a is chosen in the set $\{a,b\}$ when this set is presented in the order (a,b). Let $N(a,b)$ be the number of times $\{a,b\}$ is presented in the order (a,b). To simplify the argument, we identify probabilities and relative frequencies, in the sense that

$$p_b(a) = n[a,(a,b)]/N(a,b) .$$

The standard normalization is

$$p_b^{**}(a) = \{n[a,(a,b)] + n[a,(b,a)]\}/[N(a,b) + N(b,a)]$$

$$= (\tfrac{1}{2})[p_b(a) + 1 - p_a(b)] , \quad (49)$$

which indeed defines a balanced discrimination family $\Xi^{**} = \{p_b^{**}\}$, if $\Xi = \{p_b\}$ is a discrimination family. If we assume that both Ξ and Ξ^{**} are subtractive, then (by the remark in Section 5.6) Ξ^{**} is Fechnerian, and we must have for some continuous, strictly increasing functions u, g, F, h, and H,

$$p_b(a) = F[u(a) - g(b)] ,$$

$$p_b^{**}(a) = H[h(a) - h(b)] ,$$

which, together with Eq. (49), yields an equation of the form

$$F[u(a) - g(b)] - F[u(b) - g(a)] = K[h(a) - h(b)] \quad (50)$$

(where the constants ½ and 1 of Eq. (49) have been absorbed in the function K). An equation such as (50) is often referred to in the mathematical literature as a *functional equation*, a term suggesting that the unknowns in the equation are not numbers, as in elementary algebra, but functions (here F, u, g, K, and h). The point is that this equation severely restricts the relation between the functions u, g, and the form of the function F. In general, the normalization is ill-advised since a subtractive model will not survive it. In cases in which F is approximately linear, this normalization may not create difficulties, however.

5.7.1. Definition. In the sequel, any function $(p,p') \to \Psi(p,p')$ defined on the unit square $(0,1) \times (0,1)$, real valued, continuous, strictly increasing in the first variable, strictly decreasing in the second variable, and satisfying Eqs. (46) and (47) will be called a *balancing function*.

5.8. Examples of Subtractive Discrimination Families

Consider the family of functions $\Xi = \{p_b | b > 0\}$, defined by

$$p_b(a) = e^{-(b^{\eta}/a^{\mu})}$$

for each $a > 0$, where $\eta,\mu > 0$ are constants. This expression is closely related to a model frequently encountered in the vision literature (Green & Luce, 1975; Nachmias, 1981; Quick, 1974; see Watson, Chapter 6, and Olzak and Thomas, Chapter 7, in this handbook). It is easily checked that Ξ satisfies all the conditions of an unbalanced discrimination family, which is subtractive, since

$$p_b(a) = \exp[-e^{-(\mu \log a - \eta \log b)}] . \quad (51)$$

Let us balance Ξ. Since for every positive real number s, we have (denoting as usual by Φ the distribution function of a standard, normal random variable)

$$0 < \Phi(\log s) < 1$$

and

$$\Phi(\log s) + \Phi(\log \frac{1}{s}) = 1 ,$$

it follows that the function $(p,p') \to \Phi[\log (p/p')]$ is a balancing function. This yields the balanced family $\Xi^{**} = \{p_b^{**} | b > 0\}$, defined by

$$p_b^{**}(a) = \Phi\left\{\log\left[\frac{p_b(a)}{p_a(b)}\right]\right\} ,$$

that is

$$p_b^{**}(a) = \Phi[(ab)^{-\mu}(a^{\eta+\mu} - b^{\eta+\mu})] .$$

Since Ξ^{**} is balanced, the assumption that it is subtractive would lead (using the remark in Section 5.6), to the equation

$$(ab)^{-\mu}(a^{\eta+\mu} - b^{\eta+\mu}) = \Phi^{-1}\{G[u(a) - u(b)]\} , \quad (52)$$

in which u, G are strictly increasing, continuous functions. It is not difficult to prove that considered as a functional equation with unknown functions u, G, Eq. (52) has no solution (see

Falmagne, 1982). This shows that balancing a subtractive discrimination family does not necessarily yield a Fechnerian family.

5.8.1. Ideal Observer Model. Suppose that the background is a sample of so-called Gaussian noise, with power density b, presented for T units of time, and that the stimulus itself is also a sample of Gaussian noise of the same duration, of power density $x = b + \gamma$, with $\gamma \geq 0$, a constant. Each of the stimuli and the background is then a stochastic process which, to a good approximation (see however Levitt, 1972) admits a Fourier series representation

$$\sum_{k=1}^{WT} [\alpha_k \cos(2\pi kt/T) + \beta_k \sin(2\pi kt/T)]$$

where W is the bandwidth, and α_k, β_k are independent, normal random variables with mean 0, and variance σ_k^2 depending on the signal presented. It can be shown (e.g., Green & Swets, 1974) that in such a case the energy in the stimulus and the background are respectively distributed as

$$Wx\chi^2_{(2WT)},\ Wb\chi^{2'}_{(2WT)}$$

where $\chi^2_{(2WT)}$ and $\chi^{2'}_{(2WT)}$ are two independent chi-square random variables with $2WT$ degrees of freedom. Let us suppose that some (ideal) subject bases the decision on a comparison of the energies in the stimulus and the background. More precisely, we assume that

$$p_b(x) = \text{Prob}\{Wx\chi^2_{(2WT)} \geq Wb\chi^{2'}_{(2WT)}\} .$$

If $2WT$ is large, each of the two chi-square random variables is approximately normally distributed. Since $E(\chi^2_{(n)}) = n$ and $\text{var}(\chi^2_{(n)}) = 2n$, we obtain after simplification

$$\begin{aligned} p_b(x) &= \Phi\{WT(x - b)/[x^2WT + b^2WT]^{1/2}\} \\ &= \Phi\{(WT)^{1/2}[(x/b) - 1]/[(x/b)^2 + 1]^{1/2}\} \\ &= G(\log x - \log b) , \end{aligned}$$

with

$$G(s) = \Phi\{(WT)^{1/2}(e^s - 1)/(e^{2s} + 1)^{1/2}\} .$$

Thus $\{p_b\}$ is a subtractive family.

We recall briefly here the examples in Sections 5.1.2 and 5.1.3 concerning the frequency and latency of spike firing of a single fiber in the lateral eye of the horseshoe crab, *Limulus* (Graham & Hartline, 1933; see Sirovich & Abramov, 1977). With the logarithm of intensity in the abscissa, parallel psychometric functions were observed, which gave support to the assumption of a representation

$$p_b(a) = R[a\,\mu(b)]$$

for these psychometric functions (R, $\mu > 0$ are real-valued functions). Clearly, such representation is equivalent to a subtractive one, since it can be rewritten

$$\begin{aligned} p_b(a) &= R\{e^{[\log a - g(b)]}\} \\ &= F[\log a - g(b)] , \end{aligned}$$

with

$$F(s) = R(e^s) \quad \text{and} \quad g(b) = -\log \mu(b) .$$

In this case, parallel psychometric functions are obtained after a suitable transformation—here logarithmic—of the physical variable measuring the intensity of the stimulation. A generalization of this idea is considered in Section 5.9.

5.9. Representation of Subtractive Psychometric Families

5.9.1. Problem. Under which conditions does a psychometric family $\Xi = \{p_a | a \in I\}$ have a subtractive representation? This problem generalizes Fechner's problem, discussed in Section 3.3. Necessary conditions are not difficult to find; for example, suppose that Ξ is subtractive, with a representation (g,u,F), and that

$$p_a(x) \leq p_{a'}(x') \tag{53}$$

$$p_{a'}(y') \leq p_a(y) \tag{54}$$

$$p_b(y) \leq p_{b'}(y') \tag{55}$$

are simultaneously satisfied. Since the function F in the subtractive representation of Ξ is strictly increasing, this yields

$$\begin{aligned} u(x) - g(a) &\leq u(x') - g(a') \\ u(y') - g(a') &\leq u(y) - g(a) \\ u(y) - g(b) &\leq u(y') - g(b') . \end{aligned}$$

Adding these inequalities, we obtain

$$u(x) - g(b) \leq u(x') - g(b') ,$$

or equivalently, assuming that $x \in C_b$, $x' \in C_{b'}$,

$$p_b(x) \leq p_{b'}(x') . \tag{56}$$

5.9.2 Definition. A psychometric family $\Xi = \{p_b | b \in I\}$ satisfies *triple cancellation* iff Eqs. (53), (54), and (55) together imply (56) for all $a, a', b, b' \in I$ and $x, y \in C_a \cap C_b$ and $x', y' \in C_{a'} \cap C_{b'}$.

This condition is well known in the measurement literature (cf. Krantz et al., 1971). The above argument shows, thus, that a psychometric family has a subtractive representation only if it satisfies triple cancellation. A set of necessary and sufficient conditions, based on triple cancellation as a central axiom, was obtained by Falmagne (1982). A related result can be found in Narens and Luce (1976).

The scales u and g are usually specified up to a linear transformation. For example, the following uniqueness result follows from a slight strengthening of the conditions defining an anchored psychometric family: if (u,g,F) and (u^*,g^*,F^*) are two subtractive representations of the same psychometric family, then

$$\begin{aligned} g(t) &= \beta_0 g^*(t) + \beta_1 , \\ u(t) &= \beta_0 u^*(t) + \beta_1 + \beta_2 , \end{aligned}$$

$$F(t) = F^*\left(\frac{t - \beta_2}{\beta_0}\right),$$

for some constants $\beta_0 > 0$, β_1, and β_2.

5.10. Key References

The material in this section is based largely on a paper by Falmagne (1982), which contains a number of additional results. As far as we know, the term psychometric function is due to Urban (1907), even though the notion was in use since Fechner and Wundt. Despite its importance in psychophysical research, this notion has prompted exceptionally few theoretical investigations. Three papers by Levine (1971, 1972, 1975) deserve to be mentioned. His general approach to the analysis of a family of psychometric functions is similar to that of this section. Rather than focusing on particular models or processes, general conditions are sought that guarantee the existence and uniqueness properties of some abstract (e.g., subtractive) representation. His side conditions are somewhat different from ours, however. In his 1972 paper, Levine analyzes a problem that was not considered here, involving a generalization of the notion of a subtractive representation. In the notation of this section, this representation is symbolized by the equation

$$p_a(x) = F[k(a)u(x) - g(a)] .$$

An introduction to functional equations can be found in Aczél (1966).

6. WEBER FUNCTIONS—PSYCHOPHYSICAL METHODS

What is the smallest increment of a stimulus, on a physical continuum, which is detectable by a subject? In other terms, given a stimulus value equal to a, what is the smallest increment $\Delta(a)$ such that $a + \Delta(a)$ "just noticeably" exceeds a? This was one of the earliest questions raised by psychophysicists. This minimal increment $\Delta(a)$ is often referred to as the *just-noticeable difference* (jnd), or the *difference limen*. A variant—or rather, a special case—of this question is, What is the minimum value of a stimulus which is "just detectable" by a subject? This is called the *absolute threshold*.

Various experimental methods for the determination of $\Delta(a)$ have been designed and are described in this section. Such questions are by no means straightforward, however, since they are ambiguous. For example, what is meant by "just noticeably"? Suppose, for example, that $a + \Delta(a)$ is judged as exceeding a on 65% of the trials. Does that mean that $a + \Delta(a)$ just noticeably exceeds a? An empirical criterion is clearly involved here. In the method of constant stimuli (cf. Section 6.2.2) $\Delta(a)$ is often taken as a correct determination if $a + \Delta(a)$ is judged as exceeding a on 75% of the trials. (We are ignoring statistical issues for the moment.) The arbitrariness of this choice is troubling. This arbitrariness is less apparent, but just as critical, in the method "of limits" or in the method "of adjustments" (see Sections 6.1.1 and 6.1.2). Certainly, one would not want the general pattern of experimental results to depend critically on the choice of the criterion. In fact, as pointed out by Luce and Edwards (1958), there are theoretical difficulties involved in adopting a unique, fixed criterion. Accordingly, there is a trend in contemporary psychophysical research toward varying the value of the criterion across experimental conditions. We shall go back to this point later on.

A basic notion of this section is a function Δ of two variables,

$$(a,\pi) \mapsto \Delta_\pi(a)$$

with π, $0 < \pi < 1$, representing the value of the criterion. Thus in the particular case discussed above, $a + \Delta_{.75}(a)$ is judged as exceeding a on 75% of the trials. Notice that $\Delta_\pi(a)$ may be negative for some values of π: it is natural, for example, to expect that $a + \Delta_{.25}(a) < a$, at least in some experimental situations.

There is an obvious relation between the function of one variable $\pi \mapsto \Delta_\pi(a)$ and the psychometric function p_a, analyzed in Section 5. For instance, suppose that p_a is a psychometric function in a discrimination family Ξ (definition in Section 5.2.1), such that $p_a(a) = .5$. As illustrated in Figure 1.10, we have in such a case

$$\Delta_\pi(a) = p_a^{-1}(\pi) - a . \qquad (57)$$

In this situation, the value $p_a^{-1}(.50)$ is sometimes referred to as the *point of subjective equality*. The function Δ contains, thus, exactly the same information as the family Ξ of psychometric functions. The emphasis on this function here is justified, however. In particular, psychophysicists have found out that experimental plots of the functions Δ_π provided very revealing summaries of their data, and they use such plots routinely. Correspondingly, this function is of great theoretical interest, as we shall see in Section 7. An equally important place in our developments will be taken by the function

$$a \mapsto \xi_\pi(a) = p_a^{-1}(\pi)$$

(see Figure 1.10). Actually, it can be argued that ξ is a more central concept than Δ: ξ can always be defined from the psychometric functions, while Δ is only defined if the subtraction $p_a^{-1}(\pi) - a$ makes sense, which it does not if a is an object of a different nature than $p_a^{-1}(\pi)$. For instance, Δ would not be defined in a detection situation in which $p_a^{-1}(\pi) = x$ would specify the intensity x of a stimulus detected with a probability π, over a background of noise a, where a denotes a waveform or a spectral density function (i.e., a possibly infinite dimensional

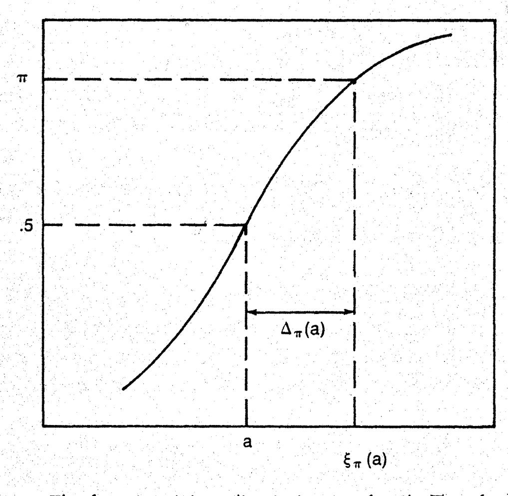


Figure 1.10. The function Δ in a discrimination family $\Xi = \{p_a\}$ satisfying $p_a(a) = .5$. We have $\xi_\pi(a) = p_a^{-1}(\pi)$ and $\Delta_\pi(a) = \xi_\pi(a) - a$.

vector). The functions Δ and ξ will be called, respectively, the *Weber function* and the *sensitivity function*.

Empirical determinations of function ξ or Δ can be achieved by a number of methods, a discussion of which is the topic of this section.

6.1. Traditional Psychophysical Methods

Since the interest of the method of limits and of the method of adjustments is mostly historical, only a brief description will be given. For more details, the reader is referred, for example, to Engen (1971) or Fechner (1860/1966). Each method involves a subject making successive comparisons of a stimulus with a standard or background. (We use the terminology of the preceding section.) In the determination of the absolute threshold, the value of the background is considered negligible.

6.1.1. The Method of Limits. The experimenter varies the value of the stimulus in small ascending or descending steps. At each step the subject reports whether the stimulus appears smaller than, equal to, or larger than the background. The experimenter records the values of the stimulus at which the subject's response shifts from one category to another. This method is used in applied situations, such as audiology, to provide a quick estimate of the point of subjective equality, $p_a^{-1}(.50)$. As pointed out by Levitt (1970), this method has serious defects from the viewpoint of efficiency (the observations may be poorly placed) and validity (the estimates may be substantially biased) (see Anderson, McCarthy, & Tukey, 1946; Brown & Cane, 1959).

6.1.2. The Method of Adjustments. The method of adjustments is similar to the method of limits. The subject adjusts the value of the stimulus, which can be varied continuously (e.g., by turning a dial), and sets it to apparent equality with the standard. Repeated applications of this procedure yield an empirical distribution of stimulus values, the variability of which is used to compute or estimate the jnd.

6.1.3. The Method of Constant Stimuli. The method of constant stimuli, which has been encountered earlier (Section 5.1.1), purports to estimate experimentally a number of suitably located points of some psychometric function p_a. If a particular mathematical expression is assumed for the psychometric functions (derived, for instance, from a mathematical model), then this expression is fitted to the experimental points. (Typically, the mathematical expression of p_a is only specified up to the values of some parameters, which have to be estimated from the data.) Finally, an estimate of the jnd is provided, for example, by Eq. (57).

In the past, a different estimate of the jnd has frequently been used, which corresponds to the equation

$$\text{jnd}(a) = (1/2)[\Delta_{.75}(a) - \Delta_{.25}(a)] \ . \tag{58}$$

The main objection to this procedure is that it is implicitly based on an assumption of symmetry between negative and positive differences, which is closely related to the balancing condition discussed in Section 5. The difficulties with such assumptions have been analyzed in Section 5.7.

If no specific mathematical model is assumed but the psychometric function appears to be approximately linear, say, between the values .20 and .80, then a straight line can be fitted to the experimental points in that interval, replacing the mathematical form used above.

In general, each of these three methods suffers from one or more of the following defects:

1. Absence of control on the criterion (Sections 6.1.1 and 6.1.2).
2. No theoretical justification for important aspects of the procedure (Sections 6.1.1 and 6.1.2).
3. The estimates may be biased (Sections 6.1.1 and 6.1.2).
4. Costs; a large amount of data is often wasted (all three methods).

In computerized laboratories, sophisticated versions of the method in Section 6.1.3 are used routinely, which we now describe. These methods are applicable when the exact mathematical form of the psychometric functions is unknown.

6.2. Adaptive Methods

Consider the problem of finding a point ξ in the domain of a psychometric function p_a, such that $p_a(\xi) = \pi$, where π is chosen arbitrarily in the range of p_a. Notice that the location of ξ depends on both a and π; ξ is thus a function of the two variables a and π. Actually, ξ is the sensitivity function introduced earlier, with $\xi_\pi(a) = p_a^{-1}(\pi)$ (Figure 1.10). In the rest of this section, we assume that the background a is fixed. We thus occasionally simplify our notation and write $\xi_\pi = p^{-1}(\pi)$.

It must be realized that the problem of estimating ξ_π with an acceptable degree of accuracy from the data is not trivial, since the exact mathematical form of the psychometric function may be unknown. A number of practical methods are described below. They differ from the methods decribed in Section 6.1 in that the course of the experiment depends critically on the data: the stimulus presented on trial n depends on the subject's responses on one or more of the preceding trials. At present, none of these methods taken by itself is completely free of defects. As indicated in Section 6.3, however, a suitable combination of methods provides an estimation procedure which seems to be reasonable for empirical applications.

From a theoretical standpoint, the sequence of stimulus–response pairs will be regarded as a stochastic process (Parzen, 1962). For the time being, we assume that the process is stationary. The following notations will be used. The stimulus presented at trial n will be denoted by $\mathbf{X}_n$, a random variable. The subject's responses will be coded:

1	If a is not judged as exceeding $\mathbf{X}_1$
0	Otherwise

Thus 0,1 are the values of a random variable, which we denote by $\mathbf{Z}_n$. We have by definition

$$\text{Prob}\{\mathbf{Z}_n = 1 | \mathbf{X}_n\} = p_a(\mathbf{X}_n) \ .$$

In the methods described below, the succession of stimuli is governed by an equation of the form

$$\mathbf{X}_{n+1} = \mathbf{X}_n + \theta(\pi, n, \mathbf{Z}_n, \mathbf{Z}_{n-1}, \mathbf{X}_{n-1}, \dots) \ , \tag{59}$$

in which θ is a function that may vary with the probability π assigned to the target value $\xi_\pi(a)$, the trial number n, the subject's response on that trial, and possibly some stimulus–response pairs on earlier trials.

6.2.1. Stochastic Approximation. Fix π, $0 < \pi < 1$ and choose a point x_1 arbitrarily, somewhere in the neighborhood of $\xi_\pi(a)$, the point to be estimated. (Since $\xi_\pi(a)$ is unknown, an educated guess has to be made. The accuracy of this guess is not crucial.) Present the pair (x_1,a) to the subject. Determine a second point x_2 by the following rule:

$$x_2 = \begin{cases} x_1 + \frac{c}{2}\pi & \text{if } \mathbf{Z}_1 = 0 \\ x_1 - \frac{c}{2}(1 - \pi) & \text{if } \mathbf{Z}_1 = 1 ; \end{cases}$$

where $c > 0$ is some constant, the choice of which is of importance, as we shall see. Thus x_2 is a value of the random variable $\mathbf{X}_2$. We have $\mathbf{X}_1 = x_1$ by convention. The above rule can be rewritten compactly as

$$\mathbf{X}_2 = \mathbf{X}_1 - \frac{c}{2}[\mathbf{Z}_1 - \pi] .$$

Next, we determine successively $x_3, x_4, \ldots, x_n, \ldots$ using the rule

$$\mathbf{X}_{n+1} = \mathbf{X}_n - \frac{c}{n}(\mathbf{Z}_n - \pi) . \tag{60}$$

This yields

$$\theta(\pi, n, \mathbf{Z}_n) = \frac{c}{n}(\pi - \mathbf{Z}_n) ,$$

in the notation of Eq. (59). The sequence of random variables $\{\mathbf{X}_n\}$ is known as a *Robbins-Monro process*. It can be shown that as n gets large, $\mathbf{X}_n$ tends to a normal random variable, with expectation equal to ξ_π and a vanishing variance. This result holds under general differentiability assumptions regarding the psychometric function p (which seem quite reasonable in the present context) and provided that the constant c is chosen appropriately. For details the reader is referred to Robbins and Monro (1951) or Wasan (1969). In practice, an estimate of ξ_π is provided by a sample value of $\mathbf{X}_n$ for some large n. This method is a substantial improvement over the preceding ones. It is not very economical, however, since a large number of trials are needed, only the last one of which is actually used. Moreover, if the number of trials is not large, the estimate of ξ_π is biased, the size and direction of the bias depending on the curvature of the psychometric function at the point to be estimated. One difficulty is that the convergence of c/n is slow, from the viewpoint of the scale of a psychophysical experiment. As suggested by Kesten (1958) and Pavel (Note 1), the convergence of the estimation process can be speeded up significantly by modifying the constant c *in Eq. (60) as a function of the subject's responses on trials preceding trial n.* For example, the value of c/n in Eq. (60) could fail to decrease in the case of a succession of identical responses. (We refer to this modification of the method as *accelerated stochastic approximation.*) Finally, it must be remembered that there are often practical limitations to the resolution of the apparatus used to generate the stimuli. In psychoacoustics, for instance, the minimum difference between distinct stimuli is often of the order of 0.25 dB or more. Even assuming that an accelerated stochastic approximation method is used, these limitations may suffice to render the estimate unacceptable. Stochastic approximation has nevertheless its use as an early component of an adaptive estimation procedure (see Section 6.3).

6.2.2. Up-Down, or Staircase, Method. This method is probably the most popular one. The essential difference with the stochastic approximation method is that on each trial the value of the stimulus is changed by a constant amount, either positively or negatively. In other terms, in Eq. (59),

$$|\theta(\pi, \mathbf{Z}_n, \mathbf{Z}_{n-1}, \mathbf{X}_{n-1}, \ldots)| = |\mathbf{X}_{n+1} - \mathbf{X}_n|$$

is constant for all trials n, the direction of the change depending on the probability π, on the subject's responses, and so on. The increments by which the stimulus is either increased or decreased are referred to as *steps.* A sequence of steps in one direction, in a realization of this process, is called a *run.* This is illustrated in Figure 1.11, in which the value of the stimulus presented at the first trial is set arbitrarily equal to 0 and the step size is equal to 1. There are eight runs, corresponding to trials 1–2, 2–5, 5–7, and so on. Three variants of the method will be described.

6.2.2.1. Simple Up-Down Method. In the *simple* up-down method, the problem is to estimate $\xi_{.5}$. As in Section 6.2.1, an educated guess is made for the initial value $\mathbf{X}_1$ of the stimulus. The successive remaining values are then obtained by the rule

$$\mathbf{X}_{n+1} = \mathbf{X}_n + \delta(1 - 2\mathbf{Z}_n) . \tag{61}$$

In words, δ is the step size, and the stimulus is increased by δ in the case of a negative response ($\mathbf{Z}_n = 0$) and decreased by δ in the case of a positive one ($\mathbf{Z}_n = 1$). In Figure 1.11, the succession of responses is "no, yes, yes, yes, no, ... " and so on. The choice of the step size is obviously important and will be commented on in a moment. Since

$$\text{Prob}\{\mathbf{Z}_n = 1|\mathbf{X}_n\} = p_a(\mathbf{X}_n) ,$$

it is apparent that Eq. (61) defines a discrete parameter Markov chain $\{\mathbf{X}_n\}$ with state space $\{x_1 \pm n\delta | n = 1,2, \ldots\}$. The states are recurrent, with a finite mean recurrent time, which implies (see, e.g., Parzen, 1962, p. 252) that the distribution of $\mathbf{X}_n$ converges as $n \to \infty$. In particular, taking expectations and limits in Eq. (61) and denoting the expectations by E, we obtain after rearranging,

$$0 = \lim_{n\to\infty} E(\mathbf{X}_{n+1}) - \lim_{n\to\infty} E(\mathbf{X}_n) = \delta[1 - \lim_{n\to\infty} E(\mathbf{Z}_n)] .$$

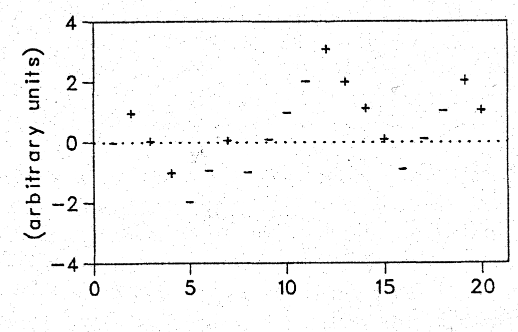


Figure 1.11. Exemplary data for the simple up-down method. The initial value is arbitrarily set at 0. The step size is equal to 1. There are eight runs, corresponding to trials 1–2, 2–5, 5–7, and so on.

Using the fact that p_a is a bounded, continuous function, this gives successively, with obvious notation,

$$.5 = \lim_{n\to\infty} E(\mathbf{Z}_n) = \lim_{n\to\infty} \text{Prob}\{\mathbf{Z}_n = 1\}$$

$$= \lim_{n\to\infty} E[\text{Prob}\{\mathbf{Z}_n|\mathbf{X}_n\}] = \lim_{n\to\infty} E[p_a(\mathbf{X}_n)]$$

$$= E[p_a(\mathbf{X}_\infty)] \approx p_a[E(\mathbf{X}_\infty)] ,$$

in which the approximation holds if we assume either that the psychometric function p_a is approximately linear in the region of concentration of the mass of $\mathbf{X}_\infty$ or that the distribution of $\mathbf{X}_\infty$ is approximately symmetric. (Indeed, in this last case, the expectation of $\mathbf{X}_\infty$ is confounded with its median $M(\mathbf{X}_\infty)$, and for any strictly increasing function f, we have $M[f(\mathbf{X}_\infty)] = f[M(\mathbf{X}_\infty)]$.) In principle, the value $\xi_{.5} \approx E(\mathbf{X}_\infty)$ can be estimated by the statistic

$$\frac{1}{k}\sum_{i=1}^{k} \mathbf{X}_{n+i} ,$$

for n sufficiently large. As pointed out by Wetherill (1963), a practical estimate of $\xi_{.5}$ is provided by averaging the peaks and the valleys in all the runs. As an illustration, the data of Figure 1.11 would yield for this estimate the value

$$\frac{1}{8}(0 + 1 - 2 + 0 - 1 + 3 - 1 + 2 + 1) = 3/8 .$$

It is easy to verify that this method amounts to considering the midpoint of every second run as an estimate of $\xi_{.5}$ and then to compute the average of these midpoints. Thus in Figure 1.11

$$\frac{1}{4}(-.5 - .5 + 1 + 1.5) = 3/8 .$$

These estimates are sometimes referred to as the *midrun estimates*. An even number of runs should be used, to reduce a bias in the estimation. The bias is then small, provided that to a reasonable approximation the psychometric function or the distribution of $\mathbf{X}_n$ satisfies the conditions indicated above. This procedure based on the midrun estimates is known to be fairly efficient. In fact for a small number of trials ($n < 30$), it is more efficient than a maximum likelihood estimate (Wetherill, Chen, & Vasudeva, 1966). There are various problems with this procedure, only some of which will be mentioned here (see Levitt, 1970.)

One problem concerns the choice of the step size δ, the value of which should be small compared to the "spread" of the psychometric function. As a rule of thumb, a good choice is to set δ equal to the slope of the psychometric function at the point to be estimated. (If we assume that the psychometric function is approximately linear in some neighborhood of the target value, then this value of δ can be shown to minimize the variance of the asymptotic distribution of the stimuli presented. See Wetherill, 1963.) Since both locations and spread are typically unknown at the early stage of experimentation, this recommendation is only of heuristic use. A frequently employed, reasonable procedure is to start the first few (say, 10) trials of each experimental session with a large step size, which is then decreased for the useful part of the data.

Another source of difficulty is that the subject may become aware of the systematic character of the stimulus changes. In turn, this may induce a strategy of anticipation of these changes that may be responsible for a bias in the responses. This is easily taken care of by "interleaving" two or more staircase processes (involving different estimates) within each experimental session. This remark also applies, obviously, to the stochastic approximation procedure.

It is clear that, as described here, the staircase procedure is only of limited use, since it only permits the estimation of the point $\xi_{.5}$.

*6.2.2.2. **Estimate of ξ.*** Following Derman (1957), the simple up-down procedure can be adapted to provide, at least in principle, an estimate of ξ_π for any choice of π. The idea is simple enough. From a given psychometric function p_a, let us define a new psychometric function p_a^* by

$$p_a^*(x) = \alpha p_a(x) ,$$

where α is a multiplicative constant, $.5 \leq \alpha \leq 1$, the role of which will be made clear in a moment. An application of the simple up-down method to p_a^* will yield a stimulus value ξ satisfying

$$\alpha p_a(\xi) = p_a^*(\xi) \approx .5$$

Thus

$$p_a(\xi) \approx 1/2\alpha ,$$

and for any π, $.5 \leq \pi < 1$, an appropriate choice of α will yield an estimate of ξ_π. In the style of Eqs. (59) and (61) this amounts to setting

$$\mathbf{X}_{n+1} = \mathbf{X}_n + \delta(1 - 2\mathbf{Z}_n\mathbf{Y}_\pi) ,$$

where $\mathbf{Y}_\pi$ is a random variable taking value 1 with probability $½\pi$ and value 0 with probability $1 - ½\pi$, and independent of the random variables $\mathbf{X}_n$'s. We have thus, clearly,

$$\alpha p_a(\mathbf{X}_n) = p_a^*(\mathbf{X}_n) = \text{Prob}\{\mathbf{Z}_n\mathbf{Y}_\pi = 1\} .$$

A similar method is used in the case of the determination of a point ξ_π with $0 < \pi < .5$. For example, we define a psychometric function

$$p_a^*(x) = (1 - \alpha)p_a(x) + \alpha ,$$

with

$$0 < \alpha = \frac{.5 - \pi}{1 - \pi} < .5 .$$

Again, applying the simple up-down procedure to p_a^* yields the required estimate of ξ_π. An objection to Derman's method is that the slope of p_a^* is smaller than the slope of p_a, a fact which may reduce the efficiency of the procedure.

*6.2.2.3. **Transformed Up-Down Method.*** The impact of this objection is less critical in the so-called transformed up-down method, where the function p_a^* is defined differently, for example, by one of the following expressions:

$$p_a^*(x) = p_a(x)^n \; ; \qquad n = 2, 3, 4, \ldots \tag{62}$$

$$p_a^*(x) = 1 - [1 - p_a(x)]^n \; ; \qquad n = 2, 3, 4, \ldots \tag{63}$$

$$p_a^*(x) = [1 - p_a(x)]p_a(x) + p_a(x) \, . \tag{64}$$

Such transformations have been used by a number of authors (see Levitt, 1970, for some references). As an illustration, we discuss the case $n = 2$ in Eq. (62). We consider the psychometric function $p_a^*(x) = p_a(x)^2$. As in the simple up-down procedure, we search for an estimate of a point ξ satisfying

$$.5 = p_a^*(\xi) = p_a(\xi)^2 \; ,$$

that is,

$$p_a(\xi) = \sqrt{.5} \approx .707 \; .$$

The case $n = 2$ in Eq. (62) is thus useful when this particular point of the psychometric function is of interest. The relevant stochastic process is defined as follows. Pick x_1 as usual. Set $x_2 = x_1$ if $\mathbf{Z}_1 = 1$ and $x_2 = x_1 - \delta$ if $\mathbf{Z}_n = 0$. For $n = 3, 4, \ldots$ we use the rule

$$\mathbf{X}_{n+1} = \mathbf{X}_n + \theta(\mathbf{Z}_n, \mathbf{Z}_{n-1}, \mathbf{X}_{n-1}) \; ,$$

in which the function θ is defined by

$$\theta(\mathbf{Z}_n, \mathbf{Z}_{n-1}, \mathbf{X}_{n-1}) = \begin{cases} \delta & \text{if } \mathbf{Z}_n = 0 \; ; \\ -\delta & \text{if } \mathbf{Z}_n = \mathbf{Z}_{n-1} = 1 \text{ and } \mathbf{X}_n = \mathbf{X}_{n-1} \; ; \\ 0 & \text{in all other cases} \; . \end{cases}$$

An example of realization of such a process is pictured in Figure 1.12(a). The point $\xi_{.5}$ of p_a^*, which is also the point $\xi_{\sqrt{.5}}$ of p_a, can be estimated by the midrun procedure. A slight adaptation of our definition of a run must be introduced however. For the data of Figure 1.12(a) a strict application of this definition leads to a count of four runs between trials 1 and 10, while we mean to have only two runs: 1–5, 5–10, with respective midpoints −1 and 0. The clearest approach is to begin by recoding the data, so as to eliminate the repetitions of a stimulus on consecutive trials. The function of this recoding is made transparent by a comparison of Figures 1.12(a) and 1.12(b). The exact definition given below is somewhat involved however. Let $\{x_n\}$ be a realization of the process $\{\mathbf{X}_n\}$. Consider the largest subsequence $\{x_{n_i}\}$ of $\{x_n\}$, such that $x_{n_i} \neq x_{n_{i+1}}$ for $i = 1, 2, \ldots$. Define $x_i^* = x_{n_i}$, for $i = 1, 2, \ldots$. The sequence $\{x_i^*\}$ will be called the *recoding* of $\{x_n\}$. An illustration of such recoding is given in Figure 1.12(b), starting from the data of Figure 1.12(a). By eliminating the repetitions, the number of trials has been reduced to 19. There are six runs: 1–3, 3–7, 11–15, and so on, with respective midpoints −1, 0, 0, and so on.

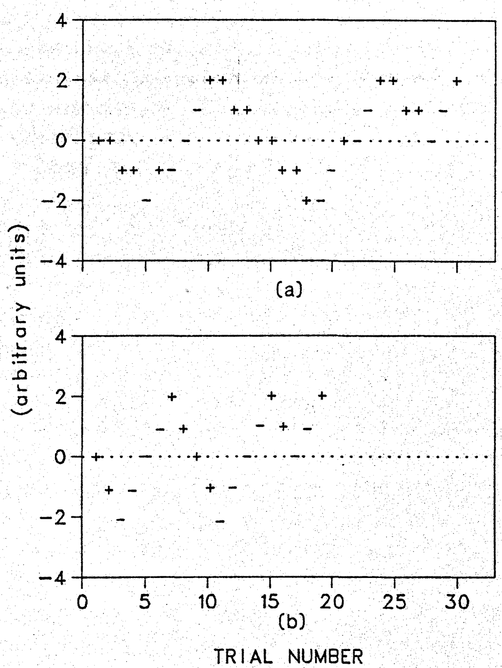


Figure 1.12. (a) Exemplary data for the transformed up-down methods, Eq. (62). The conventions regarding initial value and step size are as in Figure 1.11: $x_1 = 0$ and $\delta = 1$. (b) Recoding of the data of (a), eliminating consecutive repetitions of identical stimulus values. Six runs are obtained, corresponding to trials 1–3, 3–7, 7–10, and so on.

6.2.3. Remark. The assumption that the stochastic process $(\mathbf{X}_n, \mathbf{Z}_n)$ is stationary is critical for the procedures discussed in this section to be applicable. In some situations, the experimenter may have reasons to believe that this assumption is not warranted. An examination of the data generated by the up-down procedure may then reveal a systematic drift over time. If this happens, not only is the adaptive procedure useless for the estimation of ξ_π, but the very notion of psychometric function is of dubious value.

6.3. A Recommended Adaptive Procedure

In practice, it is advisable to adopt a combination of the methods described in Section 6.2. We recommend the following procedure. To estimate a point ξ_π satisfying $p_a(\xi_\pi) = \pi$:

Step 1. Choose $\mathbf{X}_1 = x_1$, the first stimulus to be presented, in a (conjectured) neighborhood of ξ_π.

Step 2. Determine the values of the following stimuli by accelerated stochastic approximation; for example, apply Eq. (60), modified by having c/n remaining constant in the course of a succession of identical responses. Pursue this procedure up to the limit of resolution of the stimulus continuum (e.g, 0.25 dB in psychoacoustic).

Step 3. Suppose that this limit is reached at trial n. On that trial, switch to a suitable up-down procedure, such as in Sections 6.2.2.2 or 6.2.2.3. Use the midrun estimates on the data from trial n onward to compute an estimate of ξ_π.

An example of application of this procedure is given in Figure 1.13 and Table 1.1 for some simulated data. This combined procedure avoids most of the criticisms elicited by other methods discussed in this section. We must point out, however, that it has not been investigated systematically, either from a mathematical or a practical standpoint. The last word is by no means said on the question of designing an optimal adaptive procedure, as indicated by recent activity in this field (Pavel, Note 1; Vorberg, Note 2).

6.4. Key References

A basic paper by Levitt (1970) contains a discussion of adaptive procedures geared toward psychophysical applications. A fairly complete mathematical treatment is available in the monograph by Wasan (1969).

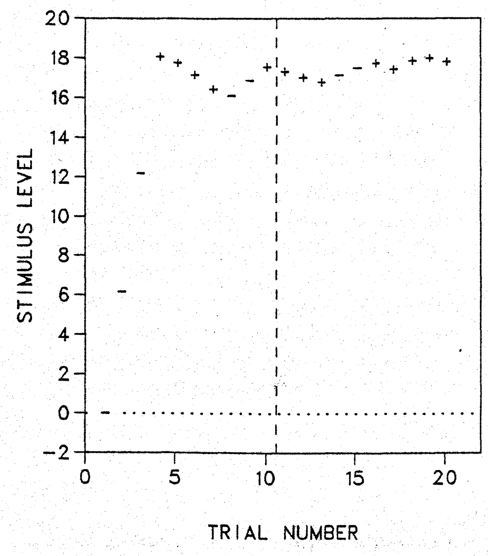


Figure 1.13. Simulated application of the adaptive procedure to estimate $\xi_{.75}$. The vertical dotted line separates the two modes of the procedure. See the Table 1.1 for details.

7. WEBER'S LAW, GENERALIZATIONS AND ALTERNATIVES

The central concept of Section 5 concerned a probability $p_a(\xi) = \pi$ for which a stimulus of intensity ξ is detected over a background denoted by a. In Section 6, the intensity ξ, rather than the probability π, was taken as the independent variable, and several practical procedures were discussed for the empirical determination of ξ, for given π and a. In general, the value of ξ will depend on both π and a; in other terms, ξ is a function of the two variables a and π. Mathematically, the function ξ contains exactly the same information as the psychometric functions of Section 5. The long-standing interest of psychophysicists in this function is well grounded, however. As we shall see, the knowledge of ξ gives a more ready access to the underlying sensory scale, a primary focus of interest for the psychophysicist. This is true at least for the most popular class of models for psychometric function data.

Table 1.1. **Simulation of the Adaptive Procedure Recommended in Section 6.3, to Estimate $\xi_{.75}$.**

Trial Number	Stimulus Value	Response ($\mathbf{Z}_n$)	$\mathbf{Y}_{.75}$	Computation of x_{n+1}
1	0	0	—	0 − 8(0 − .75) = 6
2	6	0	—	6 − 8(0 − .75) = 12
3	12	0	—	12 − 8(0 − .75) = 18
4	18	1	—	18 − (8/4)(1 − .75) = 17.5
5	17.5	1	—	17.5 − (8/4)(1 − .75) = 17
6	17	1	—	17 − (8/4)(1 − .75) = 16.5
7	16.5	1	—	16.5 − (8/4)(1 − .75) = 16
8	16	0	—	16 − (8/8)(0 − .75) = 16.75
9	16.75	0	—	16.75 − (8/8)(0 − .75) = 17.5
10	17.5	1	—	17.5 − (8/10)(1 − .75) = 17.3
11	17.25	1	—	17.25 + .25(1 − 2) = 17
12	17	1	—	17 + .25(1 − 2) = 16.75
13	16.75	1	0	16.75 + .25(1 − 0) = 17
14	17	0	1	17 + .25(1 − 0) = 17.25
15	17.25	0	1	17.25 + .25(1 − 0) = 17.5
16	17.5	1	1	17.5 + .25(1 −2) = 17.25
17	17.25	1	0	17.25 + .25(1 − 0) = 17.5
18	17.5	1	0	17.5 + .25(1 − 0) = 17.75
19	17.75	1	1	17.75 + .25(1 −2) = 17.5
20	17.50	1	1	17.5 + .25(1 −2) = 17.25

From trial 1 to 10, accelerated stochastic approximation is used, with $x_1 = 0$ and $c = 8$. (The successive values of the stimulus are obtained from Eq. (60), except that c/n does not vary in the course of repetitions of a response.) From trial 11 on, method 6.2.2.2 is used, with $\alpha = .5/.75$. The values of *$Y_{.75}$ are obtained by random sampling with Prob $\{\mathbf{Y}_{.75} = 1\} = \alpha$. The midrun estimate of $\xi_{.75}$ would be obtained by averaging 17, 16.75, 17.5; and so on.*

We consider here a number of models or properties for the function ξ and its close relative, the *Weber function* Δ, the most celebrated of which is the so-called Weber's law.

We begin with a definition recasting these two functions in the general framework of this chapter.

7.1. Basic Notions

7.1.1. Definition. Let $\Xi = \{p_a | a \in I\}$ be a psychometric family (i.e., a family of psychometric functions satisfying certain hypotheses; see the definition in Section 5.2). The *sensitivity function* of Ξ is a function ξ defined for all backgrounds (or standards) a and all probabilities π in the range of a psychometric function p_a, by the equation

$$\xi_\pi(a) = p_a^{-1}(\pi) .$$

(As usual, we write f^{-1} for the inverse of a one-to-one function f.) In words, $\xi_\pi(a)$ is the intensity of the stimulus yielding a response probability π, for the background a; that is, $p_a[\xi_\pi(a)] = \pi$. We recall that the index a in the notation p_a of a psychometric function need not always represent a real number. An example is that of a detection paradigm in which the index may denote a background noise, which in some situations may be represented by a spectral density function, that is, an infinite dimensional vector. (Thus $p_a(x)$ is the probability that the stimulus of intensity x is detected over the background a.) Let us assume that Ξ is anchored at .5 (cf. Section 5.4.3; thus .5 is a possible value for all the psychometric functions in the family). Then the Weber function of Ξ is a function Δ of two variables a and π, defined by

$$\Delta_\pi(a) = \xi_\pi(a) - \xi_{.5}(a) .$$

This was illustrated in Figure 1.10, in the particular case in which a denotes a real number (say, a physical intensity) and $\xi_{.5}(a) = a$.

It is important to distinguish in our notation the concept of Δ from that of Δ_π. The latter is a function of one variable, namely, the background, or standard. In other words, in the notation $\Delta_{.75}$, the probability .75 is implicitly assumed to be fixed. Occasionally, it will nevertheless be convenient, by abuse of language, to refer to the functions Δ_π as the Weber functions of Ξ. A similar convention will apply to the functions ξ_π, which will be called the *sensitivity functions* of Ξ. Psychophysicists often analyze their data in terms of one or more functions

$$a \mapsto \Delta_\pi(a)/a ,$$

in a situation in which division by a is legitimate. Such a function will be called the π-*Weber fraction* of Ξ, or, more simply, when no ambiguity can arise, a *Weber fraction*. Notice that since, in a case where $p_a(a) = .5$ for all intensities a, we have by definition

$$\frac{\Delta_\pi(a)}{a} = \frac{\xi_\pi(a)}{a} - 1 ,$$

any property of a Weber fraction will be almost exactly reflected in the corresponding function $\xi_\pi(a)/a$.

7.1.2. Remark. The change of notation, from $p_a^{-1}(\pi)$ to $\xi_\pi(a)$, symbolizes an important shift of focus in our analysis. The quantity π, the probability of the response, ceases to be the variable of interest and becomes the parameter. Typically, at most a couple of values of π are considered in experimental plots of Weber functions or sensitivity functions. By contrast, the effect on $\Delta_\pi(a)$ or $\xi_\pi(a)$ of the variable a is investigated in minute detail. This is in line with a long tradition in psychophysical research in which the sensory scales uncovered by the analysis of the data are deemed of central importance. This point is critical and should be discussed in some detail.

Suppose, for example, that some psychometric family $\Xi = \{p_a\}$ is subtractive, in the sense of the definition in Section 5.5.1. This means that the following representation holds for the response probabilities:

$$p_a(x) = F[u(x) - g(a)] , \tag{65}$$

in which u and F are continuous and strictly increasing functions. The psychophysicist using such a model typically interprets the functions u, g as representing a rescaling of the physical variables by the sensory mechanisms. As such, these functions are far more important than the function F, which, it is feared, may be plagued by nuisance variables of the "cognitive" type (response bias, motivation, etc.).

Let us transform Eq. (65) in terms of the sensitivity function ξ. Setting $p_a(x) = \pi$ and $F^{-1} = h$, we obtain $\xi_\pi(a) = x$, which together with Eq. (65) yields

$$\xi_\pi(a) = u^{-1}[g(a) + h(\pi)] . \tag{66}$$

Consequently, if the variable π in Eq. (66) is kept constant, the resulting equation in one variable only involves the functions u, g, which for reasons given are the interesting ones.

Such is the strategy of the psychophysicist. It relies heavily on a few assumptions. One is that the sensitivity functions ξ_π can be determined empirically with enough accuracy. A number of methods designed for this purpose have been discussed in the preceding section. Another, more critical assumption is that the rescaling functions u, g in Eqs. (65) and (66) are unaffected by nuisance (i.e., nonsensory) variables. As far as we know, there is little experimental evidence suggesting that this assumption may be invalid.

As a by-product of our discussion, we have, in any event, the following theorem.

7.1.3. Theorem. Let ξ be the sensitivity function of a psychometric family $\Xi = \{p_b | b \in I\}$. Then Ξ is subtractive (that is, Eq. (65) holds) iff there exist three functions h, u, and g, with u and h strictly increasing and continuous, such that

$$\xi_\pi(a) = u^{-1}[g(a) + h(\pi)] .$$

Indeed, we have shown that Eq. (65) implies Eq. (66), and it is clear that the reverse implication also holds. As suggested by this theorem, all the results obtained in Section 5 regarding psychometric functions could be translated in terms of sensitivity functions or, when they are defined, in terms of Weber functions. Following are a few additional examples, which may be skipped at first reading without much loss of continuity. (We omit the proofs of these results, which are easy to obtain.)

7.1.4. Theorem. Let $\Xi = \{p_a | a \in I\}$ be a psychometric family, with sensitivity function ξ. Then the following two conditions are equivalent:

1. Ξ is a parallel family in the sense of the definition in Section 5.5.1.
2. $\xi_\pi(a) - \xi_{\pi'}(a) = \xi_\pi(b) - \xi_{\pi'}(b)$, for all π, π', and a, b such that both members are defined.

This result follows readily from the definitions, as well as from the theorem in Section 5.4.2.

7.1.5. Theorem. If $\Xi = \{p_a | a \in I\}$ is an anchored psychometric family (in the sense of the definition in Section 5.4.3) with a sensitivity function ξ, then Ξ is parallel iff there exist two functions g, h with h strictly increasing and continuous, such that the equation

$$\xi_\pi(a) = g(a) + h(\pi) \tag{67}$$

holds whenever $\xi_\pi(a)$ is defined. In particular, g is defined on I.

This implies that in the situation described in the theorem, the Weber functions $a \mapsto \Delta_\pi(a)$ do not vary with a since

$$\Delta_\pi(a) = h(\pi) - h(.5)$$

for all background a. As suggested by a comparison of Sections 7.1.4 and 7.1.5, the functions g, h do not necessarily exist if the assumption of anchoring is removed.

We consider next the effect of the balancing condition (from the definition in Section 5.5.1)

$$p_a(b) + p_b(a) = 1 ,$$

on the sensitivity function of a discrimination family.

7.1.6. Theorem. A discrimination family $\Xi = \{p_b | b \in I\}$ is balanced iff its sensitivity function ξ satisfies

$$\xi_{1-\pi}[\xi_\pi(a)] = a$$

whenever the left member of this equation is defined.

These few results should suffice to familiarize the reader with the notions of the definition in Section 7.1.1. Further results along these lines can be found in Falmagne (1982).

7.2. Linear Psychometric Families—Weber's Law

So far, no assumptions were made regarding the structure of the set I of backgrounds of a psychometric family $\{p_a | a \in I\}$. Such properties as parallelism or subtractivity could be discussed while assuming that the elements $a \in I$ were just labels for the psychometric functions p_a in the family. Of particular importance in this section is the situation in which I is actually a (subset of a) vector space over the real numbers. For example, $a \in I$ may denote a spectral density function or, in the case of a discrimination family (see Section 5.5.1), a real number representing a physical intensity. What is critical here is that the multiplication

$$\lambda a$$

of a real vector a by a positive real number λ makes sense. From an empirical standpoint, the multiplication λa means that the intensity of the background has been multiplied by the factor λ. (In the case of a spectral density function, a denotes a real-valued function and λa symbolizes the fact that all the intensities of the background have been multiplied by the same constant λ.) When such a situation arises, properties can be investigated in the data, which are both strong and of central interest for psychophysical research.

7.2.1. Definition. A psychometric family $\Xi = \{p_b | b \in I\}$ is called *linear* iff the index set I is a (subset of a) vector space over the real numbers.

A special case of a linear psychometric family arises when the indices of the psychometric function denote physical intensities. This case was referred to in Section 5.5.1 as a discrimination family.

We recall that in a psychometric family $\Xi = \{p_b | b \in I\}$, the notation C_b, for any $b \in I$, refers to the domain of the psychometric function p_b, which is an open interval (see Section 5.2.1). The psychometric family Ξ will be called *positive* iff each interval C_b is positive. (This is a typical case for physical intensities.) Most results in the remainder of this section will be obtained in the framework of linear, positive psychometric families.

The definition in Section 7.2.2 will also be useful in connection with Weber's law and more general forms of this law.

7.2.2. Definition. Let V be a vector space over the real numbers. Let T be a subset of V. Let f be a real-valued function on T. Then f is said to be *homogeneous of degree* β (on T) iff for any real number $\lambda \neq 0$, whenever $a, \lambda a \in T$, then

$$f(\lambda a) = \lambda^{\beta} f(a) .$$

7.2.3. Definition. A linear, positive psychometric family $\Xi = \{p_b | b \in I\}$ satisfies Weber's law iff

$$p_a(x) = p_{\lambda a}(\lambda x) \qquad (68)$$

whenever both members of Eq. (68) are defined, with $0 < \lambda < \infty$. In other words, Ξ satisfies Weber's law iff the function p, $(a,x) \mapsto p_a(x)$ is homogeneous of degree 0. Occasionally, Eq. (68) will be referred to as Weber's law.

7.2.4. Remark. In two respects, this definition of Weber's law departs from tradition. Weber's law is usually stated in the special case in which the backgrounds are real numbers. For example, in the context of auditory detection of a stimulus embedded in noise, Weber's law would imply that the probability of a correct detection would not vary when both the stimulus and the noise are increased in intensity by the same number of decibels. We believe, however, that this prediction would apply for a fairly large set of spectral density functions specifying the noise. Such an assumption is made explicit in Section 7.2.3. Another difference is that Weber's law is most often expressed in terms of the Weber functions Δ_π. The equivalence is made clear in the theorem in Section 7.2.6. We have two reasons for adopting Eq. (68) as the defining condition of Weber's law, rather than the more customary form

$$\Delta_\pi(\lambda a) = \lambda \Delta_\pi(a) .$$

One is that Eq. (68) is more general; this equation makes sense in situations in which the Weber functions are not always defined. (The Weber function Δ_π was defined in Section 7.1.1 from the sensitivity function $\xi_{.5}$. There may be cases in which $\xi_{.5}$ is not obtainable.) Another is to stress the fact that in view of the binomial variability of the relative frequencies providing the basic data for Eq. (68), it is more readily amenable to statistical testing. In practice, however, evaluations of Weber's law are mostly based on investigating the empirical behavior of the Weber functions.

Some strengthening of our conditions will be useful for this and later results.

7.2.5. Definition. A linear psychometric family $\Xi = \{p_b | b \in I\}$ is called *solvable* iff for all $a \in I$ and all $x \in C_a$, the equation

$$p_{\lambda a}(\mu x) = p_a(x)$$

is solvable in μ for every λ and is solvable in λ for every μ. We say that Ξ has a *Weberian domain* iff for any $\lambda > 0$, $p_{\lambda a}(\lambda x)$ is defined whenever $p_a(x)$ is defined.

These strengthenings of our assumptions will occasionally be convenient but are hardly innocuous. The reader is invited to reflect on the empirical impact of these two conditions. Both of them practically entail that neither of the two physical domains spanned by the function p is bounded, obviously not a realistic assumption. Neither of these conditions is essential, but they certainly render our developments much easier. In any event, they will be used sparingly in the sequel.

In the theorem in Section 7.2.6, a central result of this section, we consider an important generalization of Weber's law, symbolized by the equation

$$p_{\lambda a}(\lambda^{\beta} x) = p_a(x) ,$$

and we establish the equivalence between this equation and some constraints on sensitivity function and Weber function data. We show that our definition of Weber's law is equivalent to the traditional one. The interpretation of the exponent β is discussed in Section 7.3.1.

7.2.6. Theorem. Let $\Xi = \{p_b | b \in I\}$ be a linear, positive, solvable psychometric family, with sensitivity function ξ. Then the following three conditions are equivalent:

1. Every sensitivity function ξ_π is homogeneous of degree $\beta > 0$:

$$\xi_\pi(\lambda a) = \lambda^{\beta} \xi_\pi(a) .$$

2. There is a constant $\beta > 0$ such that

$$p_{\lambda a}(\lambda^{\beta} x) = p_a(x) ,$$

whenever both members of this equation are defined, with $0 < \lambda < \infty$.

3. There exists a function F and a constant $\beta > 0$ such that

$$p_a(x) = F(a/x^{1/\beta}) .$$

Moreover, if Ξ is anchored at .5, then each condition 1–3 is equivalent to the assumptions that any Weber function Δ_π is homogeneous of degree $\beta > 0$. In particular, Weber's law holds iff

$$\Delta_\pi(\lambda a) = \lambda \, \Delta_\pi(a) \; ; \qquad (69)$$

that is, the Weber functions Δ_π are homogeneous of degree 1.

Proof. (1) *implies* (2). Suppose that $p_a(x) = \pi$, $p_{\lambda a}(\lambda^\beta x) = \pi'$. Then $\xi_\pi(a) = x$, and successively

$$\xi_{\pi'}(\lambda a) = \lambda^\beta x = \lambda^\beta \xi_\pi(a) = \xi_\pi(\lambda a) \, ,$$

since Ξ is solvable, and $\pi = \pi'$ follows by the strict monotonicity of $\xi_\pi(\lambda a)$ in the variable π.

(2) *implies* (3) Setting $\lambda^\beta x = K$, a constant, we obtain $\lambda = (K/x)^{1/\beta}$, yielding

$$p_a(x) = p_{\lambda a}(\lambda^\beta x) = p_{(K/x)^{1/\beta}a}(K) = F(a/x^{1/\beta}) \, ,$$

with the function F defined by $F(s) = p_{K^{1/\beta}s}(K)$. In fact, (2) and (3) are equivalent since obviously

$$F(a/x^{1/\beta}) = F[\lambda a/(\lambda^\beta x)^{1/\beta}] \, .$$

(3) *implies* (1). In view of the equivalence between (2) and (3), this is clear since, with

$$\pi = p_{\lambda a}(\lambda^\beta x) = p_a(x) \, ,$$

we have

$$\xi_\pi(\lambda a) = \lambda^\beta x = \lambda^\beta \xi_\pi(a) \, .$$

If the Weber function is defined, it follows by substitution that the sensitivity functions are homogeneous of degree $\beta > 0$ iff the Weber functions Δ_π also satisfy this condition. Finally, the equivalence between Weber's law and Eq. (69) is obtained from the case $\beta = 1$ of the equivalence between (1) and (2). ∎

7.3. Remarks

We shall see that the homogeneity equation

$$\xi_\pi(\lambda a) = \lambda^\beta \xi_\pi(a)$$

plays an important role in the analysis of data, as a substitute to Weber's law. (This equation is often referred to as the *near-miss to Weber's Law*, cf. McGill & Goldberg, 1968.) The interpretation of the exponent β in this equation must be considered carefully. There seems to be a tendency in the psychophysical community to take this exponent as representing a critical aspect of the neural coding of physical intensity. For a number of reasons, this position is open to challenge. One difficulty is indicated below. Suppose that Ξ is a discrimination family satisfying this condition, together with $\xi_{.5}(a) = a$, for all intensities a. This implies that the Weber functions are also homogeneous of degree β:

$$\Delta_\pi(\lambda a) = \lambda^\beta \Delta_\pi(a) \, . \qquad (70)$$

But we have also

$$\Delta_\pi(\lambda a) = \xi_\pi(\lambda a) - \xi_{.5}(\lambda a) = \lambda^\beta \xi_\pi(a) - \lambda a$$
$$= \lambda^\beta[\Delta_\pi(a) + a] - a \, ,$$

which together with Eq. (70) implies

$$\lambda^\beta \Delta_\pi(a) = \lambda^\beta[\Delta_\pi(a) + a] - \lambda a \, ,$$

that is,

$$a(\lambda^\beta - \lambda) = 0 \, .$$

Since $a > 0$, we must conclude that $\beta = 1$. Thus in the case of a discrimination family satisfying $p_a(a) = .5$, the assumption that the Weber functions are homogeneous of degree β implies in fact Weber's law.

The crux of the argument here is that the condition

$$p_a(a) = .5$$

or the more general balance condition

$$p_a(b) + p_b(a) = 1 \, ,$$

which implies it, results from a symmetry of the experimental paradigm which is not necessarily of a sensory nature. We shall go back to this point later in this section.

A special case of the theorem in Section 7.2.6 is of historical interest. If Weber's law holds, then $\beta = 1$ in Eq. (70) and by virtue of Condition (3) in Section 7.2.6, the choice probabilities take the form

$$p_a(x) = F(a/x) = F[e^{-(\log x - \log a)}] \, ,$$

yielding

$$p_a(x) = G(\log x - \log a) \, , \qquad (71)$$

with $G(s) = F(e^{-s})$, a strictly increasing, continuous function. Equation (71) has sometimes been given the interpretation that "the sensation grows as the logarithm of the excitation," a statement which has been named *Fechner's law*. Such interpretation has been at the center of a long controversy and should not be dismissed or accepted casually. It relies in part on some empirical evidence, Weber's law. (How well Weber's law is supported by the data is considered in Section 7.4.) It also relies on the somewhat arbitrary choice of a particular mathematical representation of such data, namely Eq. (71). Finally, it involves using a philosophically charged label such as "sensation." Each of these factors has contributed its share to the polemical aspects of the debate, a brief account of which can be found in a later section of this chapter (see Section 10.9).

7.4. Examination of the Data

As an empirical prediction, Weber's law holds reasonably well for sensory continua such as loudness discrimination of Gaussian noises, loudness discrimination of pure tones, lifted weights, and visual brightness. As mentioned in Section 7.3, the analysis of the data is sometimes based on the just-noticeable-difference (jnd) function, which can be computed from the sensitivity functions by the equation

$$\mathrm{jnd}(a) \;=\; [\xi_{.75}(a) \;=\; \xi_{.25}(a)]/2 \;.$$

The experimenter checks whether the ratio

$$\mathrm{jnd}(a)/a \qquad (72)$$

remains constant, while a varies on a chosen subset of the physical scale. (Thus a takes values in the positive reals.) We have elected to base our developments on the sensitivity function, rather than on the jnd function. For various reasons, which the reader will discover gradually, the sensitivity function is the appropriate notion to use as the cornerstone of the theory. Notice in this connection that Eq. (72) is constant if the functions

$$a \mapsto \xi_\pi(a)/a$$

are constant. More generally, the jnd function is homogeneous of degree β if the sensitivity functions ξ_π are homogeneous of degree β. It is clear that the reverse implication does not necessarily hold. Any one of the psychophysical methods discussed in Section 6 can be employed for the empirical determination of the sensitivity function. Even though these methods differ drastically from an experimental viewpoint, many believe that the overall pattern of empirical results is not seriously affected by which method has been used. This opinion is not universal, however, and we shall be cautious in this respect. (Luce and Green, 1974a, for example, analyze the data of six studies of the Weber fraction $\Delta_\pi(a)/a$ for tone intensities, with considerable discrepancy in the results.)

The experimental evidence favoring Weber's law is exemplified in Figure 1.14. For some sensory continua, the Weber fraction $\Delta_\pi(a)/a$, with $a \in \mathrm{Re}_+$ (the set of positive real numbers), remains indeed constant over a substantial portion of the domain (2–3 log units, for audition and vision), thus supporting Weber's law. A more comprehensive description of the data would emphasize the fact that the Weber fraction is initially decreasing. In fact, for audition and smell, it never increases. Finally, a conjecture which is validated by the data for all five continua in Figure 1.14 is that the Weber fraction is "convex" (i.e., it never "curves downward"). A precise definition of the convexity of a function is given in the definition in Section 7.4.1. The reader should remember these aspects of the data, which will lead to a theoretical analysis in Section 7.5.

The initial decrease of the Weber fraction is sometimes attributed to an absolute threshold of perception, while the late rise of the fraction in some cases is attributed to the sensory mechanisms reaching the limit of their operational range. However legitimate such interpretations might be, they do not necessarily justify an analysis of an empirical Weber fraction into fragments, each requiring a separate model. In this section, only models attempting a comprehensive description of the data are considered.

7.4.1. Definition. Let f be a real-valued function defined on a real interval (s,t). Then f is called *convex* iff

$$f[\lambda x + (1 - \lambda)y] \leqslant \lambda f(x) + (1 - \lambda)f(y)$$

whenever $s < x < t$, $s < y < t$, and $0 < \lambda < 1$. The function f is *strictly convex* iff the above inequality is strict. If $-f$ is convex (respectively, strictly convex), then f is said to be *concave* (respectively, *strictly concave*).

Any linear function is both convex and concave. Examples of strictly convex functions are $x \to e^x$, $x \to x^2$, for $-\infty < x < \infty$. The following results are easy consequences of the definition: any convex function is continuous; if g is an increasing, convex function and f is convex, then the composition $s \mapsto g[f(s)]$ is convex (in particular, e^f is convex, in fact, strictly convex); if the second derivative f'' of f exists, then f is convex iff $f'' \geqslant 0$. A geometrical interpretation of convexity is that any segment of a straight line joining two points of the graph of a convex function f lies above or on the graph of f (see Figure 1.15).

It is clear that the Weber fractions depicted in Figure 1.14 are convex in the sense of the definition in Section 7.3.1. A case can be made that these functions are actually strictly convex.

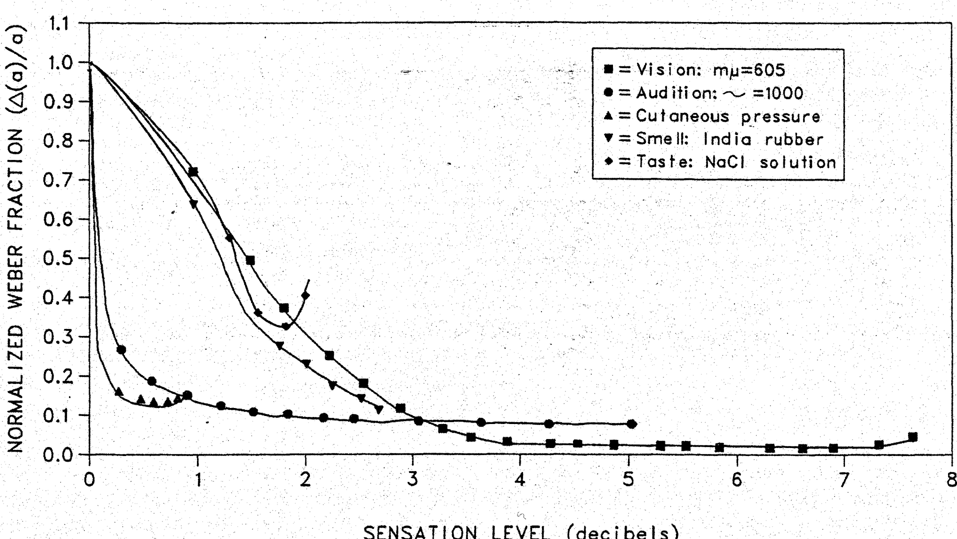


Figure 1.14. Weber fraction data, for various sensory continua. The abscissa is in decibels, sensation level. The Weber fractions have been normalized so as to be unity at threshold. (From R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology*. Copyright 1963 by John Wiley & Sons, Inc. Reprinted with permission.)

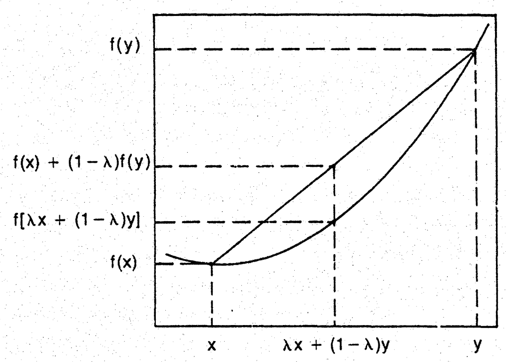


Figure 1.15. Geometrical interpretation of the convexity of a real valued function f: $[\lambda x + (1 - \lambda)y] \leq \lambda f(x) + (1 - \lambda)f(y)$.

The failures of Weber's law illustrated in Figure 1.14 prompted psychophysicists to propose various alternatives.

7.5. Alternatives to Weber's Law

One generalization of Weber's law has been encountered earlier which, in terms of the sensitivity function, is symbolized by the equation

$$\xi_\pi(\lambda a) = \lambda^\beta \xi_\pi(a) \quad (73)$$

for $0 < \lambda < \infty$ and with $\beta > 0$, a constant independent of π. When the Weber function is defined, this is equivalent to

$$\Delta_\pi(\lambda a) = \lambda^\beta \Delta_\pi(a) \; . \quad (74)$$

This prediction, which is often referred to as the near-miss to Weber's law (McGill & Goldberg, 1968), has been supported experimentally in some situations (Jesteadt, Wier, & Green, 1977; see Figure 1.16). Notice that the data in Figure 1.16 require the exponent β in Eq. (74) to be smaller than 1; the Weber fraction

$$\Delta_\pi(\lambda a)/a = \lambda^{\beta-1}\Delta_\pi(a)$$

must be decreasing. It cannot be assumed, however, that the near-miss to Weber's law holds generally, across experimental paradigms and sensory continua. For one thing, this law would fail to explain most of the data displayed in Figure 1.13. For another, it was pointed out earlier (see Section 7.3) that in the case where $\beta \neq 1$, Eqs. (73) and (74) necessarily imply the existence of some asymmetry in the paradigm: the psychometric functions cannot satisfy the condition, $p_a(a) = .5$. This condition, however, is sometimes inherent to the experimental paradigm (e.g., in a situation in visual psychophysics, where the stimuli to be compared are two spots of light symmetrically positioned

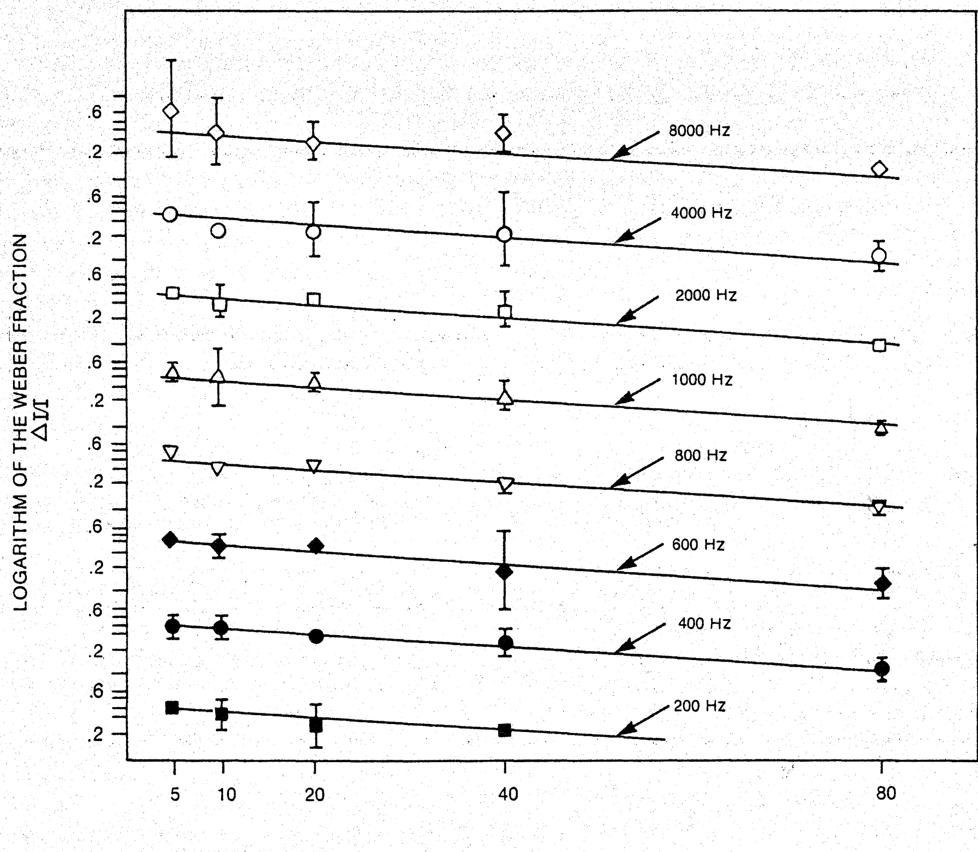


Figure 1.16. Values of the logarithm of the Weber fraction, averaged across subjects and replications, for various intensities and frequencies of pure tones. The abscissa is in decibels, sensation level. The vertical bars indicate ± 3 standard errors. A bar is omitted when its size is exceeded by that of the symbol. The same linear function has been fitted to the eight sets of data. (From W. Jesteadt, C. C. Wier, & D. M. Green, Intensity discrimination as a function of frequency, *Journal of the Acoustical Society of America*, 1977, *61*. Reprinted with permission.)

with respect to a fixation point). It is also possible to ensure that this condition holds by a "normalization" of the data. As argued in Section 5.7, we are certainly not advocating such tampering with the data. The fact is, however, that such normalizations are fairly frequent. In such cases, it is clear that Eq. (74) describes the data only if $\beta = 1$ (that is, Weber's law holds) or, possibly, if β is a function of π. This, however, is not what is intended by the near-miss to Weber's law, at least as we understand it.

In any event, the above discussion suggests a further generalization of Weber's law, which is embodied in the equation

$$\xi_\pi(\lambda a) = \lambda^{\beta(\pi)}\xi_\pi(a) , \qquad (75)$$

in which β is a function of π satisfying $\beta(.5) = 1$. (This ensures $p_a(a) = .5$.) The consequences of that assumption deserve some attention. It can be shown in particular that if the balancing condition is satisfied, then the function β must satisfy

$$\beta(\pi)\beta(1 - \pi) = 1 . \qquad (76)$$

As far as we know, this condition or, even more generally, the effect of the choice probability π on the estimated value of the exponent β in Eq. (75) has never been investigated from an experimental viewpoint. Other substitutes to Weber's law are of interest. As in the case of the near-miss to Weber's law, each of the examples below is a special case of the representation

$$\xi_\pi(a) = u^{-1}[g(a) + h(\pi)] ,$$

as can be checked without difficulty.

Let the sensitivity function ξ of a discrimination family be defined by the equation

$$\xi_\pi(\lambda a) = \lambda^\beta[\xi_\pi(a) - K] + K , \qquad (77)$$

where K is a positive constant. In the style of the theorem in Section 7.2.6, this is equivalent to the representation

$$p_a(x) = F[a(x - K)^{-1/\beta}] \qquad (78)$$

for the choice probabilities (we assume $x > K$). Provided that $\beta > 1$ and $\xi_\pi(a) > K$, the function

$$\lambda \mapsto \xi_\pi(\lambda a)/\lambda$$

is convex on the positive real numbers, with a minimum at the point

$$\lambda = [K/(\beta - 1)(\xi_\pi(a) - K)]^{1/\beta} .$$

Thus the Weber fractions are also convex in the same conditions, a fact worth noticing in connection with our discussion of the data displayed in Figure 1.14. It is clear that this model generalizes the near-miss to Weber's law. The constant K may be interpreted as a measure of a threshold value.

In the case where Ξ is a discrimination family, we also consider the representation

$$\xi_\pi(a) = [h(\pi) + \delta a^{\beta'}]^{1/\beta} \qquad (79)$$

for the sensitivity functions, involving a strictly increasing, continuous function h and three positive constants β, β', and δ. This leads immediately to the representation

$$p_a(x) = F(x^\beta - \delta a^{\beta'}) , \qquad (80)$$

with $F = h^{-1}$, for the choice probabilities. This model has been discussed by several authors (see, for example, Parker & Schneider, 1980). It is not consistent with Weber's law. However, for appropriately chosen values of the parameters, it would predict the main features (monotonicity, convexity) of data such as that pictured in Figure 1.14.

The late increase of the Weber fraction is often interpreted as resulting from a saturation of the sensory or neuronal mechanisms. In turn, this leads to the postulate that the sensory scale, for example, the function u in Eq. (65), is bounded. An example along these lines is given below. It is assumed that the sensitivity functions satisfy the equation

$$\xi_\pi(a) = K\{h(\pi)f(a)/[1 - h(\pi)f(a)]\}^{1/\beta} , \qquad (81)$$

with

$$f(a) = a^{\beta'} + K' ,$$

and $\beta, \beta', K, K' > 0$, constants. Using simple algebra, we obtain for the response probabilities the form

$$p_a(x) = G[x^\beta/(x^\beta + K^\beta)f(a)] ,$$

or equivalently, as a difference model,

$$p_a(x) = F\{\ln[x^\beta/(x^\beta + K^\beta)] - \ln f(a)\} ,$$

with $F(\ln s) = G(s)$. Thus

$$u(x) = \ln[x^\beta/(x^b + K)] < 0 .$$

Again, such a model is capable of accommodating typical Weber fraction data (cf. Alpern, Rushton, & Tori, 1970a, 1970b, 1970c).

Other models have been proposed, which differ only in details from one or the other of those discussed in this section. They are not reviewed here. Our purpose in this subsection is not to single out one particular mathematical expression as the appropriate model for the sensitivity function. In fact, it is quite conceivable that the choice of a suitable model (that would provide a good fit to the data, from a statistical viewpoint) may depend not only on the sensory continuum envisaged but also on rather specific details of the experimental paradigm. Accordingly, an effort has been made by some psychophysicists to focus the theoretical developments on aspects of the data that may perhaps be robust to minor changes of the experimental procedure (cf. Falmagne, 1977; Iverson, 1983). The results are too specialized to be included here.

7.6. Key References

Weber fraction data are compiled, for example, in Boring, Langfeld, and Weld (1948, p. 268) and Holway and Pratt (1936, p. 337) for various sensory continua. Luce and Green (1974a; see also Green, 1978, p. 257) review a number of experimental studies of the discrimination of the difference in the amplitudes of a sinusoidal tone. The data are plotted in terms of the Weber fraction. See also the Chapters in Sections II and III of this handbook. In a recent monograph, Laming (1983) gave a theoretical analysis of Weber functions, based on a large collection of data. (Unfortunately, this work came to our attention in the final stage of the writing of this chapter, and no discussion of its content could be included.)

8. SIGNAL DETECTION THEORY

Any psychophysical task has cognitive components, which covers a variety of factors, such as response bias, guessing strategy, motivation, and so on. Thus far in our approach to psychophysical theory, we have implicitly assumed that such factors could be bypassed or controlled by careful experimental design. In fact, we have ignored them. This position is not without its weaknesses. An example will make this clear.

Consider a task in which a subject is required to detect a stimulus embedded in a noisy background. On 50% of the trials the noise is presented alone. Across conditions, the intensity of the stimulus is varied, providing the basic data for a psychometric function. Two kinds of error can be made in such a task, which is often referred to as the yes-no paradigm: (1) the subject may fail to report a stimulus presented (this is called *a miss*) and (2) the subject may report a detection on a noise-alone trial (this is called a *false alarm*, or a *false positive*). A correct detection will be referred to as a *hit*. The remaining case is a *correct rejection*.

A guessing strategy is available to the subject in this situation: when not quite sure that the stimulus was presented on a trial, the subject may nevertheless claim to have detected it. Such strategy would succeed in a situation in which a miss is much more heavily penalized than a false alarm. For example, suppose that the system of rewards and penalties is the one displayed in Table 1.2, where the numbers represent monetary values. Thus each correct detection brings 3 monetary units (μs): each miss costs 3 μs, and so on. Such a table is often referred to as a *payoff matrix*. It is reasonable to suppose that the particular payoff matrix shown above would favor a guessing strategy over a conservative one. (For instance, if the subject reports a detection on every trial, whether or not the stimulus was presented, the average gain per trial is 2 μs, while the opposite strategy of responding "no detection" on every trial results in an average gain of 0 μs.) Obviously, another payoff matrix may evoke a completely different strategy. A naive experimenter may be tempted to believe that if a constant payoff matrix is used across conditions varying in stimulus intensity, the subject strategy will not change, a fact which can be tested by checking that the proportion of false alarms remains constant. Unfortunately, a subject's interpretation of a payoff matrix is largely personal, and this interpretation may change drastically from one condition to another. Needless to say, these remarks also apply when no explicit payoff matrix is used, but the subject strategy is induced by verbal instructions. The problem at hand is thus that of disentangling the purely sensory aspects of the task from those resulting from the subject's strategy.

Table 1.2. System of Rewards and Penalties in Which Numbers Represent Monetary Values

	Responses	
Stimulus	Yes	No
Yes	3 Hit	−3 Miss
No	−1 False Alarm	3 Correct Rejection

This section is devoted to a particular solution to this problem, which is usually discussed under the label *signal detection theory*, even though its applicability extends far beyond the detection of signals. Our presentation is far from exhaustive. It should however be sufficient to acquaint the reader with the most commonly used notions and techniques of signal detection theory. For an extensive treatment of this topic, see Green and Swets (1974).

8.1. Receiver Operating Characteristic (ROC) Graphs and Curves

For simplicity, we shall ignore statistical variability for the moment and identify response frequencies and probabilities.

Let us suppose that for a given stimulus intensity, three payoff matrices have been used, labeled θ_1, θ_2, and θ_3, inducing three different guessing strategies. Let s and n denote the stimulus and the noise, respectively. Let $p_s(\theta_i)$ and $p_n(\theta_i)$, $i = 1, 2, 3$, be the hit and false alarm probabilities. For concreteness, some hypothetical data follow:

	$p_n(\theta_i)$	$p_s(\theta_i)$
θ_1	.10	.35
θ_2	.40	.75
θ_3	.65	.90

A useful graphic representation of such data is often used by psychophysicists, in which each pair of response probabilities $[p_n(\theta_i), p_s(\theta_i)]$ is pictured as a point in the unit square (see Figure 1.17, but ignore the three curves for the moment). There is a consensus in psychophysics that by appropriately choosing the payoff matrix, most types of strategies can be induced in the subject, ranging from the most conservative ones (if the slightest doubt arises, say "no detection") to the most daring guessing. (High false alarm rates, however, are exceptional.) It is also reasonable to suppose that any change in a payoff matrix that would increase the probability of a false alarm would also increase (continuously) the probability of a hit. (This assumption is supported by much data.) In other words, this means that the three points in Figure 1.17 belong to the graph of a continuous function ρ mapping the interval [0,1] into itself. We have thus

$$\rho[p_n(\theta)] = p_s(\theta) ,$$

in which θ ranges in a large set of payoff matrices Θ.

8.1.1. Definition. Let Θ be a collection of payoff matrices; for each $\theta \in \Theta$, let $p_n(\theta)$ and $p_s(\theta)$ be the probabilities of a false alarm and of a hit, respectively. Then the set of points

$$\{[p_n(\theta), p_s(\theta)] \mid \theta \in \Theta\} ,$$

in the unit square is called a *receiver operating characteristic (ROC) graph* (of (n,s)). When an ROC graph is the graph of a continuous function ρ mapping the closed interval [0,1] into itself, it will be called an *ROC curve*. The function ρ will be referred to as the *ROC function*.

Three examples of ROC curves are displayed in Figure 1.17, in which the functions are increasing. It is reasonable to suppose that any change in a payoff matrix that would increase the probability of a false alarm would also increase (or at least not decrease) the probability of a hit. This assumption is supported by much data. Incidentally, the acronym ROC is borrowed

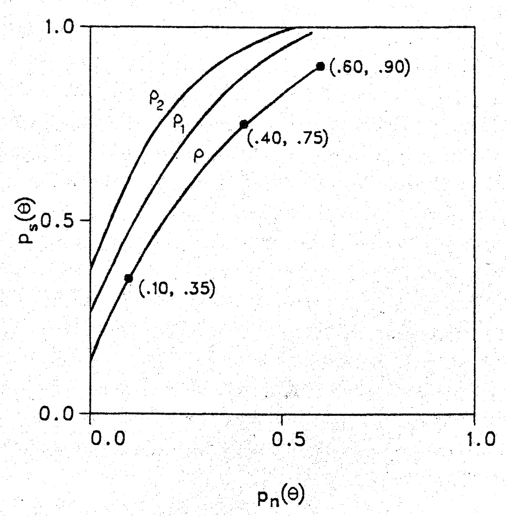


Figure 1.17. Three ROC curves and one ROC graph containing three points.

from signal detection theory in telecommunication (references are given in Section 8.11).

The basic idea of this representation is that the strategy is varying along the ROC curves, while the discriminability is varying across the ROC curves. In this framework, the three ROC curves ρ, ρ_1, and ρ_2 in Figure 1.17 correspond to increasingly detectable stimuli. A particularly illuminating interpretation of the information captured by an ROC curve is offered by the model described in Section 8.2.

8.2. A Random Variable Model for ROC Curves

Suppose that to each stimulus s corresponds a random variable $\mathbf{U}_s$, representing the activation evoked by that stimulus in some critical neural location. Similarly, let $\mathbf{U}_n$ be the activation random variable corresponding to the noise. The random variables $\mathbf{U}_s$ and $\mathbf{U}_n$ are assumed to be independent. As before, let $p_s(\theta)$ and $p_n(\theta)$ be the hit and false alarm probabilities, corresponding to a payoff matrix θ. We assume that on every trial (whether or not the stimulus is presented) a positive response occurs if the momentary (or sample) value of $\mathbf{U}_s$ or $\mathbf{U}_n$ exceeds a criterion λ_θ, the value of which is determined by the payoff matrix θ. In symbols:

$$p_s(\theta) = \text{Prob}\{\mathbf{U}_s > \lambda_\theta\} \tag{82}$$

and

$$p_n(\theta) = \text{Prob}\{\mathbf{U}_n > \lambda_\theta\} \ . \tag{83}$$

Such a model is in the spirit of the random utility, or Thurstone-type models discussed earlier in this chapter (see Sections 4.2 to 4.8). In the context of ROC curves, it entails a few interesting results. Suppose that

$$p_n(\theta) > p_n(\theta')$$

for some payoff matrices θ and θ′. Using, successively, Eqs. (83) and (82), this implies

$$\lambda_\theta < \lambda_{\theta'} \ ,$$

yielding

$$p_s(\theta) \geqslant p_s(\theta') \ .$$

This means that the ROC function $p_n(\theta) \mapsto p_s(\theta)$ must be nondecreasing, a prediction which, as indicated earlier, is consistent with much data. Another basic result is that the area under the ROC curve (the integral of the ROC function from 0 to 1) is equal to the probability that $\mathbf{U}_s$ exceeds $\mathbf{U}_n$:

$$\text{Prob}\{\mathbf{U}_s > \mathbf{U}_n\} \ .$$

The argument is spelled out below. Using the fact that the random variables in question are independent, we have

$$\text{Prob}\{\mathbf{U}_s > \mathbf{U}_n\} = \int_{-\infty}^{\infty} \text{Prob}\{\mathbf{U}_s > \lambda \mid \mathbf{U}_n = \lambda\} d\text{Prob}\{\mathbf{U}_n \leqslant \lambda\}$$

$$= \int_{-\infty}^{\infty} \text{Prob}\{\mathbf{U}_s > \lambda\} d\text{Prob}\{\mathbf{U}_n \leqslant \lambda\}. \tag{84}$$

According to this model, the response probabilities $p_s(\theta) = \text{Prob}\{\mathbf{U}_s < \lambda_\theta\}$ and $p_n(\theta) = \text{Prob}\{\mathbf{U}_n < \lambda_\theta\}$ depend on the payoff matrix θ only through the number λ_θ. There is thus no ambiguity in writing $p_s(\lambda)$ for $p_s(\theta)$ and $p_n(\lambda)$ for $p_n(\theta)$, with $\lambda = \lambda_\theta$. Consequently, using ρ to denote the ROC function, Eq. (84) yields

$$\text{Prob}\{\mathbf{U}_s > \mathbf{U}_n\} = \int_{-\infty}^{\infty} p_s(\lambda) d[1 - p_n(\lambda)]$$

$$= -\int_{-\infty}^{\infty} \rho[p_n(\lambda)] dp_n(\lambda)$$

$$= \int_{\infty}^{-\infty} \rho[p_n(\lambda)] dp_n(\lambda) \ .$$

Changing variables, from λ to $p_n(\lambda) = p$, we obtain finally

$$\text{Prob}\{\mathbf{U}_s > \mathbf{U}_n\} = \int_0^1 \rho(p)\, dp \ , \tag{85}$$

as asserted.

In the framework of this model, the area below the ROC curve appears as a reasonable measure of the detectability of the stimulus. In practice, it will often be the case that only a few points of the ROC curve have been determined experimentally. The evaluation of the area below the ROC curve may thus be prone to serious errors. One way out of this difficulty is to make specific assumptions concerning the distributions of the random variables $\mathbf{U}_s$ and $\mathbf{U}_n$. Such assumptions would determine (up to the values of a couple of parameters) the exact analytical form of the ROC curve. If the assumptions are valid, a few suitably placed points of the ROC curve will suffice to estimate the parameters of the ROC curve experimentally, and the area under the ROC curve can then be evaluated by integration.

8.3. Remarks

One may be suspicious of such a method and object that the estimated value of the area will be model bound. This objection is not as strong as it may appear. Notice in this connection that Eq. (85) was obtained without making any assumption regarding the distribution of the random variables $\mathbf{U}_s$ and $\mathbf{U}_n$. In fact, the shape of these distributions is arbitrary. For instance, let us assume that Eqs. (82), (83), and (85) hold for some random variables $\mathbf{U}_s$, $\mathbf{U}_n$. For any strictly increasing continuous function g, we have

$$p_s(\theta) = \text{Prob}\{g(\mathbf{U}_s) > g(\lambda_\theta)\} = \text{Prob}\{\mathbf{U}'_s > \lambda'_\theta\}$$

and

$$p_n(\theta) = \text{Prob}\{g(\mathbf{U}_n) > g(\lambda_\theta)\} = \text{Prob}\{\mathbf{U}'_n > \lambda'_\theta\} ,$$

with $\mathbf{U}'_s = g(\mathbf{U}_s)$ and $\mathbf{U}'_n = g(\mathbf{U}_n)$ and $\lambda'_\theta = g(\lambda_\theta)$. It is clear that the representation of the response probabilities provided by the random variables $\mathbf{U}'_s$ and $\mathbf{U}'_n$ is equivalent to that obtained with $\mathbf{U}_s$ and $\mathbf{U}_n$. In particular, the predicted ROC curve is not changed by the transformation.

Later on in this section, we will make precise hypotheses regarding the distributions of the random variables entering into Eqs. (82), (83), and (85). When evaluating these hypotheses, the reader should keep in mind the above remark pointing out the relative arbitrariness of the distributions of $\mathbf{U}_s$ and $\mathbf{U}_n$.

It must be realized that the random variable model discussed here does not necessarily describe a rational strategy. Depending on how optimality is defined and on more specific assumptions on the random variables $\mathbf{U}_s$ and $\mathbf{U}_n$, the decision rule embodied in Eqs. (82) and (83) may or may not be optimal.

To illustrate this point, let π be the probability of a stimulus trial and suppose that for some criterion value λ_o

$$\text{Prob}\{\mathbf{U}_n > \lambda_o\}(1 - \pi) > \text{Prob}\{\mathbf{U}_s > \lambda_o\}\pi . \quad (86)$$

A special case of this assumption is pictured in Figure 1.18, which is by no means unrealistic. Nevertheless, it leads to a somewhat undesirable conclusion. Namely, when reacting to an activation value exceeding λ_o, the subject reports a detection, even though the likelihood of a stimulus trial is then smaller than that of a noise trial. Such a conclusion easily follows from the above inequality. Indeed, denoting by S, as before, the stimulation at a given trial (thus $S = s$ or $S = n$), Eq. (86) holds if and only if (iff) successively

$$\text{Prob}\{\mathbf{U}_S > \lambda_o | S = n\}\text{Prob}\{S = n\} > \text{Prob}\{\mathbf{U}_S > \lambda_o | S = s\}\text{Prob}\{S = s\}$$

iff

$$\text{Prob}\{S = n, \mathbf{U}_S > \lambda_o\}/\text{Prob}\{\mathbf{U}_S > \lambda_o\} > \text{Prob}\{S = s, \mathbf{U}_S > \lambda_o\}/\text{Prob}\{\mathbf{U}_S > \lambda_o\}$$

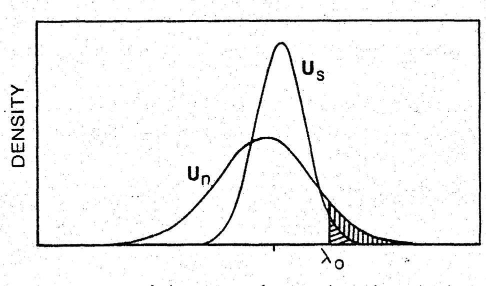


Figure 1.18. Two normal densities of $\mathbf{U}_s$ and $\mathbf{U}_n$ for which Eq. (86) holds, with $\pi = 1 - \pi = .5$.

iff

$$\text{Prob}\{S = n | \mathbf{U}_S > \lambda_o\} > \text{Prob}\{S = s | \mathbf{U}_S > \lambda_o\} .$$

In words, this last inequality means that the conditional probability of a noise trial, when observing the event $\mathbf{U}_S > \lambda_o$, is greater than that of a stimulus trial. However, according to the model, the subject will report a detection. A definition of optimality suggested by that argument would require that such a situation does not arise; that is,

$$\text{Prob}\{\mathbf{U}_s > \lambda\}\pi \geq \text{Prob}\{\mathbf{U}_n > \lambda\}(1 - \pi) ,$$

for all criterion values λ. This definition does not take into account the monetary gains or losses resulting from the strategy. Other definitions of optimality are conceivable, one of which will be considered shortly.

8.4. Axioms of the Random Variable Model

For convenient reference, the assumptions of the model are summarized in the following three axioms.

Axiom SD1. To each stimulus s and noise n correspond independent random variables, respectively, $\mathbf{U}_s$ and $\mathbf{U}_n$. The presentation of s evokes a sample value of the random variable $\mathbf{U}_s$. Similarly, the presentation of n evokes a sample value of the random variable $\mathbf{U}_n$.

Axiom SD2. For any payoff matrix $\theta \in \Theta$, the hit and false alarm probabilities satisfy the equation for $S = s,n$,

$$p_S(\theta) = \text{Prob}\{\mathbf{U}_S > \lambda_\theta\} .$$

Occasionally, we shall also assume:

Axiom SD3. The random variables $\mathbf{U}_s$ and $\mathbf{U}_n$ have densities

$$f_s \text{ and } f_n, \text{ with } f_n > 0 .$$

8.5. ROC Analysis and Likelihood Radio

The slope of the ROC curve is susceptible to an interesting interpretation. Let us assume that Axioms SD1–SD3 hold. Thus the random variables $\mathbf{U}_s$ and $\mathbf{U}_n$ have densities f_s and f_n, respectively, with $f_n > 0$. Writing, as before, ρ for the ROC function, we have successively

$$d\rho[p_n(\theta)]/dp_n(\theta) = \frac{d\text{Prob}\{\mathbf{U}_s > \lambda_\theta\}}{d\text{Prob}\{\mathbf{U}_n > \lambda_n\}}$$

$$= \frac{d[1 - \text{Prob}\{\mathbf{U}_s \leq \lambda_\theta\}]}{d[1 - \text{Prob}\{\mathbf{U}_n \leq \lambda_\theta\}]}$$

$$= \frac{-f_s(\lambda_\theta)}{-f_n(\lambda_\theta)}$$

$$= \frac{f_s(\lambda_\theta)}{f_n(\lambda_\theta)} . \quad (87)$$

In other terms, the slope of the ROC curve evaluated at the point $p_n(\theta)$ is equal to the ratio of the densities at that point.

Notice for further reference the monotonicity relation between the ratio $f_s(\lambda_\theta)/f_n(\lambda_\theta)$ and the slope of the ROC curve. Since as a consequence of Eq. (83), λ_θ decreases as $p_n(\theta)$ increases, a decrease in the slope of the ROC function in some interval corresponds to an increase in the ratio $f_s(\lambda_\theta)/f_n(\lambda_\theta)$, in the corresponding interval of the variable λ_θ. Typical data strongly support the assumption of concave ROC functions—that is, ROC functions with nonincreasing slopes. This suggests that the ratio $f_s(\lambda)/f_n(\lambda)$ should be an increasing function of λ. We shall go back to this point.

In statistical decision theory, a ratio of densities, such as the one appearing in Eq. (87), is often called a *likelihood ratio*. In fact, with a slight strengthening of our assumptions, the random variable model discussed here is consistent with a fundamental rule used in statistical decision theory. To the Axioms SD1–SD3, we shall add the following:

Axiom SD4. The likelihood ratio

$$l(x) = f_s(x)/f_n(x)$$

is a strictly increasing function of x.

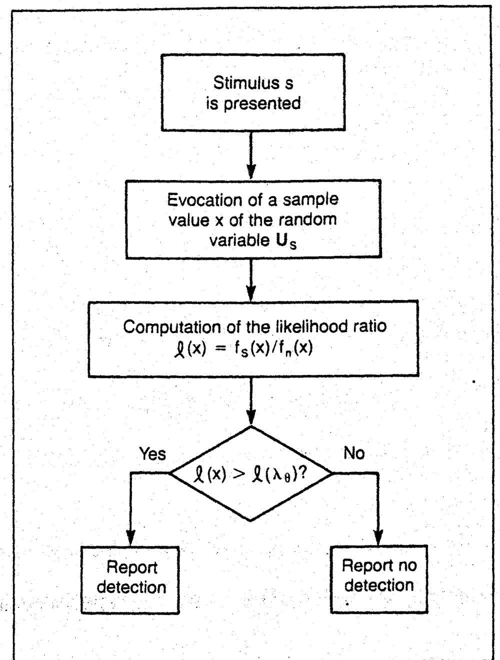


Figure 1.19. Successive stages of the decision process elicited by the presentation of the stimulus, in case of the likelihood ratio model. The diagram is identical in the case of the presentation of the noise n, except that $\mathbf{U}_s$ is replaced by $\mathbf{U}_n$.

One implication of assumptions SD1–SD4 has been indicated above: the slope of the ROC curve must be strictly decreasing. (That is, in terms of the definition in Section 7.4.1, the ROC function must be strictly concave.) Another consequence is of interest, since it suggests a drastically different interpretation of the model. By Axiom SD4, the likelihood function l is strictly increasing, which implies (see Section 8.3)

$$p_s(\theta) = \text{Prob}\{\mathbf{U}_s > \lambda_\theta\}$$

$$= \text{Prob}\{l(\mathbf{U}_s) > l(\lambda_\theta)\}$$

$$= \text{Prob}\{f_s(\mathbf{U}_s)/f_n(\mathbf{U}_s) > l(\lambda_\theta)\} .$$

By a similar argument, we also obtain

$$p_n(\theta) = \text{Prob}\{f_s(\mathbf{U}_n)/f_n(\mathbf{U}_n) > l(\lambda_\theta)\} .$$

The last two equations prompt a comparison of the subject's strategy with that of a statistician engaged in a decision task and applying an optimal decision procedure. The successive steps of the procedure are reviewed in Figure 1.19, which is self-explanatory. The statistician receiving a signal of value x must decide in some optimal fashion whether this signal is a sample of $\mathbf{U}_s$ or a sample of $\mathbf{U}_n$. We will suppose that the decision procedure maximizes the expected value of the gain, as determined by the payoff matrix θ. Let $\gamma(\theta)_{ss}$ and $\gamma(\theta)_{nn}$ be the gains resulting from a hit and a correct rejection, respectively; let $\gamma(\theta)_{ns}$ and $\gamma(\theta)_{sn}$ be the costs attached to a false alarm and a miss. Let π be the probability that a stimulus is presented on any trial. The expected value $G(\theta,\pi)$ of the gain is easily computed from the tree diagram in Figure 1.20, which displays the possible paths and their probabilities. We obtain:

$$G(\theta,\pi) = \pi\{p_s(\theta)\gamma_{ss}(\theta) - [1 - p_s(\theta)]\gamma_{ns}(\theta)\}$$
$$+ (1 - \pi)\{[1 - p_n(\theta)]\gamma_{nn}(\theta) - p_n(\theta)\gamma_{sn}(\theta)\} ,$$

which we rewrite

$$G(\theta,\pi) = \pi[\gamma_{ss}(\theta) + \gamma_{ns}(\theta)][p_s(\theta) - p_n(\theta)\beta_\theta] , \quad (88)$$

with

$$\beta_\theta = (1 - \pi)[\gamma_{sn}(\theta) + \gamma_{nn}(\theta)]/\pi[\gamma_{ss}(\theta) + \gamma_{ns}(\theta)] . \quad (89)$$

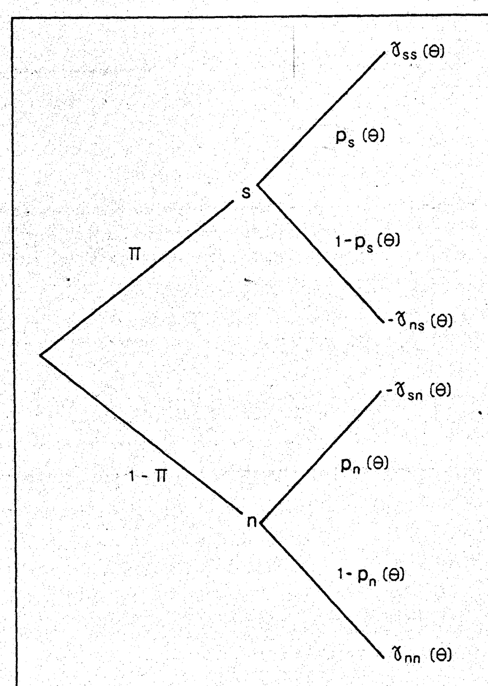


Figure 1.20. Tree diagram of the possible paths in the yes-no paradigm with their probabilities and their outcomes. See text for the definition of symbols.

The statistician decision procedure concerns the response probabilities $p_s(\theta)$ and $p_n(\theta)$, which can be manipulated via the quantity λ_θ in the equations

$$p_s(\theta) = \text{Prob}\{\mathbf{U}_s > \lambda_\theta\}$$

and

$$p_n(\theta) = \text{Prob}\{\mathbf{U}_n > \lambda_\theta\} .$$

Since $\pi[\gamma_{ss}(\theta) + \gamma_{ns}(\theta)]$ is constant, a value of λ_θ maximizes the expected gains $G(\theta,\pi)$ in Eq. (88) iff it maximizes

$$\text{Prob}\{\mathbf{U}_s > \lambda_\theta\} - \text{Prob}\{\mathbf{U}_n > \lambda_\theta\}\beta_\theta .$$

It follows that the required value of λ_θ must satisfy

$$f_s(\lambda_\theta) - f_n(\lambda_\theta)\beta_\theta = 0 ,$$

that is

$$l(\lambda_\theta) = \frac{f_s(\lambda_\theta)}{f_n(\lambda_\theta)} = \beta_\theta .$$

We conclude that the subject strategy is optimal in the sense of a maximization of the expected gain if the response probabilities satisfy the two equations

$$p_s(\theta) = \text{Prob}\{\mathbf{U}_s > l^{-1}(\beta_\theta)\}$$

$$p_n(\theta) = \text{Prob}\{\mathbf{U}_n > l^{-1}(\beta_\theta)\} ,$$

with β_θ defined by Eq. (89).

If precise assumptions are made concerning the distributions of the random variables $\mathbf{U}_s$ and $\mathbf{U}_n$, it can then be checked whether the subject's strategy is optimal in the above sense, by evaluating the fit of the above equation to the data. This comparison of the subject's strategy with that of a statistician engaged in a decision-making task was discussed in some detail, since it is an inherent part of the common wisdom in this field. It must be clear, however, that the analysis of the data in terms of ROC curves is a useful device to disentangle sensory from cognitive components of the task, whether or not the subject's strategy happens to be optimal.

This analysis is also valuable, or at least relevant, in cases of experimental procedures or paradigms somewhat different from those envisaged so far in this section. Two examples are discussed in Sections 8.6 and 8.7.

8.6. ROC Analysis and the Forced-Choice Paradigm

In the two-alternative forced-choice (2AFC) paradigm, the subject's task is to decide on every trial which of two locations, or two intervals of time, contains the stimulus. Even though the effect on performance of guessing strategies is minimized in such paradigms, an ROC analysis will be useful. In particular, the connections between the predictions in the yes-no and the 2AFC paradigms are of interest.

For concreteness, we consider as before an auditory detection situation. On every trial, the subject is presented with two successive intervals of time, of equal duration, one of which containing the stimulus (a click, say) embedded in noise, the other containing only the noise. There are thus two types of trials, depending on whether the stimulus was in the first or the second interval. We shall denote these two cases by (s,n) and (n,s), respectively. Let $p_{1,sn}$ and $p_{2,ns}$ be the corresponding probabilities of a correct response, and let $p_{2,sn}$ and $p_{1,ns}$ be the error probabilities. By design, we must have

$$p_{1,sn} + p_{2,sn} = 1$$

and

$$p_{1,ns} + p_{2,ns} = 1 ,$$

since the subject is forced to choose one of the two intervals on every trial. For the time being, let us suppose that the two probabilities of a correct response are equal,

$$p_{1,sn} = p_{2,ns} .$$

This assumption, which is not always realistic and can be rejected for some data, will be relaxed in a moment. From a purely sensory viewpoint, the 2AFC paradigm differs but little from the yes-no paradigm, and it makes sense to apply the same theoretical analysis. Let us assume that

$$p_{1,sn} = \text{Prob}\{\mathbf{U}_s > \mathbf{U}_n\} = p_{2,ns} , \qquad (90)$$

in which $\mathbf{U}_s$ and $\mathbf{U}_n$ are independent random variables with the same interpretations as in Section 8.2. If we assume that $\mathbf{U}_s$ and $\mathbf{U}_n$ are continuous, we have

$$\text{Prob}\{\mathbf{U}_s = \mathbf{U}_n\} = 0 ,$$

which implies for the probabilities of errors,

$$p_{2,sn} = 1 - p_{1,sn} = \text{Prob}\{\mathbf{U}_n > \mathbf{U}_s\} = p_{1,ns} .$$

The idea is that each of the two intervals provides a sample of one of the random variables $\mathbf{U}_s$ and $\mathbf{U}_n$, and the subject's response is based on a comparison of these samples. In the case of an (s,n) trial, for instance, if x_1 and x_2 are sample values of $\mathbf{U}_s$ and $\mathbf{U}_n$, respectively, the subject will choose interval 1 (the correct one) if $x_1 > x_2$.

Notice that, under Axiom SD4, we have

$$\text{Prob}\{\mathbf{U}_s > \mathbf{U}_n\} = \text{Prob}\{f_s(\mathbf{U}_s)/f_n(\mathbf{U}_s) > f_s(\mathbf{U}_n)/f_n(\mathbf{U}_n)\} .$$

This means that the above interpretation of the subject's decision process as based on a comparison of samples of $\mathbf{U}_s$ and $\mathbf{U}_n$ is equivalent to another, in which the subject would behave as a statistician and compare likelihood ratios.

In any event, the conclusion to be derived from Eqs. (85) and (90) is that the probability of a correct response in the 2AFC paradigm, under the assumption that $p_{1,sn} = p_{2,ns}$, is equal to the area under the ROC curve in the corresponding yes-no paradigm.

As indicated, the assumption that $p_{1,sn} = p_{2,sn}$ may be unrealistic. We shall briefly examine here the possibility that the subject may be biased toward one of the two intervals. A systematic way of inducing such bias would be to assign different probabilities to the events (s,n) and (n,s). Our random variable model for the 2AFC paradigm can be generalized as follows.

Let Θ be a set of bias-inducing conditions; let $p_{1,sn}(\theta)$ and $p_{1,ns}(\theta)$ be the two probabilities of choosing the first interval, in condition θ, for the two cases (s,n) and (n,s). We assume that the effect of a given condition $\theta \in \Theta$ is to transform the distribution of the random variable corresponding to the second interval. Specifically, we assume that the following two equations hold:

$$p_{1,sn}(\theta) = \text{Prob}\{\mathbf{U}_s > g_\theta(\mathbf{U}_n)\}$$

$$p_{1,ns}(\theta) = \text{Prob}\{\mathbf{U}_n > g_\theta(\mathbf{U}_s)\} ,$$

where g_θ is a strictly increasing, continuous function. With obvious notation, the two remaining probabilities are computed from the equations

$$p_{1,sn}(\theta) + p_{2,sn}(\theta) = 1 ,$$

$$p_{1,ns}(\theta) + p_{2,ns}(\theta) = 1 ,$$

which are inherent to the 2AFC paradigm. Let us suppose for a moment that the set of points

$$[p_{1,ns}(\theta), p_{1,sn}(\theta)] ,$$

generated by varying $\theta \in \Theta$, is an ROC curve. Under fairly general properties on the set of transformations $\{g_\theta | \theta \in \Theta\}$, it follows then that this ROC curve must be symmetric with respect to the negative diagonal of the unit square (see Figure 1.21). One such property is that if g_θ is a strictly increasing transformation, then there must be some condition $\theta' \in \Theta$ corresponding to the "opposite" transformation, $g_\theta^{-1} = g_{\theta'}$. Indeed, we have then

$$1 - p_{1,sn}(\theta) = \text{Prob}\{g_\theta(\mathbf{U}_n) > \mathbf{U}_s\}$$

$$= \text{Prob}\{\mathbf{U}_n > g_\theta^{-1}(\mathbf{U}_s)\}$$

$$= \text{Prob}\{\mathbf{U}_n > g_{\theta'}(\mathbf{U}_s)\}$$

$$= p_{1,ns}(\theta') ,$$

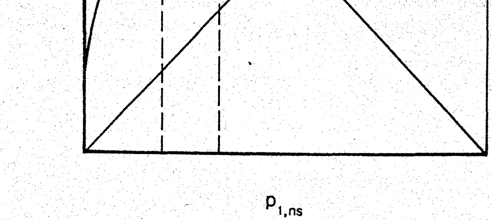


Figure 1.21. Hypothetical ROC curve symmetric with respect to the negative diagonal of the unit square, in the 2AFC paradigm. We have $p_{1,ns}(\theta') = 1 - p_{1,sn}(\theta)$.

and

$$p_{1,sn}(\theta') = \text{Prob}\{\mathbf{U}_s > g_{\theta'}(\mathbf{U}_n)\}$$

$$= \text{Prob}\{\mathbf{U}_s > g_\theta^{-1}(\mathbf{U}_n)\}$$

$$= \text{Prob}\{g_\theta(\mathbf{U}_s) > \mathbf{U}_n\}$$

$$= 1 - p_{1,ns}(\theta) .$$

The two equations,

$$1 - p_{1,sn}(\theta) = p_{1,ns}(\theta') ,$$

$$p_{1,sn}(\theta') = 1 - p_{1,ns}(\theta)$$

express the symmetry property of the ROC curve mentioned above. This situation is illustrated in Figure 1.21.

8.7. ROC Analysis of Rating-Scale Data

In the same experimental situation, consider a procedure in which, rather than giving a yes-no detection response on every trial, the subject is required to quantify the certainty that the stimulus was presented. Suppose, for example, that a six-category rating scale is used, ranging from 0 (certainty that the stimulus was not presented) to 5 (certainty that the stimulus was presented). Some hypothetical but plausible data are given in Table 1.3. Let $\mathbf{R}_s$ and $\mathbf{R}_n$ be two random variables corresponding to the ratings in the two types of trials. (For example, Prob$\{\mathbf{R}_s = 3\}$ is the probability of observing a rating of 3 on a trial when the stimulus was presented.) Since the experimental situation is unchanged except for the subject's responses, it makes sense to suppose that the same underlying activation random variables $\mathbf{U}_s$ and $\mathbf{U}_n$ are responsible for the ratings. The following model seems reasonable: an observed rating will exceed a value i ($i = 0, \ldots, 4$) only if the activation random variable exceeds a criterion λ_i, the value of which depends on the rating value considered. In symbols,

$$\text{Prob}\{\mathbf{R}_s > i\} = \text{Prob}\{\mathbf{U}_s > \lambda_i\} ,$$

$$\text{Prob}\{\mathbf{R}_n > i\} = \text{Prob}\{\mathbf{U}_n > \lambda_i\} .$$

Observe that the right members of these two equations strongly resemble those in Axiom SD2 of the yes-no procedure. This suggests an ROC analysis of the data. It is as if each possible value of the rating (with the exception of the maximal one) would implicitly define a particular payoff matrix and a recoding of the rating data into two yes-no classes. In Table 1.3, the value $i = 3$ leads to the recoding

write "yes" if $i = 4, 5$
write "no" if $i = 1, 2, 3,$

Table 1.3. Hypothetical Rating Data in a Signal Detection Task

	Rating Value					
	0	1	2	3	4	5
Noise trials	.10	.15	.35	.20	.15	.05
Stimulus trials	.05	.10	.30	.20	.25	.10

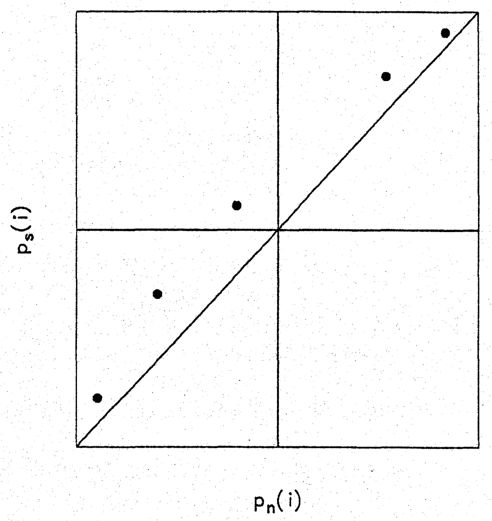


Figure 1.22. ROC graph obtained for the rating scale data.

with

proportion of false alarms	.05	+	.15	=	.20	
proportion of hits	.10	+	.25	=	.35.	

The results of this recoding for Table 1.3 are:

	Proportion of Ratings Exceeding i				
i	0	1	2	3	4
n	.90	.75	.40	.20	.05
s	.95	.85	.55	.35	.10

The corresponding ROC graph is displayed in Figure 1.22. In general, the probabilities of hits and false alarms corresponding to each rating value i up to (but not including) the maximal one are given by the equations

$$p_s(i) = \text{Prob}\{\mathbf{R}_s > i\}$$

$$p_n(i) = \text{Prob}\{\mathbf{R}_n > i\} ,$$

with the ROC function $p_n(i) \mapsto p_s(i)$.

An obvious advantage of this method is its efficiency. The subject is required to make more sophisticated responses than in the yes-no procedure, which results in a substantial economy in the collection of data. This was illustrated in our example, in which only one condition, rather than five, had to be run to obtain a five-point ROC graph.

On the negative side, it must be noted that the points of an experimental ROC graph are not independent, which may create difficulties in fitting and evaluating a model.

Finally, data collected by the rating-scale procedure, but analyzed by methods different from those discussed here, may provide a sharp test of some models. We return to this point in Section 8.10.

8.8. Gaussian Assumption

In principle, an ROC analysis is feasible without making any assumptions on the distributions of the random variables $\mathbf{U}_s$ and $\mathbf{U}_n$ (cf. Bramber, 1975). However, the application is greatly facilitated if such assumptions are made. We discuss here the case of Gaussian distributions.

8.8.1. Yes-No Paradigm. In the yes-no paradigm with a payoff matrix θ, we shall assume that

$$p_s(\theta) = \text{Prob}\{\mathbf{U}_s > \lambda_\theta\} = \Phi\left(\frac{\mu_s - \lambda_\theta}{\sigma_s}\right) , \quad (91)$$

$$p_n(\theta) = \text{Prob}\{\mathbf{U}_n > \lambda_\theta\} = \Phi\left(\frac{\mu_n - \lambda_\theta}{\sigma_n}\right) , \quad (92)$$

in which our notations are as in Section 8.2, μ_s, μ_n, and σ_s, σ_n denote the means and standard deviations of the random variables $\mathbf{U}_s$ and $\mathbf{U}_n$, and Φ is the distribution function of a standard normal random variable, that is,

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-x^2/2} dx .$$

From Eqs. (91) and (92), it is apparent that the ROC curve is determined by four parameters: the means and the standard deviations of the random variables $\mathbf{U}_s$ and $\mathbf{U}_n$. (In fact, we shall see that only two parameters are necessary.) From a practical viewpoint it will be convenient to rewrite Eqs. (91) and (92) in terms of the so-called z-scores. With

$$z_s(\theta) = \Phi^{-1}[p_s(\theta)] ,$$

$$z_n(\theta) = \Phi^{-1}[p_n(\theta)] ,$$

and dropping θ in the notations, we obtain

$$z_s = \frac{\mu_s - \lambda}{\sigma_s}$$

$$z_n = \frac{\mu_n - \lambda}{\sigma_n} .$$

Eliminating λ in these equations and solving for z_s, yields

$$z_s = z_n \sigma_n / \sigma_s + \frac{\mu_s - \mu_n}{\sigma_s} . \quad (93)$$

In other terms, when the hit and false alarm probabilities are transformed into z-scores, the ROC curve is transformed into a straight line with slope σ_n/σ_s and intercept $(\mu_s - \mu_n)/\sigma_s$. Using linear regression, these two parameters can be estimated from the response frequencies of the data. Notice that the ROC curve only specifies two of the four parameters μ_s, μ_n, σ_s, and σ_n. For example, we can assume without loss of generality, that $\mu_n = 0$ and $\sigma_n = 1$. The area under the ROC curve can be computed from the equations

$$\text{Prob}\{\mathbf{U}_s > \mathbf{U}_n\} = \text{Prob}\{\mathbf{U}_s - \mathbf{U}_n > 0\}$$

$$= \Phi\left[\frac{\mu_s - \mu_n}{(\sigma_s^2 + \sigma_n^2)^{1/2}}\right] . \quad (94)$$

It is easy to show that this model satisfies Axiom SD4 only if $\sigma_s = \sigma_n$. (If we equate the two densities of $\mathbf{U}_s$ and $\mathbf{U}_n$ and take logarithms, a quadratic equation obtains, which has a unique solution only if $\sigma_s = \sigma_n$.) We shall investigate this particular case in some detail.

8.8.2. Equal Variance Assumption. Suppose that

$$\sigma_s = \sigma_n = 1 .$$

Equation (93), which specifies the transformed ROC curve, becomes

$$z_s = z_n + (\mu_s - \mu_n)$$

Thus in the special case where $\mathbf{U}_s$ and $\mathbf{U}_n$ are independent Gaussian random variables with equal variance, the transformed ROC curves are parallel straight lines with a slope equal to 1. Only one parameter remains in the model, which is the difference $\mu_s - \mu_n$. When this model is used, a standard measure of the detectability of the stimulus is

$$d' = (\mu_s - \mu_n)/\sqrt{2} .$$

This choice has some intuitive appeal, since d' is proportional to the difference between the means of the two activation distributions. Moreover, d' is closely related to the other measure, the area under the ROC curve. Indeed, from Eq. (94) we have

$$\text{Prob}\{\mathbf{U}_s > \mathbf{U}_n\} = \Phi(d') .$$

Occasionally, it is convenient to plot the empirical ROC graphs and the theoretical ROC curves on "double-probability" paper (a two-dimensional Cartesian representation in which the coordinates are in units of the normal integral; see Figure 1.23).

8.9. Threshold Theory

A rather different interpretation of an ROC analysis of yes-no detection data is possible, in which the basic, underlying notions are not activation random variables but detection states. A number of such models have been proposed, which differ in particular by the number of (unobservable) states postulated or by the exact relation linking the states to the response probabilities or other observable quantities (e.g., response latencies, ratings). We shall discuss a simple example, due to Luce (1960, 1963a, 1963b).

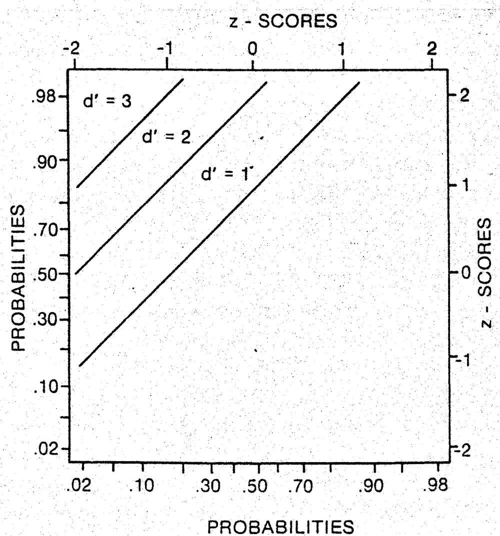


Figure 1.23. In the equal variance case, three ROC curves plotted on "double probability" paper, corresponding to the cases $d' = 1, 2, 3$.

We will assume that the presentation of the stimulus or the noise elicits one of two sensory states in the subject: either a neural threshold has been exceeded or it has not. The event that the threshold is exceeded may lead to a "yes" response (the subject reports a detection) but not necessarily so. We also assume that a given payoff matrix θ may induce one of two opposite response strategies: (1) a *conservative strategy,* in which the subject never says "yes" when the threshold has not been exceeded; when the threshold has been exceeded, the subject only says "yes" with a probability β_θ, depending on the payoff matrix, and (2) a guessing strategy, in which the subject always says "yes" when the threshold has been exceeded; when the threshold has not been exceeded, the subject says "yes" with a probability α_θ, depending on the payoff matrix. This means that the collection Θ of payoff matrices is partitioned into two classes: (1) Θ_c, the set of payoff matrices inducing a conservative strategy, and (2) Θ_g, the set of payoff matrices inducing a guessing strategy. The event that the threshold has been exceeded will be denoted $D = 1$; the complementary event will be denoted $D = 0$. Thus in the framework of a probabilistic model, D is a random variable taking values 0,1. As before, the letter S denotes the stimulation; we have two cases: $S = s$ (the stimulus is presented) and $S = n$ (only the noise is presented). The probability that the stimulation determines a neural event exceeding the threshold ($D = 1$) only depends on S and will be denoted $q(S)$. Notice that we have introduced four numerical parameters: two for the response probabilities, β_θ and α_θ, and two for the probabilities of the states, $q(s)$ and $q(n)$. In the framework of an ROC analysis, however, two of these parameters will be eliminated in the equations, leaving only $q(s)$ and $q(n)$ to be estimated from the data. Finally, we denote by Y_θ and N_θ the two events of a "yes" and a "no" response, respectively.

8.9.1. Axioms for the Threshold Theory. We provide a compact summary of these assumptions in the form of two axioms.

State Axiom T1.

$$\text{Prob}\{D = 1|S\} = q(S) , \quad \text{for } S = s, n .$$

(The probability that the threshold is exceeded is equal to $q(s)$ if the stimulus is presented and to $q(n)$ if the noise is presented, these probabilities being independent of the payoff matrix.)

Response Axiom T2. For any payoff matrix θ,

$$\text{Prob}\{Y_\theta|S,D\} = \begin{cases} (1 - D)\alpha_\theta + D & \text{if } \theta \in \Theta_g, \\ \beta_\theta D & \text{if } \theta \in \Theta_c, \end{cases}$$

independent of S.
(The probability of a "yes" response to a stimulation only depends on the payoff matrix θ and whether the threshold has been exceeded. If θ is in Θ_g, it is equal to 1 or α_θ, depending on whether $D = 1$ or $D = 0$, respectively. If θ is in Θ_c, this probability is equal β_θ or 0, again depending on whether $D = 1$ or $D = 0$.)

8.9.2. Form of the ROC Curve. As shown by a simple calculation, these axioms predict an ROC curve made of two

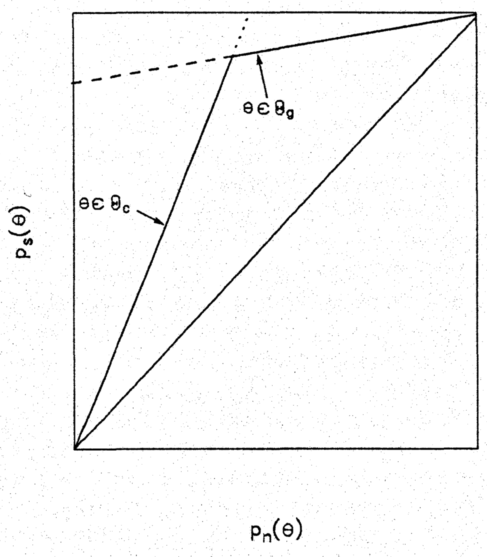


Figure 1.24. An example of an ROC curve in the two-state threshold model. The upper limb of the curve corresponds to Eq. (96), $\theta \in \Theta_g$; the lower limb corresponds to Eq. (97), $\theta \in \Theta_c$.

segments of a straight line (see Figure 1.24). The upper limb describes the guessing strategy and contains the corner (1,1) of the unit square. The points of that segment are generated by varying θ in Θ_g. The lower limb describes the conservative strategy, contains the point (0,0), and is generated by varying θ in Θ_c. Let us demonstrate this.

With our usual notations, $p_s(\theta)$ and $p_n(\theta)$ for the two probabilities of a "yes" response, we have

$$p_S(\theta) = \text{Prob}\{Y_\theta|S\}$$

$$= \text{Prob}\{Y_\theta|D = 1,S\}\,\text{Prob}\{D = 1|S\}$$

$$+ \text{Prob}\{Y_\theta|D = 0,S\}\,\text{Prob}\{D = 0|S\} .$$

Using Axiom T1, this yields

$$p_S(\theta) = \text{Prob}\{Y_\theta|D = 1,S\}q(S) + \text{Prob}\{Y_\theta|D = 0\}[1 - q(S)] . \quad (95)$$

8.9.2.1. Case $\theta \in \Theta_g$. Using Axiom T2, Eq. (95) specializes into

$$p_s(\theta) = q(s) + \alpha_\theta[1 - q(s)] ,$$

$$p_n(\theta) = q(n) + \alpha_\theta[1 - q(n)] .$$

Eliminating α_θ in these two equations, dropping θ in the notations, and solving for p_s, we obtain

$$p_s = p_n[1 - q(s)]/[1 - q(n)] + [q(s) - q(n)]/[1 - q(n)] , \quad (96)$$

a linear function containing the point (1,1). Thus as θ varies in Θ_g, the point of the ROC moves along a segment of a straight line specified by Eq. (96). Notice that in this case, we have $q(n) \leqslant p_n$.

8.9.2.2. Case $\theta \in \Theta_c$. From Eq. (95) and Axiom T2, we obtain

$$p_s(\theta) = \beta_\theta q(s) ,$$

$$p_n(\theta) = \beta_\theta q(n) .$$

Eliminating β_θ and dropping θ in the notation yields

$$p_s = p_n q(s)/q(n) , \quad (97)$$

the equation of a straight line going through the origin. Here we have $p_n \leqslant q(n)$.

Equations (96) and (97) together specify the class of ROC curves predicted by Luce's two-state threshold theory. This prediction has been shown to hold reasonably well for some data (cf. Luce, 1963a). In other cases, however, the theory is not so successful. For example, Nachmias and Steinman (1963) have shown that in some empirical situations the probability $q(n)$ that the threshold is exceeded on noise trials (as estimated from the data) has to vary with signal strength. Such a fact is obviously difficult to accommodate in the framework of the two-state threshold theory.

8.10. Rating Data and the Threshold Theory

It is natural to inquire about the predictions of the threshold theory concerning data obtained by the rating-scale procedure. Some authors have been quick to point out that rating-scale data characteristically favor an ROC function with a smooth curvature, a fact which may appear to be inconsistent with the two segments of straight lines predicted by the threshold theory (Broadbent & Gregory, 1963; Nachmias & Steinman, 1963; Swets, 1961; Watson, Rilling, & Bourbon, 1964). Actually, as stated above and in the cited papers of Luce, the theory is not relevant to rating data, and no inferences can legitimately be made in this respect. (The only "response axiom" is T2, which concerns itself specifically with the response probabilities in the yes-no paradigm.) If rating-scale data are to be predicted by the theory, a new axiom is required, and there are various candidates, one of which is briefly considered here. Our reasons for including such discussion in this chapter are twofold: (1) to show by a counterexample that the argument against the two-state theory based on the curvature of the ROC curve implied by the data does not apply and (2) to demonstrate the general vulnerability of two-state theories to a particular type of analysis of the data.

We make the reasonable assumption that the rating given by the subject on a trial only depends on the sensory state evoked by the stimulation. However, the exact value of the rating is not determined by the state. To each of the two sensory states, corresponding to the events $D = 1$, $D = 0$, corresponds a rating random variable, with distribution function G_1, G_0, respectively. In other terms, with $\mathbf{R}_s$ and $\mathbf{R}_n$ as in Eq. (88), we have the following axiom:

Axiom T3. For $S = s, n$ and $D = 0, 1$,

$$\text{Prob}\{\mathbf{R}_S \leqslant i|D\} = G_D(i) ,$$

independent of S.

Let us derive the prediction for the ROC curve. For $S = s, n$, we have

$$\text{Prob}\{\mathbf{R}_S \leq i\} = \text{Prob}\{\mathbf{R}_S \leq i|D = 1\}\text{Prob}\{D = 1|S\}$$
$$+ \text{Prob}\{\mathbf{R}_S \leq i|D = 0\}\text{Prob}\{D = 0|S\}$$
$$= G_1(i)q(S) + G_0(i)[1 - q(S)] \ .$$

Specializing this equation for the two cases $S = s$ and $S = n$, we obtain

$$\text{Prob}\{\mathbf{R}_s \leq i\} = q(s)G_1(i) + [1 - q(s)]G_0(i) \ , \quad (98)$$

$$\text{Prob}\{\mathbf{R}_n \leq i\} = q(n)G_1(i) + [1 - q(n)]G_0(i) \ . \quad (99)$$

Eliminating $G_1(i)$ in these two equations and solving for $\text{Prob}\{\mathbf{R}_s > i\}$ yields the following prediction for the ROC curve:

$$\text{Prob}\{\mathbf{R}_s > i\} = \text{Prob}\{\mathbf{R}_n > i\}q(s)/q(n)$$
$$- [1 - G_0(i)][q(s) - q(n)]/q(n) \ . \quad (100)$$

Notice that the ROC function defined by Eq. (100) depends on G_0. This implies that the corresponding ROC curve is not necessarily made of two segments of straight lines. In fact, a cursory investigation suggests that for an appropriate choice of the distribution function G_0, this equation may provide an acceptable fit to ROC data obtained from the rating-scale procedure.

On the other hand, it is doubtful that this particular version of the two-state theory is viable, since it makes extremely strong predictions concerning some other aspect of rating data. Using an argument of Falmagne (1968), Vorberg (Note 3) points out that the observed distributions of ratings should conform to a very constraining fixed-point property, stemming from the fact that, as indicated by Eqs. (98), (99), the distribution of ratings for any stimulus s (or noise n) is a "mixture" of the two latent distributions G_1 and G_0, in proportions $q(s)$ and $1 - q(s)$ (or $q(n)$ and $1 - q(n)$). This property is easily stated in words. Consider the empirical histograms of ratings obtained for s and n in some situation. Suppose that these two histograms "cross" each other at some value j (say, the proportions of ratings j are not significantly different). Then the histogram of ratings obtained for any other stimulus s' should have, except for statistical errors, the same proportion of ratings j (see Figure 1.25). The argument is as follows. Let k_s, k_n, g_0, and g_1 be the densities of $\mathbf{R}_s$, $\mathbf{R}_n$, G_0, and G_1, respectively. Thus these densities idealize the histograms mentioned above. Taking derivatives in Eqs. (98) and (99) gives

$$k_s(i) = q(s)g_1(i) + [1 - q(s)]g_0(i) \quad (101)$$

and

$$k_n(i) = q(n)g_1(i) + [1 - q(n)]g_0(i) \ . \quad (102)$$

Suppose that for some rating value j, we have

$$k_s(j) = k_n(j) \ .$$

From Eqs. (101) and (102) it follows necessarily that

$$g_1(j) = g_0(j) = k_1(j) = k_0(j) \ .$$

Consequently, if s' is some other stimulus, we must have

$$k_{s'}(j) = q(s')g_1(j) + [1 - q(s')]g_0(j)$$
$$= g_1(j)$$
$$= k_s(j) = k_n(j) \ ,$$

as predicted.

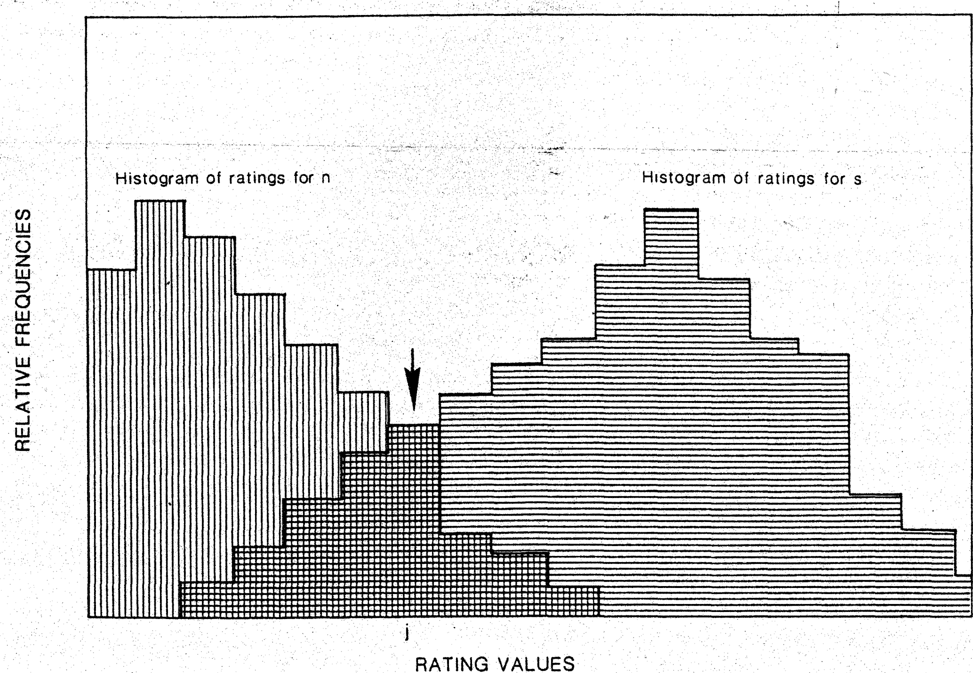


Figure 1.25. The fixed point property of the two-state threshold model applied to hypothetical rating scale data. The two histograms "cross" at the point indicated by the arrow, corresponding to rating j; any other histogram (say, of s') should go through the same point.

8.11. Key References

The notions of signal detection theory discussed in this section were selected, from a vast literature, as being the most central from the viewpoint of the analysis of psychophysical data. This theory, a very detailed account of which can be found in Green and Swets (1974), originated from an adaptation by W. P. Tanner and his coworkers at the University of Michigan, of a number of optimal procedures for the detection of signal in noise (Peterson, Birdsall, & Fox, 1954; van Meter & Middleton, 1954). In turn, these procedures are based on statistical decision theory (Neyman & Pearson, 1933; Wald, 1947, 1950). As emphasized by our presentation, in which the role of optimality is played down, signal detection theory has also a valid claim to the parentage of the law of comparative judgments (Thurstone 1927a, 1927b).

The applications of signal detection theory were first in psychophysics (e.g., Tanner & Swets, 1953, 1954a, 1954b) but were very quickly extended to other fields. The extraordinary success of the theory is evidenced by the number and variety of the papers in which it is used in some form or other. Today, applications can be found, for instance, in learning, memory, medical diagnosis, personality, reaction time, and skills (vigilance). A large sample of the early papers is collected in Swets (1964). Green and Swets (1974), the basic reference on this topic, contains a very extensive bibliography. As indicated in the text, various forms of the theory are obtained depending on specific assumptions made on the distributions of the random variables $\mathbf{U}_s$ and $\mathbf{U}_n$. A discussion of these special cases is provided in Egan (1975). A number of versions of the threshold theory are examined in Krantz (1969).

For applications of the basic notions of signal detection theory to other paradigms, see Sperling and Dosher, Chapter 2.

9. PSYCHOPHYSICS WITH SEVERAL VARIABLES OR CHANNELS

We consider here a number of paradigms and models designed to analyze how a subject integrates the information flowing from different sensory inputs. Examples of how this may arise have been encountered earlier in this chapter. For instance, in the yes-no paradigm discussed in Section 8, the subject had to detect a stimulus s embedded in a masking noise n. The subject's responses were regarded as resulting from some operation combining, on the sensory side, the effect of both s and n on the organism and, on the cognitive side, factors affecting decision making.

Another example, which this section treats in some detail, is offered by an auditory detection situation in which a stimulus is presented binaurally. The intensity in the two auditory channels may be manipulated independently, and the resulting performance may be investigated. This section is devoted to a general study of such situations, various cases of which will be given. Our purpose is not to provide an extensive survey. Rather, our selection of examples aims at familiarizing the reader with a collection of useful tools.

The word *channel* is of standard usage in psychophysics. As far as we know, however, no satisfying, generally accepted definition has been given for this term, even though several have been proposed. Depending on the context, *two channels* may mean that two sensory modalities are involved, or two neurophysiological locations, or two psychophysical variables, or even the same psychophysical variable but with different intensities. For the time being, we urge the reader to use the term intuitively and to check any ambitious drive toward rigor or consistency.

9.1. A General Model for Two-Channel Detection

9.1.1. Detection of Binaural Stimuli. In a version of the yes-no paradigm, the stimulus is a binaural, 1000-Hz tone (a,x) embedded in a masking noise (n,n'). The letters a, x denote the intensities of the stimulus in the left and right auditory channels, respectively; n and n' stand for the intensities of the noise in the two channels. As in the standard yes-no paradigm, the noise is presented alone on some proportion of the trials. To evaluate a possible response bias, the experimenter varies the payoff matrix across conditions. (See Section 8 for a discussion of payoff matrices.) Let us denote by $p_{ax}(\theta)$ and $p_{nn'}(\theta)$ the two probabilities of a "yes" response on a stimulus trial and on a noise trial, respectively, with a payoff matrix θ. (Our notation is slightly misleading. A more explicit but much heavier notation for these response probabilities would be $p_{ax,nn'}(\theta)$, $p_{oo,nn'}(\theta)$.)

This paradigm can obviously be transposed to other experimental situations (e.g., binocular perception as in Arditi, Chapter 23). From a theoretical viewpoint, the problem is to provide an explanation for the typical data: the presentation of the stimulation through two channels results in an improvement of detection performance.

9.1.2. The Model. In a natural extension of the signal detection model discussed in Section 8.2, we assume that the presentation of a stimulus of intensity a in the left auditory channel evokes some activity in a specific neural location, the level of which is represented by a random variable $\mathbf{U}_{1,a}$. Correspondingly, the presentation of x in the other channel generates a sample of a random variable $\mathbf{U}_{2,x}$. On noise trials, samples are taken from two "noise" random variables $\mathbf{V}_{1,n}$ and $\mathbf{V}_{2,n'}$. We assume that $(\mathbf{U}_{1,a},\mathbf{U}_{2,x})$ and $(\mathbf{V}_{1,n},\mathbf{V}_{2,n'})$ are pairs of independent random variables. On a trial where the stimulus (a,x) is presented, the information available to the organism is thus a sample of the pair of random variables $(\mathbf{U}_{1,a},\mathbf{U}_{2,x})$. We suppose that $\mathbf{U}_{1,a}$ and $\mathbf{U}_{2,x}$ are combined or pooled in some way, resulting in a random variable $\mathbf{Q}_{ax}$. The subject reports a detection if $\mathbf{Q}_{ax}$ exceeds a criterion λ_θ, the value of which depends on the payoff matrix θ. In other terms, we assume that there is some function F of two variables, the form of which is left unspecified for the moment, such that

$$\mathbf{Q}_{ax} = F(\mathbf{U}_{1,a},\mathbf{U}_{2,x}) .$$

The subject reports a detection if

$$\mathbf{Q}_{ax} > \lambda_\theta .$$

(See Figure 1.26.) Similar assumptions hold for the noise trials, with the same function F operating on the pair of random variables $(\mathbf{V}_{1,n},\mathbf{V}_{2,n'})$. The model is thus specified by the two equations

$$p_{ax}(\theta) = \text{Prob}\{F(\mathbf{U}_{1,a},\mathbf{U}_{2,x}) > \lambda_\theta\} , \qquad (103)$$

$$p_{nn'}(\theta) = \text{Prob}\{F(\mathbf{V}_{1,n},\mathbf{V}_{2,n'}) > \lambda_\theta\} . \qquad (104)$$

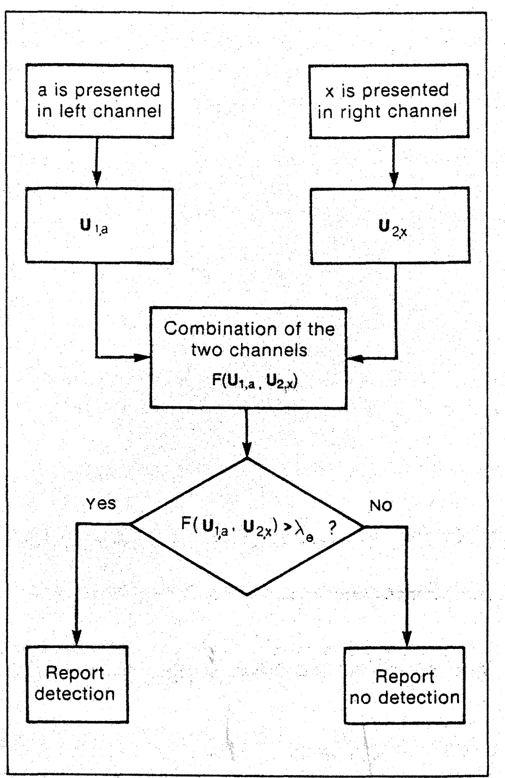


Figure 1.26. A general random variable model for the pooling of information from two sensory channels. Special cases of the model correspond to specifications of the function F.

It is clear that the data collected by varying the payoff matrix θ are amenable to a receiver operating characteristic (ROC) analysis (cf. Section 8). Applying the argument used in Section 8.2, we obtain as a measure of the area under the ROC curve, which, you will recall from Section 8.2, is a measure of performances independent of the effects of response criterion,

$$\text{Prob}\{F(\mathbf{U}_{1,a},\mathbf{U}_{2,x}) > F(\mathbf{V}_{1,n},\mathbf{V}_{2,n'})\} . \qquad (105)$$

Several particular cases of this model, which is also discussed by Olzak and Thomas in Chapter 7, are considered. These cases correspond to special forms of the function F in Eqs. (103), (104), and (105).

9.2. Probability Summation

Probability summation covers a class of models in which the improvement of performance resulting from having the stimulation delivered to two or more channels is attributed to chance alone. For an analogy, consider a group of $n \geq 2$ observers, watching the same visual display. Suppose that the probability p of detecting a faint stimulus is the same for all observers and that the group reports a detection if at least one of the n observers claims to have detected the stimulus. Assuming that the observers' responses are independent, the detection probability of the group is

$$1 - (1 - p)^n \geq p .$$

The application of this idea in psychophysics can be traced back to Pirenne (1943) and plays an important role in current theorizing, especially in visual perception. We shall limit our discussion to a two-channel situation. (For the case of a large number of channels, see Watson, Chapter 6, and Olzak and Thomas, Chapter 7.)

In the framework of the general model discussed in Section 9.1.2, this notion leads to the assumption that the subject reports a detection if at least one of the two activation random variables exceeds the criterion λ_θ. (Sometimes, different criteria are postulated for the two channels. This assumption seems to be more general. See, however, Section 9.3.) This means that the function F has the form

$$F(s,t) = \max\{s,t\} ,$$

in which *max* stands for the maximum in the set of numbers $\{s,t\}$. (Thus $\max\{s,t\} = s$ iff $s \geq t$.) Using the assumption of independence of the random variables, we obtain for the stimulus trials,

$$\begin{aligned} p_{ax}(\theta) &= \text{Prob}\{\max\{\mathbf{U}_{1,a},\mathbf{U}_{2,x}\} > \lambda_\theta\} \\ &= 1 - \text{Prob}\{\lambda_\theta \geq \max\{\mathbf{U}_{1,a},\mathbf{U}_{2,x}\}\} \\ &= 1 - \text{Prob}\{\lambda_\theta \geq \mathbf{U}_{1,a},\ \lambda_\theta \geq \mathbf{U}_{2,x}\} \\ &= 1 - \text{Prob}\{\lambda_\theta \geq \mathbf{U}_{1,a}\}\text{Prob}\{\lambda_\theta \geq \mathbf{U}_{2,x}\} . \qquad (106) \end{aligned}$$

Similarly, for the noise trials,

$$p_{nn'}(\theta) = 1 - \text{Prob}\{\mathbf{V}_{1,n} \leq \lambda_\theta\}\text{Prob}\{\mathbf{V}_{2,n'} \leq \lambda_\theta\} . \qquad (107)$$

A basic assumption is implicit in Eqs. (106) and (107), a clear statement of which is critical at this point.

9.2.1. Weak Criterion Invariance. Within a yes-no paradigm involving a stimulus (a,x) and a noise (n,n'), the criterion value λ_θ only depends on the payoff matrix θ. (This value is thus the same on stimulus trials and on noise trials.) This assumption is inherent to an ROC analysis and practically inescapable.

At this level of generality, it is not clear that the predictions of the model are sufficiently constraining to be rejected by available data. The negative evidence, some of which is reviewed by Blake and Fox (1973), is mostly circumstantial, by which we mean that it has no direct bearing on the predictions formally derivable from the assumptions. However, for (implicitly) fixed θ, Eq. (106) has been checked by various authors. Some refinements of the assumptions, considered below, lead to useful empirical tests.

9.2.2. Strong Criterion Invariance. Consider the following strengthening of the criterion invariance. The criterion value λ_θ only depends on the payoff matrix θ. In particular, for a given payoff matrix, this value is constant over conditions varying the intensities of the stimulus (a,x) and of the noise (n,n').

This assumption is frequently made (explicitly or not). It lends itself to a straightforward empirical test. For example, consider an application of Eq. (106) to a situation in which two payoff matrices θ_1 and θ_2 have been used, together with four values of the variable a and six values of the variable x. The data consist of $2 \times 4 \times 6 = 48$ empirical frequencies of "yes" responses, to be explained with $2 \times 4 + 2 \times 6 = 20$ parameters. This leads to a standard chi-square (or likelihood ratio) test, with $48 - 20 = 28$ degrees of freedom. In the framework of the strong criterion invariance, this is essentially a test of the independence of the random variables $\mathbf{U}_{1,a}$ and $\mathbf{U}_{2,x}$. To the best of our knowledge, no such test has been performed.

Notice that signal detection theory is not used or even needed here (the data only concern Eq. (106)). In fact, signal detection theory was introduced explicitly to deal with situations in which an assumption such as the strong criterion invariance does not hold.

A rejection of this model could thus be attributed either to a failure of the strong criterion invariance or to a failure of the assumption of independence of the random variables. Dropping the strong criterion invariance, the model can be strengthened in a different way, by making specific assumptions regarding the distributions of the random variables. Obviously, there are numerous possibilities, each of which leads to a specific form of the ROC curves. We shall not enter here the details of such assumptions.

9.3. Remarks

As indicated above, the probability summation model defined by Eqs. (106) and (107) takes different forms depending on specific assumptions regarding the distributions of the activation random variables. The arbitrariness of the choice of the distributions should not be a cause of excessive concern (cf. Section 8.3). Indeed, suppose that a particular version of the model involves the four random variables $\mathbf{U}_{1,a}$, $\mathbf{U}_{2,x}$, $\mathbf{V}_{1,n}$, and $\mathbf{V}_{2,n'}$. Nothing changes in the predictions if these random variables are subjected to a strictly increasing transformation, provided that this transformation is the same for all variables. For example, let g be an arbitrary, strictly increasing function. Starting from Eq. (105), with F as the maximum function, we have

$$\mathrm{Prob}\{\max\{\mathbf{U}_{1,a},\mathbf{U}_{2,x}\} > \max\{\mathbf{V}_{1,n},\mathbf{V}_{2,n'}\}\}$$

$$= \mathrm{Prob}\{g(\max\{\mathbf{U}_{1,a},\mathbf{U}_{2,x}\}) > g(\max\{\mathbf{V}_{1,n},\mathbf{V}_{2,n'}\})\}$$

$$= \mathrm{Prob}\{\max\{g(\mathbf{U}_{1,a}),g(\mathbf{U}_{2,x})\} > \max\{g(\mathbf{V}_{1,n}),g(\mathbf{V}_{2,n'})\}\} \,. \tag{108}$$

Thus the prediction for the area under the ROC curve is unaffected by the transformation g. Notice that this relative "robustness" of the prediction with regard to the particular form of the distributions cannot be extended to other combination rules, that is, when the function F is different from the maximum function. Two examples of such combination rules will be briefly considered in Section 9.4.

As specified by Eqs. (106) and (107), probability summation assumes that the same criterion λ_θ is used for both channels. According to Eq. (106), for instance, a "yes" response occurs following the presentation of a stimulus (a,x) if

$$\text{either } \mathbf{U}_{1,a} > \lambda_\theta \quad \text{or} \quad \mathbf{U}_{2,x} > \lambda_\theta \,. \tag{109}$$

Occasionally (e.g., Nachmias, 1981), a model is used in which Eqs. (106) and (107) are replaced by the forms

$$p_{ax}(\theta) = 1 - \mathrm{Prob}\{\mathbf{U}_{1,a} \leq \lambda_{\theta,1}\}\mathrm{Prob}\{\mathbf{U}_{2,x} \leq \lambda_{\theta,2}\} \,, \tag{110}$$

$$p_{nn'}(\theta) = 1 - \mathrm{Prob}\{\mathbf{V}_{1,n} \leq \lambda_{\theta,1}\}\mathrm{Prob}\{\mathbf{V}_{2,n'} \leq \lambda_{\theta,2}\} \,. \tag{111}$$

Thus for a given payoff matrix θ, the criteria $\lambda_{\theta,1}$ and $\lambda_{\theta,2}$ corresponding to each channel may be different. The extra generality is only apparent, however. This must be understood as follows. Let Θ be the set of all payoff matrices. To each payoff matrix θ in Θ correspond two criteria $\lambda_{\theta,1}$ and $\lambda_{\theta,2}$. In other terms, there are two functions $\theta \mapsto \lambda_{\theta,1}$, $\theta \mapsto \lambda_{\theta,2}$, each of which maps Θ onto a real interval. It is reasonable to suppose that even though these functions may be different, they generate the same order on the set Θ of payoff matrices. That is, for any two θ and θ' in Θ, we must have

$$\lambda_{\theta,1} < \lambda_{\theta',1} \quad \text{iff} \quad \lambda_{\theta,2} < \lambda_{\theta',2} \,.$$

By a simple mathematical argument, this means that there exists a continuous, strictly increasing function g, such that $g(\lambda_{\theta,2}) = \lambda_{\theta,1}$. But then Eq. (110) implies

$$p_{ax}(\theta) = 1 - \mathrm{Prob}\{\mathbf{U}_{1,a} \leq \lambda_{\theta,1}\}\mathrm{Prob}\{g(\mathbf{U}_{2,x}) < g(\lambda_{\theta,2})\}$$

$$= 1 - \mathrm{Prob}\{\mathbf{U}_{1,a} \leq \lambda_{\theta,1}\}\mathrm{Prob}\{g(\mathbf{U}_{2,x}) < \lambda_{\theta,1}\} \,.$$

Similarly Eq. (111) yields

$$p_{nn'}(\theta) = 1 - \mathrm{Prob}\{\mathbf{V}_{1,n} \leq \lambda_{\theta,1}\}\mathrm{Prob}\{g(\mathbf{V}_{2,n'}) \leq \lambda_{\theta,1}\} \,.$$

Thus after transforming $\mathbf{U}_{2,x}$ into $g(\mathbf{U}_{2,x})$ and $\mathbf{V}_{2,n'}$ into $g(\mathbf{V}_{2,n'})$, the criteria are identical for both channels. We conclude that the two models are equivalent. Obviously, the distributions of the random variables $\mathbf{U}_{2,x}$ and $\mathbf{V}_{2,n'}$ may be modified by the transformation g. For example, if both $\mathbf{U}_{2,x}$ and $\mathbf{V}_{2,n'}$ are normal, $g(\mathbf{U}_{2,x})$ and $g(\mathbf{V}_{2,n'})$ are normal only if g is a linear function. This means that if particular forms of distributions are imposed by the model, the above equivalence does not necessarily hold.

The notion of probability summation is often formalized differently (e.g., Nachmias, 1981), in terms of a two-state threshold model in the spirit of Luce (1960, 1963a, 1963b) which we discussed in Section 8.9. This model is defined by the two equations

$$p_{ax}(\theta) = 1 - (1 - p_{1,a})(1 - p_{2,x})[1 - \gamma(\theta)] \tag{112}$$

$$p_{nn'}(\theta) = 1 - (1 - p_{1,n})(1 - p_{2,n'})[1 - \gamma(\theta)] \,. \tag{113}$$

and is sometimes referred to as the *high-threshold model*. In Eq. (112) $p_{1,a}$ and $p_{2,x}$ are two parameters specifying the probabilities, when stimulus (a,x) is presented, that the thresholds are exceeded in channels 1 and 2, respectively. A "yes" response is given if the threshold is exceeded in at least one of the two channels. A "yes" response may also result from a guess, in a case in which neither of the two thresholds is exceeded. The probability of this positive guess is $\gamma(\theta)$, the value of which may vary with the payoff matrix. A similar interpretation holds for Eq. (113), which corresponds to the noise trials and introduces two additional parameters $p_{1,n}$, and $p_{2,n'}$.

The apparent popularity of this model is difficult to justify since it makes the inescapable but unlikely prediction that the ROC curves in the binaural situation are straight lines. For visual contrast detection data, this model was rejected convincingly by Nachmias (1981) in the framework of particular assumptions on the parameters $p_{1,a}$, $p_{2,x}$, $p_{1,n}$, and $p_{2,n'}$.

Further discussion regarding probability summation models can be found in Watson, Chapter 6, and Olzak and Thomas, Chapter 7.

9.4. Two Additive Combination Rules

For the same two-channel paradigm, we consider here two other possibilities for the form of the function F of the general model defined by Eqs. (103) and (104).

9.4.1. An Additive, Equal Variance, Gaussian Model. We assume that F is a binary addition, namely

$$F(s,t) = s + t .$$

The area under the two-channel ROC curve, expressed by Eq. (105), becomes

$$\text{Prob}\{\mathbf{U}_{1,a} + \mathbf{U}_{2,x} > \mathbf{V}_{1,n} + \mathbf{V}_{2,n'}\} . \tag{114}$$

Let $\mu_{1,a}$, $\mu_{2,x}$, $\mu_{1,n}$, and $\mu_{2,n'}$ be the expectations of the respective random variables, and suppose that their common variance is equal to 1. Assume moreover that all four random variables are normally distributed. From Eq. (114) we obtain

$$\text{Prob}\{\mathbf{U}_{1,a} + \mathbf{U}_{2,x} - \mathbf{V}_{1,n} - \mathbf{V}_{2,n'} > 0\}$$
$$= \Phi[(\mu_{1,a} + \mu_{2,x} - \mu_{1,n} - \mu_{2,n'})/2] = \Phi(d'_{1,2}) .$$

The last equation defines a detectability index $d'_{1,2}$, for the two-channel situation, consistent with that introduced in Section 8.3 for the one-channel situation. Let d'_1 and d'_2 be the detectability indices in the 2 one-channel situations. That is,

$$\text{Prob}\{\mathbf{U}_{1,a} > \mathbf{V}_{1,n}\} = \Phi[(\mu_{1,a} - \mu_{1,n})/\sqrt{2}] ,$$
$$= \Phi(d'_1) ,$$
$$\text{Prob}\{\mathbf{U}_{2,x} > \mathbf{V}_{2,n'}\} = \Phi[(\mu_{2,x} - \mu_{2,n'})/\sqrt{2}]$$
$$= \Phi(d'_2) .$$

By simple algebra, it follows that

$$d'_{1,2} = (d'_1 + d'_2)/\sqrt{2} , \tag{115}$$

a prediction which can be tested by methods discussed in Section 8.

9.4.2. Integration Model. Let f_{ax} be the joint density of $\mathbf{U}_{1,a}$ and $\mathbf{U}_{2,x}$, and let $f_{nn'}$ be the joint density of $\mathbf{V}_{1,n}$ and $\mathbf{V}_{2,n'}$; let $f_{1,a}$, $f_{2,x}$, $f_{1,n}$, and $f_{2,n'}$ be the densities of $\mathbf{U}_{1,a}$, $\mathbf{U}_{2,x}$, $\mathbf{V}_{1,n}$, and $\mathbf{V}_{2,n'}$, respectively (e.g., Green & Swets, 1974). As in Section 8.5, we suppose that the subject behaves as a statistician and bases the decision on the computation of likelihood ratios. In other terms, we assume that the function F has the form

$$F(s,t) = \frac{f_{ax}(s,t)}{f_{nn'}(s,t)} = f_{1,a}(s)f_{2,x}(t)/f_{1,n}(s)f_{2,n'}(t) , \tag{116}$$

by the independence of the random variables. Thus when a stimulus (a,x) is presented, the subject reports a detection if

$$f_{1,a}(\mathbf{U}_{1,a})f_{2,x}(\mathbf{U}_{2,x})/f_{1,n}(\mathbf{U}_{1,a})f_{2,n'}(\mathbf{U}_{2,x}) > \lambda_\theta .$$

The same decison rule holds for the noise trials, based on a sample of $(\mathbf{V}_{1,n},\mathbf{V}_{2,n'})$. Green and Swets (1974, p. 271) show that if all four random variables are Gaussian, and in addition

$$\text{Var}(\mathbf{U}_{1,a}) = \text{Var}(\mathbf{V}_{1,n}) ,$$
$$\text{Var}(\mathbf{U}_{2,x}) = \text{Var}(\mathbf{V}_{2,n'}) ,$$

then

$$d'_{1,2} = [(d'_1)^2 + (d'_2)^2]^{1/2} . \tag{117}$$

Applications of this model to visual perception data are discussed in Kristofferson and Dember (1958) and Green and Swets (1974). Another combination rule

$$d'_{1,2} = d'_1 + d'_2$$

is also considered there.

9.5. Additive Conjoint Measurement

A central notion in a number of models in this chapter is that the sensory system of the subject, when confronted with a multidimensional stimulus, performs a simple arithmetical operation (e.g., addition, multiplication, subtraction). Often, this operation is at the kernel of a process modeling other aspects of the subject's performance (e.g., probabilistic or cognitive), such as in the models introduced in Section 9.4. The analysis of such operations, to the extent that they can model aspects of scientific data, is the concern of measurement theory, a case of which was discussed in Section 2. This subsection is devoted to a discussion of an important special case, in which the effect on the organism of a two-dimensional stimulus (a,x) is captured by an addition of two numbers, $f(a) + g(x)$.

Consider a two-alternative forced-choice (2AFC) paradigm. On each trial, the subject is presented with two stimuli (a,x) and (b,y). For concreteness, suppose that as earlier in this section, these are pure tones presented binaurally. Thus a and b are the intensities of the tone in the left auditory channel, and x and y are the intensities in the right auditory channel. The subject is asked which of (a,x) and (b,y) seems loudest. If (a,x) is chosen, the experimenter writes

$$by < ax ,$$

as the data for the trial. It is assumed that the effect of component a of stimulus (a,x) can be represented by some number, denoted by $f(a)$. Similarly, the effect of component x is represented by a number $g(x)$. These numbers can be interpreted as measuring the intensities of the activations evoked by the stimulus at some neural locations. The model, however, is noncommittal in that respect. The basic assumption is that

$$by < ax \quad \text{iff} \quad f(b) + g(y) < f(a) + g(x) . \tag{118}$$

This model is in the spirit of those discussed in Section 9.4, except that, somewhat unrealistically, it is deterministic: the presentation of (a,x) always evokes the same number $f(a) + g(x)$. (By comparison, in the model of Section 9.4.1, each presentation of (a,x) determines a sample of a random variable $\mathbf{U}_{1,a} + \mathbf{U}_{2,x}$.) This implies that each presentation of a pair of stimuli (a,x) and (b,y) results in the same choice by the subject, a prediction which may be reasonable for some carefully selected set of stimuli but would certainly not be acceptable in general. It is not assumed that the numbers $f(a)$, $f(b)$, $g(x)$, and so on, are accessible to direct investigation. It may not be immediately clear that this model imposes strong constraints on the data, but it does. Suppose indeed that the experimenter observes

$$bz < ay \quad \text{and} \quad cy < bx \ .$$

According to the model, this can arise only if

$$f(b) + g(z) < f(a) + g(y)$$

and

$$f(c) + g(y) < f(b) + g(x) \ .$$

Adding these two inequalities and canceling appropriately yields

$$f(c) + g(z) < f(a) + g(x) \ ,$$

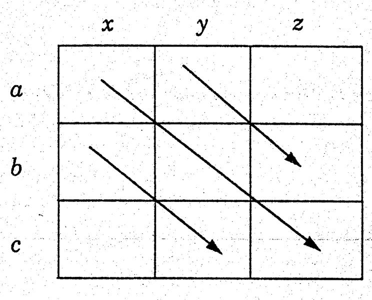
which in turn predicts that

$$cz < ax \ .$$

Summarizing this argument, we see that the model specified by Eq. (118) holds only if

$$\text{whenever } bz < ay \text{ and } cy < bx, \text{ then } cz < ax \ .$$

In the measurement literature, this is known as the *double-cancellation condition*. It is illustrated thus:



In addition, a pair of independence conditions are easily shown to be necessary:

1. $ax < bx$ iff $ay < by$.
2. $ax < ay$ iff $bx < by$.

(The verification of the necessity is left to the reader.) The double-cancellation condition and the two independence conditions are the key axioms of a model that, measurement theory tells us, implies the existence of the two scales f and g satisfying Eq. (118). We will not go into the details of this model, which are quite technical (see, e.g., Roberts, 1979; or Krantz et al., 1971). It suffices to remember that if the data are to be explained by the additive model specified by Eq. (118), then the double-cancellation condition and the two independence conditions must be satisfied.

An illustration of an experimental test of these conditions, in binaural perception, can be found in a paper by Levelt, Riemersma, and Bunt (1972). However, Levelt and colleagues' positive conclusions have recently (and rightly, in the opinion of this writer) been criticized by Gigerenzer and Strube (1983).

The appeal of measurement models of this kind is that they offer, at least in principle, the possibility of getting at the essential determinants of the subject's performance, from a psychophysical viewpoint: the scale or scales transforming the physical input or inputs, the basic operation or operations performed by the sensory system. A serious weakness of such models is that they are ill-equipped to deal with data variability, which characteristically results from psychophysical experimentation. In Sections 9.6 and 9.7 we discuss some probabilistic versions of the additive conjoint measurement model considered here.

9.6. Random Conjoint Measurement

We begin with a slight modification of the binaural loudness paradigm.

9.6.1. Matching Task. As in the 2AFC paradigm, the subject is first presented with a binaural stimulus (a,x), followed by another stimulus (b,y). The task is to modify the intensity of b (for example, by turning a dial) until, by successive approximations, the two stimuli appear equally loud. This final value of b is recorded. Typically, this value varies across trials (for fixed a, x, and y).

9.6.2. The Model. Let us write $\mathbf{U}_{xy}(a)$, a random variable, for the final value of b yielding a match. This notation seems appropriate since this value depends on a, x, and y. (The reason for the asymmetry in the notation—x, y as indices and a in parentheses—will become clear in a moment.) In the deterministic framework of additive conjoint measurement, (a,x) should appear as loud as (b,y) iff

$$f(a) + g(x) = f(b) + g(y) \ ,$$

or, equivalently,

$$f(b) = g(x) - g(y) + f(a) \ .$$

If b is replaced by the random variable $\mathbf{U}_{xy}(a)$, it seems reasonable to balance the above equation by adding an error term in the right member, which gives

$$f[\mathbf{U}_{xy}(a)] = g(x) - g(y) + f(a) + \boldsymbol{\epsilon}_{xy}(a) \ . \qquad (119)$$

The error term $\boldsymbol{\epsilon}_{xy}(a)$ is assumed to be a random variable with a (uniquely defined) median equal to 0. This model is in the spirit of the additive conjoint measurement model discussed in this section but may be applied to noisy data.

Since the scales f and g are unknown, one may ask, How is Eq. (119) constraining the data? Or, in other terms, under which conditions (necessary or sufficient) do scales f and g exist satisfying Eq. (119)? It turns out that Eq. (119) imposes strong, highly testable constraints on the medians of the random variables $\mathbf{U}_{xy}(a)$. A simple argument demonstrating this fact is given in Section 9.6.3.

9.6.3. Some Necessary Key Conditions. If $\mathbf{T}$ is a random variable having a unique median ν, we write $M(\mathbf{T}) = \nu$. The following fact will be useful: if h is any real, strictly increasing function, then $M[h(\mathbf{T})] = h[M(\mathbf{T})]$. (This follows immediately from the definition of the unique median of $\mathbf{T}$.) For simplicity, we shall adopt the abbreviation

$$m_{xy}(a) = M[\mathbf{U}_{xy}(a)]$$

for the median of the matching random variable $\mathbf{U}_{xy}(a)$. Taking medians on both sides of Eq. (119) yields

$$M\{f[\mathbf{U}_{xy}(a)]\} = f\{M[\mathbf{U}_{xy}(a)]\} = g(x) - g(y) + f(a)$$

or, equivalently,

$$m_{xy}(a) = f^{-1}[g(x) - g(y) + f(a)] , \quad (120)$$

in which f^{-1} is the inverse of the scale f. From this equation, the following condition is easily derived.

9.6.3.1. Cancellation Rule.

$$m_{xz}(a) = m_{xy}[m_{yz}(a)] ,$$

whenever all three medians are defined.

This condition, which is illustrated in Figure 1.27, is the counterpart in this probabilistic framework of the double-cancellation condition encountered in additive conjoint measurement. It has an elegant, compact expression but appears somewhat abstract at first. A good grasp of this condition requires careful study. To begin with, notice that it only concerns the "observable" medians of the matching random variables (the unknown scales f and g have been eliminated). Let us show how the cancellation rule follows from Eq. (120). Successively,

$$\begin{aligned} f\{m_{xy}[m_{yz}(a)]\} &= g(x) - g(y) + f[m_{yz}(a)] \\ &= g(x) - g(y) + f\{f^{-1}[g(y) \\ &\quad - g(z) + f(a)]\} \\ &= g(x) - g(y) + g(y) - g(z) + f(a) \\ &= f[m_{xz}(a)] . \end{aligned}$$

We conclude that

$$f\{m_{xy}[m_{yz}(a)]\} = f[m_{xz}(a)] ,$$

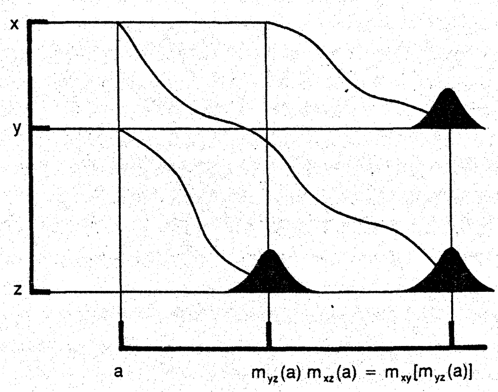


Figure 1.27. Cancellation rule. The three distributions of the figure are those of $U_{xz}(a)$, $U_{xy}[m_{yz}(a)]$ and $U_{yz}(a)$. The three curves are the "isoloudness curves" of (a, x), $[m_{yz}(a)$, x] and (a, y).) (See also section 9.6.4.) (From J. C. Falmagne, Random conjoint measurement and loudness summation, *Psychological Review, 83*. Copyright 1976 by American Psychological Association. Reprinted with permission.)

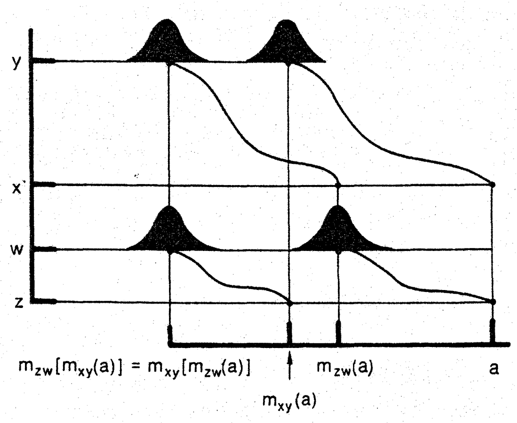


Figure 1.28. Commutativity rule. The conventions are similar to those of Figure 1.27. The four distributions are those of $U_{xy}(a)$, $U_{zw}(a)$, $U_{xy}[m_{zw}(a)]$, and $U_{zw}[m_{zy}(a)]$. The four curves are the "isoloudness curves" of (a, x), (a, z), $[m_{zw}(a)$, x], and $[m_{xy}(a)$, w].) (From J. C. Falmagne, Random conjoint measurement and loudness summation, *Psychological Review, 83*. Copyright 1976 by American Psychological Association. Reprinted with permission.)

which is equivalent to the cancellation rule, since f is a one-to-one function. A further understanding of this condition will be obtained from a discussion of how it can be tested. (See Section 9.6.4.) Using similar methods, another condition can also be shown to be necessary.

9.6.3.2. Commutativity Rule.

$$m_{xy}[m_{zw}(a)] = m_{zw}[m_{xy}(a)] ,$$

whenever all four medians are defined.

In other terms, if we pick one intensity a in the left channel and four intensities x, y, z, and w in the right channel and take the two medians $M[\mathbf{U}_{zw}(a)] = m_{zw}(a) = b_1$ and $M[\mathbf{U}_{xy}(a)] = m_{xy}(a) = b_2$ and, next, the two medians $M[\mathbf{U}_{xy}(b_1)] = m_{xy}(b_1)$ and $M[\mathbf{U}_{zw}(b_2)] = m_{zw}(b_2)$, then the two medians $m_{xy}(b_1)$ and $m_{zw}(b_2)$ should be equal. The commutativity rule is illustrated in Figure 1.28.

It can be shown that if continuity assumptions are made, then the implication can be reversed: the cancellation rule and the commutativity rule together imply that Eq. (120) holds for some scales f and g. A proof of this fact can be found in Falmagne (1976).

9.6.4. A Test. A test of the cancellation rule could proceed as follows.

Step 1. Choose one intensity a in the left channel and three intensities x, y, and z in the right channel. Have the subject find an intensity b in the left channel, so that (b, z) matches (a,y) in loudness. Repeat $2p$ times. Order these $2p + 1$ values $b_1 \leq b_2 \leq ... \leq b_{2p+1}$. Then b_{p+1} is an estimate of $m_{yz}(a)$.

Step 2. Have the subject find an intensity c such that (c, y) matches (b_{p+1},x) in loudness. Repeat q times. The obtained empirical distribution is denoted by D.

Step 3. Have the subject find an intensity d, such that (d,z) matches (a,x). Repeat k times. The obtained empirical distribution is denoted D'.

Step 4. Test whether $\mathbf{U}_{xy}(b_{p+1})$ and $\mathbf{U}_{xz}(a)$ have the same median, for example, by performing a median test comparing D and D'. (This test is known to be reasonably robust to a difference in the shape of the distributions; cf. Pratt, 1964.)

A similar test can be designed for the cancellation rule. A discussion regarding the soundness of such procedures can be found in Falmagne (1976).

Gigerenzer and Strube (1983) have applied this model to binaural loudness data. The hypothesis that the two auditory channels are additive, in the sense of Eq. (119), is convincingly rejected. The data favor a model in which one channel dominates when its intensity sufficiently exceeds that of the other.

9.7. Probabilistic Conjoint Measurement

As reported in a number of papers, Falmagne and his coworkers have investigated another way of injecting statistical considerations into additive conjoint measurement (Falmagne, 1978, 1979; Falmagne & Iverson, 1979; Falmagne, Iverson, & Marcovici, 1979). Suppose that, in a 2AFC paradigm, the subject must select one of the 2 two-component stimuli (a,x) and (b,y). As before, we assume that a, b, x, and y are numbers denoting physical variables. Let $P_{ax,by}$ be the probability that (a,x) is chosen over (b,y). A general additive model is embodied in the equation

$$P_{ax,by} = F[f(a) + g(x), f(b) + g(y)] \ , \qquad (121)$$

in which the real-valued functions F, f, and g in the right member are unspecified, except for monotonicity and continuity properties: all three functions are continuous, F is strictly increasing in the first variable and strictly decreasing in the second variable, and f and g are strictly increasing. We also assume that the function F in Eq. (121) satisfies the following balance property (see the definition in Section 3.5.1):

$$F(s,t) = .5 \quad \text{iff} \quad s = t \ . \qquad (122)$$

The connections between this model and that previously discussed under the label *random conjoint measurement* must be appreciated. Consider a situation in which the experimenter, an expert in adaptive methods (see Section 6), fixes the values of a, x, and y in Eq. (121) and has the subject's performance converging over trials—say, using stochastic approximation—to a point β satisfying

$$P_{ax,\beta y} = F[f(a) + g(x), f(\beta) + g(y)] = .5 \ .$$

The estimated value of β is actually a random variable, the distribution of which depends on a, x, and y. Under reasonable differentiability assumptions (see Section 6.2.1), the asymptotic distribution of this random variable is normal and has an expectation equal to β. Let us denote this asymptotic random variable by $\mathbf{V}_{xy}(a)$. Notice that, using Eq. (122),

$$f(a) + g(x) = f(\beta) + g(y) \ ,$$

and thus

$$E[\mathbf{V}_{xy}(a)] = f^{-1}[g(x) - g(y) + f(a)] \ , \qquad (123)$$

which is, for all practical purposes, equivalent to Eq. (120) constraining the medians, in our discussion of random conjoint measurement. (Indeed, for normal distributions the expectation and the median are equal.) The model specified by Eq. (121) can thus be tested by checking whether the cancellation and commutativity rules are satisfied empirically by the expectations $E[\mathbf{V}_{xy}(a)]$. A number of special cases of this model are of interest, the defining equations of which are listed below. Note that the function k in Eq. (124) is assumed to be strictly increasing and continuous.

$$P_{ax,by} = F\{k[f(a) + g(x)] - k[f(b) + g(y)]\} \ . \qquad (124)$$

$$P_{ax,by} = F[f(a) + g(x) - f(b) - g(y)] \ . \qquad (125)$$

$$P_{ax,by} = F\{[f(a) + g(x)]/[f(b) + g(x)]\} \ . \qquad (126)$$

$$P_{ax,by} = F[f(a)g(x) - f(b)g(y)] \ . \qquad (127)$$

Diagnostic properties permitting one to sort out these models have been developed (Falmagne, 1979). The behavior of the function $a \mapsto P_{ax,bx}$ is particularly instructive in this respect. Assuming that Eq. (124) holds, it can be shown, for example (Falmagne et al., 1979), that for $a > b$, the function

$$k \text{ is} \begin{cases} \text{linear} \\ \text{strictly convex} \\ \text{strictly concave} \end{cases} \text{iff} \quad P_{ax,bx} \text{ is} \begin{cases} \text{independent of } x; \\ \text{strictly increasing in } x; \\ \text{strictly decreasing in } x. \end{cases}$$

(We recall that *strictly convex* means curved upward and *strictly concave* means curved downward; cf. Section 7.4.) Important examples of (strictly) convex and concave functions are the logarithmic and exponential functions. Observe in this connection that each of the Eqs. (125), (126), and (127) follows from Eq. (124) by assuming that k is a linear, a logarithmic, or an exponential function, respectively. (Obviously, a change of notations vis-à-vis functions F, f, and g is taking place between Eqs. (124) and (125), (126), and (127).

These models have been applied by Falmagne and colleagues (1979) to binaural loudness data collected in a series of experiments, using the 2AFC paradigm. A special case of Eq. (126) was found to yield a good fit. (See, however, Gigerenzer & Strube, 1983.) More will be said about this study in Section 9.8.

9.8. Homogeneity Laws

There is a class of empirical laws that deserves serious consideration by the psychophysicist. (Let us avoid both misunderstanding and a philosophical trap. By "law" we mean an important equation purporting to explain a body of data. The equation derives its importance, and thus the label "law," from that of the data to be explained, from the consequences of the equation regarding feasible theories, and possibly also from the simplicity of its form. Scientific usage indicates that complete accuracy of the prediction is not a major requirement, e.g., the failure of Boyle's law at low temperature.) Examples of laws in that class are provided by two forms of Weber's law encountered in Section 7. As defined in Section 7.2.3, it constrains the psychometric functions and takes the form

$$p_{\lambda a}(\lambda x) = p_a(x) \ . \qquad (128)$$

In words, a psychometric function is invariant under multiplication of the intensities of the standard and the stimulus by the same constant $\lambda > 0$. Equivalently (theorem in Section 7.2.6), Weber's law concerns the Weber functions and states that

$$\Delta_\pi(\lambda a) = \lambda \Delta_\pi(a) . \qquad (129)$$

We recall that a real-valued function h of n real variables is *homogeneous of degree* β iff

$$h(\lambda x_1, \lambda x_2, ..., \lambda x_n) = \lambda^\beta h(x_1, x_2, ..., x_n) ,$$

for all $\lambda > 0$ (see Section 7.2.2). Thus Eqs. (128) and (129) mean that the functions p and Δ are homogeneous of degree 0 and 1, respectively.

A couple of additional examples of homogeneity laws will be discussed. They show that such laws are typically easy to verify experimentally and tend to have strong implications on theorization. If Weber's law is any indication, they may have a more durable impact than specific process models, a prospect that justifies the space allocated here to this topic.

No proof of any of the results discussed below will be given. Incidentally, we mention that the arguments used to derive the theoretical consequences of homogeneity laws often appeal to results from a field of mathematics called *functional equations*, an introduction to which the reader can find in Aczél (1966).

9.8.1. The Conjoint Weber Laws. Let us go back to the 2AFC paradigm used by Falmagne and colleagues (1979), in which the subject was required to compare binaural stimuli (a,x) and (b,y). A test of the following generalization of Weber's law was performed:

$$P_{(\lambda a)(\lambda x),(\lambda b)(\lambda y)} = P_{ax,by} .$$

That is, using the decibel scale, the choice probability does not vary when the same number of decibels is added to all four intensities. This prediction, called the *conjoint Weber law*, was found to be well supported by the data, at least for the relatively modest range of stimulus intensities considered in the experiment. The importance of this result from a theoretical standpoint should not be underestimated. Researchers in this field are concerned with the hypothesis that the two auditory channels may be additive. As indicated in Section 9.7, a possible formalization of this notion lies in Eq. (121).

$$P_{ax,by} = F[f(a) + g(x), f(b) + g(y)] ,$$

in which the functions F, f, and g are unspecified, except for continuity and monotonicity properties. Falmagne and Iverson (1979) show that if both the conjoint Weber law and Eq. (121) hold, then the choice probabilities must have one of the following three forms:

$$P_{ax,by} = G[(a^\beta + \delta x^\beta)/(b^\beta + \delta y^\beta)] ; \qquad (130)$$

$$P_{ax,by} = G[a^\beta x^\gamma / b^\beta y^\gamma] ; \qquad (131)$$

$$P_{ax,by} = Q[a/x, b/y] , \qquad (132)$$

in which β and γ are constants, G is strictly increasing and continuous, and Q is continuous, strictly increasing in the first variable and strictly decreasing in the second variable. These three equations are easy to discriminate experimentally. For instance, Eq. (132) can be eliminated immediately, as a model for binaural loudness, since it predicts that $P_{ax,by}$ is decreasing in x and increasing in y. A different way of separating these equations leads us to introduce two other homogeneity laws, each of which is a strengthening of the conjoint Weber law:

Strong Conjoint Weber Law Type I (SCWI)

$$P_{ax,by} = P_{(\lambda a)(\lambda x),(\tau b)(\tau y)}$$

Strong Conjoint Weber Law Type II (SCWII)

$$P_{ax,by} = P_{(\lambda a)(\tau x),(\lambda b)(\tau y)} .$$

These equations are assumed to hold for all positive a, b, x, y, λ, and τ. These two laws provide a sharp method to distinguish between Eqs. (130), (131), and (132) from an experimental standpoint. In particular, it is easy to prove that SWCI is equivalent to Eq. (132). It can also be shown that the additive form Eq. (121), together with SCWII, is equivalent to Eq. (131). A useful conclusion follows: if the conjoint Weber law holds, but both SCWI and SCWII fail, then the additive form Eq. (121) has necessarily the form of Eq. (130).

This example shows how the experimental testing of homogeneity laws (with positive or negative outcomes) may result in a considerable strengthening of the hypotheses of a model. Following is another example, along the same lines but with a different motivation and a different paradigm.

9.8.2. Shift Invariance in Loudness Recruitment. A tone embedded in noise does not appear as loud as the same tone in quiet. As the intensity of the tone increases, however, the subjective difference tends to disappear. In psychoacoustics, this phenomenon is known as *loudness recruitment*. Let us denote by $\varphi(x,n)$ the intensity of a tone in quiet matching an intensity x of the same tone embedded in a noise of intensity n. These matching functions were recently investigated by Iverson and Pavel (1981a, 1981b; see also Pavel, 1980), who demonstrated that to an excellent approximation the following property was satisfied by the data: for some $\theta > 0$ and all $\lambda > 0$

$$\varphi(\lambda x, \lambda^\theta n) = \lambda \varphi(x,n) . \qquad (133)$$

They investigated the theoretical consequences of this property, which they called *shift invariance*. The choice of this name is justified by a geometrical interpretation of Eq. (133), an illustration of which is given in Figure 1.29. Shift invariance can be seen as a homogeneity property under a slight disguise: defining the function ψ,

$$\psi(x,y) = \varphi(x, y^\theta) ,$$

it follows that

$$\psi(\lambda x, \lambda y) = \varphi(\lambda x, \lambda^\theta y^\theta) = \lambda \varphi(x, y^\theta) = \lambda \psi(x,y) .$$

That is, ψ is homogeneous of degree 1.

As in the preceding example, it may be asked whether shift invariance may be assumed in conjunction with some general, reasonable model, with the effect of strengthening the model

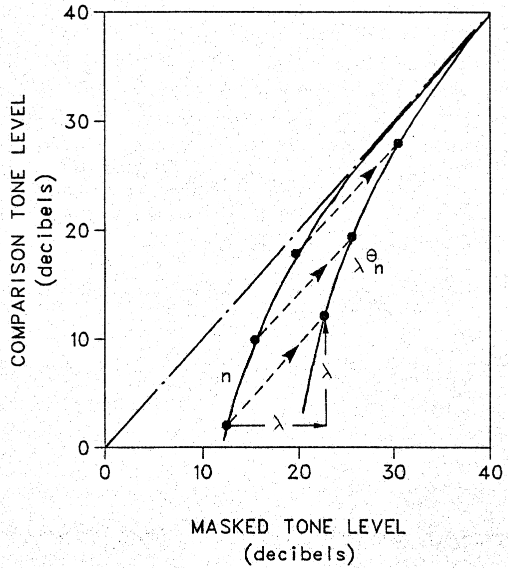


Figure 1.29. The property of shift invariance illustrated by two hypothetical loudness matching curves. The right curve representing loudness matching with noise $\lambda^{\theta}n$ can be obtained by a rigid shift of the left curve (generated by loudness matches with noise n) along the first bisector. (From M. Pavel, *Homogeneity in complete and partial masking*. Unpublished doctoral dissertation, New York University, 1980.)

in a useful way. Iverson and Pavel (1981a) assume that the matching function ρ satisfies the gain control equation

$$\varphi(x,n) = F\{g(x)/[h(x) + k(n)]\} , \qquad (134)$$

with the functions g, h, k, and F being subjected only to natural monotonicity and continuity conditions. They show then that in the presence of shift invariance Eq. (134) can take only one of two forms:

$$\varphi(x,n) = [Ax^{\alpha}/(x^{\alpha'} + Kn^{\alpha'/\theta})]^{1/(\alpha-\alpha')} \qquad (135)$$

or

$$\varphi(x,n) = A[x^{\alpha}(x^{\alpha'} - Kn^{\alpha'/\theta})]^{1/(\alpha+\alpha')} , \qquad (136)$$

where A, K, α, α', and θ are appropriately chosen constants. Note that a special case of Eq. (136) was proposed by Lochner and Burger (1961) as an extension of the power law (cf. Section 10), incorporating the effects of a masking noise. Objections to Eq. (136) as a possible model for recruitment can be found in Scharf (1978). A plot of a least-square fit of Eq. (135) to some data of Stevens and Guirao (1967) is presented in Figure 1.30.

Another example of homogeneity law arises in our discussion of the bisection method in Section 9.9.

9.9. Bisection

This name refers to a class of paradigms in which, on each trial, a subject is presented with a pair (a,b) of stimuli and is required to "produce" (in a way depending on the experimental conditions) a stimulus appearing "midway" between a and b. As before in this section, a and b are physical intensities. We shall denote by $B(a,b)$ the midway intensity produced by the subject. In some situations, the subject may be asked to adjust a dial; $B(a,b)$ may then be estimated by averaging over trials. In other cases, $B(a,b)$ may be obtained by applying an adaptive procedure (cf. Section 6.2).

A frequently proposed model for the resulting data is formalized by the equation

$$u[B(a,b)] = \frac{u(a) + u(b)}{2} , \qquad (137)$$

in which the function u is assumed to be strictly increasing and continuous but is otherwise arbitrary. The idea behind this representation is that the subject performs the task by computing the arithmetic average of a and b. This computation, however, is not (necessarily) carried out in the physical scale, but may involve instead some unknown psychophysical scale, represented in Eq. (137) by the function u. It may seem that since u is unspecified, Eq. (137) is not saying very much. But this is not so. Equation (137) is telling us that B is an operation which must behave essentially like an arithmetic average. In fact, this model puts severe constraints on the data. A simple example is the *commutativity* equation

$$B(a,b) = B(b,a) , \qquad (138)$$

which immediately results from Eq. (137) by observing that the terms $u(a)$ and $u(b)$ commute in the right member. Equation (137) also implies that B must be *idempotent*; that is, we must have

$$B(a,a) = a \qquad (139)$$

for all stimuli a. This follows from the fact that

$$u[B(a,a)] = [u(a) + u(a)]/2 = u(a)$$

which yields Eq. (139) since u is a one-to-one function. A less obvious consequence of Eq. (137) is the condition

$$B[B(a,b),B(c,d)] = B[B(a,c),B(b,d)] , \qquad (140)$$

which is often referred to as *bisymmetry*. The easy proof that Eq. (137) implies bisymmetry is left to the reader. This implication can be reversed; under general continuity and monoton-

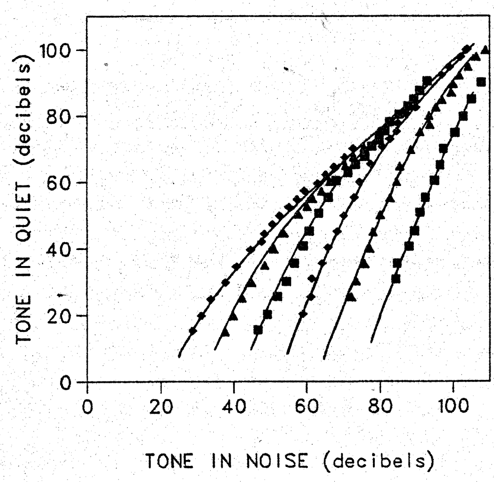


Figure 1.30. Best fit of Eq. (135) to scale data of Stevens and Guirao (1967). (From G. J. Iverson & M. Pavel, On the functional form of partial masking functions in psychoacoustics, *Journal of Mathematical Psychology*, 1981, *24*. Reprinted with permission.)

icity properties of the midway function B, idempotent, commutativity, and bisymmetry together imply the existence of a function u satisfying Eq. (137) (Krantz et al., 1971).

An experimental test of bisymmetry in auditory perception can be found in Cross (1965; cited by Coombs, Dawes, & Tversky, 1970). Bypassing such tests, it is also possible to "search" directly for a function u satisfying Eq. (137). This is done by Weiss (1975) and Anderson (1976, p. 107, 1981, p. 37). A good fit is obtained for a power function $u(a) = \lambda a^{\beta}$. As we shall see in this section, this form of the scale u is of particular interest.

In some cases, commutativity may not hold. Consider a situation in which a and b are two intensities of a pure tone presented monaurally and successively. It is conceivable (in fact, likely) that the produced midway value will depend on which of a and b is presented first: the midway operation has to be performed between two stimuli, one of which is being kept in memory for some time and thus subject to the effects of a possible decay. The idempotent property may also fail. In such cases, Eq. (137) may be generalized as follows. If bisymmetry and idempotent hold, but not (necessarily) commutativity, then the appropriate model is

$$u[B(a,b)] = \alpha u(a) + (1 - \alpha) u(b)$$

with $\alpha > 0$, a constant. If neither idempotent nor commutativity is assumed to hold, but bisymmetry is satisfied, then we have the still more general model

$$u[B(a,b)] = \alpha u(a) + \gamma u(b) + \delta$$

with $\alpha > 0, \gamma > 0$.

Bisection provides an additional example of a homogeneity law. In a classic application of this type of paradigm, Plateau (1872) gave a pair of painted disks, one white and one black, to each of eight artists and instructed them to return to their respective studios and paint a gray disk midway between the two. The resulting gray disks, reported Plateau, were almost identical for all eight artists, in spite of the variation in the illumination conditions under which they were produced. Let us suppose that such results would hold for any pair of gray disks. A possible formalization of this circumstance would be as follows. Let (a,b) denote a pair of gray disks, in a specified viewing condition in Plateau's laboratory. Let a and b denote the luminance of the disks in conventional units. Let $B(a,b)$ denote the midway gray disk in the same viewing condition. The artist, however, has performed the task in a studio, in different illumination conditions, that is, with the pair $(\lambda a, \lambda b)$ (where λ is a positive constant equal to the ratio of the illumination in the artist's studio to that of Plateau's laboratory). By hypothesis, the resulting midway disk is independent of the illumination. As a consequence, the following equation must hold:

$$B(\lambda a, \lambda b) = \lambda B(a,b) . \qquad (141)$$

Indeed, this means that the midway disk obtained in Plateau's laboratory is identical to that produced by the artist in the studio, when seen under the same conditions. In other terms, B is homogeneous of degree 1 (cf. Section 7.2.2). But if both the averaging model, Eq. (137), and the homogeneity property, Eq. (141), are assumed to hold, then the possible forms of the function u are very limited; u must be either a power function,

$$u(a) = \alpha a^{\beta} + \gamma ,$$

or a logarithmic function,

$$u(a) = \alpha \log a + \delta$$

(α, β, γ constants). No other forms exist which would satisfy both Eqs. (137) and (141). This was noted by Krantz (Note 3), who also remarked that essentially the same argument applies to the basic equal-spacing principle underlying the construction of the Munsell system (Munsell, 1929). We recall in this connection the results of Weiss (1975) and Anderson (1976, 1981) who, using a different method, also obtained a power function for their bisection data.

The bisection models discussed here are deterministic, which renders their application to data delicate. Fortunately, probabilistic versions of such models can be developed, which are similar in spirit to the models discussed in Sections 9.6 and 9.7 for additive conjoint measurement. For the sake of illustration, one possibility is outlined here.

We begin by replacing the operation yielding a stimulus midway between a and b by a random variable $\mathbf{B}(a,b)$. Eq. (137) becomes

$$u[\mathbf{B}(a,b)] = [u(a) + u(b)]/2 + \epsilon(a,b) , \qquad (142)$$

in which $\epsilon(a,b)$ is an error random variable with a unique median equal to 0. Let $m(a,b)$ be the median of the random variable $\mathbf{B}(a,b)$. By a simple argument along the lines of that used in Section 9.6.3, Eq. (142) implies

$$u[m(a,b)] = [u(a) + u(b)]/2 ,$$

an equation which has exactly the form of Eq. (137), with the median m replacing the deterministic operation B. This means that the conditions of idempotency, commutativity, and bisymmetry must be satisfied by the medians. In other terms, we must have

$$m(a,a) = a ,$$

$$m(a,b) = m(b,a) ,$$

$$m[m(a,b), m(c,d)] = m[m(a,c), m(b,d)] .$$

Similarly, the homogeneity condition uncovered in Plateau's experiment is formalized by the equation

$$m(\lambda a, \lambda b) = \lambda m(a,b) .$$

As in the case of the random conjoint measurement model, nonparametric tests can be used to evaluate the empirical validity of these conditions.

9.10. Key References

Preoccupations with the role of chance in the improvement of performance observed in multichannel perception were expressed early (Dawson, 1913). The first explicit formalization of probability summation in the sense of this section is attributed to Pirenne (1943). A review can be found in Blake and Fox (1973). Chapter 9 in Green and Swets (1974) is devoted to models for multichannel perception, including the integration model.

Tests of a specific model of probability summation are described in a recent paper by Nachmias (1981). The applications of probability summation considered in this chapter were limited to a two-channel situation. The ideas developed are easily extended to n-channels. However, when a large number of channels is involved, a new situation arises, in which some convergence theorems of probability theory are applicable. These issues are considered in chapters by Watson; Arditi; Olzak and Thomas; Regan, Kaufman, and Lincoln; Ginsburg; and Treisman in this handbook.

Additive conjoint measurement is a standard topic of measurement theory, a detailed exposition of which is contained in Krantz and colleagues (1971) and Roberts (1979). The introduction of probabilistic models in additive conjoint measurement is in the spirit of the models encountered in probabilistic choice theory (Luce & Suppes, 1965). The models discussed here were developed by Falmagne and his coworkers (Falmagne, 1976, 1978, 1979; Falmagne & Iverson, 1979; Falmagne, Iverson, & Marcovici, 1979).

Implicitly, the study of homogeneity laws has been part of psychophysics since its inception (Weber's law is a homogeneity law). Many psychophysical laws or models are instances of or at least are consistent with some homogeneity law. A systematic investigation of homogeneity laws and their impact on psychophysical theorizing has recently been undertaken by Falmagne, Iverson, and Pavel (Falmagne & Iverson, 1979; Falmagne, Iverson, & Marcovici, 1979; Iverson & Pavel, 1980, 1981a, 1981b; Pavel, 1980). An introduction to the functional equation techniques used in these papers can be found in Aczél (1966).

A treatment of bisection, from the viewpoint of measurement theory, is contained in Krantz and colleagues (1971; see also Pfanzagl, 1968).

There was a substantial amount of arbitrariness in the choice of topics covered in this section. The reader may be surprised, for example, that only a passing reference was made to the work of Anderson and his collaborators (Anderson, 1970a, 1970b, 1974, 1976, 1981]. Actually, the organizing principle for this section was to include multivariable models or techniques only if they were a natural extension of "classical" psychophysics. Scaling models or techniques are covered in Section 10.

10. SCALING

Scaling covers a collection of models, procedures, and empirical analyses, purporting to provide a representation of some data in terms of one or more numerical scales. Such is, of course, also the aim of measurement theory, a field in which, typically, axiom systems are given justifying specific methods of scale construction (cf. Section 2). In the work usually classified under the scaling label, however, acquiring the scales is often regarded as an end in itself, and the theoretical underpinnings are of secondary importance. Objections have deservedly been made to that state of affairs. The uses of a scale without a firm theoretical foundation are restricted. In particular, if the *type* of a scale (see Section 10.1) is unknown, it may be difficult to decide whether a given model or a mathematical expression employing that scale makes sense from a certain logicophilosophical viewpoint (see Section 10.10).

After an introduction to scale types, the most common unidimensional scaling methods and data will be reviewed. Two theoretical approaches will then be considered: the Shepard-Krantz relation theory and the functional measurement procedures, introduced by Norman Anderson. A brief discussion of the issue of the psychophysical scale and the measurement of sensation will follow. Finally, the notion of meaningful psychophysical laws will be brought to the attention of the reader. (Note that the so-called multidimensional scaling techniques are considered by Wyszecki, Chapter 9.)

10.1. Common Types of Scales

In most cases, numerical scales constructed from (and explaining) some empirical data are not defined uniquely. It is usually agreed, for example, that the numerical scale used for the measurement of length is a *ratio scale*, which means that the numerical values assigned to the objects are defined up to a multiplication by a positive constant (e.g., a change of units from centimeter to meter is admissible). In this exemplary case, the exact degree of arbitrariness of the scale is a consequence derivable from a completely axiomatized theory. One assumes that the data satisfy the axioms of the theory, which in turn provides a procedure for the construction of the scale and specifies the degree of arbitrariness of such construction. How this applies in the case of length has been discussed in detail in Sectio. 2. The degree of arbitrariness of the scale is referred to as its *type*. Despite the variety of forms of data, only a few types of scales are actually used in scientific practice. The reasons for this scarcity are not very well understood (see, however, Narens, 1981). The most commonly used types of scales are listed in Table 1.4.

10.2. Overview of (Unidimensional) Scaling Methods

The psychophysical procedures discussed in earlier sections of this chapter (such as that used in the yes-no paradigm) were rather painstaking. In a typical experiment, several hundred observations per point are collected for each subject. By contrast, the methods considered here may use only a few observations per point (sometimes as few as one or two observations per subject). However, the subject's responses tend to be much more elaborate. For example, the subject may be asked to identify the stimulus presented, using a label previously attached to that stimulus (as in the *absolute identification* method) or be required to evaluate the stimulus numerically, according to some rule (as in *magnitude estimation*). There are a number of such scaling methods, and ways of classifying them. In the next four sections, we classify the methods by the type of response required from the subject. Each subsection contains a brief de-

Table 1.4. The Most Commonly Used Types of Scales.

Scale Type	Admissible Transformations	Examples
Absolute	Identity: $x \mapsto \phi(x) = x$	Counting
Ratio	Similarity: $x \mapsto \phi(x) = \alpha x$, with $\alpha > 0$	Length, mass
Interval	Affine: $x \mapsto \phi(x) = \alpha x + \beta$, with $\alpha > 0$	Temperature
Log-interval	$x \mapsto \phi(x) = \alpha x^{\beta}$, with $\alpha > 0$ and $\beta > 0$	Density

Each type is defined by the class of admissible transformations of the scale; for example, the ratio scale type is that defined by all the transformations of the form $x \mapsto \alpha x$, with $\alpha > 0$. The case for density and other fundamental physical quantities to be log-interval scales is made by Krantz, Luce, Suppes, and Tversky (1971).

scription of the procedures and of the typical experimental results.

Under the impetus of S. S. Stevens, an impressive array of experimental results were collected, which generally support the contention that through subjective judgments the sensory continua are related to each other and to the number continuum by power laws (at least to a first approximation). The *power law* was offered by Stevens as a substitute for the logarithmic relation of Fechner (cf. Section 7.3). The merit of this proposal is discussed in Section 10.10. In this connection, the reader should bear in mind that the psychophysical methods discussed in this section were often introduced in a spirit of criticism of the "classical" methods, such as the yes-no paradigm and its close relatives. These were thought to lack realism, to the extent that the data were focusing on "local" effects (e.g., the discrimination of neighboring stimuli), while the natural environment involves the simultaneous apprehension of a large collection of widely distributed stimuli. The terms *local* and *global* are sometimes used to denote the two classes of procedures.

10.3. Absolute Identification

In a preliminary period, the subject is trained to associate one of n labels (say, the numbers from 1 to $n = 10$) to each of n stimuli. During the main phase of the experiment the subject is presented a stimulus on each trial and is required to produce the appropriate label. The subject's response is recorded. The succession of stimuli is random. Occasionally, immediate repetitions are avoided.

A straightforward analysis of the data is in terms of the proportion of correct responses. Another measure of performance, not often used now, is the *average information transmitted by the responses* (cf. Coombs, Dawes, & Tversky, 1970; Garner, 1962; Miller, 1953). More recently, a measure based on the index d' of signal detection theory (cf. Section 8) has been proposed (Luce, Green, & Weber, 1976).

For stimuli varying along one sensory continuum, the main finding is that the maximum number of stimuli that can be identified perfectly by an untrained subject is between five and nine, depending on the continuum, for example, Pollack (1952) and Garner (1953). (See Miller, 1956, for a review of the facts. Obviously, specialists, such as professional musicians for pitch identification, may score much better than that.) This result is regarded as puzzling since it appears to be at variance with the data of local studies. For instance, only stimuli that are very close on the physical scale (say, less than a couple of decibels apart in auditory discrimination) are ever confused in a two-alternative forced-choice (2AFC) paradigm. An extrapolation would lead to predict a perfect identification of several dozen suitably located stimuli in an absolute identification experiment. The discrepancy may be due to the fact that efficient guessing strategies can be used in a 2AFC situation, which are no longer available in absolute identification.

At first, the absolute identification paradigm may seem straightforward. Actually, the data are plagued with a variety of sequential effects and "anchoring" effects that render the analysis extremely difficult. Regarding anchoring, or edge, effects, see, for example, Berliner and Durlach (1973), Berliner, Durlach, and Braida (1977), Braida and Durlach (1972), Durlach and Braida (1969), Gravetter and Lockhead (1973), Lippman, Braida, and Durlach (1976), and Weber, Green, and Luce (1977). Sequential effects in absolute identification have been explored, for example, by Holland and Lockhead (1968), Jesteadt, Luce, and Green (1977), Purks, Callaghan, Braida, and Durlach (1980), Ward (1972), and Ward and Lockhead (1970, 1971).

10.4. Category Rating

As with absolute identification, in *category rating* one stimulus from a sensory continuum is presented at each trial. The subject is instructed to assign each stimulus to one of m-ordered categories, for example, the numbers 1 to m. These categories are assumed to be subjectively "equally spaced"; that is, the subjective distance between category 3 and 4 is identical to that between 10 and 11. The number m of categories is often smaller than the number of stimuli and may vary from a few (5–7) to several dozen.

In variations of this method, pairs of stimuli are presented at each trial, and the subject is required to rate (that is, to assign a category to) subjective "differences" or "ratios" of these stimuli. In a rather extreme version, four stimuli are presented simultaneously, and the subjects are asked to make very sophisticated judgments, such as rating the ratio of differences or the difference of ratios (Birnbaum, 1978).

The most startling result is perhaps that the subjects not only are capable of performing such tasks without major difficulties but also provide surprisingly regular data: the average rating values often appear to vary smoothly with stimulus intensity. The data may be analyzed in various ways. In an exemplary case, the subjects rate subjective differences between stimuli, and it is assumed that the average rating $D_{x,y}$ corresponding to physical intensities x, y satisfies a Fechnerian-type relation

$$D_{x,y} = F[u(x) - u(y)], \qquad (143)$$

in which u and F are strictly increasing functions. In some cases, F is shown to be well approximated by a linear function. Unfortunately, the subject's performance in these tasks varies markedly with the context. For example, the value of $D_{x,y}$ in Eq. (143) strongly depends on the distribution of all the stimuli used in the experiment, specifically, the range, the spacing, and the frequencies of these stimuli. Such facts, which are well documented, may create problems for psychophysical theorizing, depending on the locus of the effects. To pursue our example, a major concern is which of the two functions u, F in the right member of Eq. (143) is affected by the context. The available evidence points out that only F is affected (Parducci, 1963, 1965, 1974; Parducci & Perrett, 1971; see Birnbaum, 1982, for a general discussion and further references). Since the function u is a candidate for the psychophysical scale, the key invariant, this would leave open the possibility of a general theory.

10.5. Magnitude Estimation

In the widely used method of *magnitude estimation*, the main advocate of which was S. S. Stevens, the subject is required to provide "direct" numerical estimates of the magnitude of the sensation evoked by the stimulation. Two variants of the method have been employed.

In one, the subject is initially presented with a stimulus (the standard) and told that the sensory magnitude of that stimulus is assigned a certain value (modulus), say, 100. Other stimuli are then presented in random order, and the subject is instructed to estimate their sensory magnitude so as to preserve ratios. For instance, if the second stimulus presented seems to have a sensory magnitude which is half that of the standard, its sensory magnitude should be estimated to be 50. Typically, only a couple of observations are taken from each subject, and the data of all subjects are combined by computing the median or the geometrical mean.

The second variant has the favor of many investigators. No standard and no modulus are provided. The subject is simply told to assign to any stimulus presented any number that seems suitable as an estimate of the sensation magnitude.

Interestingly, the results are very similar for the two methods. For intensive continua, the mean or median response $\phi(x)$ is approximately a power function of the physical intensity x:

$$\phi(x) = \alpha x^{\beta} . \qquad (144)$$

In log–log coordinates, Eq. (144) becomes the equation of a linear function with slope β, which can be fitted to the data by linear regression. As exemplified in Figure 1.31, this prediction holds reasonably well for much data, at least for moderate to large intensities (see Marks, 1974, or S. S. Stevens, 1975, for a presentation of the evidence). A better overall fit may be obtained, at the cost of one extra parameter, by forms such as

$$\phi(x) = \alpha x^{\beta} + \gamma$$

or

$$\phi(x) = \alpha(x - \gamma)^{\beta} ,$$

both of which are capable of handling the data at low intensities (cf. Ekman, 1956, 1961; Fagot, 1963; Galanter & Messick, 1961; Luce, 1959a; S. S. Stevens, 1959a).

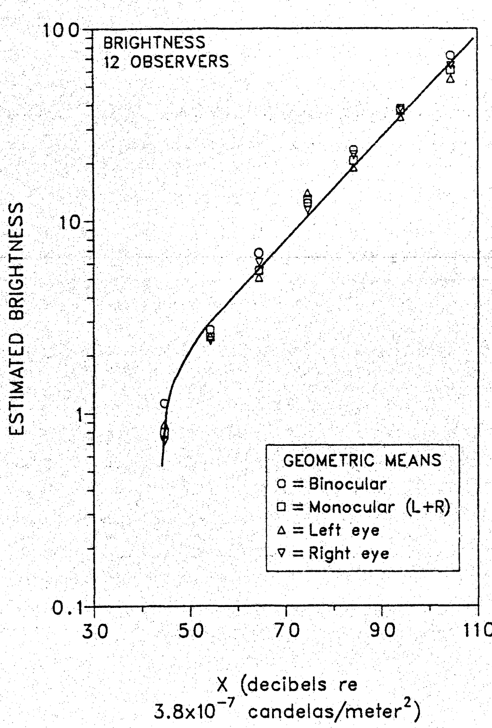


Figure 1.31. Magnitude estimates of brightness. In abscissa, the luminance of the stimuli in decibels re 10^{-6} Lambert. (Adopted from J. C Stevens, Brightness function: Binocular versus monocular stimulation, *Perception and Psychophysics*, 1967, *2*. Reprinted with permission.)

The magnitude estimation procedure is also used in other paradigms. For example, in *ratio estimation*, the observer is asked to evaluate the subjective "ratio" of two stimuli. At least to a first approximation, the experimental results are consistent with those reported for the magnitude estimation of single stimuli. More is said about such consistency in Section 10.7.

It was strongly argued by S. S. Stevens (1957, 1959a, 1961a, 1961b, 1961c) that Eq. (144) should be taken as the fundamental psychophysical law, rather than the logarithmic Fechnerian form derived from Weber's law together with the difference representation for choice probabilities (cf. Section 7.3). Accordingly, serious consideration is given to the estimated value of the exponent β in Eq. (144), which some believe could be a measure of some basic feature of the subject's sensory system. Several dozen sensory continua were investigated by Stevens and others, and the values of the exponent β were tabulated (see, for example, Table 1 in S. S. Stevens, 1975). The claim that the exponent in Eq. (144) is of fundamental importance for psychophysical theory encounters difficulties with various data, however, which indicate that its estimated value strongly depends on the experimental conditions or even on the instructions given to the subject. Among other studies, we mention Teghtsoonian (1971), who shows that the estimated values of β are correlated with the range of the set of stimuli used in the experiment, and Robinson (1976), who demonstrates how the instructions can systematically affect this exponent. A review of some of these effects can be found in Poulton (1968). In the light of available evidence, it is clear that no single, basic sensory factor is responsible for the variations of the exponent. In particular, as argued by Green and Luce (1974), its value may reflect some aspects of the subject's decision-making process.

10.6. Production and Matching Methods

In *production and matching methods*, the subject is requested to react to the stimulation by "producing" a value of a sensory variable, for example, by turning a dial. There are several commonly used procedures, some of which have been encountered earlier in this chapter. The bisection method described in Section 9.9 belongs to that category.

The *magnitude production* reverses the procedure used in magnitude estimation. The subject is given a number and asked to produce a matching intensity. As in magnitude estimation, a power law can be fitted to the data. However, as observed by many investigators, the estimated exponent tends to be larger (see S. S. Stevens & Greenbaum, 1966, for a summary of the data).

In the *ratio production* method, the observer is instructed to adjust the intensity of the stimulus in such a manner that it appears to be a particular multiple or fraction of a standard. (In this last case, the term *fractionation* method is also used.) For example, the subject may be required to produce a tone intensity appearing half as loud as the standard tone of the same frequency. These methods have a long but scattered history and were regarded with some suspicion until Stevens's major contribution to the field. By and large, the data are similar to those obtained with magnitude estimation. (For details, see Marks, 1974 or S. S. Stevens, 1975).

A rather startling prediction may be obtained for the data of the so-called cross-modality matching method. Suppose that for two sensory continua, denoted below as 1 and 2, the magnitude estimation data are adequately summarized by the two power laws

$$\phi_1(x) = \alpha_1 x^{\beta_1}, \quad \phi_2(y) = \alpha_2 y^{\beta_2}. \quad (145)$$

For concreteness, suppose that the two sensory continua are loudness and brightness. Imagine now that in a third experiment the subject, rather than matching physical quantities to numbers as in a magnitude estimation experiment, is requested to match the values directly from one sensory continuum to the other, say, from loudness to brightness. At first, this instruction may seem rather bizarre. Actually, not only are the subjects capable of performing such a task without undue hardship, but, once again, they provide reasonably regular data. Assuming that the matching of brightness to loudness is achieved by equating the values of the two psychophysical scales, that is, the two right members in Eq. (145), we obtain

$$\alpha_1 x^{\beta_1} = \alpha_2 y^{\beta_2}.$$

Writing $\phi_{1,2}$ for the cross-modality matching function (thus $\phi_{1,2}(x) = y$) and rearranging, yields

$$\phi_{1,2}(x) = \alpha_{1,2} x^{\beta_{1,2}}, \quad (146)$$

a power law with

$$\beta_{1,2} = \beta_1/\beta_2, \quad (147)$$

and

$$\alpha_{1,2} = \alpha_1/\alpha_2. \quad (148)$$

The prediction that the cross-modality matching function is a power law has been verified by several authors, for many continua, and it holds rather well (cf. Figure 1.32). For a number of reasons, the verification of the specific relation linking the exponents in cross-modality matching and magnitude estimation is not as straightforward as it may seem. While S. S. Stevens (1975) and Marks (1974) conclude that Eq. (147) is well supported by the facts, doubt has been expressed by others, based on their analysis of their own data (Baird, Green, & Luce, 1980; Mashour & Hosman, 1968).

10.7. Krantz-Shepard Theory

Despite the limitations, the array of results collected by Stevens and his followers, and summarized in the last two subsections, contains enough regularities to require a systematic explanation.

The *relation theory* outlined below represents the most satisfactory effort made to account for a substantial part of the data. Some seminal ideas were first proposed by Shepard, in an unpublished manuscript, and were then elaborated and axiomatized by Krantz (1972; see also Shepard, 1981). In presenting this theory, we make a number of idealizations. We omit the fact that the data are noisy, are the locus of important contextual and sequential effects, and so forth. To simplify and shorten the exposition, we also specify the theory by properties actually derivable from more abstract axioms in Krantz's paper. (To some extent, our presentation "trivializes" the theory but hopefully renders key notions more transparent.)

The data concern n sensory continua, numbered 1, 2, ..., n. We begin by tightening up the notations. The letters x, y, ... (or sometimes $x_i, y_i, ..., 1 \leq i \leq n$, to avoid ambiguities) will stand for positive real numbers representing physical intensities of the stimuli (energy level). We denote by:

$N_i(y|x,p)$. The magnitude estimation of stimulus y, with standard x and modulus p, in the sensory continuum i, $1 \leq i \leq n$.

$P_i(x,y)$. The ratio estimation of the pair (x,y) in the sensory continuum i.

$C_{ji}(y_j|x_j,x_i)$. The cross-modality matching value of stimulus y_j from sensory continuum j into sensory continuum i, with modulus (x_j,x_i).

In Krantz's system, the cross-modality matching modulus may be taken to be the stimulus–response pair of the preceding trial. Six axioms, labeled RT1–RT6, specify the theory.

Axiom RT1. For every sensory continuum i, $1 \leq i \leq n$, there is a function $(x,y) \to l_i(x,y)$ mapping the pairs of stimuli onto a subset of the positive reals (independent of i). These functions

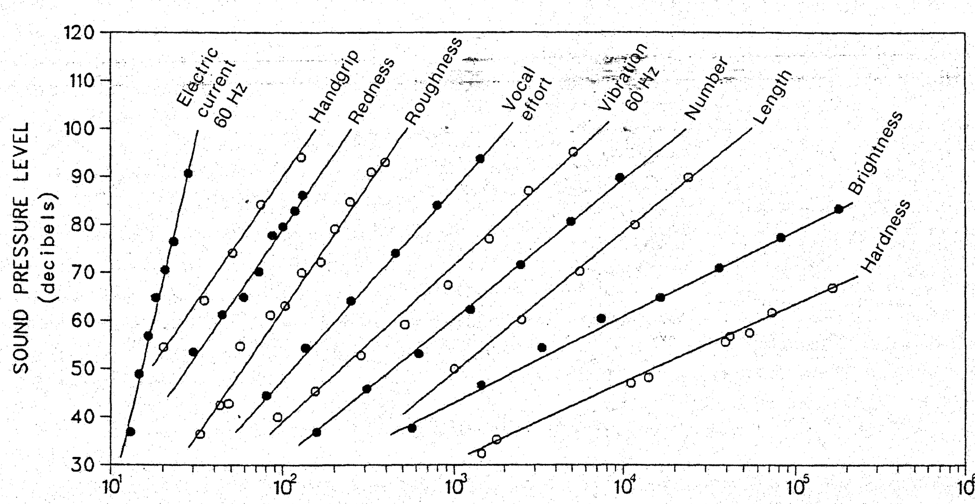


Figure 1.32. Cross-modality matching data between loudness and 10 other sensory continua. (From S. S. Stevens, Matching function between loudness and ten other continua, *Perception and Psychophysics*, 1966, *1*. Reprinted with permission.)

are continuous, strictly increasing in the first variable and strictly decreasing in the second variable. Moreover, the functions l_i are assumed to satisfy the two conditions:

1. $l_i(x,y) \geq l_j(z,w)$ implies $l_j(w,z) \geq l_i(y,x)$.
2. If $l_i(x,y) \geq l_j(x',y')$ and $l_i(y,z) \geq l_j(y',z')$, then $l_i(x,z) \geq l_j(x',z')$.

This is the basic notion. Every pair (x,y) in a sensory continuum i is mapped into a *sensation continuum* by the function l_i. We shall see that the two conditions (1) and (2) ensure that the quantities $l_i(x,y)$, $l_j(z,w)$, ..., and so forth, behave in a certain sense like arithmetical ratios (see Eq. (149)). In the sequel, we shall refer to $l_i(x,y)$ as the *sensation "ratio"* of (x,y). Any estimation or production task is then carried out through the mediation of the sensation "ratios" of the pairs of stimuli involved. Examples are given in the next two axioms.

Axiom RT2. There is a positive-valued, strictly increasing function H such that for every sensory continuum i,

$$P_i(x,y) = H[l_i(x,y)] .$$

In words, the ratio estimates are strictly increasing with the sensation "ratios."

Axiom RT3. For every pair (j,i) of sensory continua,

$$C_{ji}(y_j|x_j,x_i) = y_i \text{ implies } l_j(y_j,x_j) = l_i(y_i,x_i) .$$

In words, with cross-modality matching modulus (x_j,x_i), y_j is matched to y_i only if the sensation "ratios" of (y_j,x_j) and (y_i,x_i) coincide.

The next two axioms emphasize the special role played by one sensory continuum, arbitrarily numbered 1.

Axiom RT4. For the sensory continuum 1,

$$P_1(x,y) \cdot P_1(y,z) = P_1(x,z) .$$

Axiom RT5. If $l_i(y,x) = l_1(z,w)$, then

$$N_i(y|x,p) = p_1^P(z,w) .$$

The special continuum is assumed to be length. Axiom RT4 states essentially that mental estimation of length ratios behaves like physical measurement, an assumption which, Krantz argues, is supported by the fact that the estimated exponents of the power law for judgments of distance are often close to 1. (Some would question that fact. We postpone criticism at this point.) Axiom RT5 is consistent with a mechanism in which magnitude estimation in any sensory continuum i is obtained through computation in the length continuum.

Axiom RT6. For any sensory continuum i and any positive real numbers x, y, and λ,

$$l_i(\lambda x, \lambda y) = l_i(x,y) .$$

Note that this last axiom, which will procure the power law, has the form of Weber's law but applies also to discriminable stimuli. These six axioms have a number of consequences for psychophysical judgments, examples of which follow.

From Axiom RT1 it can be derived that for any sensory continuum i

$$l_i(x,y) = G[f_i(x)/f_i(y)] , \qquad (149)$$

for strictly increasing, continuous functions G and f_i. Combining this result and Axiom RT2, we obtain

$$P_i(x,y) = H\{G[f_i(x)/f_i(y)]\} . \qquad (150)$$

By a standard functional equation argument, applying Eq. (150) and Axiom RT4 results in the function $H[G(s)]$ having the form

$$H[G(s)] = s^{\gamma} \qquad (151)$$

for some positive constant γ. From Eq. (150) and Axiom RT6, we deduce

$$f_i(\lambda x)/f_i(\lambda y) = f_i(x)/f_i(y) ,$$

a functional equation which (in the conditions of monotonicity or continuity of f_i) has only the solution

$$f_i(x) = \alpha_i x^{\beta_i} ,$$

for some constant α, $\beta > 0$. From Eq. (149), we obtain thus

$$l_i(x,y) = G[(x/y)^{\beta_i}] . \qquad (152)$$

Replacing the sensation magnitudes in Axioms RT2, RT3, and RT5 by their expressions as given by Eq. (152) and using also Eq. (151) gives the expected predictions:

$$P_i(x,y) = (x/y)^{\beta_i \gamma} ;$$

$$N_i(y|x,p) = p(x/y)^{\beta_i \gamma} ;$$

$$C_{ji}(y_j|x_j,x_i) = x_i(y_j/x_j)^{\beta_j/\beta_i} .$$

Notice that the cross-modality matching exponents β_j/β_i can be predicted by the ratio of the magnitude estimation exponents of the preceding equation.

Various criticisms can be made against this theory. In particular, (1) it is deterministic, while the data are highly variable, within or across observers; (2) it omits important sequential and contextual effects, which some believe to be important enough to bias the picture seriously; and (3) the special role of the length continuum can be questioned, specifically the contention that the estimated exponent of the power law is approximately equal to 1 (Baird, 1970).

In our opinion, even though the predictions of relation theory may not be fully supported by the data, they certainly represent useful approximations. If nothing more, relation theory may be taken as a good summary of the way a sizable part of the psychophysical community idealizes data, still a serviceable device.

10.8. Functional Measurement

For some psychophysicists, the data of magnitude estimation and production are hopelessly biased by uncontrollable nuisance effects and such methods should be abandoned. Such is the position of Anderson, who advocates an alternative collection of procedures and models which he calls *functional measurement* (see Anderson 1974, 1976, 1981, for numerous references).

In a typical application of functional measurement, the subject is presented with stimuli varying along several dimensions or aspects, in a factorial design, and is required to produce a rating value, say, on a 20-category rating scale. In one experiment, for example, designed to assess the so-called size–weight illusion, subjects were asked to rate the subjective heaviness of cubical blocks varying in weight and size (Anderson, 1970a). One or more algebraic models are then applied, symbolizing different combination rules for the factors. Let r_{ij} stand for the (average) rating in cell (i, j) of a two-factor design. The most frequently used models are:

The Adding Model $\quad r_{ij} = \alpha_i + \beta_j \; ;$

The Averaging Model $\quad r_{ij} = \dfrac{w_i \alpha_i + w_j \beta_j}{w_i + w_j} \; ;$

The Multiplying Model $\quad r_{ij} = \alpha_i \cdot \beta_j \; .$

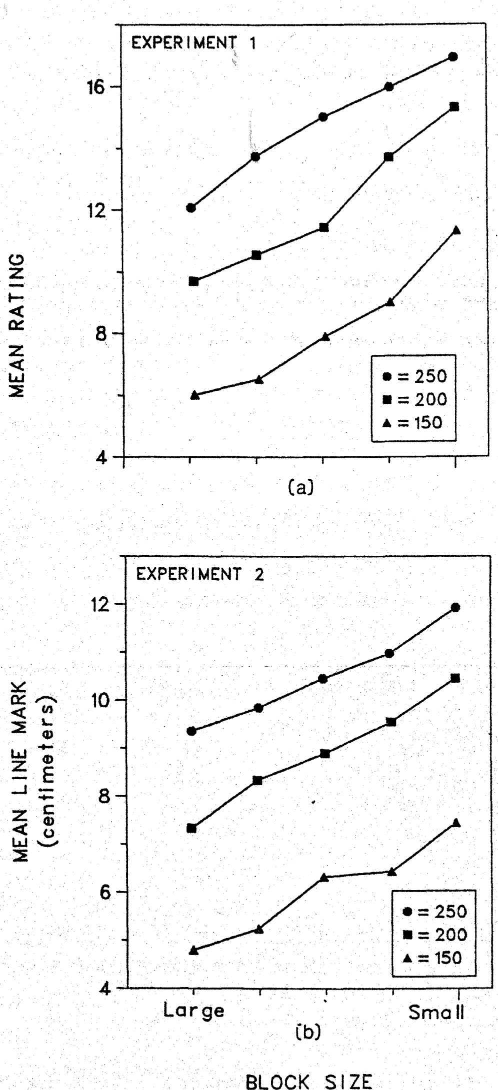


Figure 1.33. Anderson's data for the size-weight illusion. Subjects lift and judge heaviness of cubical blocks in a 3 × 5, gram weight × block size design. Verbal rating response plotted in upper graph (a); graphic rating response in lower graph (b). (From M. H. Anderson, Averaging model applied to the size-weight illusion, *Perception and Psychophysics*, 1970, *8*. Reprinted with permission.)

Assuming appropriate distributions for the variances of the errors, these models can be tested through standard analysis of variance techniques. A graphic plot of the data is also used to validate a model. In the case of the adding model, for instance, since $r_{ik} - r_{jk} = \alpha_i - \beta_j$, independent on k, a check of "parallelism" can be made. This is illustrated in Figure 1.33 for the size–weight illusion experiment mentioned above (Anderson, 1970a, 1981). This analysis favors a model in which subjective heaviness (as evaluated by the ratings) is represented as the sum of subjective weight and appearance.

Occasionally, the standard models cannot be fitted to the rating data. A monotonic rescaling of the ratings is then carried out by numerical techniques. When the fit of a model is taken to be acceptable, the estimated values of the parameters α_i, β_j can be plotted against the corresponding physical measure. The resulting relation is called the *psychophysical law* for that sensory continuum. It is assumed, or hoped, that this relation will hold across situations varying the experimental design, the instructions, and the model but involving the same sensory continuum.

Over the years, Anderson and his followers have applied functional measurement methods to a large body of data in psychophysics and elsewhere and have often succeeded in parsing out the effects of the factors on the ratings, through one or the other of the standard models.

A number of criticisms of Anderson's approach have been made however. The major point of contention concerns the rating response used and, in particular, the status of that response measure with respect to scale type. The mathematical form of the adding and averaging models is invariant under affine transformations (i.e., transformations $x \mapsto \gamma x + \delta$). This property led Anderson to argue that when one such model is found to fit some data, it can be concluded that the rating response and the estimated parameters are interval scales. The objections to this controversial claim are reviewed in Birnbaum (1982).

10.9. Measurement of Sensation

It is impossible in a few pages to do justice to the diversity of positions concerning the measurement of sensation and the form of the "psychophysical scale," that is, the mathematical function relating physical intensity to sensation magnitude. (For a recent sample, see Warren, 1981, and the comments following the article.) These positions go from a rejection of the issue (the measurement of sensation is a hopeless enterprise; e.g., Tumarkin, 1981) to a strongly held opinion that given appropriate experimental control, a particular method yields the desired psychophysical scale (e.g., Anderson, 1981; or S. S. Stevens, 1975). A consensus is not in sight; it has been helpful to distinguish two classes of sensible positions.

1. *Category 1.* Given a large collection of psychophysical data considered important by the psychophysical community, a psychophysical scale should be *adopted* that renders simple or convenient the numerical expression of these data and of the models explaining them. In line with such a position, it is recognized that there is typically a degree of arbitrariness in the choice of a scientific scale and that models and data can usually be recoded if a monotonic rescaling is taking place. Exemplars of this position are Luce and Galanter (1963a), Ellis (1966), and Falmagne (1974). In this connection, we note that there is an overwhelming tendency to plot psychophysical data in logarithmic coordinates and that many models currently in

use have their variables in decibel units or could easily be recast in such terms. From this viewpoint, the Fechnerian logarithmic scale would yet appear—notwithstanding all the attacks—as a reasonable choice for the psychophysical scale.

One objection to this admittedly utilitarian position is that there is no foreseeable agreement regarding what constitutes the bulk of important psychophysical data.

2. *Category 2.* The psychophysicists in this second category consider some particular data to be of primal value in *uncovering* the psychophysical law. The basic idea is that stimulus intensities have a *numerical* representation in the subject's organism, which can be accessed *directly* if the right response is elicited from the subject in the right paradigm. In the same vein, the logarithmic scale is rejected by observing, for example, that pairs of stimuli which are equidistant on the logarithmic scale do not appear to be equidistant subjectively or by showing that this scale differs from that obtained by the selected direct method. Many examples of tenants of such a position may be found among Stevens followers. The belief in the existence, within the organism, of a numerical representation of sensory intensities may perhaps strike a philosopher as a severe case of reification. However, the surprising consistency of the results reported by different laboratories using the same direct method prevents a casual dismissal of the notion. As if some analog device were available to them, the subjects are indeed able to make sense of descriptions of stimuli, such as "half as loud" or "twice as bright," or to provide regular magnitude estimation or rating data.

The difficulty for the advocates of a particular direct method is, again, that there is no agreement in the psychophysical community regarding the choice of such a method. This is both understandable and justified, since the regularity and consistency of the data generated by any direct method (however surprising they may be) are not such that these data could provide the foundation for a scientific scale.

In our opinion, the choice of a psychophysical scale is in part a matter of scientific strategy, with unavoidable political overtones. What should be accomplished with such a scale? It is easily conceivable that no scale could usefully serve the dual purpose of (1) determining a convenient numerical notation of scientific psychophysical facts and models and (2) providing a medium of communication with a naive public on practical questions involving subjective impressions of sensory intensities.

10.10. A Note on Meaningful Psychophysical Laws

One might suppose that the choice of a mathematical formula to represent some data, say, in the form of a scientific law, is solely a matter of goodness of fit. Of course, routine precautions must be taken when evaluating the fit, such as accounting for the number of parameters. This can often be done by standard statistical methods, such as likelihood ratio or minimum chi-square. Granted a proper statistical analysis, the best-fitting formula or model should be chosen, or so it may seem.

Actually, the above scheme is not completely accurate, and considerations of a completely different nature may enter into the selection of a formula. In particular, depending on the type of the scale or scales involved, a given formula may or not be a sensible choice. Suppose, for example, that in an application of the 2AFC paradigm, the binary choice probabilities are represented by the equation

$$P_{a,b} = F\left(\frac{a + 1.83}{b + 1.83}\right), \tag{153}$$

in which a and b are stimulus intensities expressed in some units of a standard ratio scale (say, sound pressure, weight, or length), F is a strictly increasing continuous function, and 1.83 is a constant. Equation (152) can be objected to on the grounds that it conveys little information if the particular units of the variables a and b are not mentioned. One might ask, Why not mention the units? It turns out that all the scientific laws of importance satisfy the property that they can be quoted without mentioning the units of the scales. Curiously, this is a statement of fact, not a regulation. To illustrate, according to Coulomb's law, "The force in a homogeneous isotropic medium of infinite extent between two point charges is proportional to the product of their magnitude, divided by the square of the distance between them" (Gray, 1957).

Note that this statement of Coulomb's law remains true no matter which units are adopted for the scales entering in the formulation of the law. This statement is thus unambiguous. Numerous similar examples could be given in physics and other fields. By contrast, the form of Eq. (152) is not invariant with admissible transformations of the scales. A better formulation for the lawfulness that Eq. (152) was attempting to capture would be

$$P_{a,b} = F\left(\frac{a + \lambda}{b + \lambda}\right)$$

in which λ is a scale-dependent constant. In the technical jargon, those mathematical formulas having a form invariant of the units of the scales are called *meaningful*. As noted by Falmagne and Narens (1983), the strong liking of scientists for meaningful formulas to represent laws is probably due to a combination of practical and theoretical reasons. From a practical viewpoint, the adoption of nonmeaningful formulas would almost certainly introduce chaos into scientific communication. From a theoretical viewpoint, meaningfulness appears to lead to coherent systems of units (cf., Luce, 1959(b)). Our example involving Eq. (152) may suggest that these matters are relatively trivial and that, with some care, considerations of meaningfulness are easy to apply. Actually, this is only true in the case of very simple mathematical forms.

The space available here only permits us to alert the reader to this question, a full discussion of which would take many rather technical pages. For an introduction to the issue of meaningfulness, see Suppes and Zinnes (1963), Roberts (1979), or Falmagne and Narens (1983). Applications in psychophysics can be found in Luce (1959b).

10.11. Key References

The field of scaling is among those covered regularly in the *Annual Review of Psychology*; for example, Ekman and Sjoberg (1965), Zinnes (1969), Cliff (1973), Carroll and Arabie (1980). The last paper reviews the developments in multidimensional scaling techniques.

The notion of the type of a measurement scale is analyzed in basic measurement papers or books (Ellis, 1966; Krantz et al., 1971; Roberts, 1979; Suppes & Zinnes, 1963).

Techniques, data, and philosophy of direct scaling are discussed in great detail in the books by Marks (1974) or S. S. Stevens (1975).

Introductions to functional measurement procedures are contained in a number of papers and in one book by Anderson (e.g., 1970a, 1970b, 1974, 1976, 1981). An axiomatic analysis of functional measurement, from a measurement standpoint, has been given by Luce (1981).

A recent discussion of some controversial issues in psychological measurement, with a special emphasis on rating scales (including functional measurement methods), can be found in Birnbaum (1982).

Since Fechner, numerous discussions of the issue of measuring sensation have been published, few of which are really enlightening. The last section of Krantz (1972)—although written in a rather terse style—is useful reading in this connection.

Space limitation forced us to consider in detail only two theoretical viewpoints on psychophysical scaling, namely, the Krantz-Shepard relation theory and Anderson's functional measurement procedures, both of which were chosen in view of the amount of data concerned by the theories. This selection may give a distorted view of the field. Among other regrettable omissions, we mention Green and Luce's theory (1974; see also, e.g., Green, Luce, & Duncan, 1977, and Levine, 1974).

REFERENCE NOTES

1. Pavel, M. Personal communication, 1982.
2. Vorberg, D. *Bayesian estimation of arbitrary points on psychometric functions.* Paper presented at the Thirteenth Annual Mathematical Psychology Meeting, University of Wisconsin, August 1980.
3. Krantz, D.H. Personal communication.

REFERENCES

Aczél, J. *Lectures on functional equations and their applications.* New York: Academic Press, 1966.

Alpern, M., Rushton, W. A. H., & Tori, S. The attenuation of rod signals by backgrounds. *Journal of Physiology*, 1970, *206*, 209–227. (a)

Alpern, M., Rushton, W. A. H., & Tori, S. The signals from cones. *Journal of Physiology*, 1970, *207*, 463–475. (b)

Alpern, M., Rushton, W. A. H., & Tori, S. The size of rod signals. *Journal of Physiology*, 1970, *206*, 193–208. (c)

Anderson, N. H. Averaging model applied to the size-weight illusion. *Perception and Psychophysics*, 1970, *8*, 1–4. (a)

Anderson, N. H. Functional measurement and psychophysical judgement. *Psychological Review*, 1970, *77*, 153–170. (b)

Anderson, N. H. Information integration theory: A brief survey. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology, Vol. 2: Measurement, psychophysics, and neural information processing.* San Francisco: W. H. Freeman, 1974.

Anderson, N. H. Integration theory, functional measurement and the psychophysical law. In H. G. Geissler, & Y. U. M. Zabrodin (Eds.), *Advances in psychophysics.* Berlin: VEB Deutscher Verlag, 1976.

Anderson, N. H. *Foundations of information integration theory.* New York: Academic Press, 1981.

Anderson, T. W., McCarthy, P. I., & Tukey, I. W. Staircase methods of sensitivity testing. *Navord Reports*, 1946, March 21, 46–65.

Baird, J. C. *Psychophysical analysis of visual space.* Oxford: Pergamon Press, 1970.

Baird, J. C., Green, D. M., & Luce, R. D. Variability and sequential effects in cross-modality matching of area and loudness. *Journal of Experimental Psychology: Human Perception and Performance*, 1980, *6*, 277–289.

Berliner, J. E., & Durlach, N. I. Intensity perception IV. Resolution in roving-level discrimination. *Journal of the Acoustical Society of America*, 1973, *53*, 1270–1287.

Berliner, J. E., Durlach, N. I., & Braida, L. D. Intensity perception VII. Further data on roving-level discrimination and the resolution and bias edge effects. *Journal of the Acoustical Society of America*, 1977, *61*, 1577–1585.

Birnbaum, M. H. Differences and ratios in psychological measurement. In F. Restle & N. J. Castellan, Jr. (Eds.), *Cognitive theory* (Vol. 3). Hillsdale, N.J.: Erlbaum, 1978.

Birnbaum, M. H. Controversies in psychological measurement. In B. Wegener (Ed.), *Social attitudes and psychophysical measurement.* Hillsdale, N.J.: Erlbaum, 1982.

Blake, R., & Fox, R. The psychophysical inquiry into binocular summation. *Perception and Psychophysics*, 1973, *14*, 161–185.

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

Bock, R. D., & Jones, L. V. *The measurement and prediction of judgement and choice.* San Francisco: Holden-Day, 1968.

Boring, E. G., Langfeld, H. S., & Weld, H. P. *Foundations of psychology.* New York: Wiley, 1948.

Bradley, R. A. Incomplete block rank analysis: On the appropriateness of the model for a method of paired comparisons. *Biometrics*, *10*, 1954, 375–390. (a)

Bradley, R. A. Rank analysis of incomplete block designs. II. Additional tables for the method of paired comparisons. *Biometrika*, *41*, 1954, 502–537. (b)

Bradley, R. A. Rank analysis of incomplete block designs. III. Some large sample results on estimation and power for a method of paired comparisons. *Biometrika*, *42*, 1955, 450–470.

Bradley, R. A., & Terry, M. E. Rank analysis of incomplete block designs. I. The method of paired comparisons. *Biometrika*, *39*, 1952, 324–345.

Braida, L. D., & Durlach, N. I. Intensity perception. II. Resolution in one-interval paradigms. *Journal of the Acoustical Society of America*, 1972, *51*, 483–502.

Bramber, D. The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *Journal of Mathematical Psychology*, 1975, *12*, 387–415.

Broadbent, D. E., & Gregory, M. Vigilance considered as a statistical decision. *British Journal of Psychology*, 1968, *54*, 309–323.

Brown, J., & Cane, V. R. An analysis of the limiting method. *British Journal of Statistical Psychology*, 1959, *12*, 119–126.

Carroll, J. B., & Arabie, P. Multidimensional scaling. *Annual Review of Pscyhology*, 1980, *31*, 607–649.

Cliff, N. Scaling. *Annual Review of Psychology*, 1973, *24*, 473–506.

Coombs, C. H., Dawes, R. M., & Tversky, A. *Mathematical psychology: An elementary introduction.* Englewood Cliffs, N.J.: Prentice-Hall, 1970.

Cornsweet, T. N., & Pinsker, H. M. Luminance discrimination of brief flashes under various conditions of adaptation. *Journal of Physiology*, 1965, *17*(6), 713–719.

Cramer, H. *Mathematical methods of statistics.* Princeton, N.J.: Princeton University Press, 1963.

Cross, D. V. An application of mean value theory to psychological measurement. In *Progress Report No. 6* (Report No. 05613-3-P). Ann Arbor: The Behavioral Analysis Laboratory, University of Michigan, 1965.

Dawson, S. Binocular and uniocular discrimination of brightness. *British Journal of Psychology*, 1913, *6*, 78–108.

Debreu, G. Topological methods in cardinal utility theory. In S. Karlin & P. Suppes (Eds.), *Mathematical methods in the social sciences.* Stanford: Stanford University Press, 1960.

Derman, C. Non-parametric up-and-down experimentation. *Annals of Mathematical Statistics*, 1957, *28*, 795–797.

Doignon, J. P., & Falmagne, J. C. Difference measurement and simple scalability with restricted solvability. *Journal of Mathematical Psychology*, 1974, *11*(4), 473–499.

Durlach, N. I., & Braida, L. D. Intensity perception: I. Preliminary theory of intensity resolution. *Journal of the Acoustical Society of America*, 1969, *46*, 372–383.

Egan, J. P. *Signal detection theory and ROC analysis*. New York: Academic Press, 1975.

Ekman, G. Subjective power functions and the method of fractionation. *Report from the psychological laboratory (No. 34)*. Stockholm: University of Stockholm, 1956.

Ekman, G. Methodological note on scales of gustatory intensity. *Scandinavian Journal of Psychology*, 1961, *2*, 185–190.

Ekman, G., & Sjoberg, L. Scaling. *Annual Review of Psychology*, 1965, *16*, 451–474.

Ellis, B. *Basic concepts of measurement*. London: Cambridge University Press, 1966.

Engen, T. Psychophysics: Discrimination and detection. In J. W. Kling & L. A. Riggs (Eds.), *Experimental psychology*. New York: Holt, Rinehart & Winston, 1971.

Fagot, R. F. On the psychophysical law and estimation procedures in psychophysical scaling. *Psychometrika*, 1963, *28*, 145–160.

Falmagne, J. C. Note on a simple property of binary mixtures. *British Journal of Mathematical & Statistical Psychology*, 1968, *21*(1), 131–132.

Falmagne, J. C. The generalized Fechner problem and discrimination. *Journal of Mathematical Psychology*, 1971, *8*, 22–43.

Falmagne, J. C. Foundations of Fechnerian psychophysics. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology, Vol. 2. Measurement psychophysics and neural information processing*. San Francisco: W. H. Freeman, 1974.

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

Falmagne, J. C. Note: Weber's inequality and Fechner's problem. *Journal of Mathematical Psychology*, 1977, *16*, 267–271.

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

Falmagne, J. C. On a class of probabilistic conjoint measurement models: Some diagnostic properties. *Journal of Mathematical Psychology*, 1979, *19*, 73–88.

Falmagne, J. C. A probabilistic theory of extensive measurement. *Journal of Philosophy of Science*, 1980, *47*(2), 277–296.

Falmagne, J. C. Psychometric functions theory. *Journal of Mathematical Psychology*, 1982, *25*(1), 1–50.

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

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

Falmagne, J. C., & Narens, L. Scales and meaningfulness of quantitative laws. *Synthese*, 1983, *55*(3), 287–326.

Fechner, G. T. *Elements of psychophysics*. D. H. Howes and E. C. Boring (Eds.), (H. E. Adler, trans.). New York: Holt, Rinehart & Winston, 1966. (Originally published, 1860).

Fisher, R. A., & Tippett, L. H. C. Limiting forms of the frequency distributions of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Society*, 1928, *24*, 180–190.

Fréchet, M. Sur la loi de probabilité de l'écart maximum. *Annales de la Société Polonaise de Mathématiques (Cracow)*, 1927, *6*, 93.

Galambos, J. *The asymptotic theory of extreme order statistics*. New York: Wiley, 1978.

Galanter, E., & Messick, S. The relation between category and magnitude scales of loudness. *Psychological Review*, 1961, *68*, 363–372.

Garner, W. R. An informational analysis of absolute judgements of loudness. *Journal of Experimental Psychology*, 1953, *46*, 373–380.

Garner, W. R. *Uncertainty and structure as psychological concepts*. New York: Wiley, 1962.

Gescheider, G. A. *Psychophysics: Method and theory*. Hillsdale, N.J.: Erlbaum, 1976.

Gigerenzer, G., & Strube, G. Are there limits to binaural additivity of loudness? *Journal of Experimental Psychology: Human Perception and Performance*, 1983, *9*(1), 126–136.

Gnedenko, B. V. Sur la distribution limite tu terme maximum d'une serie aleatoire. *Annals of Mathematics*, 1943, *44*, 423–453.

Gravetter, F., & Lockhead, G. R. Criterial range as a frame of reference for stimulus judgments. *Psychological Review*, 1973, *80*, 203–216.

Gray, D. E. (Ed.). *American Institute of Physics handbook*. New York: McGraw-Hill, 1957.

Green. D. M. *An introduction to hearing*. Hillsdale, N.J.: Erlbaum, 1978.

Green, D. M., & Luce, R. D. Variability of magnitude estimates: A timing theory analysis. *Perception and Psychophysics*, 1974, *15*, 291–300.

Green, D. M., & Swets, J. A. *Signal detection theory and psychophysics*. New York: Krieger, 1974.

Green, D. M., & Luce, R. D. Parallel psychometric functions from a set of independent detectors. *Psychological Bulletin*, 1975, *82*, 483–486.

Green, D. M., Luce, R. D., & Duncan, J. E. Variability and sequential effects in magnitude production and estimation of auditory intensity. *Perception and Psychophysics*, 1977, *22*, 450–456.

Gumbel, E. G. *Statistics of extremes*. New York: Columbia University Press, 1958.

Helmholtz, H. V. Zahlen und Messen erkenntnis-theoretisch betrachtet, Philosophische Aufsutze Eduard Zeller gewidmet, Leipzig, 1887. (Reprinted in Gesammelte Abhandl., 1895, *3*, 356–391. (English translation by C. L. Bryan), *Counting and measuring*. Princeton, N.J.: Van Nostrand, 1930.

Holland, M. K., & Lockhead, G. R. Sequential effects in absolute judgments of loudness. *Perception and Psychophysics*, 1968, *3*, 409–414.

Holway, A. H., & Pratt, C. C. The Weber-ratio for intensive discrimination. *Psychological Review*, 1936, *43*, 322–340.

Iverson, G. J. Note: Conditions under which Thurstone Case III representations for binary choice probabilities are also Fechnerian. *Journal of Mathematical Psychology*, 1979, *20*(3), 263–271.

Iverson, G. J. Weber's inequality and asymptotic representations of binary choice probabilities. Submitted to *Journal of Mathematical Psychology*, 1983.

Iverson, G. J., & Falmagne, J. C. Statistical issues in measurements. *Mathematical Social Sciences*. In press.

Iverson, G. J., & Pavel, M. Invariant properties of masking phenomena in psychoacoustics and their theoretical consequences. *SIAM-AMS Proceedings*, 1980, *13*, 17–24.

Iverson, G. J., & Pavel, M. On the functional form of partial masking functions in psychoacoustics. *Journal of Mathematical Psychology*, 1981, *24*, 1–20. (a)

Iverson, G. J., & Pavel, M. Invariant characteristics of partial masking: Implications for mathematical models. *Journal of the Acoustical Society of America*, 1981, *69*, 1126–1131. (b)

Jesteadt, W., & Bilger, R. C. Intensity and frequency discrimination in one- and two-interval paradigms. *Journal of the Acoustical Society of America*, 1974, *55*, 1266–1276.

Jesteadt, W., Luce, R. D., & Green, D. M. Sequential effects in judgments of loudness. *Journal of Experimental Pscyhology: Human Perception and Performance*, 1977, *3*, 92–104.

Jesteadt, W., & Sims, S. L. Decision processes in frequency discrimination. *Journal of the Acoustical Society of America*, 1975, *57*, 1161–1168.

Jesteadt, W., Wier, C. C., & Green, D. M. Intensity discrimination as a function of frequency and sensation level. *Journal of the Acoustical Society of America*, 1977, *61*, 169–177.

Johnson, N. I., & Kotz, S. *Distributions in statistics: Discrete distributions*. Boston: Houghton Mifflin, 1969.

Johnson, N. I., & Kotz, S. *Distributions in statistics: Continuous univariate distributions* (Vols. 1 & 2). Boston: Houghton Mifflin, 1970. (a)

Johnson, N. I., & Kotz, S. *Distributions in statistics: Continuous multivariate distributions*. New York: Wiley, 1970. (b)
Kesten, H. Accelerated stochastic approximation. *Annals of Mathematical Statistics*, 1958, *29*, 41–59.
Kling, J. W., & Riggs, L. A. *Experimental psychology* (3rd ed.). New York: Holt, Rinehart & Winston, 1971.
Krantz, D. H. Threshold theories of signal detection. *Psychology Review*, 1969, *76*, 308–324.
Krantz, D. H. Integration of just noticeable differences. *Journal of Mathematical Psychology*, 1971, *8*, 591–599.
Krantz, D. H. A theory of magnitude estimation and cross-modality matching. *Journal of Mathematical Psychology*, 1972, *9*, 168–199.
Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. *Foundation of measurement* (Vol. 1). New York: Academic Press, 1971.
Kristofferson, A. B., & Dember, W. N. *Detectability of targets consisting of multiple small points of light*. University of Michigan: Vision Research Laboratories, Technical Report No. 2144-298-T, 1958.
Laming, D. R. J. *Sensory analysis*. (Technical report), Department of Experimental Psychology, Cambridge University, 1983.
Levelt, W. J. M., Riemersma, J. B., & Bunt, A. A. Binaural additivity in loudness. *British Journal of Mathematical and Statistical Psychology*, 1972, *25*, 51–68.
Levine, M. V. Transformations that render curves parallel. *Journal of Mathematical Psychology*, 1971, *7*, 410–444.
Levine, M. V. Transforming curves into curves with the same shape. *Journal of Mathematical Psychology*, 1972, *9*, 1.
Levine, M. V. Geometric interpretations of some psychophysical results. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology, Vol. 2, Measurement, psychophysics, and neural information processing*. San Francisco: W. H. Freeman, 1974.
Levine, M. V. Additive measurement with short segments of curves. *Journal of Mathematical Psychology*, 1975, *12*, 212–224.
Levitt, H. Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, 1970, *49*, 467–476.
Levitt, H. Decision theory, signal-dectection theory, and psychophysics. In E. E. David & P. D. Denas (Eds.), *Human communication: A unified view*. New York: McGraw-Hill, 1972.
Lim, J. S., Rabinowitz, W. M., Braida, L. D., & Durlach, N. I. Intensity perception VIII: Loudness comparisons between different types of stimuli. *Journal of the Acoustical Society of America*, 1977, *62*, 1256–1267.
Lippman, R. P., Braida, L. D., & Durlach, N. I. Intensity perception. V. Effect of payoff matrix on absolute indentification. *Journal of the Acoustical Society of America*, 1976, *59*, 121–134.
Lochner, J. P., & Burger, J. F. Form of the loudness function in the presence of masking noise. *Journal of the Acoustical Society of America*, 1961, *33*, 1705–1707.
Luce, R. D. *Individual choice behavior: A theoretical analysis*. New York: Wiley, 1959. (a)
Luce, R. D. On the possible psychophysical laws. *Psychological Review*, 1959, *66*(2), 81–95. (b)
Luce, R. D. Detection thresholds: A problem reconsidered. *Science*, 1960, *132*, 1495.
Luce, R. D. A threshold theory for simple detection experiments. *Psychological Review*, 1963, *70*, 61–69. (a)
Luce, R. D. Detection and Recognition. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 2). New York: Wiley, 1963. (b)
Luce, R. D. The choice axiom after twenty years. *Journal of Mathematical Psychology*, 1977, *15*, 215–233. (a)
Luce, R. D. Thurstone discriminal processes fifty years later. *Psychometrika*, 1977, *42*(4), 461–498. (b)
Luce, R. D. Axioms for the averaging and adding representations of functional measurement. *Mathematical Social Sciences*, 1981, *1*, 144–139.
Luce, R. D., & Edwards, W. The derivation of subjective scales from just noticeable differences. *Psychological Review*, 1958, *65*(4), 222–237.
Luce, R. D., & Galanter, E. Discrimination. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 1). New York: Wiley, 1963. (a)
Luce, R. D., & Galanter, E. Psychophysical scaling. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology*. New York: Wiley, 1963. (b)
Luce, R. D., & Suppes, P. Preference utility and subjective probability. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 3). New York: Wiley, 1965.
Luce, R. D., & Green, D. M. A neural timing theory for response times and the psychophysics of intensity. *Psychological Review*, 1972, *79*, 14–57.
Luce, R. D., & Green, D. M. Neural coding and psychophysical discrimination data. *Journal of the Acoustical Society of America*, 1974, *56*, 1554–1564. (a)
Luce, R. D., & Green, D. M. The response ratio hypothesis for magnitude estimation. *Journal of Mathematical Psychology*, 1974, *11*(1), 1–14. (b)
Luce, R. D., Green, D. M., & Weber, D. M. Attention bands in absolute identification. *Perception and Psychophysics*, 1976, *20*, 49–54.
Manski, C. F. The structure of random utility models. In G. L. Eberlein, W. Kroeber-Reil, W. Leinfellner, & F. Schick (Eds.), *Theory and decision*. Dordrecht, Holland: D. Reidel, 1977.
Marks, L. E. *Sensory processes*. New York: Academic Press, 1974.
Marschak, J. Binary choice constraints on random utility indicators. In K. E. Arrow, S. Karlin, & P. Suppes (Eds.), *Standard symposium on mathematical methods in the social sciences*. Stanford, Calif.: Stanford University Press, 1960.
Mashour, M., & Hosman, J. On the new "psychophysical law": A validation study. *Perception and Psychophysics*, 1968, *3*, 367–375.
McFadden, D., & Richter, M. K. *Revealed stochastic preferences*. Department of Economics, University of California, Berkeley, 1970.
McFadden, D., & Richter, M. K. *On the extension of a set function to a probability on the Boolean algebra generated by a family of events, with applications*. (Working paper No. 14), Mathematical Social Science Board Workshop on the Theory of Markets under Uncertainty, Department of Economics, University of California, Berkeley, unpublished, 1971.
McFadden, D. Quantal choice analysis: A survey. *Annals of Economic and Social Measurement*, 1976, *5*(4), 363–390.
McGill, W. J., & Goldberg, J. P. Pure-tone intensity discrimination and energy detection. *Journal of the Acoustical Society of America*, 1968, *44*, 576–581.
Miller, G. A. What is information measurement? *American Psychologist*, 1953, *8*, 3–11.
Miller, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 1956, *63*, 81–97.
Mises, R. von. La distribution de la plus grande de n valeurs. *Revue mathématique de l'union interbalkanique*, 1939, *I*(1), 141–160. (Reprinted in *Selected papers II*). American Mathematical Society, Providence, R.I., 1954, pp. 271–294.
Munsell. *Munsell Book of Color*. 1929.
Nachmias, J. On the psychometric function for contrast detection. *Vision search*, 1981, *21*, 215–224.
Nachmias, J., & Steinman, R. M. Study of absolute visual detection by the rating-scale method. *Journal of the Optical Society of America*, 1963, *53*, 1206–1213.
Naka, K. I., & Rushton, W. A. H. S-Potentials from colour units in the retina of fish (*Cyprinidae*). *Journal of Physiology, London*, 1966, *185*, 536–555. (a)
Naka, K. I., & Rushton, W. A. H. An attempt to analyse color perception by electrophysiology. *Journal of Physiology, London*, 1966, *185*, 556–586. (b)
Naka, K. I., & Rushton, W. A. H. S-Potentials from luminosity units in the retina of fish (*Cyprinidae*). *Journal of Physiology, London*, 1966, *185*, 587–599. (c)
Narens, L. On the scales of measurement. *Journal of Mathematical Psychology*, 1981, *24*, 249–275.

Narens, L., & Luce, R. D. The algebra of measurement. *Journal of Pure and Applied Algebra*, 1976, *8*, 197–233.

Neyman, J., & Pearson, E. S. On the problem of the most efficient tests of statistical hypothesis. Royal Society London, Series A, 1933, p. 289.

Parducci, A. Range-frequency compromise in judgment. *Psychological Monographs*, 1963, *77* (2, Whole No. 565).

Parducci, A. Category judgment: A range-frequency model. *Psychological Review*, 1965, *72*, 407–418.

Parducci, A. Contextual effects: A range-frequency analysis. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of perception* (Vol. 2). New York: Academic Press, 1974.

Parducci, A., & Perrett, L. F. Category rating scales: Effects of relative spacing and frequency of stimulus values. *Journal of Experimental Psychology*, 1971, *89*, 427–452.

Parker, S., & Schneider, B. Loudness and loudness discrimination. *Perception and Psychophysics*, 1980, *28*(5), 398–406.

Parzen, E. *Stochastic processes*. San Francisco, Calif.: Holden-Day, 1962.

Pavel, M. *Homogeneity in complete and partial masking*. Unpublished doctoral dissertation, New York University, 1980.

Peterson, W. W., Birdsall, T. L., & Fox, W. C. The theory of signal detectability. (Trans.) *IRE Prof. Group on Info. Theory*, PGIT-4, 1954, 171–212.

Pfanzagl, J. *Theory of measurement* (2nd ed.). New York: Wiley, 1968, 1971.

Pirenne, M. M. Binocular and uniocular thresholds for vision. *Nature*, 1943, *153*, 698–699.

Plateau, M. J. Sur la mesure des sensations physiques, et sur la loi qui lie l'intensité de ces sensations à l'intensité de la cause excitante. *Bull. Acad. Royale Belge*, 1872, *33*, 376–388.

Pollack, I. Information in elementary auditory displays. *Journal of the Acoustical Society of America*, 1952, *24*, 745–750.

Poulton, E. C. The new psychophysics: Six models for magnitude estimation. *Psychological Bulletin*, 1968, *69*, 1–19.

Pratt, J. W. Robustness of some procedures for the two-sample location problem. *Journal of the American Statistical Association*, 1964, *59*, 665–680.

Purks, S. R., Callaghan, D. J., Braida, L. D., & Durlach, N. I. Intensity perception. X. Effect of preceding stimulus on identification performance. *Journal of the Acoustical Society of America*, 1980, *67*, 634–637.

Pynn, C. T., Braida, L. D., & Durlach, N. I. Intensity perception. III: Resolution in small-range identification. *Journal of the Acoustical Society of America*, 1972, *51*, 559–566.

Quick, R. F. A vector magnitude model of contrast detection. *Kybernetic*, 1974, *16*, 65–67.

Robbins, H., & Monro, S. A stochastic approximation method. *Annals of Mathematical Statistics*, 1951, *22*, 400–407.

Roberts, F. S. Measurement theory. In Gian-Carlo Rota (Ed.), *Encyclopedia of mathematics and its applications* (Vol. 7): *Mathematics and the social sciences*. Reading, Mass.: Addison-Wesley, 1979.

Robinson, G. H. Biasing power law exponents by magnitude estimation instructions. *Perception and Psychophysics*, 1976, *19*, 80–84.

Scharf, B. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of perception. Vol. 4: Hearing*. New York: Academic Press, 1978, 187–242.

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

Sirovich, L., & Abramov, I. Photopigments and pseudo-pigments. *Vision Research*, 1977, *17*, 5–16.

Stevens, J. C. Brightness function: Binocular versus monocular stimulation. *Perception and Psychophysics*, 1967, *2*, 189–192, 452.

Stevens, S. S. On the psychophysical law. *Psychological Review*, 1957, *64*, 153–181.

Stevens, S. S. Cross-modality validation of subjective scales for loudness, vibration, and electric shock. *Journal of Experimental Psychology*, 1959, *57*, 201–209. (a)

Stevens, S. S. Review of L. L. Thurstone, "The measurement of values." *Contemporary Psychology*, 1959, *4*, 388–389. (b)

Stevens, S. S. Procedure for calculating loudness: Mark VI. *Journal of the Acoustical Society of America*, 1961, *33*, 1577–1585. (a)

Stevens, S. S. The psychophysics of sensory function. In W. A. Rosenblith (Ed.), *Sensory communication*. New York: Wiley, 1961, 1–33. (b)

Stevens, S. S. To honor Fechner and repeal his law. *Science*, 1961, *133*, 80–86 (c)

Stevens, S. S. Matching functions between loudness and ten other continua. *Perception and Psychophysics*, 1966, *1*, 5–8. (a)

Stevens, S. S. A metric for the social consensus. *Science*, 1966, *151*, 530–541. (b)

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

Stevens, S. S., & Greenbaum, H. B. Regression effect in psychophysical judgment. *Perception and Psychophysics*, 1966, *1*, 439–446.

Stevens, S. S., & Guirao, M. Loudness functions under inhibition. *Perception and Psychophysics*, 1967, *2*, 459–465.

Strackee, G., & van der Gon, J. J. D. The frequency distribution of the difference between two Poisson variables. *Statistica Neerlandica*, 1962, *16*, 17–23.

Suppes, P., & Zinnes, J. L. Basic measurement theory. In *Handbook of mathematical psychology* (Vol. 1). New York: Wiley, 1963.

Swets, J. A. Detection theory and psychophysics: A review. *Psychometrika*, 1961, *26*, 49–63.

Swets, J. A. (Ed.) *Signal detection and recognition by human observers: Contemporary readings*. New York: Wiley, 1964.

Tanner, W. P., Jr., & Swets, J. A. *A new theory of visual detection*. (Technical Report No. 18), University of Michigan: Electric Defense Group, 1953.

Tanner, W. P., Jr., & Swets, J. A. The human use of information: I. Signal detection for the case of the signal known exactly. (Trans.) *IRE Prof. Group on Info. Theory*, PGIT-4, 1954, 213–221. (a)

Tanner, W. P., Jr., & Swets, J. A. A decision-making theory of visual detection. *Psychological Review*, 1954, *61*, 401–409. (b)

Teghtsoonian, R. On the exponents in Stevens' law and the constant in Ekman's law. *Psychological Review*, 1971, *78*, 71–80.

Thompson, W. A., Jr., & Singh, J. The use of limit theorems in paired comparison model-building. *Psychometrika*, 1967, *32*, 255–264.

Thurstone, L. L. A law of comparative judgment. *Psychophysical Review*, 1927, *34*, 273–286. (a)

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

Tumarkin, A. A biologist looks at psychoacoustics. A commentary to: Warren, R. M. Measurement of sensory intensity. *The Behavioral and Brain Sciences*, 1981, *4*, 175–223.

Urban, F. M. On the method of just perceptible differences. *Psychological Review*, 1907, *14*, 244–253.

Van Meter, D., & Middleton, D. Modern statistical approaches to reception in communication theory. (Trans.) *IRE Prof. Group on Info. Theory*. PGIT-4, 1954, 119–141.

Wald, A. *Sequential analysis*. New York: Wiley; London: Chapman and Hall, 1947.

Wald, A. *Statistical decision functions*. New York: Wiley, 1950.

Wandell, B., & Luce, R. D. Pooling peripheral information: Average versus extreme values. *Journal of Mathematical Psychology*, 1978, *17*(3), 220–235.

Ward, L. M. Category of judgments of loudness in the absence of an experimenter-induced identification function: Sequential effects of power-function fit. *Journal of Experimental Psychology*, 1972, *94*, 179–184.

Ward, L. M., & Lockhead, G. R. Sequential effects and memory in category judgments. *Journal of Experimental Psychology*, 1970, *84*, 27–34.

Ward, L. M., & Lockhead, G. R. Response system processes in absolute judgment. *Perception and Psychophysics*, 1971, *9*, 73–78.

Warren, R. M. Measurement of sensory intensity. *The Behavioral and Brain Sciences*, 1981, *4*, 175–223.

Wasan, M. T. *Stochastic approximation*. Cambridge: University Press, 1969.

Watson, C. S., Rilling, M. E., & Bourbon, W. T. Receiver-operating characteristics determined by a mechanical analog to the rating scale. *Journal of the Acoustical Society of America*, 1964, *36*, 283–288.

Weber, D. L., Green, D. M., & Luce, R. D. Effects of practice and distribution of auditory signals on absolute identification. *Perception and Psychophysics*, 1977, *22*, 223–231.

Weiss, D. J. Quantifying private events: A functional measurement analysis of equisection. *Perception and Psychophysics*, 1975, *17*, 351–357.

Wetherill, G. B. Sequential estimation of quantal response curves. *Journal of Royal Statistical Society*, 1963, *B25*, 1–48.

Wetherill, G. B., Chen, H., & Vasudeva, R. B. Sequential estimation of quantal response curves: A new method of estimation. *Biometrika*, 1966, *53*, 439–454.

Wilks, S. S. *Mathematical statistics*. New York: Wiley, 1962.

Yellot, J. L., Jr. The relationship between Luce's choice axiom, Thurstone's theory of comparative judgment, and the double exponential distribution. *Journal of Mathematical Psychology*, 1977, *15*, 109–144.

Zinnes, J. L. Scaling. *Annual Review of Psychology*, 1969, *20*, 447–478.