Chapter 3

Psychophysics

Seeing and hearing have a comparable function: transforming types of ubiquitous physical energy into a format that brains can use in support of the organisms they are a part of. Tasting and smelling rely on the interaction of organic or inorganic chemicals in the environment with special chemical sensors in the corresponding organs. Other, specialized sensations, such as certain feelings of pain or pleasure, presumably result from the generation of electrochemical signals entirely within the body. Psychophysical methods are empirical attempts to discover the connection between stimuli and the resulting sensory experiences by measurement of reported percepts or performance. By categorizing experiences and analyzing the connection between stimuli intensities and the resulting experiences, psychophysics attempts to discover how sensory psychology operates and to help establish linking propositions between neurobiology and psychology.[1] One might expect sensations to be linearly related to stimuli, but this is rarely the case as will become apparent. The mathematical underpinnings of psychophysics are related to measurement theory, a branch of applied mathematics. A postulate of measurement theory is that measurements are not the same as the attribute being measured. By measuring the difference between two colors, we have not revealed anything about the nature of the two colors. W. S. Sarle (1995) defines measurement of an attribute of a set of things as "the process of assigning numbers or other symbols to the things in such a way that relationships of the numbers or symbols reflect relationships of the attribute being measured. A particular way of assigning numbers or symbols to measure something is called a *scale* of measurement." The mathematics of measurement are complex and have been

---

Color Space and Its Divisions: Color Order from Antiquity to the Present, by Rolf G. Kuehni
ISBN 0-471-32670-4

described in the three-volume work *Foundations of Measurement* (Krantz et al., 1971, 1989, 1990).

3.1 FUNDAMENTS OF PSYCHOPHYSICS

The problem at the heart of psychophysics is lack of understanding, and of a theory about, how our feelings and experiences are generated. It is what the philosopher D. J. Chalmers referred to as "the hard problem" of consciousness (Chalmers, 1996). Given the lack of such knowledge attempts to create a fundamental theory of psychophysics are speculative. Fechner faced critique from his contemporary Kries that sensation magnitude is the result of hidden processes (that sensation is cognitively impenetrable) and that, as a result, numbers applied to sensation are not numerical in the mathematical sense but merely convenient labels. Numbers may erroneously suggest that quantitative measurement in the sense of physical measurement has taken place, implying a precision that is not there.

Psychophysics has not answered this attack in a fundamental way. It needs to be seen as an essentially empirical, pragmatic attempt to connect intensity of stimulation with magnitude of sensation or, rather, perception. Measuring or scaling psychological magnitudes or differences involves making judgments. Judgment is defined in the relevant meaning as "the process of forming an opinion or evaluation by discerning and comparing" (*Webster's Collegiate Dictionary*, 10th ed.). The processes by which judgments are formed and their neurological expression are unknown. If S is the experienced sensation and R the judged numerical response to the stimulus intensity I, there is an implicit psychophysical transformation F_1 as follows:

$$S = F_1(I). \tag{3-1}$$

On the other hand, the response of observers, R, is related to intensity by the judgment function, F_2:

$$R = F_2(I). \tag{3-2}$$

The two transformation functions F_1 and F_2 are normally conflated to arrive at the result

$$R = F(I). \tag{3-3}$$

Without knowing the values of either component it is impossible to determine the other (Marks and Algom, 1998). Equation (3-3) is a mathematical expression of the assumption underlying psychophysics.

A recent attempt, among several, to develop a fundamental psychophysical theory has been made by K. H. Norwich (1993). He based it on information theory. More intense stimuli have greater information content than less intense ones. According to Norwich sensation provides a measure of the information content in the stimuli:

$$S = kH, \tag{3-4}$$

where S is a perceptual variable (the sensation magnitude), k is a positive constant, and H is the stimulus information available for conversion. S is taken to be the result of several separate stimuli, and H is calculated as a function of the probability of each of the stimuli involved. While the stimuli may in some cases be easy to define in a physical sense, the sensors (e.g., the cones of human vision) are believed to continually undergo changes at the microscopic level. The Weber-Fechner law as well as Steven's power law have been shown to be special cases of Norwich's law, with the Weber-Fechner law presumably applying at higher levels of information and the power law at lower. By itself Norwich's law is of course not an explanation of the "hard problem."

Psychophysical measurements can be made at different points in the seeming continuum of experience: (1) absolute threshold, (2) difference threshold (just noticeable differences), and (3) scaling of the full continuum. Absolute and difference thresholds have classically been determined by Fechner's methods, described below. In the 1960s the signal detection theory was developed (Green and Swets, 1966), and it breaks Fechner's idea of a discrete sensory threshold into two components: (1) a neurophysiologically determined basic component (usually called discriminability, d', and (2) a cognitively determined decision process (response bias, β). As a result there are four response types possible: hits, misses, false alarms, and correct rejections.

Fechner believed that sensory scales can be derived on basis of discrimination and that just noticeable differences expressed in terms of stimulus increments or decrements can add up to sensory magnitudes. Stevens, on the other hand, postulated that sensory magnitudes can be directly assessed. According to him measurement is the assignment of numbers to objects according to certain rules and results in empirical relations. There are experimental data in support of either approach. Psychophysicists today seem to lean toward the idea that the fundamental process involves taking differences. Establishing sensory scales may be said to involve adjusting distances between stimuli to match an internal or external standard, or to make numerical estimates of sensory differences.

Psychophysical judgments are known to be subject to various contextual effects. For example, the results of the scale halving method are known to depend on if the stimulus samples are presented in bottom-up or top-down sequence (hysteresis effect). For color scaling as for many other types of perception, the level of adaptation strongly affects the resulting scale. As will be seen in Chapters 4 and 5, gray scales depend strongly on the lightness level of the surround.

Stimulus sequence has also been shown to affect scaling results. Among the possible reasons are changes in the adaptation level based on the previously seen stimulus. The range of stimuli displayed and the level of stimulus are also known to affect the result, particularly in case of magnitude estimates. Among midgets and basketball players the meaning of small, medium, and large involves absolute values that are much different.

Context effects are well known. How the lightness of a surface is assessed depends not only on the adaptation level but also on other contextual information such as is it seen in shadow or in direct illumination. Similar effects apply to chromatic surfaces.

Despite its lack of a solid foundation psychophysics persists because it has considerable pragmatic value. The continuing interest in psychophysics may also be due to what L. E. Marks called "the metaphorical imperative" (1978). Accordingly we have a need to express our inner world in terms of metaphors that are more easily comprehended by other humans. Scaling and magnitude determinations of perceptions can be seen as an expression of the metaphorical imperative.

When viewed from the perspective of Fechner's outer and inner psychophysics, mentioned in Chapter 2, the conventional psychophysical enterprise can be represented in the schematic sketch of Fig. 3-1. The relationship between stimulus and conscious experience represents classical psychophysics. The relationship between nerve excitation and inner sensation is in the domain of neurophysiology and as such subject to process variability that can be

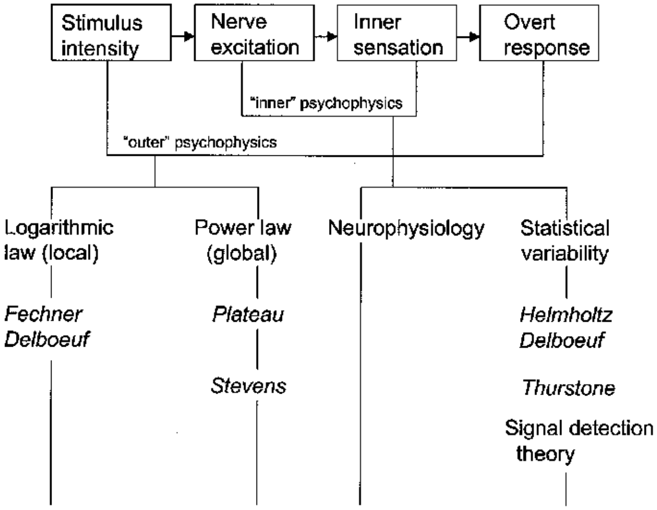

Fig. 3-1 *Conventional psychophysical enterprise as viewed from the perspective of Fechner's inner and outer psychophysics (modified from Murray, 1993). Outer psychophysics attempts to connect the stimulus magnitude with the response magnitude. Inner psychophysics attempts to connect nerve excitation with nonconscious "sensation." It is quantified by recording from visual system cells. Statistical variability and signal detection are taken to be aspects of inner psychophysics.*

treated in different ways, as we have seen. The classical view of psychophysics is that humans react to stimuli. The modern view is that people act on stimuli by interpreting them in a way that contributes toward achieving the goals of the individual.

3.2 CATEGORIES

Categorization is a fundamental trait of humans and perhaps of all animals. To distinguish between edible and poisonous plants, meaning to put them into categories, was an early and imperative task for plant eaters, for example. An important human method of categorization is naming. The process by which we establish categories and sort stimuli into the established categories is clearly complex and at this time largely unknown. Several different theories have been established and vie for recognition.

Stimulus dimensions available for classification are either taken to be continuous (e.g., color) or discrete (four-sided vs. three-sided figures). The structure of categories is either overlapping (hue and chroma) or nonoverlapping (color and shape). In the former case perfect categorization is difficult or impossible while in the latter it is likely. The accuracy of response in categorization experiments is fundamentally either deterministic, meaning equal stimuli result in an equal categorization, or probabilistic: the response of the observer is always based on a (more or less informed) guess. It is obvious that already the stimulus is usually probabilistic. In case of visual stimuli, for example, light is reflected probabilistically off the surface of the object as well as off the eye's cornea, and once the remaining stimulus enters the eye, there is probabilistic activity at all levels of the vision system. Most categorization theories therefore treat categorization data in terms of variability.

Among the theories of category access (what process shapes our category decisions) are those of the necessary and sufficient condition (NSC), the prototype theory, and, related to it, the exemplar theory. According to NSC, categories are described by a series of necessary and sufficient conditions, and the observer tests the sample to see of it meets these conditions. In some situations (e.g., separating squares from triangles), NSC is clearly applicable, but in many others, it is not (e.g., it is difficult to define the necessary and sufficient conditions for a letter symbol to represent an *a*). The prototype theory proposes a (kind of Platonic) prototype for a category whereby objects are classified by their resemblance to the prototype. It implies that we only have the prototype stored in memory and that determining resemblance to it is a cognitive task. It simply delegates the critical categorization to another process. To get out of this difficulty, the exemplar theory was proposed. Accordingly we have all examples of past experiences of the category item stored in memory and compare new examples to the stored ones to categorize them. It seems doubtful, however, that observers need to have seen all reddish colors before they can categorize them as such, for example.

Visual classification is believed to take data from the visual area at the back of the brain to the inferior temporal cortex of the brain located near the temples. This region appears to analyze the data for certain features. It exchanges data with the prefrontal cortex area in the front of the brain an area that may contain category border codes. Categorization is achieved jointly between the latter two areas (Hasegawa and Miyashita, 2002).

In regard to color classification the fundamental question is what process results in setting classification boundaries. If given a thousand randomly different Munsell color chips and asked to categorize them, how would we do it if we had not been exposed to theories and examples before? An obvious attribute for categorization is hue. But it has taken until the seventeenth century to recognize that hue forms a closed circular continuum. Newton has separated spectral colors into seven categories and saw that bluish red and purple colors, not found in the spectrum, can close the spectral array into a circle. Rational classification of hued colors according to whiteness and blackness or lightness and chroma only took place in the nineteenth and early twentieth centuries.

There is the question of the cause of human color categorization in language. As briefly mentioned in Chapter 2, there is a theory by Berlin and Kay that basic color terminology in human languages has followed the same pattern. They identified eleven basic color terms, aside from the six Hering fundamental colors: brown, purple, pink, orange, and gray. The list is notable for its idiosyncrasy: all four chromatic colors are in the yellow-red region of a color circle. In addition the English term pink had in the sixteenth and seventeenth centuries the words yellow or brown attached to it (Merrifield, 1967), and the color of English hunting jackets is called pink. There are no terms for yellow-green, blue-green, and their dark, desaturated versions olive and navy. While there should have been a good number of such colors in natural surroundings in many environments, it has taken systematic categorization of the hue circle to recognize these colors as categories. In C.D.'s extended color circle they are listed as yellowish green and sea blue (see Fig. 2-12). A convincing theory of color categorization that explains the historical development of basic color terms seems to be a long way off.

3.3 DIFFERENCES VERSUS MAGNITUDES

The Weber-Fechner law assumes that just noticeable differences can be considered units of perception. It states that the increment required to result in a just noticeable perceived difference is a constant percentage or fraction of the stimulus magnitude (see Chapter 2). The Weber-Fechner law has been found to be context sensitive. Typical experimental Weber fractions are given in Table 3-1 and an example of the Weber-Fechner law is illustrated in Fig. 3-2.

On a parallel but initially less conspicuous track runs the idea of a power relationship between stimulus and response. Power relationships were

TABLE 3-1 Representative values of Weber fractions for different senses

Sense	Weber Fraction, $\Delta I/I$
Vision (brightness, white light)	1/60
Kinesthesis (lifted weights)	1/50
Pain (thermally aroused skin)	1/30
Audition (tone of moderate loudness)	1/10
Pressure (applied cutaneously)	1/7
Smell (odor of India rubber)	1/4
Taste (table salt)	1/3

Source: Geldard (1962).

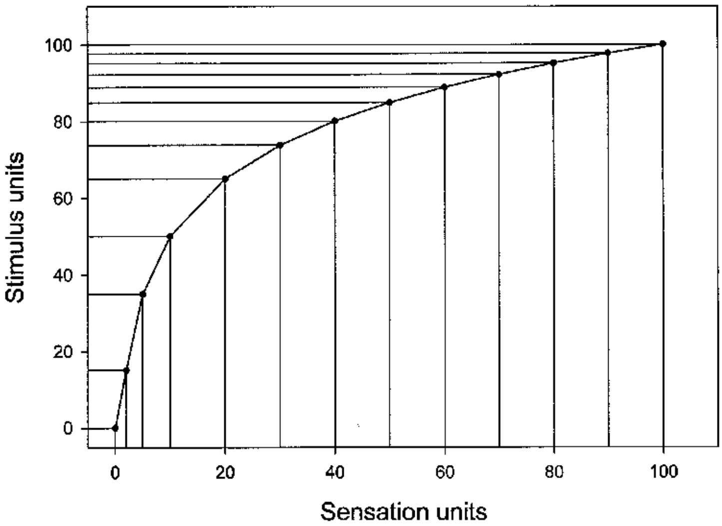

Fig. 3-2 Relation between stimulus and sensation according to the Weber-Fechner law. The logarithmic function illustrates the relationship between an arithmetic (sensation) and a geometric (stimulus) scale.

originally proposed by the mathematicians G. Cramer and D. Bernoulli in their consideration of the utility (subjective value) of money (Bernoulli, 1738). They concluded that the subjective value of an increment of given size of money is larger if it is added to a small amount than when added to a large amount, a relationship that can be expressed with a power function. A power function was proposed by Plateau to express his range partition results in which he had several painter friends paint samples of the psychological midpoint between a black and a white painted samples. From the remarkably uniform results and Plateau's assumption that there were ratios of stimuli and response involved, he concluded that a power relationship could explain them. However, wanting to establish a more detailed gray scale, he asked his friend Delboeuf to develop

TABLE 3-2 Representative exponents of power functions relating subjective magnitude to stimulus magnitude

Sensory Continuum	Measured Exponent	Stimulus Condition
Brightness	0.5	Point source
Brightness	0.5	Brief flash
Brightness	1.0	Point source briefly flashed
Lightness	1.2	Reflectance of gray papers
Visual length	1.0	Projected square
Redness (saturation)	1.7	Red-gray mixture
Taste	1.4	Salt
Smell	0.6	Heptane
Thermal pain	1.0	Radiant heat on skin
Heaviness	1.45	Lifted weights
Electric shock	3.5	Current through fingers

Source: Abbreviated from Stevens (1975). Reprinted by permission of the publisher.

more steps. The results were in better agreement with the Weber-Fechner law, and the power law was largely forgotten until it found its champion in Stevens.

Before Stevens it was J. P. Guilford who proposed in 1932 the first general power law of psychophysics:

$$\Delta I = c_a I^n, \tag{3-5}$$

where I is the stimulus intensity, c_a is the Weber fraction, and n is an exponent. In this equation if $n = 1$, the relationship is logarithmic. Beginning in the 1930s Stevens investigated the relationship between sensory and stimulus magnitudes of many kinds. He concluded that a power law connects all of them, with the power variable anywhere from 0.25 (fourth root) to 1.5:

$$V = bx^p, \tag{3-6}$$

where V is the sensory magnitude, b is a constant, x is the intensity of the stimulus, and p is an exponent. The implication is that the stimulus increments proportional to equal sensory magnitude increments vary less than by equal percentages as predicted by the Weber-Fechner law. Table 3-2 reproduces selected power exponent values from a larger table by Stevens, and Fig. 3-3 graphically illustrates power functions with different exponents.

Stevens and other researchers found that subjects agree quite well on the size of absolute magnitudes of sensory experiences as related to stimulus magnitude. This can be done by direct judgment or magnitude production via cross-modality matching. In the latter case the sensory magnitude is expressed in another medium, for example, by comparing brightness of lights to loudness of sounds or by squeezing a hand dynamometer to express the perceived brightness magnitude.

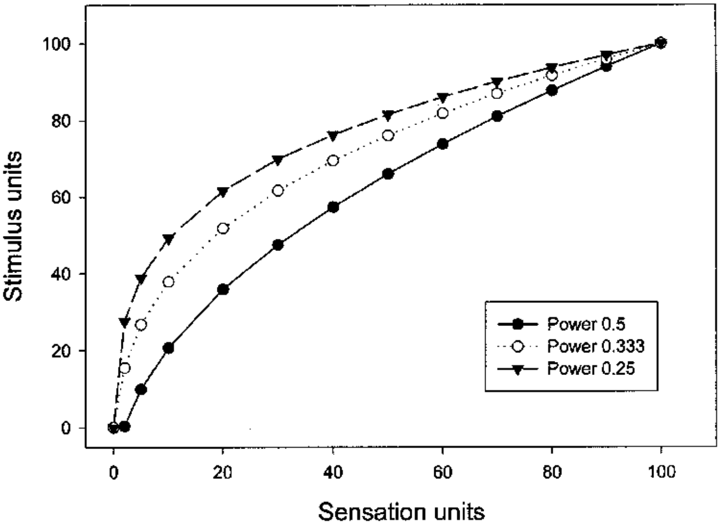

Fig. 3-3 *Relation between stimulus and sensation based on three different power modulations.*

From analysis of his own work and that of other researchers, Stevens concluded that magnitude estimates of sensory experiences tended to have an exponent near 1. In magnitude judgments the observer is free to assign any chosen number to express the magnitude of the perceived stimulus. His own results of magnitude judgments of the steps of a gray scale resulted in a power exponent of 1.2 (Stevens and Galanter, 1957). When making category judgments and difference judgments well above threshold, the power exponent declined in size, and when making difference judgments near threshold, it approached zero and the Weber-Fechner law.

Category scales have set numbers or adjectives and the results of category judgments are usually nonlinearly related to those of magnitude judgments. A gray scale based on equisection or category judgments, namely on differences, results typically in an exponent of 0.25 to 0.5. In his own multiple bisection experiment of brightness Stevens (1953) found an exponent of 0.26. Thresholds are category measurements related to the uncertainty of the response, with a relationship to the stimulus difference most accurately expressed as a logarithmic function. The power function can also be closely matched with an hyperbolic function, first implicitly used by Hering (1874):

$$V = \frac{ax}{1+kx}, \tag{3-7}$$

where a and k are constants and x is the intensity of the stimulus. In a plot with a linear ordinate and a log abscissa the resulting function has an S shape. Thresholds, and the Weber-Fechner law approximating them represent

TABLE 3-3 Effect of judgment type and size of difference on applicable stimulus modulation

Type of Evaluation	Applicable Power	Example
Magnitude	0.8 and larger	Lightness of gray papers
Paired comparison	0.05–0.75	Small color differences, Munsell chroma scales
Thresholds	0.01–0.33	Color thresholds

(in Fechnerian terms) local psychophysics; the power law is, in terms of differences or magnitudes, representative of more global psychophysics. Table 3-3 provides a comparison.

It is apparent that neither the logarithmic nor the power law provides an explanation for the causes of the psychophysical relationships. They are merely mathematical models. In the case of visual scaling the explanation for the apparent signal compression is buried in the complexity of the visual system and the possible impact of empiricism. While some details about the visual system are known at this time, they are not nearly enough to develop a fully detailed model predictive of all known effects. J. C. Baird has developed a general simulation model that attempts to explain the effects described by the laws from the aggregate action of neural cells with varying thresholds (sensory aggregate model; Baird, 1997). It is not evident that this model applies to visual scaling. K. Richter (1996) has proposed a model that attempts to explain visual psychophysics only in terms of cone receptor saturation. It seems more likely that post-receptor effects also play a role.

Psychophysics in the classical sense depends on so-called linking propositions. These are propositions of how neurophysiology and psychophysics might be connected. Linking propositions can be strong or weak, depending on the level of evidence supporting the proposition. A typical linking proposition is that chromatic perception is supported by neurons with opponent response in the visual area of the brain. Linking propositions related to the concept of a uniform color space are still weak. The modern view of vision is that the brain constructs what we consciously experience from the output of several of its visual modules and using many (and probably many as yet unknown) decision rules to construct the experience from generally ambiguous input at the retinal level. As S. E. Palmer expressed it: “The job of visual perception is to combine external and internal information to make meaningful facts about the environment available to the organism” (Palmer, 1999). Simple links between neural sensitivities to stimuli and psychological measurements or judgments are therefore unlikely except in highly relativized situations.

3.4 PSYCHOPHYSICAL SCALING: LEVELS OF MEASUREMENT

An important aspect of psychophysics is the scaling methods used to assess local and global scaling. Stevens (1946) proposed that onedimensional scales may be placed into a hierarchy with each subsequent type having greater explanatory power than the previous one. This hierarchy is called levels of measurement.

Nominal Scales

They are at the lowest level and refer to names or identifications of items only. The same symbol is assigned to two things if they have the same value of the attribute. An applicable sample for color is names. Colors can be grouped, for example, into blues, browns, pinks, purples, and grays. An appropriate statistic is the number of cases.

Ordinal Scales

Numbers are assigned to things in a way that the order of the numbers reflects the order of the attribute. Items are placed into ascending or descending rank order depending on some kind of magnitude. For colors a typical example is a series of grays that can be ordered according to the concept of blackness. The color nearest to black (sample E in Fig. 3-4) has ordinal scale value 1, that closest to it but lighter (sample A) has value 2, and so on. A series of chromatic colors of varied chromaticness can also be placed into an ordinal chroma scale. Ordinal scales do not contain any information about the sensory magnitude of the steps between the grades on the scale. In Fig. 3-4 a random

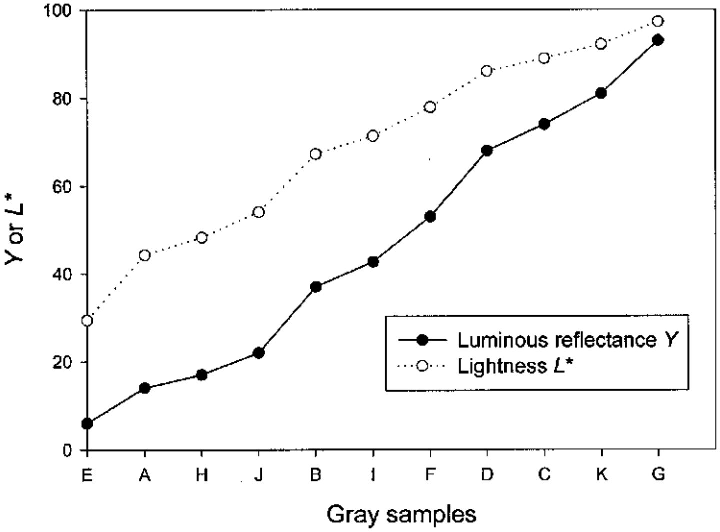

Fig. 3-4 Twelve random samples of gray papers (A–K) are placed into an ordinal scale based on their luminous reflectance value Y *and on an interval scale based on lightness* L*.

collection of gray samples has been ordered according to the luminous reflectance scale (Y). Luminous reflectance places the samples into an ordinal scale where the scale values do not provide any information about the perceptual magnitude of the steps between the grades of the scale. An ordinal scale is subject to logical operations: equal to, greater than, less than; and appropriate statistics are median, percentiles, or rank-order correlation.

Interval Scales

At the next level interval scales provide quantitative information concerning the distances or differences between grades. Here things are assigned numbers so that differences between numbers reflect differences of the attribute. In an interval scale two grades differing by three interval units at the lower end of the scale, and two grades at the higher end of the scale also differing by three interval units are equally distant (e.g., the Celsius and Fahrenheit temperature scales). In interval scales the distance between two percepts is represented with a number according to

$$a = mb + c_0, \tag{3-8}$$

where a is the scale number representing a grade, m is any positive number, b is the distance of the grade from the neighboring grade or the origin, and c_0 is any finite number. Addition of and multiplication with constants are permissible transformations and statistics include mean, standard deviation, or correlation coefficient. Color scales usually are interval scales. In Fig. 3-4 the twelve random gray samples have also been ordered according to the interval scale L^*. The perceptual difference between samples B and I is about half as large as that between samples F and D.

Ratio Scales

These are interval scales that have a natural origin. In this case things are assigned numbers such that differences and ratios between the numbers reflect differences and ratios of the attribute. While in interval scales two numbers could be applied arbitrarily to the scale (m and c_0) only one number can be arbitrarily assigned:

$$a = mb. \tag{3-9}$$

Examples are lengths in meters, duration in seconds, and temperature in degrees Kelvin. As mentioned before, ratio scales in regard to color are controversial. Ratio scales can be multiplied with a constant only and statistics include percent variability.

Absolute Scales

Here things are assigned numbers such that all properties of the numbers reflect analogous properties of the attribute. These measurement levels are part of a continuum of order, with the level of order being weakest for nominal scales and strongest for absolute scales. In practice, depending on the measurement technique used, the same scale can have properties of two measurement levels.

3.5 SCALING METHODS

Many different methods of scaling have been developed in psychophysics. For absolute or difference thresholds the classical methods discussed in this section are those described by Fechner (1860).

Method of Adjustment or Method of Average Error

In this method the observer can adjust the stimulus magnitude and does so until she observes a just noticeable difference. On the surface this is a simple and straightforward method. However, it requires equipment on which the stimulus magnitude (under control of the observer) can be continuously adjusted. The results are direct but also subject to cognitive adjustments and biases.

Method of Limits or Method of Minimal Change

Here the experimenter presents changes in the stimulus magnitude in preset small increments in ascending or descending order, starting with an imperceptible increment in the former case and a clearly perceptible one in the latter. This continues until the observer indicates the perception of a difference in the former case or the absence of a difference in the latter. The JND is taken to be represented by the average of ascending and descending trials.

Method of Constant Stimuli

In this method the experimenter selects several constant pairs of stimuli ranging from below threshold to above threshold. These are then presented several times randomly to the observer. The observer responds with a yes or no, and the JND limit is determined at the 50% or perhaps at the standard deviation level. Alternatively, the experiment is set up so that the observer must respond to a forced choice between two pairs of stimuli. This method has been further varied in the so-called staircase procedure. Here a forced choice is imposed at the lowest level. If the observer responds wrongly, the stimulus difference is increased for another forced choice. The JND limit is approached from both sides in such experiments.

Matching Method

A well-known experiment, believed related to color thresholds, is the determination of color matching error by MacAdam (1942). He constructed an apparatus in which color stimuli could be varied along selected lines in the CIE chromaticity diagram at constant luminosity. A standard stimulus was displayed and the observer adjusted a test field until the two fields matched, that is, had identical appearance. From the variability of repeated tests MacAdam calculated the matching error that he thought to be related to the difference threshold.

These methods are known under the general rubric of confusability scaling. Thurstone stated a law of comparative judgment in 1927. This law considers every perceptual magnitude judgment as a variable datum from a discriminable process that he took to be normally distributed. In this manner the full power of statistics of normal distribution can be applied. Under specific conditions such statistical treatment results in confirmation of the Weber-Fechner law, other conditions result in applicability of Steven's power law.

3.6 UNIDIMENSIONAL SCALING METHODS

Unidimensional scaling involves perceptions that have one attribute only. Several physical dimensions may be involved in generating the attribute. Over the century and a half since the beginning of psychophysics several scaling methods for unidimensional scaling of perceptual distances have been developed. Only those used in color scaling will be briefly discussed.

Partition Scaling

First used in the form of the equisection method by Plateau, partition scaling is a direct estimation method. In the equisection method the perceptual distance between two different stimuli may be halved in several steps. Another version determines equal-appearing intervals.

Ratio Production and Ratio Estimation

As the name implies, the method results in a ratio scale. In ratio production the observer adjusts a magnitude of a perceptual attribute until it equals the perceived double magnitude (or other ratio) of a reference. In ratio estimation the observer estimates the ratio between reference and test stimuli. For reasons discussed previously, this method is rarely applied in color scaling.

Magnitude Production and Estimation

In production the observer is given a stimulus and a number that corresponds to it and is asked to produce a stimulus that is representative of another number. Thus a certain sound may be said to represent the value 10, and the observer is asked to modify the sound stimulus so that the resulting loudness of the sound represents the value 15, say. In magnitude estimation the observer estimates the perceived magnitude of experiences from stimuli and assigns numbers to them. For color scaling a large number of stimuli or the capability of stepwise or continuous stimulus adjustment is required.

Category Scaling

Here the observer is asked to separate large numbers of experiences into categories. The corresponding samples must be similar enough so different observers arrive at different categorization. The variability in judgment by different observers is assumed to follow a standard normal distribution from which an interval scale can be constructed. A typical example is acceptability or pass-fail judgments of small color differences in which the observer is asked to determine if a given sample meets or fails a criterion of acceptability.

Paired Comparison

In this method all samples are presented to the observer in all possible pairs or in all pairs of test against a reference sample. The proportion of times a given sample is judged greater in magnitude of a given attribute is determined. Interval scales are derived from the results under the assumption of statistically normal distribution.

In recent years a form of interval judgment for the purpose of suprathreshold color scaling has been in wide use. In this method sample pairs exhibiting small differences are compared against a reference difference, usually in form of an achromatic pair with a perceptual difference of similar magnitude to those of the sample pairs under estimation. Alternately, an International Standards Organization (ISO) type gray scale that displays achromatic pairs with varying perceptual differences has been used. In such situations what are multidimensional color differences (usually involving hue, chroma, and lightness differences at the same time) are evaluated as if they were unidimensional. The surround lightness and chromaticness affects the perceptual magnitude of the reference pair(s), as well as of the test pairs making the result a function of the surround.

Minimally Distinct Border Scaling

A novel method of scaling color differences was developed in the 1960s and 1970s by R. M. Boynton and P. K. Kaiser (Boynton, 1983). Their idea was that with two perfectly juxtaposed color fields at equal luminance the strength of the resulting border should be an indicator of the chromatic difference in the constant luminance plane. The method was found to be applicable only where the two fields differed in activation of the L and M cones.

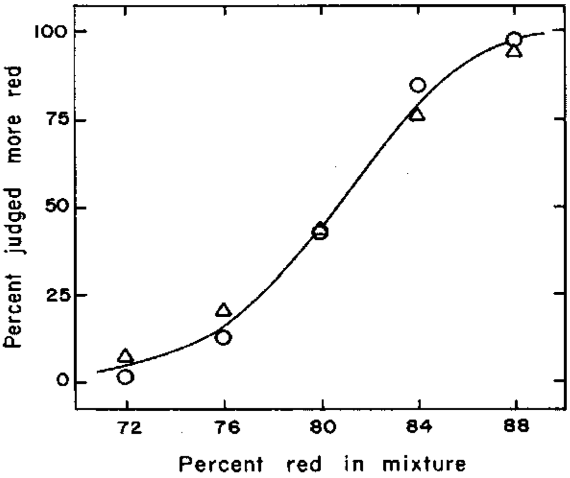

Fig. 3-5 *Example of a psychometric function. Samples contain different amounts of a red and a gray stimulus (by spinning disk mixture). Observers judged if a sample contained more red than the one shown previously. Circles indicate results when a reference sample containing 80% red was shown immediately before the test sample, triangles where no reference was shown. From Stevens (1975).*

3.7 PSYCHOMETRIC FUNCTION

Psychometric functions are the functions expressing the relationship between subjective and objective measures. The usually normal statistical distribution of individual judgments generally results in nonlinear psychometric functions. An example is pass-fail judgments of color difference. Here the relationship between percent acceptability (% pass) and calculated color difference was found to have a typical S-shaped (sigmoidal) form (see Fig. 3-5 for another example). In such cases it is necessary to linearize the visual scale by using an appropriate method (e.g., Indow and Morrison, 1991). A function that has been used in recent years for this purpose is the cumulative normal distribution (probit) function (e.g., see Berns et al., 1991).

3.8 MULTIDIMENSIONAL SCALING

Multidimensional scaling (MDS), originally developed by W. S. Torgerson in 1952 (1958) and R. N. Shepard (1962), creates geometric models based on similarities, dissimilarities, or proximity. Mathematical analysis is performed on spatial or distance data. However, MDS can also be applied to nondistance data. It can address the nature of the metric of the multidimensional stimulus space implicit in the data, meaning it can extract the unknown dimensionality. Because psychological data usually contain considerable variability differ-

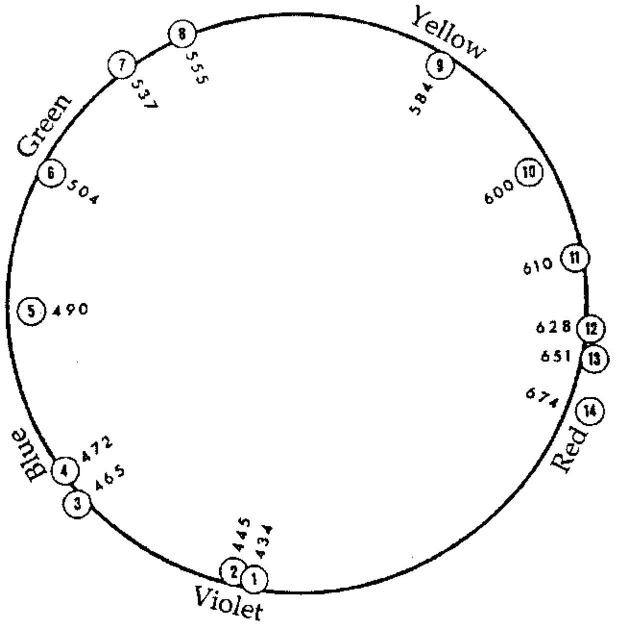

Fig. 3-6 Configuration of MDS analysis of judged differences between fourteen spectral colors (circles with numbers and identified by spectral wavelength in nanometers. A circle segment (closed to a full circle) has been fitted to the points. Presumably on the segment between spectral colors 1 and 14 extraspectral red and purple colors mixed from various ratios of 1 and 14 would be located. From Shepard (1964).

ent MDS techniques often find several different models, all with comparable accuracy of fit. A widely used methodology of MDS is INDSCAL (individual difference scaling; Carroll and Chang, 1970). It is based on euclidean geometry, but evidence has been found that it can "adequately" recover sample configurations even if the true metric is non-euclidean.

In one version of MDS the pairwise distance (in terms of a visual estimate or any other distance measurement) between all items used in the test is established. From these data a similarity or dissimilarity matrix is created as input to the MDS module. Several parameters or assumptions can be changed in MDS and the results are subject to interpretation.

An interesting finding of MDS is that when perceptual distances between fourteen spectral colors are judged by observers the result of the MDS analysis of the data is a two-dimensional structure that is fit best by a circle segment (Fig. 3-6; Shepard, 1994).

MDS has been applied extensively to Munsell data by T. Indow and coworkers (Indow, 1988) and others. Using judgments of magnitude of color difference between many samples of the Munsell system as input into the MDS analysis, Indow recovered the psychological diagram of the Munsell system with good accuracy (see Fig. 4-4 in the next chapter). In principle, scaling of small complex color differences (involving more than one attribute) might be analyzed by MDS to determine the implicit dimensions. This does not appear to have been done so far, perhaps because the dimensionality of the color experience is believed known.

The MDS method is not without its critics (see e.g., Saunders and van Brakel, 2002). Some criticism harks back to Kries original doubts about psychophysics, namely are magnitudes and differences distances that can meaningfully be expressed in numerical or geometrical form.

3.9 PSYCHOLOGICAL AND PSYCHOPHYSICAL SPACES

Complex psychological phenomena such as color vision are usually illustrated in geometrical spaces. Such spaces presuppose an isomorphism between color experiences and the selected geometrical model. The isomorphism may be based on the logic of the four unique hues and their mixtures as well as of hues, black and white or on equality of differences. In most color order systems, as seen in Chapter 2, white is placed on top producing the isomorphic association lighter equals higher. Greener means closer to the location of unique green, and so on. Closer to the centerline means grayer. The specific isomorphism depends on the distance criterion used. In a perceptually uniform (for a specific set of conditions) space equal distances between points represent equal perceived differences. There are two kinds of isomorphism in regard to color space: (1) between the perceptual experiences and the psychological color space and (2) between the psychological and the psychophysical space.

As seen in Chapter 2, historically color solids had simple geometrical forms with euclidean geometry (cones, spheres, pyramids, cubes). The Munsell system indicated that for a "uniform" system the underlying structure might be simple (cylindrical) but the surface of the solid complex. As will be seen in Chapters 7 and 8, a euclidean form of a uniform color space can be ruled out, at least at the global level.

The concept of space is of unlimited complexity. While most humans cannot imagine a space with more than three dimensions mathematicians have developed a number of categories of spaces. C. F. Gauss and N. I. Lobachevski invented hyperbolic geometry and G. F. B. Riemann the elliptic geometry named after him. Riemannian geometry has been used in the past in connection with color scaling (e.g., MacAdam, 1981; Völz, 1998b). In the hyperbolic geometry the sum of the angles in a triangle is less than 180° degrees, in the elliptic geometry greater than 180° degrees. In string theory modern physicists have proposed a universe of nine dimensions.[2]

As will be shown, a uniform color space based on small color differences requires an elliptic geometry. Psychophysical spaces that have no claim on uniformity, such as the Rösch-MacAdam space, fit into euclidean geometry without difficulty. It is evident that any kind of three-dimensional euclidean space form and associated color solid can be used as a regular (but not uniform) psychophysical color space. Modern color difference formulas dealing with some of the complexities of uniform color space currently are based on mathematical modifications of a euclidean space. The exact type of geometry applicable to a uniform color space under well-defined conditions of observation and magnitude of differences remains to be determined.

3.10 PSYCHOPHYSICAL SCALING AS A BASIS OF COLOR SPACE

Color space can be represented in a continuous fashion only with mathematical formulas that, for most people, lack direct comprehension in terms of color experiences. Historically color solids have been illustrated with two- or three-dimensional arrangements of color samples in form of atlases or "color trees." Here psychological and psychophysical scaling is essential if psychological uniformity relative to specific conditions is the goal. Psychophysical scaling has the advantage of providing physical measurement support for the stimuli. On the other hand, they require a psychophysical model that accurately reflects psychological results. A major impetus for psychophysical scaling of threshold and small suprathreshold differences and related mathematical models has been the desire for objective quality control of colored materials.

Psychophysics as a methodology is essential for the construction of a uniform color space. Psychophysical data are subject to considerable variation based on the physics of measurement and observer variability as well as variability in observational context. In the case of color attribute and color difference scaling, it is not yet clear what the major contributors to variability are. Careful experimentation is required to determine the effects of individual contributors.