TOWARD A UNIFIED THEORY OF PSYCHOPHYSICS

Robert L. McBride B.Sc.

Commonwealth Scientific and Industrial Research Organization
(CSIRO), Food Research Laboratory, and
School of Behavioural Sciences, Macquarie University

A Thesis submitted for the degree of Doctor of Philosophy
at Macquarie University

June 1982

ACKNOWLEDGMENTS

I am indebted to Dr J. H. B. Christian, Chief, CSIRO Division of Food Research, and Dr A. R. Johnson, Officer in Charge, CSIRO Food Research Laboratory, for permission to undertake this degree. I thank my Supervisor, Dr R. P. Power, for his support and encouragement; also Dr B. V. Chandler for his meticulous, but always constructive, criticism of my writing.

Mary Willcox and John Best helped with the statistical analyses, and Arthur Kuskis provided invaluable assistance in the laboratory. I express appreciation to my many colleagues at the Food Research Laboratory who, as subjects, withstood the rigours of the experimental work without complaint; also to Ian Mathieson, without whose help I should not have been able to construct this Thesis on a somewhat temperamental word processor.

I am most grateful to my wife Pamela for support throughout my candidature - especially for her cheerful endurance of my many hours of preoccupation.

CONTENTS

INDEX TO FIGURES

SUMMARY

This Thesis proposes a means of unifying psychophysics. No new concepts are invoked; the model is a rearrangement of existing ideas and principles. First, it is shown that magnitude scales of sensation fail to comply with rigorous validity criteria. It is argued this occurs because the number continuum, in magnitude tasks, is perceived in a logarithmic manner. This explanation offers a means of resolving the discrepancy between magnitude and category scales.

A re-evaluation of the psychophysical law suggests that a valid psychophysical function may be derived from two theoretical premises: the empirical Weber function (not Weber's law), and Fechner's original assumption that just noticeable differences (JNDs) are subjectively equal. These premises specify the obsolete JND (or DL) scale. The present model also predicts, however, that a valid psychophysical function may be obtained by direct interval estimation techniques, e.g., category rating. The concomitant prediction is that, for a given modality, the psychophysical function obtained by direct interval estimation should be isomorphic with the function derived by cumulating JNDs. This isomorphism is shown to be supported

by published work in a number of sensory modalities, and also to be consistent with the properly validated findings of functional measurement analysis.

All experiments in the Thesis were conducted in the taste modality. First, the predicted JND scale-category scale isomorphism is confirmed for taste stimuli representative of the four basic tastes: sweet, acid (sour), salty, and bitter. Methodological bias in the category rating of taste intensity is investigated and found not to be a serious problem in the present approach; nevertheless, a procedure for avoiding contextual bias is suggested and tested experimentally.

A further experiment offers some support for the contention of the present model that rating scales are valid because they involve subjects matching sensation to the position on a line. Finally, in two experiments, the interaction of the sweeteners sucrose and fructose is explored using the functional measurement paradigm. Support for sweetness additivity at low concentrations provides a properly validated estimate of the psychophysical function for sucrose. This function is found to correspond well with the JND and category scales for sucrose obtained earlier, and also corroborates Fechner's assumption of the subjective equality of JNDs.

DECLARATION

I, Robert Lemon McBride, declare that the work contained in this Thesis is original and my own work except where acknowledged in the text. This Thesis has not been submitted to any other university or institution.

RLMcBride

A traveller hired an ass to convey him to a distant place. The day being intensely hot, and the sun shining in its strength, the traveller stopped to rest and sought shelter from the heat under the shadow of the ass. As this afforded protection for only one, and as the traveller and the owner of the ass both claimed it, a violent dispute arose between them as to which of them had the right to it. The owner maintained that he had let the ass only, and not his shadow. The traveller asserted that he had, with the hire of the ass, hired its shadow also. The quarrel proceeded from words to blows and while the men fought the ass galloped off.

Aesop c. 550B.C.

1

PSYCHOPHYSICS AND MEASUREMENT

It often comes as a surprise to psychologists to learn that measurement in psychology has never been formally sanctioned; at least not to everyone's satisfaction. In fact physicists have claimed that measurement, in any true sense, is impossible in psychology. It is one matter glibly to assign numbers to subjective events; it is altogether another to know that these numbers reflect the subjective magnitude to which they are matched in a linear, or additive, manner.

According to Stevens (1975, p. 44), it was the polymath Helmholtz who, in 1887, first tied the concept of measurement to the formal rules of addition. "Fundamental measurement" was deemed to occur only when the empirical operations could be mirrored in the mathematical laws of additivity. Later, in 1932, the British Association for the Advancement of Science appointed a distinguished committee to investigate this same measurement issue and its implications for psychology. But after seven years the committee had still not reached a consensus and, as Reese (1943) observed, it is clear from their final report that dispute over the additivity axiom presented the major stumbling block: "any law purporting to express a quantitative relation between sensation intensity and stimulus intensity [i.e., the psychophysical law] is not merely false but is in fact meaningless unless and until a meaning can be given to the concepts of addition as applied to sensation" (Final Report, 1940, p. 345). Nevertheless, the report concludes on a more optimistic note: "some members, perhaps all, admit that their opinion might change if new facts were established" (p. 334).

The aim of this Thesis is to present some new facts and thereby cast further light on this "additivity criterion" which continues to haunt psychophysics. However, instead of agonizing over the logical requirements of measurement, as did the British committee, the Thesis takes the more functional approach of seeking to resolve the nature of the psychophysical law. Apart from its bearing on the measurement problem, resolution of the psychophysical law is of substantial practical importance and would, according to Gescheider (1976), "have far reaching implications for scholars in a variety of fields" (p. 125).

A history of psychophysics is essentially a history of two competing laws: the logarithmic (log) law and the power law (Warren, 1981 presents a useful historical perspective). In reviewing the historical development of these two laws, Stevens (1975, p. 2) notes that evidence of the conflict dates from 1738, which means that the debate is much older than psychology itself. Briefly, the log law of Fechner stood unchallenged for almost a century before it was discredited, whereupon the power law (Stevens, 1957) came to prominence. It appears, however, that the power law, like Fechner's law, is not to stand the test of time, since it has failed to comply with rigorous validity criteria (e.g., Anderson, 1970, 1972, 1975).

This failure has left contemporary psychophysics without direction. It is one matter to demonstrate that a law is invalid; it is another to suggest a viable alternative. Stevens (1957) remarked: "the lesson of history is that a bold and plausible theory that fills a scientific need is seldom broken by the impact of contrary facts and arguments. Only with an alternative theory can we hope to displace a defective one" (p. 153). The main aim of this Thesis is to propose such an alternative theory. It should be stressed at the outset, however, that no new concepts will be invoked: the theory proposed is merely a rearrangement of already existing ideas.

The following dissertation will not be concerned with philosophical objections to psychophysics (e.g., Savage, 1970). The author must confess to sharing S.S. Stevens' disdain for these arguments: "such polemics concern meaning more than substance ... and the scientist finds it thin to try to nourish his understanding by the ingestion of semantic disputation" (Stevens, 1975, p. 58). Indeed, this view is confirmed in a recent article (Warren, 1981) and its accompanying peer review. Questions such as "does sensation exist?" are metaphysical, and beyond scientific resolution. So, in terms of broad philosophical stance, this Thesis is similar to the later, and much less operational, position taken by Stevens (e.g., 1971, 1975), in which human subjects are considered to be capable of psychological measurement. A mediating sensation scale is an integral part of the process and concepts such as "subjective JND" are seen to be entirely legitimate. Baird and Noma (1978, p. 95) describe such a stance as "Subjectivism".

Furthermore, the author shares the view held by both Fechner and Stevens that, for a given modality, there is a single correct scale of sensation. This is an appealing view and in accord with the law of parsimony, viz., that it is the aim of science to present the laws of nature in the simplest and most economical conceptual formulations. Note, though, such a stand does not necessarily presume that this "single scale of sensation" conforms to a simple mathematical formulation (cf. Falmagne, 1974, p. 129).

Finally, there is one aspect of the philosophy of science which <u>is</u> of central importance to this Thesis - fact and theory, or, as the introductory allegory would have it, substance and shadow. The overriding guideline in the development of the following work was that strict regard be paid to the raw empirical evidence in psychophysics. Bronowski wrote of science: "it has made its way not secretly but by sticking to the plain facts - never mind who discovered them - who challenges them" (1948/1977, p. 4).

2

THE POWER LAW AND VALIDITY CRITERIA

S.S. Stevens first proposed the power law, somewhat tentatively, in 1953; a more comprehensive statement followed a few years later (Stevens, 1957). In its simplest form the power law states that $S=kI^n$, where S is sensation (or subjective) magnitude, I is stimulus intensity, n is an exponent whose value is modality dependent, and k is a constant. Thus, the basic principle underlying the power law is that "equal stimulus ratios produce equal sensation ratios" (Stevens, 1975, p. 36).

Certainly there is a good prima facie case for a power law of sensation, and even the most cursory inspection of the psychophysical literature will bear testimony to its popularity. Such popularity is understandable. When obtained by the response technique of magnitude estimation, the psychophysical functions of most continua approximate to straight lines in log-log plots. This property, together with the slope of the line reflecting the exponent value, is conceptually attractive, especially to the psychologists who crave greater scientific respectability for their discipline - a table of power function exponents bears some resemblance to a table of physical constants in physics. Accordingly, the power law rapidly became the status quo in psychophysics; only more recently has its credibility been threatened (e.g., Anderson, 1972).

Inadequate Validity Criteria

Perhaps the main reason that the status of the power law has remained intact is that stringent tests of validity have seldom been applied. Traditionally, data from magnitude scaling have been presented almost exclusively in the form of log-log plots, without error bars, and the goodness of fit assessed by visual inspection. Now, while it is true that many continua are well fitted by power functions in log-log plots (e.g., loudness, brightness), there are other continua for which the fit is poor (e.g., sweetness, Stevens, 1969; apparent intensity of electric shock, Stevens, Carton, & Shickman, 1958).

An example. Stevens et al. (1958) used magnitude estimation to scale the apparent intensity of electric shock. Median responses were plotted in log-log coordinates and from three experiments gave an average slope (exponent) of around 3.5, indicating that apparent intensity of electric shock is a strongly positively accelerating function of electric current. This exponent value is considerably larger than those for most other sensory continua and, according to Stevens et al., "suggests a rather basic difference in the physiological mechanisms involved" (p. 332). However, when the raw data from Stevens et al. are replotted in linear coordinates (without any transformation or correction) and the line of best fit is drawn by eye, it is apparent that the function is not positively accelerating at all.

Figure 1 shows that the magnitude estimates actually form a slightly sigmoid function of electric current, deviating markedly from the best fitting power function (broken line): evidence for a "basic difference" in physiological mechanisms has evaporated. Why the discrepancy? The blatant lack of agreement between the two curves suggests that, in this case at least, the psychophysical function is not genuinely a power function. If it were, then the broken line and solid line would more closely correspond.

Another example. In the sensory scaling of taste intensity, Stevens (1969) presents power functions for the two artificial sweeteners saccharin and Sucaryl (Sucaryl is a mixture of calcium cyclamate and calcium saccharin). Magnitude estimation gave .8 and 1.9 as the sweetness exponents for saccharin and Sucaryl respectively, suggesting that the sweetness of saccharin is a negatively accelerating function of concentration, while for Sucaryl, sweetness increases at a faster rate than does concentration. Taken at face value this finding might be seen to imply that two different sensory mechanisms are operating (cf. Teghtsoonian, 1971).

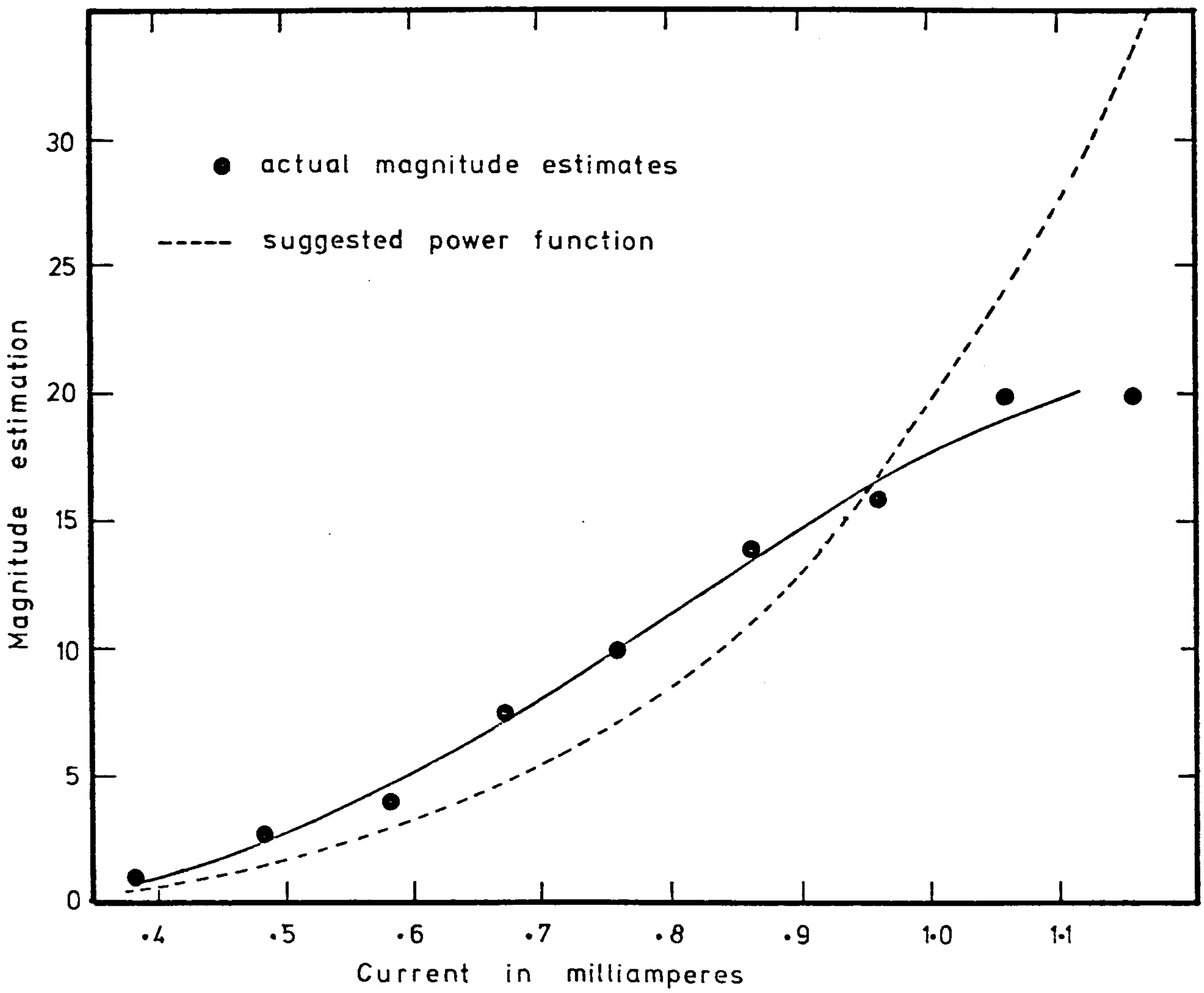

Figure 1. Magnitude estimation of the apparent intensity of electric shock (data from Stevens et al., 1958, Experiment I) replotted in linear coordinates. The filled points are the median magnitude estimates and the solid curve was fitted by eye. The best fitting power function departs markedly from the raw data.

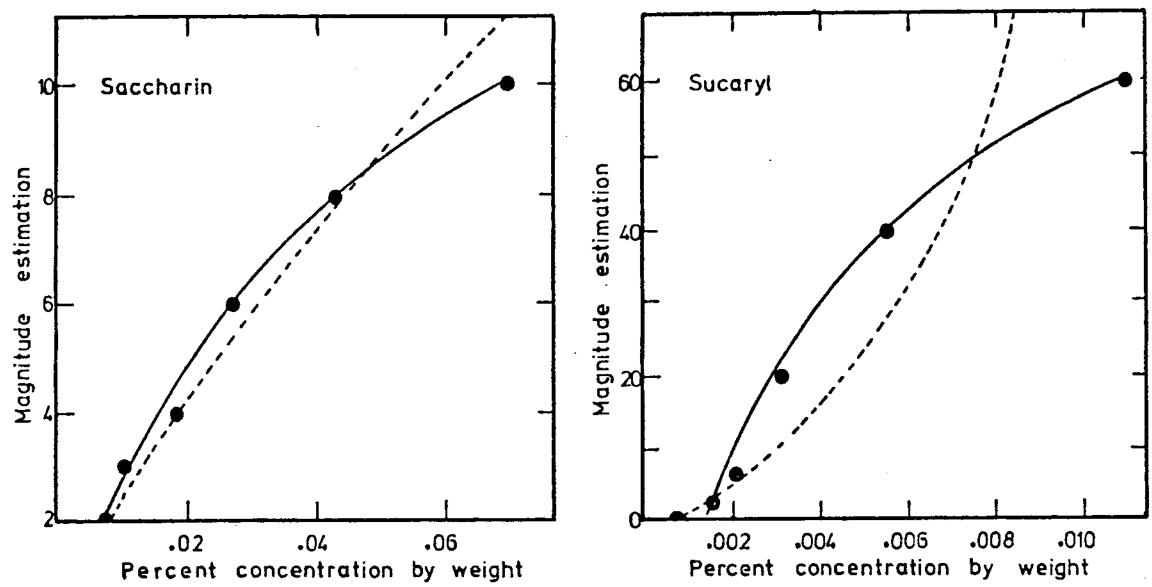

Figure 2. Magnitude estimates of the taste intensity of saccharin and Sucaryl (data from Stevens, 1969, Figures 4 & 6) replotted in linear coordinates. The median magnitude estimates in the two panels describe psychophysical functions of similar shape (solid curves fitted by eye), while the best fitting power functions (broken lines) specify distinctly different shapes.

But, when the raw data are replotted in linear coordinates, the basis for any such interpretation vanishes. Figure 2 shows the empirical data for saccharin and Sucaryl together with their proposed power functions: the lines of best fit have been drawn by eye, and the coordinates have been adjusted linearly to facilitate comparison of the shapes of the curves. In linear coordinates it becomes clear that the data curves are, in fact, similar in shape - both negatively accelerating. The positively accelerating power function for Sucaryl, specified by the exponent of 1.9, is markedly discrepant from the empirical data.

When a power function is fitted to data which fundamentally do not conform to a power function, the exponent can only reflect the ratio of log response range to log stimulus range (i.e., the exponent is the slope of the regression line fitted to log response and log stimulus values; cf. Poulton, 1967, 1979). This is demonstrated first in Figure 1, where the best fitting power function actually has an exponent of 2.7. Here the response range is moderate (1.3 log units) and the stimulus range small (.5 log units), hence the large exponent. In Figure 2 the two stimulus ranges are comparable (1.0 log units for saccharin, 1.2 log units for Sucaryl), but the difference between response ranges dictates the divergence between the best fitting power functions. It follows, then, that here the numerical value of the exponent cannot possibly have any fundamental significance, such as reflecting the nature of the sensory tranducers (Stevens, 1960; Teghtsoonian, 1971), since it does not even describe the raw data from which it is derived.

Even if a power function does provide a good description of the raw data, it does not necessarily follow that the data fundamentally (mathematically) conform to a power function. For example, Uttal (1973, Figure 6.11) showed that a polynomial and a trigonometric function, neither of which are power functions, both approximate to straight lines when plotted in log-log coordinates. Uttal concludes: "while all power functions do plot up as straight lines on log-log scales, all curves that are relatively good fits to straight lines on such scales do not necessarily represent power functions" (p. 267). The pertinence of this statement will be demonstrated again later, with reference to loudness measurement (Chapter 7). In arriving at a similar conclusion, Nihm (1976) suggests, somewhat facetiously, that one of the most attractive features of the power law is its capacity of fitting a large variety of curves; of providing a reasonable approximation to almost any data that might be obtained in psychophysical experiments. Anderson (1981, p. 341) and Weiss (1981) comment along similar lines.

Still on this point, the practice of using the correlation coefficient as an index of goodness of fit has further unjustifiably enhanced the face validity of the power law (cf. Anderson, 1977a). For instance, despite the obvious discrepancy in Figure 1 between the suggested power function for electric shock and the data from which it was derived, the fit is still reasonable (Pearson $\underline{r}$=.98).

So visual inspection of log-log plots, even when augmented by the correlation coefficient, does not constitute a rigorous validity criterion. Nor for that matter does cross-modality matching (CMM), a criterion much vaunted by Stevens (e.g., 1959). Treisman (1964a) clearly demonstrated that cross-modality matching does not uniquely validate the power law at all. In fact, it equally well "validates" Fechner's log law. This very important point will be dealt with further in the next Chapter. (Baird & Noma, p. 89 argue that the finding of transitivity in empirical CMM studies does go some way toward validating the power law; however, strictly speaking, as others besides Treisman have noted, e.g., Anderson, 1972, Teghtsoonian 1974, CMM cannot validate the power law.)

Toward Proper Validity Criteria

Torgerson (1958) appears to have been the first to stipulate stringent validity criteria in scaling. According to Torgerson: "we have no basis for concluding that the scale possesses the properties attributed to it from the data gathered in the scaling process itself" (p. 113). One of the criteria proposed by Torgerson is that of internal consistency, i.e., the scale generated should be independent of the particular ratios (or intervals) used in its construction. Apart from isolated studies which have used factorial stimulus presentation (e.g., Comrey, 1950), this criterion appears to have gone unheeded until the advent of functional measurement (Anderson, 1970).

The functional measurement approach, which relies on the use of factorial stimulus designs, has been described comprehensively elsewhere (Anderson, 1970, 1972, 1974a, 1974b, 1975, 1979a, 1979b, 1981). Functional measurement falls within the general framework of cognitive algebra and is not specifically concerned with psychophysics; nevertheless, it is of substantive value to psychophysical research because success of the algebraic integration model validates the response scale, and at the same time provides the valid psychophysical function. This must be seen as a breakthrough for traditionally univariate psychophysical measurement, which has for so long lacked a proper validational base. The question of response measures and the question of the psychophysical law are different issues, and each will be dealt with in turn.

There are two distinct types of response measure commonly used in sensory measurement. First, there are the ratio or magnitude scaling methods (e.g., magnitude estimation); second, there are the interval or partition methods (e.g., category rating, graphic rating). Anderson (1970, 1972, 1975) has frequently emphasized that functional measurement is essentially neutral in the controversy between these two types of response measure, and that both have equal opportunity to comply with the internal consistency criterion. But, whereas rating methods ordinarily satisfy this criterion, magnitude estimation usually fails. This finding contradicts the stand of Stevens (e.g., 1961, 1971, 1975), who claimed the rating methods are biased and nonlinear. This dispute can be resolved, however, by invoking the guideline proposed earlier, viz., that of attending strictly to the empirical evidence. Only Anderson's approach satisfies this requirement. In fact it appears that Stevens was not concerned with proper validity criteria: "which scale best measures what it is that we want to measure? Since it is that kind of question, the answer becomes a matter of opinion - a value judgment" (Stevens, 1971, p. 448).

Failure of magnitude estimation to meet the requirements of a linear response scale means that even with the sometimes obscurative effects of the log-log plot eliminated, as in Figures 1 and 2, the psychophysical function will not be valid. And, even when replotting data in linear coordinates does not reveal any discrepancy of the nature shown in Figures 1 and 2, the psychophysical function will nevertheless still not be valid.

Could Both Types of Response Scale Be Valid?

In an attempt to accomodate both types of scale, Marks (1974) proposed that magnitude estimation and ratings may both be valid, but that they tap different perceptual qualities. In this schema, magnitude estimates are seen as measures of sensory intensity, ratings as measures of sensory dissimilarity. But Anderson (1975) pointed out difficulties with this notion. Magnitude estimation does not usually comply with the simple adding or averaging models of functional measurement; therefore, if it is to remain a valid response measure, then the adding or averaging tasks themselves must somehow induce perception of sensory dissimilarity rather than sensory magnitude. This is an awkward corollary of the Marks proposal. In a task involving the addition of stimuli, it is difficult to see why sensory dissimilarity should be perceived rather than sensory magnitude. As Anderson (1975) reasoned, in at least some cases the perception of magnitude must surely be prerequisite to the perception of dissimilarity.

3

SUBJECTIVE NUMBER

This Chapter will be concerned with <u>why</u> the ratio response techniques, such as magnitude estimation , fail to comply with proper validity criteria.

It has been noted by many authors that the category scale is approximately the log of the magnitude scale (Baird, 1970; Baird & Noma,p. 86; Eisler, 1965; Galanter & Messick, 1961; Montgomery, 1975; Poulton, 1968; Stevens & Galanter, 1957; Torgerson, 1960, 1961). So, if the category scale is linear with sensation, as the functional measurement validity criterion suggests, then it follows that in magnitude scaling tasks it must be the logs of the responses, rather than the responses themselves, which are linearly related to sensation. In other words, the perceived, or subjective, magnitude of number in free number matching tasks is logarithmically related to actual (objective) number.

This statement may be disconcerting at first, but upon reflection it does fit in with our everyday use of the number system; of perceiving constant ratio (geometric) increments to be subjectively equal rather than constant distance (arithmetic) increments. For instance, the subjective jump between the numbers 10 and 11 is seen as more nearly equivalent to the jump between the numbers 100 and 110, than it is to the jump between 100 and 101. Attneave (1962) invokes a similar argument; Anderson (1974a) describes it as a "Weber law for numbers"; Banks and Hill (1974) claim that the perceived magnitude of numbers is approximately logarithmic from 10 to 1000; and Poulton (1979) talks of a "logarithmic bias". It should be noted that some investigators (e.g., Attneave, 1962) have used power functions with exponent <1 to describe subjective number. While power functions may provide a good description of the data (cf. Chapter 2), they lack the more plausible rationale of the logarithmic relationship.

The above reasoning suggests that the catchphrase of magnitude scaling, "equal stimulus ratios produce equal sensation ratios", is not justified. Equal stimulus ratios may produce equal response ratios, but these responses are not linearly related to the underlying sensation. Besides functional measurement, there are a number of approaches which lend support to this interpretation.

Apparent Length

Because of its special properties, it is difficult to investigate the number continuum as a psychophysical continuum directly. When asked to assess the magnitude of numbers most subjects will, almost involuntarily, perform arithmetic calculations, thereby precluding the true assessment of subjective magnitude.

One way around this problem is to investigate another continuum which free number matching studies have shown to be linearly related to the number continuum. Apparent length is such a continuum (Stevens & Galanter, 1957; Stevens & Guirao, 1963; Teghtsoonian & Teghtsoonian, 1970). On this point, proponents of magnitude scaling (e.g., Marks, 1974; Stevens, 1975, p. 109) have taken the linear relationship between apparent length and magnitude estimation responses as providing confirmation of the linearity of subjective number. But this is a non sequitur: a linear relation between magnitude estimates and apparent length does not necessarily imply that both of these psychophysical continua are linearly related to their respective physical correlates. It simply means that, as psychophysical continua, they vary in the same manner (in power funtion parlance they have the same exponent, but this exponent is not necessarily equal to 1.0).

Many of the studies which have not used magnitude scaling show apparent length to be a negatively accelerating, approximately logarithmic function of actual length. Parker, Schneider, and Kanow (1975) found this to be so in a thorough nonmetric study, although these authors actually fitted power functions and obtained an exponent of .5. Also, although Anderson (1977b) found that his data failed the additivity test in a bisection of length task, when the data were transformed to be as additive as possible (Anderson's Figure 4), the apparent length function turned out to be negatively accelerating, reasonably consistent with a logarithmic function for apparent length. Similarly, Krueger (1970), who used a "scale-free" matching technique, suggested a negatively accelerating function for apparent length.

Moreover, category rating studies support a logarithmic function for apparent length (Eisler, 1963; Stevens & Galanter, 1957). A few studies do not (e.g., Stevens & Guirao, 1963) but, as Stevens and Guirao themselves point out, this is almost certainly the result of unintentionally providing subjects with a frame of reference and thereby allowing judgment of apparent position rather than apparent length. Eisler (1963) also makes this point.

Thus, a substantial amount of empirical work suggests that apparent length is a logarithmic function of actual length. Since apparent length and magnitude estimates are linearly related it follows that, in magnitude scaling, the subjective magnitude of the responses must be a logarithmic function of the actual responses.

The Two-Stage Model of Magnitude Judgment

This model was proposed by Curtis, Attneave, and Harrington (1968), following earlier work by Attneave (1962). It is structured in terms of a sensory input transformation ("perception") and an output transformation ("response"). Thus, the exponent obtained in scaling by magnitude estimation is seen to comprise the product of a sensory input exponent, typical of the modality scaled, and an output exponent whose value depends on the way in which people use numbers.

The general outcome of a number of investigations which have employed the two-stage model (e.g., Curtis, 1970; Curtis et al., 1968; Curtis & Fox, 1969) is that, when category rating is used as the response measure, the output transformation is linear, whereas with magnitude estimation the output transformation is nonlinear, suggesting a negatively accelerating function for subjective number. So, although the two-stage model is framed in terms of power function exponents, in general terms its findings are consistent with a logarithmic function for subjective number.

Other Approaches

In an extensive study, Banks and Hill (1974) investigated subjective number by having subjects generate random numbers. They found that, for numbers between 1 and 1000, the subjective number function could be equally well described by a power function with exponent of approximately 2/3, or by a partly linear, partly logarithmic function (the A-L function). The linear section of the A-L function applies only to the numbers less than 10. As noted earlier, this Thesis takes the view that the mainly logarithmic function is more plausible theoretically.

In their work on "ratios" versus "differences", Birnbaum and co-workers (e.g., Birnbaum, 1980; Birnbaum & Elmasian, 1977; Birnbaum & Veit, 1974) have repeatedly found scales derived by "ratio" judgment to be logarithmically related to scales obtained by "difference" judgments. Birnbaum (e.g., 1980) claims that "ratio" judgments may be exponentiated differences. This contrasts with Rule and Curtis (1980), who propose that "ratios" are power functions of subjective ratios. The present approach has elements of both approaches, viz., that reported "ratios" are exponentiated subjective ratios. As Rule and Curtis (1980) note, at present the evidence is equivocal (for example, close inspection of Figure 4 in Birnbaum, 1980 shows a few instances of nonmonotonicity not predicted by the Birnbaum model), and there is a need for further experiments using a much wider range of subjective values.

Various other attempts have been made to investigate number as a psychophysical continuum. Some (e.g., Attneave, 1962;

Birnbaum, 1974; Ekman & Hosman, 1965) have required that subjects directly estimate the magnitude of numbers; others have used conjoint measurement (e.g., Rule & Curtis, 1973). Information about subjective number has also been derived indirectly from comparison of ratio and interval scales (e.g., Garner, 1954). All of this work is reasonably consistent with a logarithmic function for subjective number.

However, Anderson (1972) noted some important exceptions to this rule. On those continua termed metathetic (Stevens & Galanter, 1957), the scales derived by magnitude estimation and category rating are linearly related. Here at least, the subjective number function must be linear, and these exceptions will be discussed in some detail later in Chapter 6.

Implications of Subjective Number

Recognition of the nonlinearity of subjective number has far reaching consequences. If the number continuum as used in magnitude estimation approximately obeys Weber's law, then the wealth of support for magnitude scales as power functions is to be expected. In these cases, the number continuum is no different from any other physical continuum used in cross-modality matching; and, if both continua approximate to Weber's law, they will give a straight line in a log-log plot. It does not follow, however, that the psychophysical law is a power function, because sensation is not linearly related to the magnitude estimation responses.

Recognition of subjective number offers a solution to the time-honoured discrepancy between category scales and magnitude scales. The concave downward curve which results when the category scale is plotted against the magnitude scale, described as "one of the most reliable findings in experimental psychology" (Engen, 1971, p. 82), all but disappears when the magnitude scale is transformed logarithmically to correct for subjective number.

This is demonstrated in Figure 3, which presents data from the category scaling (Schutz & Pilgrim, 1957a) and magnitude estimation (Moskowitz, 1970) of the sweetness of sucrose over the concentration range .06-.50 Molar. The left hand panel illustrates the typical concave downward relation between the two scales. In the right hand panel the magnitude estimates have been plotted on a log scale, and the broken line represents the relationship that would obtain if subjective number were perfectly logarithmic. Clearly, the log transformation is a reasonable approximation, but if anything it is a little too severe: the deviation of the empirical curve from the log function suggests there is still a tendency for the number continuum to be perceived as linear (cf. the A-L function of Banks & Hill; see also Figure 4 of Poulton, 1979).

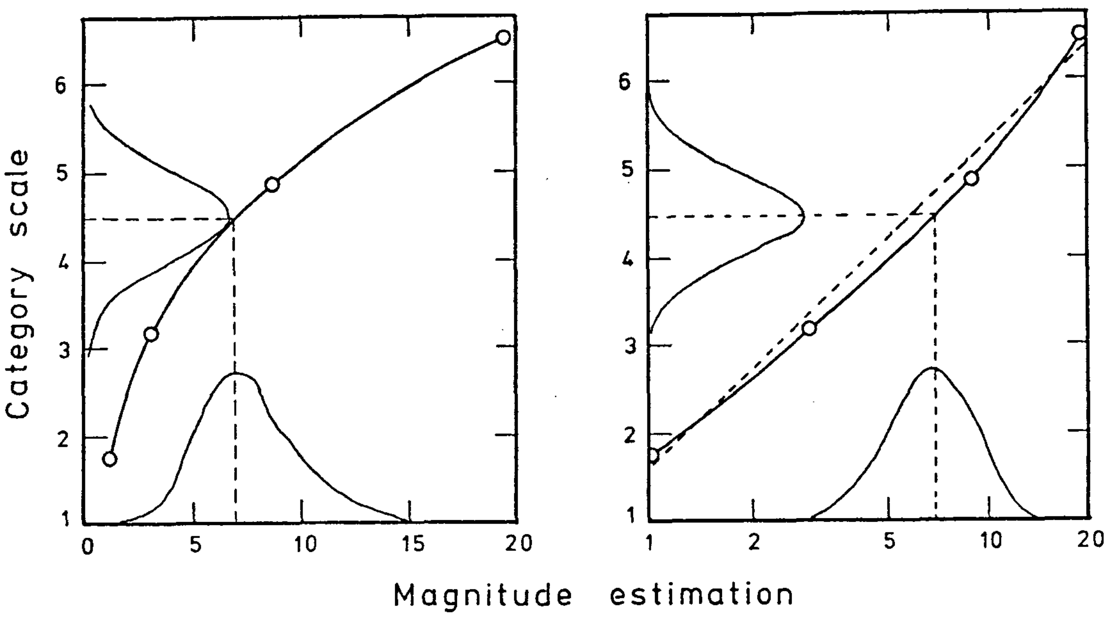

Figure 3. Category ratings (data from Schutz & Pilgrim, 1957a) vs. magnitude estimates (data from Moskowitz, 1970) of the sweetness of sucrose. The coordinates in the left and right hand panels are linear and semilog, respectively. The broken line in the right hand panel specifies a log relationship between the two scales. Note that the log-normal distribution of magnitude estimates becomes normal after a log transformation.

This example was taken from the taste modality, but it does not matter whether the continuum in question is loudness, brightness, the strength of attitudes, or the roughness of sandpaper - the log transform will approximately restore linearity.

But there is another implication which is fundamentally even more important. Not only does a log transformation bring the median magnitude estimates into line with the mean category responses: as Eisler (1965) showed, it does the same for the variability. S.S. Stevens often noted that, in log units, the variability of magnitude estimates is approximately constant over the entire stimulus range (e.g., Stevens, 1969). It therefore follows that, after a log transformation for subjective number, the variability of the transformed magnitude estimates will be constant, regardless of subjective magnitude. This is illustrated schematically in Figure 3.

Recognition of nonlinear subjective number in magnitude scaling offers resolution of the dispute between response scales, but how does this affect the psychophysical law? As noted previously, functional measurement can provide valid psychophysical functions, but these functions do not appear to adhere to either a log law or a power law (e.g., the loudness scale of Carterette & Anderson, 1979). This indeterminacy suggests a re-evaluation of the foundations, the fact and theory of psychophysics.

4

THE PSYCHOPHYSICAL LAW

Treisman (1964a) showed that both the log law and power law consist of two separate theoretical premises. The first premise, known as Weber's law, relates to the stimulus (physical) continuum. The second premise relates to the subjective (psychological) continuum. Fechner's law and Stevens' law are based on the same first premise, but different formulations of the second premise dictate a log law on the one hand and a power law on the other. The theoretical structure of these two laws is summarized in Figure 4. (These laws correspond to cases II and IV respectively in the "fourfold way" of Baird & Noma, p. 60, who present algebraic derivation; see also Laming, 1973, Chapter 3).

While it is not uncommon to see Fechner's log law derived from first principles (Guilford, 1954, p. 38; Torgerson, 1958, p. 149), only seldom (Baird & Noma, p. 62) is the power law similarly derived. Few psychophysicists (e.g., Teghtsoonian, 1974) seem to have recognized the theoretical dependence of the power law on Weber's law, and this may explain why, despite the considerable effort that has gone into the study of the psychophysical law, relatively little attention has been directed at Weber's law.

Theoretical Foundations of Psychophysics

(1) Weber's law $\frac{\Delta I}{I} = k_1$

(2) Fechner's assumption $\Delta S = k_2$

(3) Brentano's assumption $\frac{\Delta S}{S} = k_3$

(1) and (2) give

Fechner's law $S = k_4 \log I$

(1) and (3) give

Stevens' law $S = k_5 I^n$

Figure 4. Theoretical structure of the psychophysical laws.

The First Premise - Weber's Law

Traditionally defined, Weber's law asserts that the size of the just noticeable difference (JND or ΔI) is proportional to stimulus intensity (I), i.e., $\Delta I = kI$. This type of definition is a legacy of classical psychophysics. Alternatively, the JND may be viewed as just another measure of sensitivity, in which case Weber's law would state that sensitivity (or error) is relative on the stimulus continuum ("The 'Objective' View"; Torgerson, 1958, p. 135).

At this point it is worth mentioning that the JND has been variously interpreted. Stevens often castigated Fechner's use of the JND as a unit of measurement because he considered it to be "Unitizing Error" (Stevens, 1961, p.80). This is certainly true. The JND is estimated by a confusability paradigm and can, therefore, be conceptualized as a measure of error, noise, dispersion, or variability. But it is also true that in the confusability paradigm error is the reciprocal of sensitivity; a small JND indicates good sensitivity. When the JND is regarded as a measure of sensitivity, Stevens' argument loses impact. Indeed, a scale based on sensitivity seems intuitively reasonable.

Weber's law is usually regarded as a good first order approximation (Guilford, 1954, p. 24; Stevens, 1957; Teghtsoonian, 1974), but this is a somewhat generous view. Reservations about Weber's law began with Fechner himself:

"I must confess finding, after a preliminary survey, that we are far from achieving a thorough verification, not to say proof of this law" (Fechner, 1860/1966, p. 114). Cobb (1932) noted that Weber's law is the exception rather than the rule, and Treisman (1964b) stated "the Weber functions (the relations between ΔI and I) which are determined experimentally for the different modalities are not well described by Weber's law" (p. 314). Many other authors echo the same sentiments (e.g., Christman, 1971, p. 41; Corso, 1967, p. 272, 284, 294; Hecht, 1924; Mueller, 1975, p. 16).

Closer inspection of empirical data does suggest a certain invariance, not so much in the Weber fraction but in the shape of the Weber function. Holway and Pratt (1936) demonstrated that Weber functions typical of brightness, loudness, somesthesis, taste and olfaction have a characteristic shape: as stimulus intensity increases, the Weber fraction decreases, approaching some minimal value. Volkmann (1974) observed: "Weber-law plots have usually looked to me like reciprocal functions, having in common with them an initial sharp drop and a long approach to some asymptote" (p. 177).

Figure 5 illustrates three Weber functions. Curves 1 and 3 are hypothetical, curve 2 is empirical. Curve 1, a straight line parallel to the abscissa, occurs when sensitivity on the stimulus continuum is <u>relative</u>, i.e., Weber's law holds and $\Delta I/I = K$. In contrast, Curve 3 is a reciprocal function, obtained when sensitivity on the stimulus continuum is absolute (constant), i.e., $\Delta I = K'$. Curve 2 is the empirically determined Weber function for the sweetness intensity of sucrose over the concentration range .015 - .250 Molar (data from Schutz & Pilgrim, 1957b).

The empirical Weber law plot lies between curves 1 and 3. The sharp initial drop in curve 2 suggests that, at low stimulus intensities, sensitivity may be absolute rather than relative. However, as stimulus intensity increases, sensitivity tends to become more relative. This relativity is shortlived on some continua (e.g., pitch), because the Weber fraction increases again at high stimulus intensities. Evidence for Weber's "law" in these instances is tenuous indeed.

In recognizing that empirical Weber functions deviate from Weber's law, especially at low stimulus intensities, many workers (Engen, 1971, p. 18; Guilford, 1954, p. 40; Hecht, 1924; Miller, 1947; Treisman, 1964b) have invoked a correction factor which, they claim, accounts for the "sensory noise" near threshold. This has led to the generalized form of Weber's law: $\Delta I = kI + c$, where k and c are constants. But it is not clear just what purpose is served by this correction factor, or for that matter, why the deviation should necessarily be interpreted as "sensory noise". A correction factor may bring the empirical data

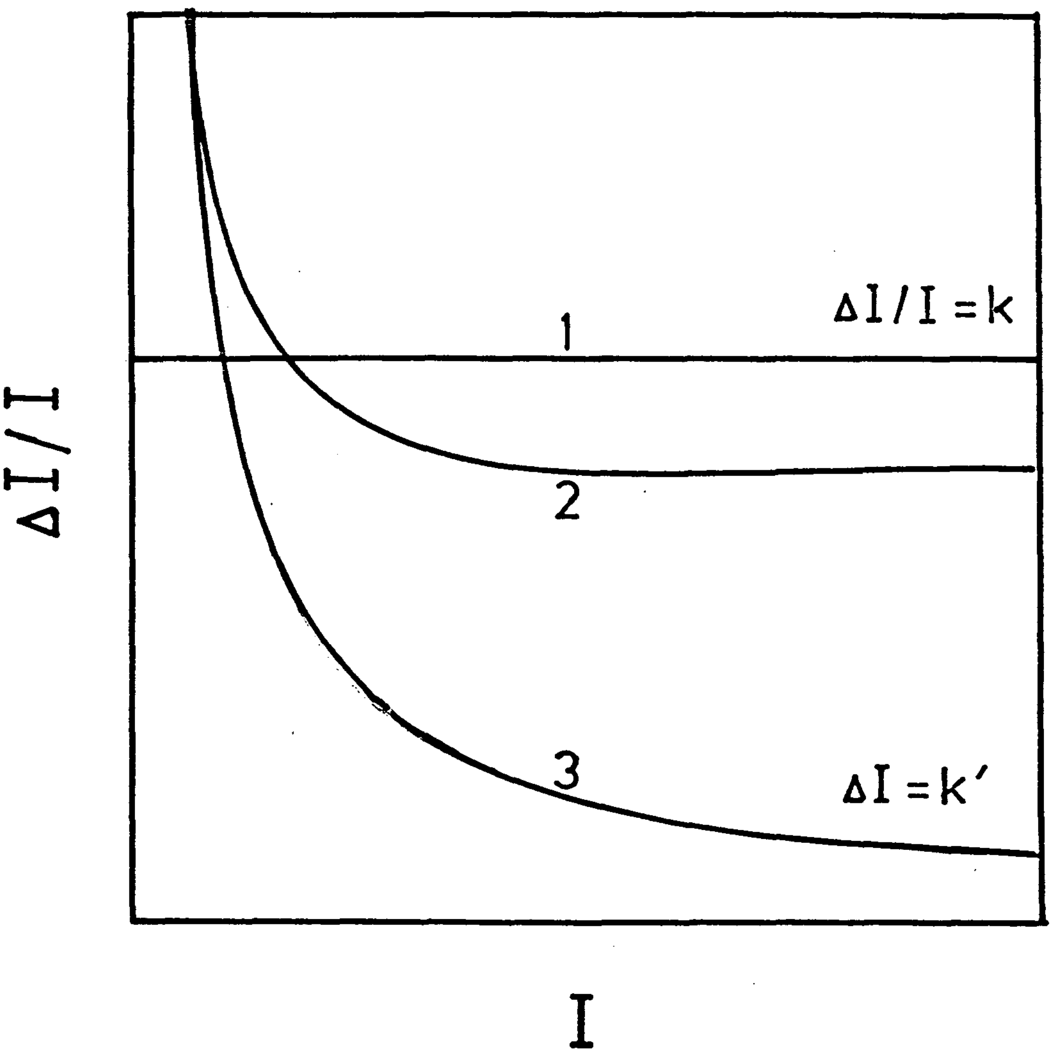

Figure 5. Three different Weber functions. Curve 1: Weber's law; sensitivity on the stimulus continuum is relative. Curve 2: a typical empirical Weber function. Curve 3: the result when sensitivity on the stimulus continuum is absolute (constant) rather than relative.

(fact) and Weber's law (theory) into closer agreement, but in so doing it obscures an important point. The sharp drop in the Weber fraction at low stimulus intensities occurs systematically on all sensory continua; it is not a random perturbation. Furthermore, as noted above, this deviation from Weber's law can have psychophysical significance. It will be argued later that the empirical Weber function alone uniquely determines the form of the psychophysical function for a given modality, therefore it should not be arbitrarily modified.

The Second Premise - Subjective Size of the JND

The subjective size of the JND is a "stubborn and vexatious problem" (Stevens, 1957, p. 172) of central importance to psychophysics, since it is this issue on which Fechner and Stevens diverge. Fechner assumed all JNDs to be subjectively equal, whereas Stevens adopted the earlier assumption of Brentano (see Stevens, 1961) that the subjective size of the JND is proportional to the subjective magnitude (Brentano's assumption is also known as Ekman's law; Stevens, 1966). Weber's law and Fechner's assumption together specify a log law; Weber's law and Brentano's assumption together specify a power law (see Figure 4).

Fechner's Assumption

Fechner's assumption has been further developed by Thurstone (1927a, 1927b, 1927c, 1927d, 1927e, 1932). Thurstone's models are well summarized elsewhere (Torgerson, 1958, chap. 8). Briefly, by beginning with the assumption that variability on the subjective continuum is normally distributed (an assumption ultimately verified by Luce, 1977), a measure of such variability produced by a given stimulus (the discriminal dispersion) can be determined empirically.

Thurstone (1927c) was careful to point out that the subjective continuum is qualitative and independent of the stimulus continuum; that the discriminal dispersion is not in any way a "subjective analog" of the JND. Rather, variability on the subjective continuum is seen to reflect the amount of psychological ambiguity generated by a stimulus. Thurstone (1927a, 1927d, 1932) repeatedly stressed that within a homogeneous stimulus series, such as a psychophysical continuum, stimuli do not vary with respect to ambiguity, therefore variability should be constant across the subjective continuum (i.e., Fechner's assumption should hold; Thurstone's Case V).

Using a Thurstonian paired comparison analysis of lifted weights, Guilford (1954, p. 157) reported response variability to be almost constant over the weight range tested. Similarly, in a successive intervals analysis of loudness, Galanter and Messick (1961) found that, apart from some deviation at low stimulus intensity, response variability was nearly constant, consistent with Fechner's assumption.

With all this said, however, it must now be pointed out that Eisler (1965) showed there is a monotonic indeterminacy in Thurstonian analysis; that in Thurstonian tasks subjects generate only ordinal level data, and exactly how these data are to be used is a decision made by the experimenter, not the subject.

What evidence there is available from category rating studies tends to support Fechner's assumption. Unfortunately, despite the widespread use of category rating, the response variability is seldom even measured, let alone reported. In a study of apparent length Eisler (1963) found that, except at the extremes of the range, the variability of the category ratings was roughly constant. Likewise, in the category scaling of the sweetness of sucrose, Moskowitz and Vaisy Genser (1977, p. 35) showed that the variability in ratings was approximately constant, regardless of stimulus concentration.

Brentano's Assumption

It should first be pointed out that much of the so called evidence for Brentano's assumption (Ekman, 1956, 1959; Harper & Stevens, 1948; Stevens, 1936, 1957) is not legitimate because it is derived a posteriori from the power law (Garner, 1958; Teghtsoonian, 1974). Brentano's assumption is mathematically a necessary component of the power law, therefore any "evidence" so derived must be tautological and consequently inadmissible. How could this faulty reasoning have been overlooked? An early paper (Stevens, 1936) suggests that Stevens was, a priori, much enamoured of ratios and ratio scaling, and this enchantment may have blinded him to the circularity of the argument. Also, since the basis for the power law was empirical, not theoretical (Stevens, 1975, p. 19), investigators at the developmental stage may not have been fully aware of the theoretical role of Brentano's assumption.

Still, other evidence is claimed. If Brentano's assumption is to hold, then response variability should be proportional to subjective magnitude. In the scaling of loudness, Eisler (1962) found such a linear relation between magnitude estimates and their standard deviations. Also, inspection of the log-log plots in which magnitude scaling studies are usually summarized reveals that the error bars are approximately constant in length, in agreement with Brentano's assumption (e.g., see Stevens, 1969; Stevens, 1975, Figures 10 & 101).

But this evidence cannot be taken at face value. As noted by Teghtsoonian (1974), it rests upon the assumption that magnitude estimates are linear with sensation, an assumption shown earlier in Chapter 3 to be unjustified. After magnitude estimates have been transformed logarithmically to correct for subjective number, their error behaviour becomes consistent with Fechner's assumption. In other words, the variability of raw magnitude estimates does not reflect fundamental error behaviour on the subjective continuum; rather, it results from the use of a nonlinear matching continuum. This reasoning suggests that Fechner's assumption is valid but Brentano's assumption is not. Contrary to common belief (e.g., Torgerson, 1961), this conclusion can be supported empirically.

5

FECHNER'S ASSUMPTION AND THE JND SCALE

The study of loudness occupies a special place in the history of psychophysics. Stevens (1957) claimed it was failure of Fechner's law to account for loudness data that provided the necessary impetus for development of a power law. Not long after the decibel (dB) scale was adopted, it was noticed that "equal steps on the logarithmic (decibel) scale do not behave like equal steps, for a level 50 dB above threshold does not sound at all like half of 100 dB as Fechner's law implies it should" (Stevens, 1957, p. 163). Stevens deduced that the failure of Fechner's law must be due to failure of Fechner's assumption that JNDs are subjectively equal. Invoking Brentano's assumption, Stevens reasoned that the loudness of a 50 dB sound is less than half that of a 100 dB sound because the subjective size of the JNDs in the 0-50 dB interval is less than the subjective size of the JNDs in the 50-100 dB interval. But there is a simpler explanation: <u>failure of Fechner's law may be due to failure of Weber's law</u>, and not to a breakdown of Fechner's assumption.

Figure 6 shows two versions of the Weber function for the loudness of a 1000 Hertz (Hz) tone (Jesteadt, Wier, & Green,

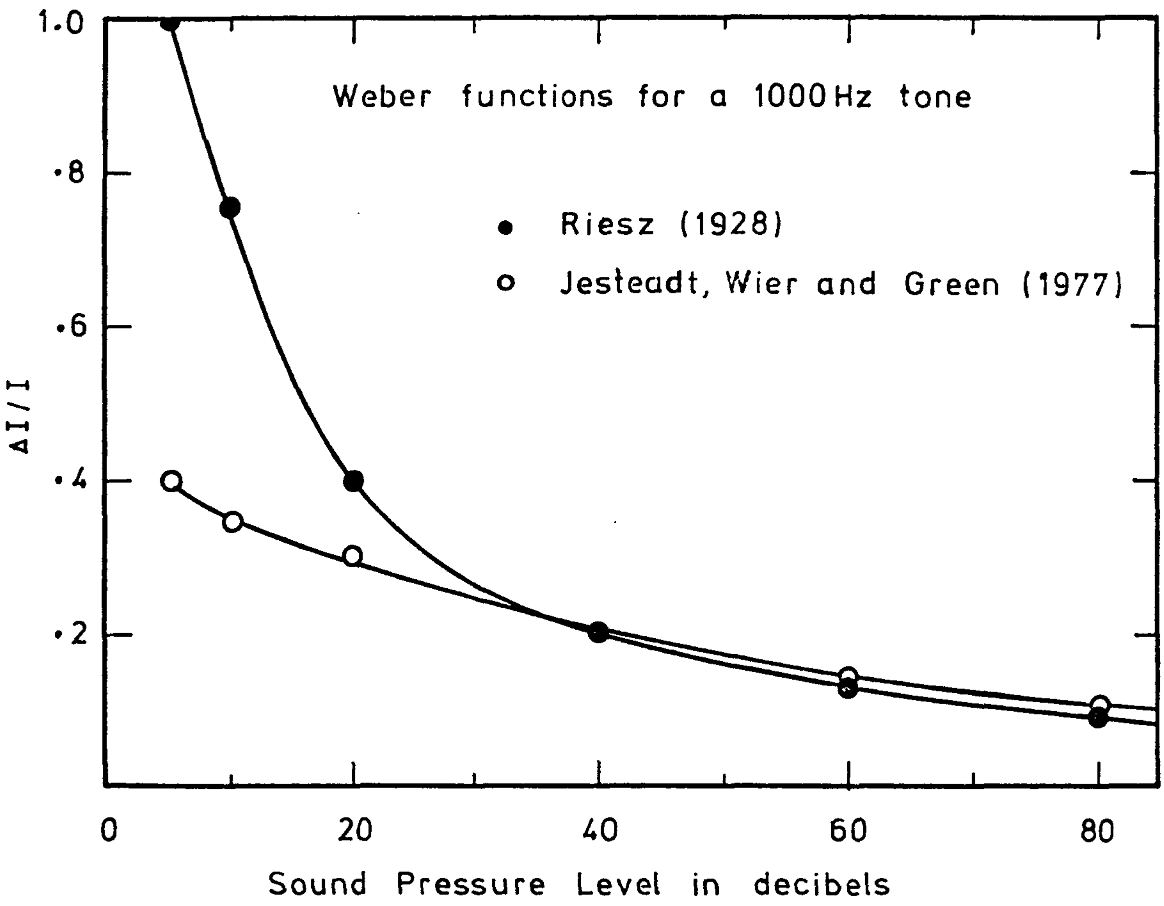

Figure 6. Weber functions for the loudness of a 1000 Hz tone over the range 0 - 80 dB SPL.

1977, Figure 4; Riesz, 1928, Figure 4). Except for the deviation near threshold the curves are in good agreement, especially considering they were obtained by different experimental methods. It is evident from Figure 6 that the Weber fraction decreases monotonically with increasing stimulus intensity. From a cursory inspection of the plot it may be tempting to conclude that this decrease is insignificant; that after an initial drop at low stimulus intensity (cf. curve 2 in Figure 5) the Weber fraction is approximately constant. Indeed, this is what most psychophysicists have done (e.g., Baird & Noma, p. 53) implicitly, if not explicitly. (The Baird & Noma Figure 4.2 replot of the Jesteadt et al. results does not closely follow the original data.) But this "constancy" is deceptive. An apparently small change in the Weber fraction can substantially affect the size of the JND.

The Weber fraction 25 dB above threshold is approximately .3, whereas the Weber fraction 75 dB above threshold is closer to .1. Thus, the size of the JND (in decibels) 25 dB above threshold is approximately three times as large as the size of the JND 75 dB above threshold. Since 25 dB and 75 dB are the midpoints of the 0-50 dB and 50-100 dB ranges respectively, it follows that there are approximately three times as many JNDs in the 50-100 dB interval as in the 0-50 dB interval. So, the loudness at 50 dB above threshold sounds less than half that at 100 dB above threshold <u>because</u>

the 0-50 dB interval contains fewer JNDs, not because the JNDs in the 0-50 dB interval are subjectively smaller. In other words, instead of accounting for the discrepancy in terms of apparent variation in the subjective size of the JND, as Stevens proposed, the discrepancy can be accounted for in terms of real, observable variation in the size of the JND in stimulus units - there is no need to revoke Fechner's simple assumption that JNDs are subjectively equal. Arguing from a quite different angle, Parker and Schneider (1980) reach a similar conclusion: "Fechner's mistake may not have been in his assumption, but rather in his acceptance of Weber's law" (p. 404).

Empirical support for this argument is presented in greater detail in Chapter 7: at this point it is important to note the ramifications of accepting Fechner's assumption.

The JND Scale Resurrected

An empirically determined Weber function and Fechner's assumption that JNDs are subjectively equal together specify the JND scale, sometimes referred to as the DL scale (Engen, 1971, p. 49; Gescheider, 1976, p. 92). Despite its long history in psychophysics, this method has been used only sporadically in scaling (Hardy, Wolff, & Goodell, 1947; Riesz, 1933; Troland, 1930/1969, p. 77, 215, 220), probably because it lacks general theoretical appeal. Anderson (1975) remarked: "Fechner's method of cumulating just noticeable differences would probably have attracted minor interest except that Weber's empirical law led to a theoretical logarithmic function" (p. 479).

JND scales do lack a single, simple mathematical form; nevertheless, their shapes reflect a pattern which corresponds to the Weber functions in Figure 5. Curves 1, 2, and 3 in Figure 7 represent the JND scales which result from cumulating in the Fechnerian tradition the correspondingly numbered Weber functions in Figure 5. In panel A the coordinates are linear, while in panel B they are semilogarithmic. There has been some concern about the status of this "cumulation"; whether or not it constitutes integration (e.g., Eisler, 1963; Luce & Edwards, 1958). However, as pointed out by Baird and Noma (p. 59), to the experimentalist this debate is of no practical consequence. JND scales can always be obtained graphically regardless of their status mathematically.

Curve 1 is the result of cumulating the Weber function when sensitivity measured in stimulus units is relative, i.e., Weber's law holds. This condition specifies a distinctly concave downward ("logarithmic") curve in panel A, but a straight line (Fechner's law) in panel B. Curve 3 results from cumulating the Weber function when absolute sensitivity measured in stimulus units is constant. Obversely, this gives a straight line in panel A, but a decidedly concave

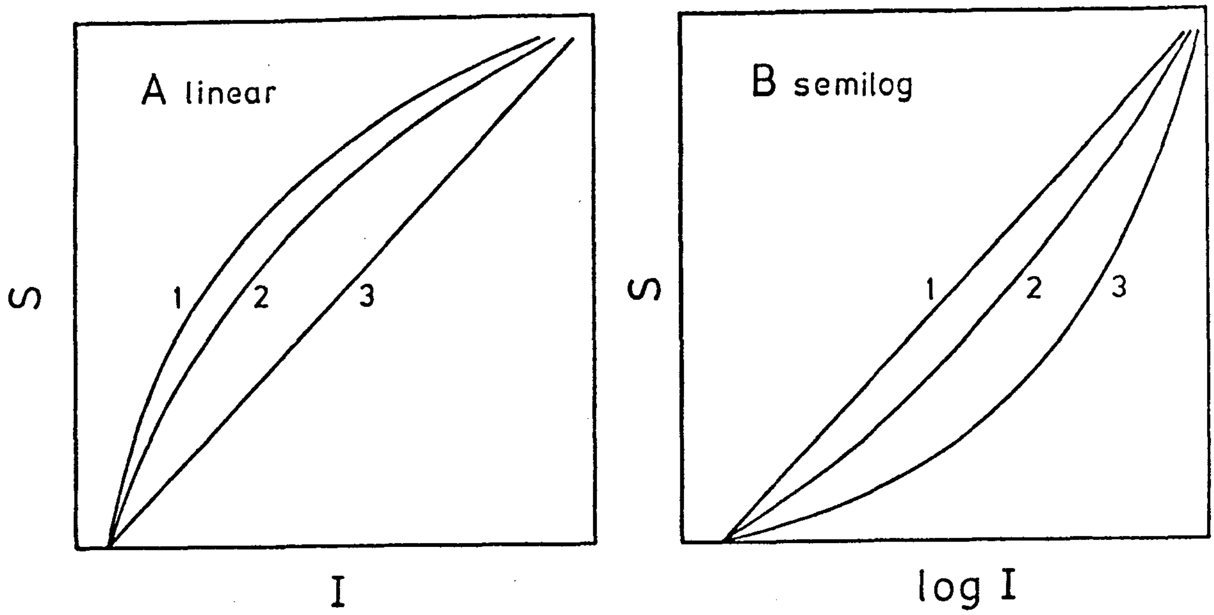

Figure 7. Psychophysical functions that result from the Fechnerian cumulation of the correspondingly numbered Weber functions in Figure 5. Coordinates in panel A are linear, in panel B semilog; S is subjective magnitude, I stimulus intensity. Curve 1: the psychophysical function obtained when discrimination on the stimulus continuum is relative (i.e., Weber's law holds); this is recognizable as Fechner's law in panel B. Curve 2: a typical empirical psychophysical function. Curve 3: the psychophysical function obtained when discrimination on the stimulus continuum is absolute (constant). Note that curve 2 lies between curves 1 and 3.

upward curve in panel B. (Curves 1 & 3 correspond to cases II and I respectively in Baird & Noma, p. 60.) Curve 2, the empirical function, lies between curves 1 and 3 as it did in Figure 5. While the exact shape of curve 2 will vary from one sensory continuum to the next, the present theoretical model dictates that all psychophysical functions may be conceptualized as lying between curves 1 and 3 in Figure 7. For example, if a modality has a Weber function which approximates well to Weber's law, then its psychophysical function would approximate to curve 1. On the other hand, if the Weber ratio decreases away from threshold, as happens on most continua, the psychophysical function would resemble curve 2. If, in yet another case, the Weber ratio first decreases then increases again at high stimulus intensity (e.g., pitch; Harris, 1952), the psychophysical function would flatten out at high stimulus intensity, i.e., would appear sigmoid when plotted in panel B.

Psychophysics and Neurophysiology

One notable prediction of this model is that it would be extremely unlikely, if not impossible, to obtain a positively accelerating psychophysical function. To obtain such a function the size of the JND in stimulus units would actually have to decrease with increasing stimulus intensity, and it appears that no sensory or perceptual continuum behaves this way.

This prediction is important because it brings psychophysical functions into line with the available neurophysiological evidence. To date, no plausible explanation has been offered for the apparent discrepancy between psychophysical and neurophysiological data. Uttal (1973) observed that neurophysiological functions always seem to be compressed (negatively accelerating), and the apparent existence of positively accelerating psychophysical functions (i.e., magnitude scales with exponents >1) led him to conclude that "there is no direct linear relationship between the two" (p. 336). However, as was shown earlier in Chapter 2, replotting magnitude scales in linear coordinates often reveals that the positive acceleration is an artifact of the log-log plot; and, if there is still any residual positive acceleration in linear coordinates, this will certainly disappear when the magnitude responses are transformed logarithmically to correct for subjective number. With the problem of positively accelerating functions removed, there is new hope for a linear correspondence between psychophysical and neurophysiological data. While not essential to sensory coding theory, such a correspondence is attractive and economical, and obviates the need to invoke more complex processing at the psychological level.

JND Scales and Category Scales

If the present model is correct and traditional JND scaling is valid, it still has severe practical limitations. No psychophysicist would relish returning to this tedious, time consuming approach, especially after a taste of direct scaling.

Fortunately there is no need. Psychophysical scales generated by direct interval judgment (e.g., category rating) are also similar in shape to curve 2 in Figure 7. This is not coincidence. If JND scales are valid, and if, as functional measurement has found, category rating is a valid response measure, then such an isomorphism is to be expected. Although it has usually been claimed that JND scales and category scales produce different measures of sensation (e.g., Stevens, 1957; Stevens & Stone, 1959), the present model predicts that, provided they are carefully obtained (i.e., free from methodological bias), scales derived by direct interval judgment and by cumulating JNDs should be one and the same. This prediction will be investigated later.

6

ON PROTHETIC AND METATHETIC CONTINUA

Stevens (1957) introduced the terms <u>prothetic</u> and <u>metathetic</u> to differentiate between two apparently dissimilar types of psychophysical continua. These terms were supposed to reflect the "size vs. sort" dichotomy in psychophysics (see Stevens & Galanter, 1957). The terms have, for better or worse, become imbued with a good deal of significance over the years, and it is therefore important that the present model can offer an explanation for them.

An exhaustive list of the differences between the two types of continua is given elsewhere (Stevens, 1975, Table 9); however, the two most salient differences relate to the subjective size of the JND, and to the relationship between magnitude scales and category scales. Stevens claimed that, with prothetic continua, the subjective size of the JND is proportional to subjective magnitude (i.e., Brentano's assumption holds) and the category scale is concave downward when plotted against the magnitude scale; with metathetic continua, the subjective size of the JND is constant and the category-magnitude scale relationship is linear. But the present model maintains that JNDs are always subjectively equal, irrespective of continuum type; therefore Stevens' main criterion for distinguishing prothetic and metathetic continua is no longer tenable.

Prothetic Continua

As shown earlier in Figure 3, the apparent growth of subjective variability with subjective magnitude in magnitude scaling (i.e., empirical support for Brentano's assumption) can be explained by nonlinear subjective number; support for this premise disappears once magnitude estimates are corrected with a log transform. And, as demonstrated in the previous Chapter, with some continua (e.g., loudness) real variation in the physical size of the JND serves to compound the illusion of variation in the subjective size of the JND. Therefore, those sensory continua labelled prothetic (e.g., loudness) will have Weber functions similar to curve 2 in Figure 5. Accordingly, the psychophysical functions of prothetic continua will be similar in shape to either curve 2 in Figure 7A or, if the stimulus range is large and a logarithmic abscissa more appropriate (e.g., loudness), curve 2 in Figure 7B. On prothetic continua an approximately linear relationship between category and magnitude scales should be obtained after the magnitude scale has been corrected for subjective number.

Metathetic Continua

Visual position and visual angle are both examples of metathetic continua (Stevens & Galanter, 1957). Pitch, the psychophysical continuum held to be prototypically metathetic (Stevens, 1975, Table 9), is not metathetic - or at least it does not behave like visual position and visual angle. The reason that pitch was classified metathetic will be discussed later in Chapter 7.

The psychophysical functions for visual position and visual angle are both linear with stimulus intensity, regardless of the scaling method employed (see Stevens & Galanter, Figure 13). In the present model, this type psychophysical function is exemplified by curve 3 in Figure 7A. Inspection of its underlying Weber function (curve 3 in Figure 5) reveals that such a psychophysical function obtains when absolute sensitivity is constant on the stimulus continuum (i.e., JND or $\Delta I = k$). This fits Stevens' description of metathetic continua being qualitative, rather than quantitative: with visual position and visual angle, stimulus intensity "increases" only inasmuch as the spatial orientation of the stimulus changes, therefore the difficulty of discrimination remains constant across the stimulus continuum (a similar explanation is proposed by Baird & Noma, p. 76).

This interpretation is further supported by Ono (1967) in an experiment on discrimination of line length. Ono found that when the experimental conditions allowed subjects to discriminate on visual position rather than length, the size of the JND remained constant regardless of stimulus length.

Visual position and visual angle also possess a unique property: they both have inherent anchors or "natural landmarks" (Stevens & Galanter, p. 403) which are always readily accessible to the observer. In the case of visual position the ends of the line serve as anchors, while in visual angle there are implicit boundaries at 0° and 180° (90° might also serve as an anchor). Now, when a response continuum is anchored to a stimulus continuum at two distinct points, ratio judgments, in the way they are made in magnitude estimation or free number matching, become impracticable. The constant presence of a second anchor necessarily converts the response task from one of ratio judgment to one of interval judgment. In a fractionation task, for instance, where the observer is required to select a stimulus which is subjectively half as intense as the standard, the moment a bottom stimulus anchor is introduced and designated a value on the response continuum, the observer's strategy switches from one of fractionation to one of bisection (cf. Stevens & Volkmann, 1940).

For this reason the "magnitude estimation" of metathetic continua is not magnitude estimation in the usual sense of free number matching. The present model proposes that "magnitude estimates" of visual angle are linearly related to category ratings (Stevens & Galanter, Figure 13D), and to graphic ratings (Weiss & Anderson, 1972), because they are themselves interval-type judgments.

A Revised Distinction

In the present framework, then, the distinction between prothetic and metathetic continua may be summarized as follows: on "prothetic" continua, sensitivity measured in stimulus units is approximately relative; on "metathetic" continua, sensitivity measured in stimulus units is absolute regardless of stimulus intensity. The crucial difference between these definitions and those of Stevens is that, here, sensitivity varies only in observable stimulus units.

Given this fundamental difference, is there any point in retaining the prothetic/metathetic dichotomy in the present model? According to the revised criterion, virtually all continua are "prothetic". Except for the usual deviation near threshold (see curve 2 in Figure 5), on most continua sensitivity is approximately relative - only one or two continua are truly "metathetic". There would seem to be little point in persisting with this somewhat arcane distinction; however, if a distinction is still seen to be necessary, then perhaps "metathetic" continua could simply be relabelled linear, and "prothetic" continua, nonlinear.

Why Is the Category Scale Valid?

The above re-evaluation suggests a theoretical explanation for the validity of the category scale and the graphic rating scale. In using these response scales, subjects are essentially marking a position on a line: in other words, position on a line is the matching continuum. As just noted, the psychophysical function for position on a line is linear and, in the present framework, described by curve 3 in Figure 7A. Since it is itself linear, position on a line (i.e., the rating scale) is perfectly appropriate as a matching continuum in direct scaling: it linearly reflects the subjective magnitude to which it is matched, without the need for any further transformation or correction. This reasoning will be put to empirical test later in the Thesis. However, it is also extremely important to note that certain precautions are necessary to exclude methodological bias from rating scales, and this matter will also be covered later in Chapter 10.

Number - An Amphoteric Continuum

Why do the numbers which are customarily attached to category scales constitute a linear scale, when the numbers used in magnitude estimation are perceived in a nonlinear fashion?

In category scaling, the numbers (e.g., 1-7) constitute a finite continuum, analogous to position on a line as a psychophysical continuum: the numbers serve as positional markers, and do not have "subjective magnitude" in the usual sense of the term. In this case, subjective number, like position on a line, has a psychophysical function described by curve 3 in Figure 7A. This contrasts with magnitude scaling where, as proposed earlier, the number continuum is unrestricted and subjective number is logarithmically related to the number itself. In magnitude scaling, the psychophysical function for subjective number would be approximated by curve 1 in Figure 7A. As noted earlier, Banks and Hill (1974) found that the numbers below 10, commonly used on category scales, are perceived as linear. Poulton (1979) also makes the point that logarithmic bias (such as occurs in magnitude estimation) can be precluded in category rating by using only single digit numbers on the scale.

This dual status of number, its capacity to be either nonlinear in magnitude scaling, or linear in category scaling, suggests it is <u>amphoteric</u> as a psychophysical continuum. Depending upon the experimental conditions, number can have a Weber function which resembles either curve 1 or curve 3 in Figure 5.

7

EMPIRICAL EVIDENCE

With the salient points of the model outlined, it remains to be seen if the theory can accomodate the fact. The main prediction of the model is clear. For any given modality, the JND scale should be consistent with those scales derived by interval estimation (e.g., category rating, category production, graphic rating, bisection, equisection); however, these interval scales should be systematically at variance with scales derived by ratio procedures, e.g., magnitude estimation, magnitude production, ratio estimation, fractionation. This systematic discrepancy should be (approximately) removed when a log transformation is applied to the ratio scales to correct for nonlinear subjective number. A further prediction is that the JND and interval scales should correspond more closely to the available neurophysiological evidence than magnitude scales.

Citing empirical support for a particular theory always raises the vexed question of selectivity of evidence. To a certain extent all empirical evidence is selective, especially in the formative stages of a theory. Rare indeed would be the scientist who, after proposing a new theory, were then promptly to squash it with a bundle of contrary evidence. However, in an attempt to avoid the accusation of selectivity, and as a check on the "generality criterion", this Thesis draws from psychophysical work in a number of modalities. Where there is dispute, for example about the exact shape of an empirical Weber function for a given modality, no single set of data has been taken as veridical; rather, the trend of the data is reported without any detailed quantitative predictions.

Loudness

Figure 8 depicts five different loudness scales plotted against sound intensity over the range 40-90 dB SPL. Units on the ordinate are linear, but arbitrary. The curves were obtained by first equalizing the response ranges of the scales (e.g., setting the response at 40 dB equal to 0, the response at 90 dB equal to 100), then, for each scale, calculating the response values corresponding to the intervening levels of 50, 60, 70, and 80 dB. This is strictly a linear transformation and cannot affect the shapes of the psychophysical functions. The five curves were then equispaced vertically to facilitate comparison of their shapes.

The top curve is the JND scale: the JNDs were cumulated from the empirical values of the Weber fraction for a 1000 Hz tone obtained by Jesteadt et al. Proceeding downward, the next three curves were independently obtained by:

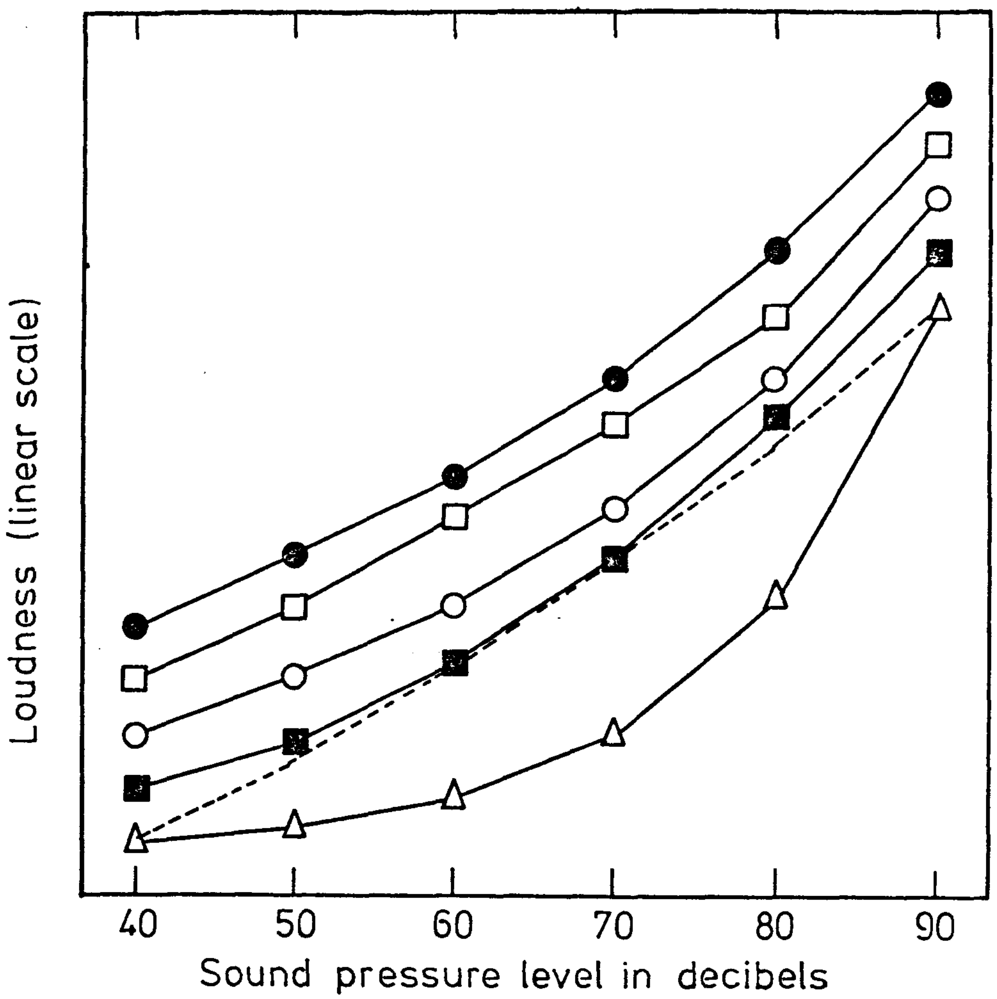

Figure 8. Five different loudness scales vs. sound intensity in decibels. These were obtained by: cumulating JNDs from empirical values of the Weber fraction of a 1000 Hz tone (filled circles); bisection of noise stimuli, within the framework of functional measurement analysis (open squares); both equisection and fractionation of a 1000 Hz tone (the lambda scale, open circles); category scaling of noise stimuli where precautions were taken to exclude bias (filled squares); and free number matching of a 1000 Hz tone (triangles).

bisection, within the framework of functional measurement analysis (Carterette & Anderson, 1979); equisection (the lambda scale, Garner, 1954); and category rating, where precautions were taken to preclude methodological bias (Stevens, 1975, Figure 50). Clearly, there is reasonable agreement between these four curves, and support for a JND scale-interval estimation isomorphism. For instance, the decibel level corresponding to a loudness halfway between that at 40 dB and that at 90 dB is, for each of these four scales in turn, 71.3, 71.1, 73.2, and 72.3 dB.

The JND scale-interval estimation isomorphism is impressive in this instance; but scale agreement in itself would not constitute a rigorous validity criterion were it not for the curve second from the top. The close agreement between the two top curves is of particular significance. The curve second from the top was found to comply with the stringent linearity (additivity) criterion of functional measurement analysis, and therefore provides corroborative support for Fechner's assumption that the JND constitutes a valid unit of sensation. A loudness scale derived by scale-free nonmetric methods (Parker & Schneider, 1974; not shown in Figure 8) conforms to this same shape, as does the pure category scale of Eisler (1962).

The bottom curve in Figure 8 was obtained by free number matching (Stevens, 1975, Figure 10), and is typical of the magnitude scales of loudness obtained in this way (it is almost identical to the sone scale, the most commonly known magnitude scale of loudness). This curve deviates markedly from the upper four but, as predicted, approximate agreement is obtained after application of a log transformation to correct for subjective number (broken line).

Analysis of loudness data illustrates once again the need to distinguish between curve-fitting as a purely descriptive technique, and use of a mathematical function as a fundamental explanation of the data (cf. Chapter 2). For instance, it is customary to fit a power function to data obtained by free number matching (bottom scale in Figure 8). Indeed, this gives a good fit (exponent=.64, r>.99), and if goodness of fit alone is taken as the validity criterion then there can be no argument. But a power function cannot account fundamentally for loudness data because, apart from any other reason, one of its underlying premises (Weber's law) is not met by the data. A threshold correction (Stevens, 1975, p. 182) will not help matters, because the departure of the empirical Weber function from Weber's law is not a threshold effect - it continues at high sound intensity. Failure of Weber's law means that, in loudness, equal stimulus ratios do not produce equal sensation ratios, invalidating both the power law and the less well known physical correlate theory (Warren, 1981), which also relies on this principle.

As another demonstration of its seductive descriptive ability, it can be noted that a power function also provides an excellent description of the JND scale in Figure 8 (top curve; exponent=.23, $\underline{r}$>.99). However, the JND scale is certainly not a genuine power function, since it is derived from the empirical Weber function (not Weber's law) and Fechner's assumption (not Brentano's).

In their determination of Weber fractions for tones of different frequency, Jesteadt et al. found, contrary to the earlier work of Riesz (1928), that in fact the Weber fraction at a given decibel level did not change with stimulus frequency. This is an important result. First, it contradicts early work which purported to invalidate Fechner's assumption. Using a loudness matching task across different frequencies and basing calculations on the Weber fraction values of Riesz (1928), Newman (1933) claimed that JNDs could not be subjectively equal. With the benefit of the more recently and rigorously obtained data of Jesteadt et al., Newman's assertion can be overruled. Second, the Jesteadt et al. result implies that the JND scale for, say, a 1000 Hz tone, should be much the same as the JND scale for a 4000 Hz tone: that stimulus frequency should have no effect on the shape of the loudness function. This prediction is supported by the agreement between loudness functions of pure tone and noise stimuli in Figure 8, and also by the observation of Marks (1981) that, except for stimuli below 400 Hz, variation in frequency has no effect on the psychophysical function for loudness.

The empirical evidence for loudness vindicates Garner (1954, 1958), who claimed that loudness scales based on a discriminability criterion are valid, whereas scales based on a fractionation task cannot be taken at face value. The shape of the top four curves in Figure 8 support Garner's assertion that the original decibel scale (Fechner's law) more closely approximates the true loudness function than does the sone scale (Stevens' law).

Luce (1977) reported difficulty reconciling neural auditory pulse data with the magnitude (sone) scale of loudness. The observed range of neural responsiveness is lower that that predicted by the magnitude scale, and this induced Luce to speculate that two separate intensity codes might be operating. But recognition of subjective number obviates the need for any such hypothesis. Once corrected for subjective number with a log transformation, the response range of the magnitude scale falls into line with the top four scales in Figure 8.

<u>Pitch</u>

Harris (1952) pointed out that methodological difficulties preclude precise specification of the Weber fraction at high frequencies; consequently, there is some variation between published Weber functions for pitch. Nonetheless, a pattern exists. In a summary of investigations, Stevens (1954) showed the Weber function for pitch to be similar in form to the Weber function for loudness, except that the Weber fraction tends to increase again at high frequencies (above 4000 Hz; confirmed more recently by Wier, Jesteadt, & Green, 1977). This would produce a JND scale which is sigmoid in shape, i.e., like the JND scale in Figure 8 but flattened out at the top end.

Figure 9 illustrates two pitch scales over the range 125-10000 Hz. Units on the ordinate are linear, but arbitrary, and the response ranges have been equalized by linear adjustment. The top curve is the JND scale: the JNDs were cumulated from the Weber function for a 40 dB stimulus (Wier et al). The lower curve is the revised mel scale of pitch (stimulus about 55 dB; Stevens & Volkmann, 1940) obtained by equisection. The predicted agreement is reasonable, although the JND scale is more obviously sigmoid. In addition, Stevens and Galanter (1957, Figure 15C) showed that the category scale for pitch is linearly related to the revised mel scale, in accord with the JND scale-interval estimation isomorphism. Stevens (1954) also argues for an isomorphism between the JND and revised mel scales of pitch.

But the pitch modality has been accredited with a distinctive feature: Stevens and Volkmann (1940) claimed to obtain the same function for pitch (the revised mel scale)

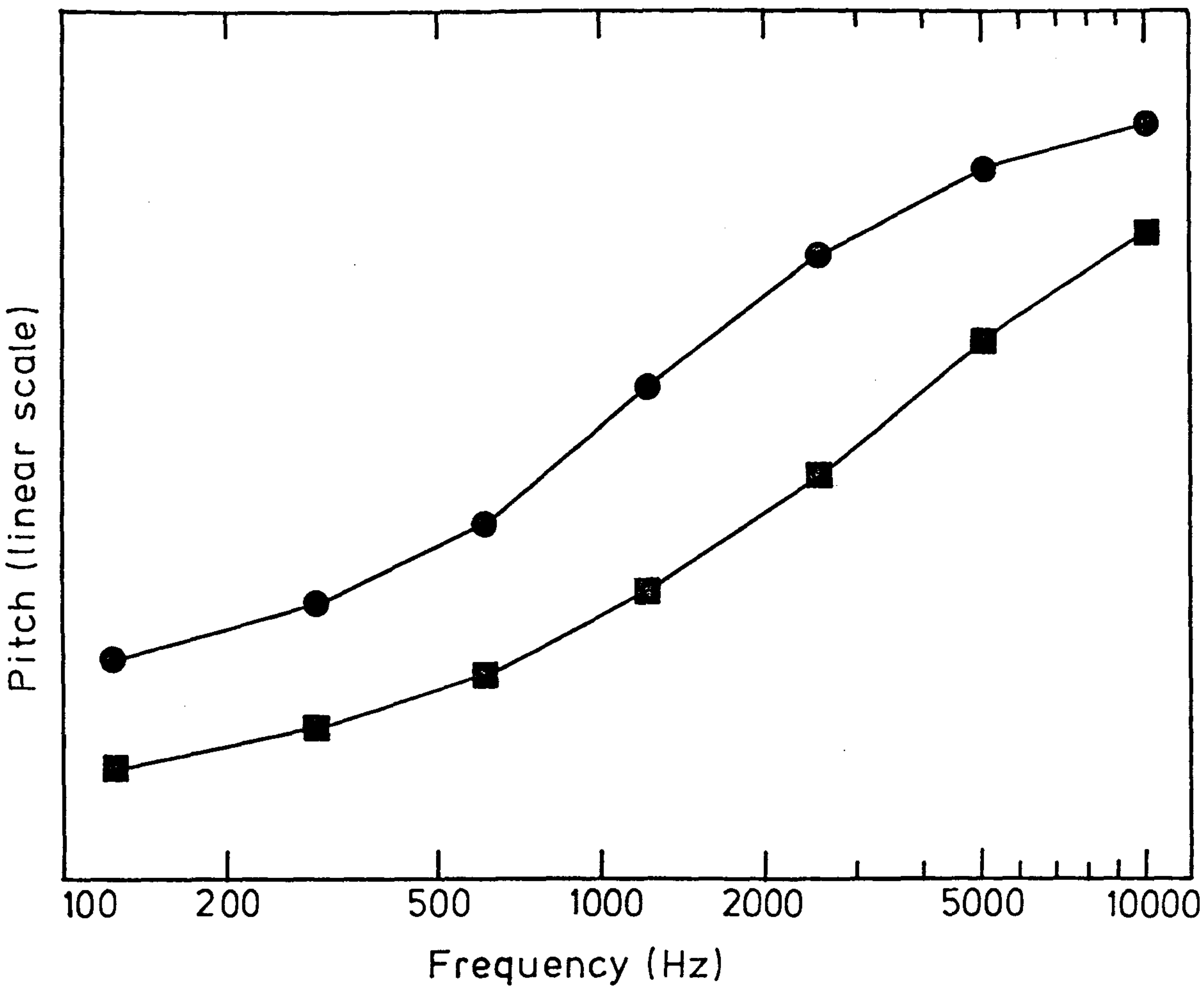

Figure 9. Pitch scales vs. frequency in Hertz. These were obtained by: cumulating JNDs from the empirical Weber function for a 40 dB stimulus (filled circles); and equisection of stimuli of varying frequency at about 55 dB (filled squares).

either by the interval method of equisection or the ratio method of fractionation. This led to the proposal (see Stevens, 1975, p. 165) that pitch is a metathetic continuum, fundamentally different from loudness. The following argument suggests it is not.

The apparent agreement between the two types of scale occurs because the "ratio" scale was not derived by genuine ratio judgment at all. In constructing the revised mel scale by "fractionation", a bottom anchor of 40 Hz was constantly available to the subjects. According to Stevens and Volkmann, the subjects referred to this anchor two or three times in the course of each judgment. So, in actuality, the response task was one of bisection, not fractionation (cf. the argument proposed in Chapter 6). In fairness to Stevens and Volkmann they did concede: "it might be objected that the introduction of the 40~ tone transforms the experiment on fractionation into one on bisection" (p. 340). Regrettably, this critical shortcoming in the experimental procedure seems to have been overlooked in all the subsequent theorizing. However, the <u>original</u> mel scale (Stevens, Volkmann, & Newman, 1937) is a scale genuinely derived by the ratio method of fractionation (no bottom anchor presented) and it is different from the revised mel scale. A log transformation can largely remove the discrepancy.

<u>Imaginary anchors</u>. There is yet another interesting aspect

of this work on pitch measurement. Stevens and Volkmann (1940) reported that earlier, in construction of the original mel scale by fractionation (Stevens et al., 1937), the responses from two subjects had deviated systematically from the others. Post-experimental reports revealed that these two subjects had contemplated the problem of "zero pitch", and had decided to imagine it as the "lowest note on the organ". Now, by imagining a bottom anchor and using it in the judgment process, these subjects effectively employed a bisection strategy. The other subjects apparently thought little about the concept of zero pitch and therefore judged ratios. It was subsequently discovered in the Stevens and Volkmann (1940) study that the responses from the two subjects who had imagined "zero pitch" were well correlated with the responses from subjects in the equisection task where a 40 Hz bottom anchor was physically available. The intriguing implication of this finding is that, in certain situations, imagining an anchor may be just as effective as the physical presentation of an anchor, thereby converting a fractionation response task to one of equisection.

With regard to neurophysiological data, Stevens and Volkmann (1940, Figure 2) found a close correspondence between the human pitch function (revised mel scale) and the data from studies on the guinea pig which located the positions on the basilar membrane stimulated by tones of different frequencies.

Brightness

The many methodological problems in the investigation of brightness (e.g., state of adaptation, size of target, intensity of background illumination, contrast effects) make comparison between studies hazardous. The array of units used in brightness studies (e.g., candelas, photons, millilamberts, decibels) further confuses the issue. However, it appears that the empirical Weber function for brightness (Graham & Bartlett, 1940; Hecht, 1924; Herrick, 1956; Keller, 1941; Mueller, 1951) is similar in shape to that for pitch, although whether or not the Weber function increases again at high stimulus intensity depends on the stimulus duration (Keller, 1941). Therefore, depending upon the experimental conditions, the JND scale for brightness would be similar to either that for loudness (Figure 8) or for pitch (Figure 9; cf. Troland, 1930/1969, Figure 38). Both category scales (e.g., Marks, 1968; Stevens & Galanter, 1957) and bisection scales (Stevens, 1975, Figure 56) of brightness conform to the general shape of the JND scale in Figure 8.

The magnitude scale of brightness (the bril scale) closely resembles the magnitude (sone) scale for loudness; therefore a logarithmic transformation for subjective number brings the magnitude scale for brightness into approximate agreement with the JND and interval scales of brightness.

The brightness paradox. The present model offers an alternative explanation for the enigmatic bisection-fractionation paradox in brightness judgment (Stevens, 1975). Stevens' explanation for this phenomenon involves the concept of a "virtual exponent" and is somewhat ad hoc. In a brightness bisection task, the subject is required to adjust a variable stimulus until its brightness lies halfway between a very bright stimulus (top standard) and a dim stimulus (bottom standard). In a fractionation task, on the other hand, the top stimulus would remain the same, but no bottom standard would be presented; the subject would be required to adjust the variable stimulus to a brightness half that of the top standard. These two seemingly equivalent tasks give different results. Stevens (1975) observed:

> In those two experiments, bisection and fractionation, we encounter a curious paradox. When I turned off the bottom stimulus, instead of lowering the level of the middle or halfway stimulus, the observer raised the level of the halfway stimulus. It is as though halfway from the top stimulus to "zero" were somehow less than halfway from the top to a value well above zero. (p. 157)

Since fractionation is a ratio procedure, it is susceptible to the complications of logarithmic subjective number; bisection is not. The subject's estimate of "1/2" is therefore not genuinely 1/2, but proportional to its logarithm - closer to 2/3 in effect (Banks & Hill, 1974 found the subjective number function for numbers <1 also to be negatively accelerating). Hence, in a fractionation task, the stimulus adjusted to be "half" as bright is more intense than the stimulus which is veridically adjusted to lie halfway in a bisection task.

Taste

Taste is not a single modality like loudness or brighness in that it consists of four qualitatively distinct submodalities. Nevertheless, a consistent pattern predominates. Part of the Weber function for sucrose, the most common sweet stimulus, was shown in Figure 5. Weber functions for stimuli representative of the other three basic tastes - sodium chloride (salty), citric acid (sour or acid), and caffeine (bitter) - are similar in shape to that for sucrose; in some cases, however, there is evidence that the Weber fraction increases again at high concentration (Schutz & Pilgrim, 1957b).

Figure 10 illustrates four different sweetness scales vs. sucrose concentration over the range .0625-.5000 Molar. As with the loudness scales in Figure 8, the response ranges have been equalized and the curves spaced vertically to

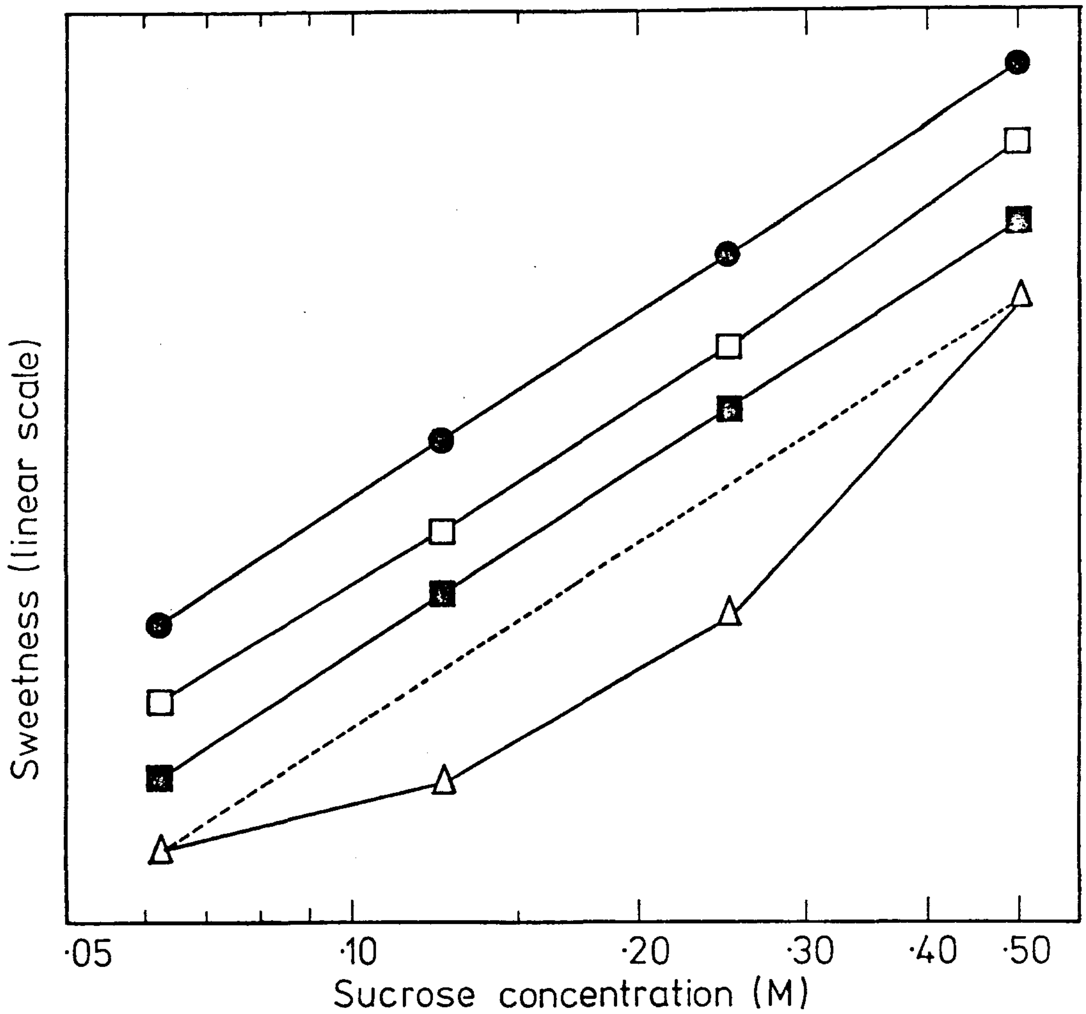

Figure 10. Sweetness scales vs. molar (M) sucrose concentration. Scales derived by cumulating JNDs (filled circles), bisection (open squares), and category rating (filled squares) are in good agreement. The scale obtained by free number matching is discrepant, but conforms to the other curves after a log transformation.

allow comparison. The top curve is the JND scale: the JNDs were cumulated from the empirical values of the Weber fraction (Schutz & Pilgrim, 1957b). The next curve down was obtained by a variant of bisection (MacLeod, 1952). The third curve was obtained by category rating (Schutz & Pilgrim, 1957b). Once again, there is evidence of a JND scale-interval estimation isomorphism and, in this particular modality, support for Fechner's law.

Does the magnitude scale for sucrose sweetness also conform to Fechner's law after the log transformation for subjective number? The answer to this question depends on what is taken to be the magnitude scale. Meiselman (1972) found published values of the sucrose exponent to range from .46 to 1.80, with a mean value of .93. However, this survey included studies which used the "flow method" (technique described by Meiselman) and which customarily give lower exponents than the conventional "sip and spit" technique. It is probably fair to say that most "sip and spit" magnitude studies of sucrose sweetness give exponents around 1.0, or slightly above, suggesting that magnitude estimates of sweetness are near linear with sucrose concentration. Similarly, the ratio technique of fractionation also suggests a linear relationship (Beebe-Center & Waddell, 1948; Lewis, 1948). Assuming linearity, the magnitude scale is concave upward in semilog coordinates (bottom curve in Figure 10), but a log transformation to correct for subjective number renders the magnitude scale consistent with the other three, in accordance with the prediction of the present model.

The misleading nature of the log-log plot, demonstrated earlier in Chapter 2, is again evident with the magnitude scale for sucrose. Several magnitude estimation studies (e.g., Moskowitz, 1970, 1971; Stevens, 1969) have claimed the exponent for sucrose to be 1.3, implying that the sweetness of sucrose is a positively accelerating function of concentration. If this were so, it would mean that the effect of adding a constant amount of sucrose to a solution would increase with the basal concentration of the solution, i.e., the increment in sweetness from 10 to 11% would be greater than from 1 to 2%. But this prediction is contradicted by discrimination studies. Schutz and Pilgrim (1957b) found the Weber fraction for sucrose to be reasonably constant at .15. Therefore, the sweetnesses of 1 and 2% sucrose would be readily discriminable because their discrepancy exceeds the Weber fraction several times over, but 10 and 11% sucrose could not be discriminated at all by most people, because their discrepancy falls below the value of the Weber fraction.

Figure 11 contains data from the magnitude estimation of sucrose (Stevens, 1969, Figure 1). In the left hand panel the data are plotted in customary log-log coordinates, giving an approximation to a straight line with slope

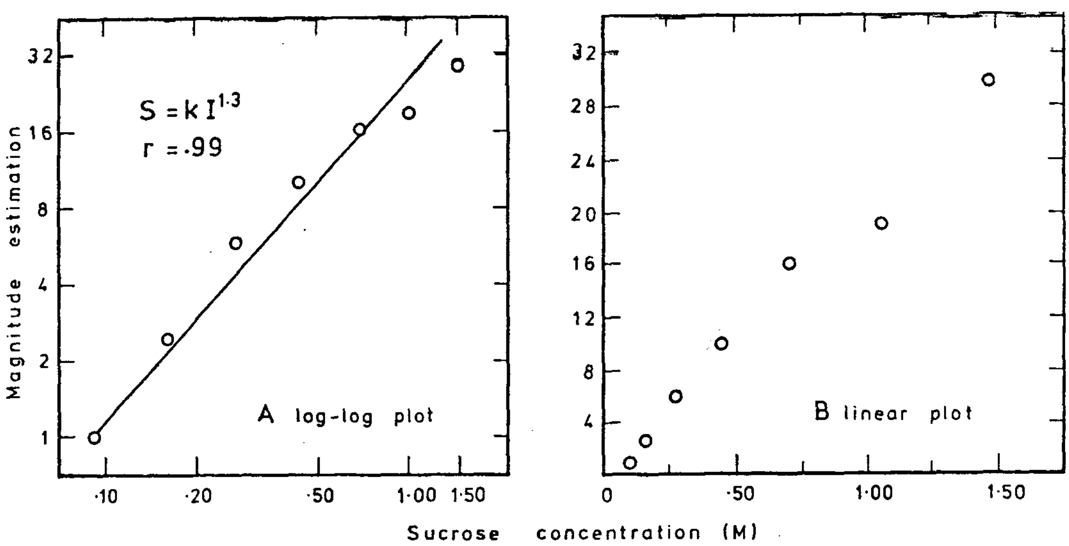

Figure 11. Magnitude estimation of the sweetness of sucrose (data from Stevens, 1969, Figure 1). In a log-log plot (left hand panel) the data have been taken to conform to a power function with exponent 1.3, but when replotted in linear coordinates (right hand panel) there is no evidence of positive acceleration.

(exponent) around 1.3. In the right hand panel, the same data are replotted in linear coordinates, without any transformation or correction. Clearly, there is no evidence of positive acceleration: the data curve is near linear up to .500 Molar and slightly concave downward at the higher concentrations. Why? The explanation is the same as that offered for the discrepant Figures 1 and 2 in Chapter 2. When a power function is fitted to data which fundamentally do not conform to a power function, the exponent simply reflects the ratio of log response range to log stimulus range; here the ratio is >1. So, when the obscurative effect of the log-log plot is removed, as in the right hand panel, there is further evidence that the magnitude scale is linear with sucrose concentration (at least over the concentration range presented in Figure 10). In other words, subjective number varies with actual number in the same manner as sucrose sweetness varies with sucrose concentration - both continua approximate to Fechner's law.

Like the previous sensory modalities, taste scaling has also been confused by the subtle bisection-fractionation distinction. For example, the response technique used to obtain the scale second from the top in Figure 10 was actually termed "fractionation" by the author. However, closer inspection reveals that the experiment was, in one crucial respect, similar to the Stevens and Volkmann study on pitch: a bottom anchor was presented. The response task was actually bisection, not fractionation; subjects were required to judge intervals, not ratios.

Neurophysiological correspondence. The taste sense is unique in that it affords neurophysiological data from human subjects. Borg, Diamant, Strom, and Zotterman (1967) presented two subjects with varying concentrations of sucrose during middle ear surgery, and were thus able to measure the summated electrical response in the chorda tympani nerve. The neural response curve for sucrose is a reasonable approximation to Fechner's law (slightly concave upward at low concentration), similar to the JND and category scales in Figure 10. The neural response curves for sodium chloride and citric acid obtained by Borg et al. are similar to that for sucrose, consistent with their respective Weber functions. It is also interesting to note that the neural response curves of the rat to sucrose (Hagstrom & Pfaffmann, 1959) and sodium chloride (Beidler, 1953) are consonant with the human data.

Further to this point, the approximation of the neural data to Fechner's law is not apparent from the original presentation of results. Borg et al. actually reported their data in the form of power functions in log-log plots and, for sucrose, claimed an exponent of 1.1. However, this function is near linear when replotted in semilog coordinates. It appears most investigators in taste research are not aware of the possible deception which results from use of the power function (e.g., Van der Wel, 1980).

Olfaction

Most of the Weber functions for olfactory stimuli reported in the literature (e.g., Holway & Pratt, 1936; Stone & Bosley, 1965; Wenzel, 1949) approximate to the same general shape as the Weber function for loudness shown in Figure 6. As a consequence, JND scales of olfactory intensity would approximate to the shape of the top curves in Figure 8. Regrettably, the large variation in methodology among studies prohibits any quantitative inference. There appears to have been no consistency in the choice of stimuli in olfactory studies, and the situation is further confused by the variety of techniques of stimulus presentation (e.g., sniff bottles of different volume, olfactometers of different design). There is little published on the category scaling of odour intensity; however, Piggott and Harper (1975, Figure 1) show the perceived intensity of butanol odour to be related to butanol concentration in a manner consistent with the present model, i.e., the shape of the response curve is similar to the top curves in Figure 8.

Lifted Weights

The Weber function for lifted weights (Engen, 1971, Figure 2.1; Oberlin, 1936) is similar to the Weber functions of taste stimuli, so the JND scale for lifted weights would be similar to the JND scale for sucrose (Figure 10). The category scale for lifted weights (Figure 7 of Stevens & Galanter, 1957) is similar in shape to curve 2 in Figure 7A, which means that, when replotted in semilog coordinates (cf. curve 2 in Figure 7B), it, too, will be similar in shape to the JND scale for sucrose, as predicted by the present model. The psychophysical functions for lifted weights obtained by functional measurement studies (Anderson, 1972; Birnbaum & Veit, 1974) are also negatively accelerating in linear coordinates. The magnitude scale of subjective weight (the veg scale; Stevens & Galanter, 1957), positively accelerating with a power function exponent of 1.45, becomes consistent with the JND and category scales after a log transformation for subjective number. As with the sucrose curve in Figure 11, the positive acceleration suggested by the exponent of 1.45 is largely an artifact of the log-log plot and disappears in linear coordinates.

Apparent Length

Ono (1967) found reasonable evidence for Weber's law in the judgment of apparent length, implying that the JND scale for apparent length would approximate to Fechner's law. And, as reported earlier in Chapter 3, there is evidence from a number of different approaches (including nonmetric studies and category rating) that the psychophysical function for apparent length is "logarithmic" and conforms to Fechner's law. The magnitude scale conforms similarly after a log transformation.

8

THE EXPERIMENTS

In discussing the necessary requirements of a sensory scale, Galanter (1962) proposed that a scaling method should provide a consistently repeatable result, unaffected by "manipulation of ostensibly nonessential characteristics of the experiment" (p. 147). So, although the present model appears to account for empirical data from several sensory modalities, important questions remain. Can an empirical Weber function be precisely replicated? How invariant is the psychophysical scale obtained by category rating? What are the necessary and sufficient conditions for a valid category scale? Can the proposed JND scale-category scale isomorphism withstand more rigorous empirical scrutiny? These methodological questions are no less important than the foregoing theory, and they will now be tackled in the experimental component of this Thesis.

The main tenet of this Thesis is that JND scales and category scales provide the same psychophysical function, and the experiments largely focus on this prediction. As a consequence, most of the experimental work is univariate; however, factorial designs are invoked as a final validity check.

All experiments were conducted in the taste modality. The dearth of information is alone sufficient justification for further work in taste. Taste is an interesting modality in that it comprises four qualitatively different "submodalities": sweet, acid, salty, and bitter. (It is commonly recognized that these are the only taste primaries, but there is some dispute on this issue; McBurney & Gent, 1979; Schiffman & Erickson, 1980.) Furthermore, on an applied note, just as psychophysical work in loudness has had application in acoustics, psychophysical research in taste has practical application both for product development in the food and beverage industry, and also in food acceptance research.

Experiments I and II investigate the predicted JND scale-category scale isomorphism for each of the four basic tastes. Experiments III-VI explore the effect of methodological bias in the category rating of taste intensity. If, as the present model predicts, the category scale is a valid means of measuring sensation, then it is important to determine its limitations in practice. Experiment VII investigates the experimental conditions necessary for a valid rating scale. Experiments VIII and IX, in the functional measurement paradigm, employ a stimulus integration task as a means of validating the foregoing univariate work.

Experimental Work - General

The experiments were carried out in the Sensory Evaluation Laboratory at the CSIRO Food Research Laboratory complex. The CSIRO Division of Food Research has a tradition of research in the sensory evaluation of food (Bastian, McBean, & Smith, 1979).

Facilities. The sensory laboratory has been described in detail elsewhere (Christie, 1964, 1966). Basically, the laboratory consists of a room divided into two parts: one is a kitchen/preparation area, the other contains individual compartments (tasting booths, see Figure 12) which allow eight subjects to be tested simultaneously. The laboratory is air conditioned and maintained at a temperature between 20° and 25° C.

Subjects. The subjects (men and women, age range 25-58 years) were drawn from the 160 employees on the site of the Food Research Laboratory. Although largely inexperienced in psychophysical tasks before the experiments commenced, most subjects had some previous knowledge of the sensory evaluation of food. Subjects were not paid for their services: it has long been accepted that participation in sensory testing, while voluntary, is expected of laboratory employees. The sensory laboratory has been fortunate in maintaining good cooperation with staff members in this

Figure 12. A taste panel in the Sensory Evaluation Laboratory.

regard. One advantage of this situation over using students on the university campus is that replicate judgments can easily be obtained from the same subjects over a considerable time period. This facility allowed investigation of a methodological approach, described later, not available to all experimenters. Some experimenters submit vast numbers of taste stimuli to the same subject at a single session, inadequately controlling for the considerable problem of adaptation in taste.

Stimuli. Testing was conducted in the mornings only. Stimuli consisted of colourless, reagent grade chemicals dissolved in distilled water. To allow equilibration and thereby prevent the possible influence of mutarotation effects with sugars (Cameron, 1947), all solutions were made up in 1 litre flasks at least 24 hours before testing and stored at 5°C. Subjects were presented with solutions in standard size (120 ml) glass tumblers, coded for identification purposes. Stimulus volume was always 30 ml, and stimuli were allowed to equilibrate at ambient temperature before testing.

Responses. In all cases subjects responded by marking a questionnaire form. Except for thorough rinsing of the mouth between samples, which was mandatory when subjects judged more than one stimulus at a single session, the experimental procedure in the psychophysical tasks (e.g., interstimulus intervals, tasting duration) was not rigorously controlled. Previous work at this laboratory (e.g., McBride & Laing, 1979) has indicated that subjects perform just as well, and are more at ease, in a <u>laissez faire</u> condition.

9

JND AND CATEGORY SCALES OF TASTE INTENSITY

Torgerson (1958, p. 150) presents a thoughtful discussion on the validity of the JND scale. Torgerson notes that since the interval properties are obtained by definition, not experimentation, then "there are few experimental tests that are directly relevant in testing the 'validity' of the scale" (p. 151). With the advent of functional measurement analysis, however, such tests did become available. As shown earlier with loudness (Figure 8), the fact that the JND scale is parallel to the properly validated bisection scale provides confirmation, albeit indirect, of the equal interval properties of the JND scale. Nevertheless, Torgerson's comment on scale reliability and uniqueness is still pertinent, e.g., is the form of the Weber function affected by the psychophysical method used to obtain it? Torgerson points out that the absolute size of the JND obtained under different experimental conditions is unimportant: the relevant consideration is whether or not the Weber functions derived by different methods are determined to within a linear transformation.

For loudness it would appear this criterion is satisfied, since Jesteadt et al. point out that their results are in good agreement with several other published studies. But, as noted earlier, the situation is less well defined in other modalities. For example, Meiselman (1972) states that virtually no work has been done on indirect scales of taste: the Schutz and Pilgrim (1957b) Weber functions remain unreplicated. Schutz and Pilgrim used the method of single stimuli (Pfaffmann, 1935) in their investigation; as a check on Torgerson's linear transformation criterion, the first experiment in the present study entailed redetermining the Weber function for sucrose using the method of constant stimuli, generally held to be the most sensitive of the classical psychophysical methods (Guilford, 1954, p. 118).

Category scaling is prone to methodological bias: range effects, frequency effects, sequential effects (Parducci, 1974; Poulton, 1979). These issues will be covered more thoroughly in Chapter 10. But perhaps the most frequently cited drawback of category scaling in psychophysics is its susceptibility to stimulus spacing bias: the shape of the psychophysical function obtained by category rating is dependent on the spacing of the stimulus intensities (Stevens & Galanter, 1957). Fortunately, though, Pollack (1965) and Stevens (1975, p. 141) have described an iterative procedure for eliminating the bias due to arbitrary stimulus spacing, a procedure investigated theoretically elsewhere (Anderson, 1975). Essentially, the procedure consists of first presenting subjects with an arbitrarily spaced set of stimuli. The psychophysical scale obtained from this initial presentation then determines the stimulus spacing for the next iteration, and so on, until the scales dictated by consecutive iterations converge, i.e., the scale last obtained suggests the same spacing as that used to obtain it. Such a scale is then said to be in "pure" or unbiased form.

When the iterative technique is employed and the pure category scale is finally obtained, the mean category ratings will be equally spaced (arithmetically) on the ordinate. Conversely, obtaining mean responses that are equally spaced on the ordinate implies that the scale must be in pure form. So, if geometrically (or logarithmically) spaced sucrose concentrations produce mean category ratings which are equally spaced arithmetically, it at once follows that the category scale is free from stimulus spacing bias and also conforms to Fechner's law. The category scale of Schutz and Pilgrim (1957a) satisfies these conditions. As a check on this finding, the first experiment in the present study also entailed deriving the psychophysical function for sucrose sweetness by category rating, using both geometric and logarithmic spacing. As a further check on uniqueness, the sucrose sweetness function was derived by applying the iterative technique to arithmetically spaced sucrose stimuli.

EXPERIMENT I

Method

Determination of the Sucrose Weber Function

Subjects. Twenty subjects (age range 25-58) participated. Overall, there were approximately equal numbers of responses from men and women.

Stimuli. Stimuli consisted of reagent grade sucrose (cane sugar) in distilled water. Six concentration levels (standards) were used: .025, .050, .100, .200, .300, and .500 Molar. These levels approximately match those of Schutz and Pilgrim (1957b), thereby allowing a comparison with this study. For one hour a day over a three week period, six subjects (none of whom were used in the experiment proper) performed informal, preliminary tasting, in order to guide the selection of appropriate comparison stimuli. Six arithmetically spaced comparison stimuli were eventually chosen. At the lowest level (.025 Molar), the comparison stimuli were .50, .70, .90, 1.10, 1.30, and 1.50 times the standard; while for the upper five levels the preliminary testing suggested greater sensitivity and the comparison stimuli were .70, .82, .94, 1.06, 1.18, and 1.30 times the standard.

Procedure. The procedure employed was a forced-choice variant of the method of constant stimuli where the standard is not disclosed (cf. Harrison & Harrison, 1951). The design was similar to that of Lundgren, Pangborn, Pikielna, and Daget (1976), but fewer stimuli were presented at each testing session to minimize disruption to subjects' work routine.

There were 60 experimental sessions, 10 sessions per concentration level; two sessions were run per day (at 0900 & 1130 hrs). Only one concentration level was assessed per day, other than that the order of assessment of levels was random. At each session, subjects were presented with three coded pairs of solutions, each pair consisting of the standard and a comparison stimulus. Order of tasting both between and within pairs was randomized. Subjects were instructed to taste and expectorate the solutions within a pair, and to identify the sweeter solution. Thorough rinsing with distilled water was mandatory between pairs. The 20 sets of three pairs presented at each session comprised all possible combinations of the six comparison stimuli taken three at a time. The 60 responses from each session thus provided as overall total of 3600 responses, 100 responses for each of the 36 standard-comparison pairs. The testing was completed within 10 weeks.

Data treatment. The percentage of "sweeter" responses was calculated for each of the six comparison stimuli at each concentration level. These data (given in Table 1) were then analyzed in two different ways. In the first and more traditional analysis, the percentages were converted to normal deviates and fitted against log sucrose concentration (the phi-log-gamma hypothesis, Rubin, 1976; Thurstone, 1928) by the least squares procedure (Guilford, 1954, p. 125). (In practice the untransformed stimulus values are more commonly used, e.g., Lundgren et al., even though, strictly speaking, it is more correct to use logs when there is an approximation to Weber's law.) In this type of analysis the normal deviates are sometimes further corrected by the use of the Müller-Urban weights (Guilford, p. 129), and this was in fact done; however, the correction had barely any effect on the outcome. As Guilford remarks, in practice such corrections are of dubious value; considering the experimental error in the discrimination task, the use of weights is hardly justified.

In view of this observation, in the second analysis arc sine transforms of the percentages ($\sin^{-1}\sqrt{p/100}$, where p is the percentage) were fitted against sucrose concentration by the method of least squares. As a variance stabilizing measure (Box, Hunter, & Hunter, 1979, p. 132), the arc sine transform allows estimation of the standard errors of the JNDs, and hence the error of the cumulated JND scale. The approximate standard errors were calculated from the formula for the variance of a function (Kendall & Stuart, 1977, p.

Table 1

Percentages of "Sweeter" Responses in the Paired Comparison Discrimination Task of Experiment I

Molar concentration of standard	Ratio of comparison stimulus to the standard*					
	.70	.82	.94	1.06	1.18	1.30
.0250	14	14	29	59	74	98
.0500	8	20	38	57	78	86
.1000	8	25	35	73	82	94
.2000	6	12	36	67	75	91
.3000	6	20	31	61	88	89
.5000	6	14	37	62	81	88

* At the .025 Molar concentration the six comparison stimuli were actually .50, .70, .90, 1.10, 1.30, and 1.50 times the standard.

246). To facilitate comparison with the Schutz and Pilgrim (1957b) study, the JND (ΔI) was computed in traditional fashion in each case, by halving the difference between the concentrations corresponding to the 75% and 25% points on the ordinate. The relevant calculations were as follows:

let $$T = \sin^{-1}\sqrt{p/100}$$

and $$T = a + bx$$

where a and b are the intercept and slope respectively of the regression equation, and x is sucrose concentration.

rearranging $$x_{75} = (T_{75} - a)/b$$

similarly $$x_{25} = (T_{25} - a)/b$$

by definition $$\text{JND } (\Delta I) = (x_{75} - x_{25})/2$$

substituting $$= (T_{75} - a)/2b - (T_{25} - a)/2b$$

$$= (T_{75} - T_{25})/2b$$

inserting values for T $$= (1.0472 - .5236)/2b$$

therefore $$\Delta I = .2618/b$$

it follows from the formula for the variance of a function that

$$\text{var}(\Delta I) = (-.2618/b^2)^2 \text{ var}(b)$$

and $$\text{SE}(\Delta I) = (.2618/b^2) \text{ SE}(b)$$

where var and SE represent variance and standard error, respectively.

In this and subsequent experiments where statistical analyses were necessary, the data were processed by either Mary Willcox or John Best, both of the CSIRO Division of Mathematics and Statistics, using the GENSTAT system (Rothamsted Experimental Station, U.K.) on a CDC Cyber 76 computer.

Category Scaling

Subjects. The same 24 subjects were used in each part of the category scaling investigation.

Stimuli. Stimuli were reagent grade sucrose in distilled water. In Experiment Ia the concentrations were .0625, .1250, .2500, and .5000 Molar (4 stimuli, geometric spacing); in Experiment Ib .0625, .0947, .1436, .2177, .3299, and .5000 Molar (6 stimuli, log spacing); and in Experiment Ic .0625, .2083, .3542, and .5000 Molar (4 stimuli, arithmetic spacing). In each case the concentration range is the same as used in the JND determination.

Response scale. The scale comprised 13 categories with five equidistant verbal descriptors and is shown in Figure 13. As such, it was in accord with the suggestion (Anderson 1974b, p. 232) that between 10 and 20 categories is desirable, and also with Bendig and Hughes (1953) who found that verbal anchors increase the amount of information transmitted by a scale. Note, however, that stimulus end anchors were not presented, nor were the categories numbered on the subjects' response sheet. This type of scale has been used extensively in the routine sensory evaluation of food at this laboratory.

Sweetness Intensity

Date__________ Time________ Name________________________________

Please taste these sucrose solutions in the order specified below. After you have tasted each sufficiently, spit it out in the paper cup provided, <u>then rinse thoroughly with water</u> before proceeding to the next sample.

	Order of tasting			
Extremely sweet				
Very sweet				
Moderately sweet				
Slightly sweet				
No sweetness at all				

Figure 13. The response scale used in Experiment I.

Procedure. Experiments Ia and Ib each consisted of two replicate sessions, the replicate sessions within each part being held on consecutive days. Stimuli were presented simultaneously and subjects were required to taste and expectorate each in the order specified and to rate each sweetness on the 13 point scale. Thorough rinsing with distilled water was mandatory between stimuli. All 24 possible permutations of presentation order were used in Experiments Ia and Ic; order was randomized in Experiment Ib.

In Experiment Ic the procedure was identical, except that arithmetic stimulus spacing was used at the first session. The sweetness scale from the initial session suggested revised concentrations for the second session on the following day, and so on, until the sweetness scales converged at the fourth session (Pollack, 1965 provides details of the methodology).

All category rating was completed within 3 weeks.

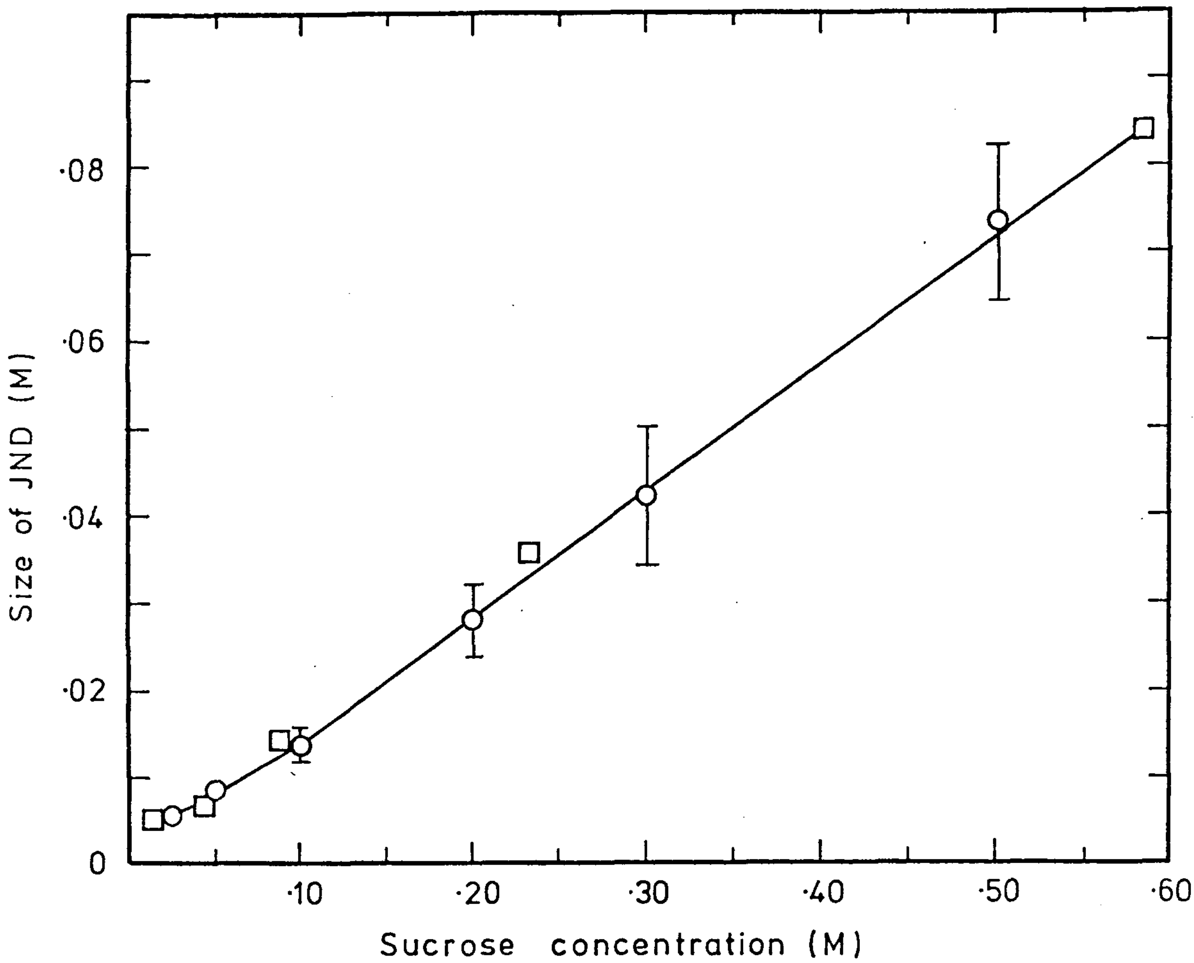

Figure 14. The size of the JND vs. sucrose concentration in molar (M) units. The line of best fit was drawn by eye. Data from the present study (circles, ± 2 SE; error bars for the two lowest concentrations are within the data points) and those of Schutz and Pilgrim (1957b) (squares) are in close agreement. Except for the low concentrations near threshold there is good approximation to Weber's law.

Results and Discussion

The Sucrose Weber Function

Figure 14 shows the JND estimates from the present study obtained by the arc sine transform method (circles ±2 SE) and those of Schutz and Pilgrim (1957b) (squares), plotted against sucrose concentration. Converting the present data to normal deviates and fitting against log concentration, as in the first analysis described, gave an almost identical result: for all but the highest concentration (.5000 Molar), for which the estimate was discrepant by 4%, the JND values coincided with the circles in Figure 14. Clearly, Torgerson's linear transformation criterion is satisfied: in fact, the correspondence between the present data and those of Schutz and Pilgrim is sufficiently close to permit representation by a common line of best fit (drawn by eye). According to Torgerson, such replication alone "imparts a certain amount of scientific respectability to the scale" (1958, p. 151).

Apart from the characteristic deviation at low concentration, it is apparent from Figure 14 that the size of the JND is proportional to sucrose concentration, as dictated by Weber's law. Weber's law may break down again at very high concentration, as it does for sodium chloride (Holway & Hurvich, 1937; Schutz & Pilgrim, 1957b), but this cannot be ascertained using a discrimination paradigm. The viscosity of sucrose solutions increases markedly above .5000 Molar: the confounding of viscosity and sweetness would leave the investigator uncertain as to which cue serves in the discrimination (cf. MacLeod, 1952).

JND Scale

Figure 15 illustrates the JND scale that results when the JND estimates in Figure 14 are cumulated in the Fechnerian tradition. As expected from Figure 14, there is good support for Fechner's law except for the deviation near threshold (taken here as .015 Molar; cf. Pfaffmann, 1959). The envelope (dotted lines) surrounding the JND scale represents approximately ± 2 *SE*. The standard error at any one point on the JND scale was taken as the square root of the sum of the squared standard errors of the JNDs cumulated up to this point (following Kendall & Stuart, p. 246). Here, of course, the standard errors lie horizontally on the JND scale; by fiat there can be no error in the vertical dimension.

Category Scale

For each replicate in Experiments Ia and Ib separately, and for data from the final session in Experiment Ic, category ratings were fitted against log sucrose concentration by linear regression. There was no significant difference

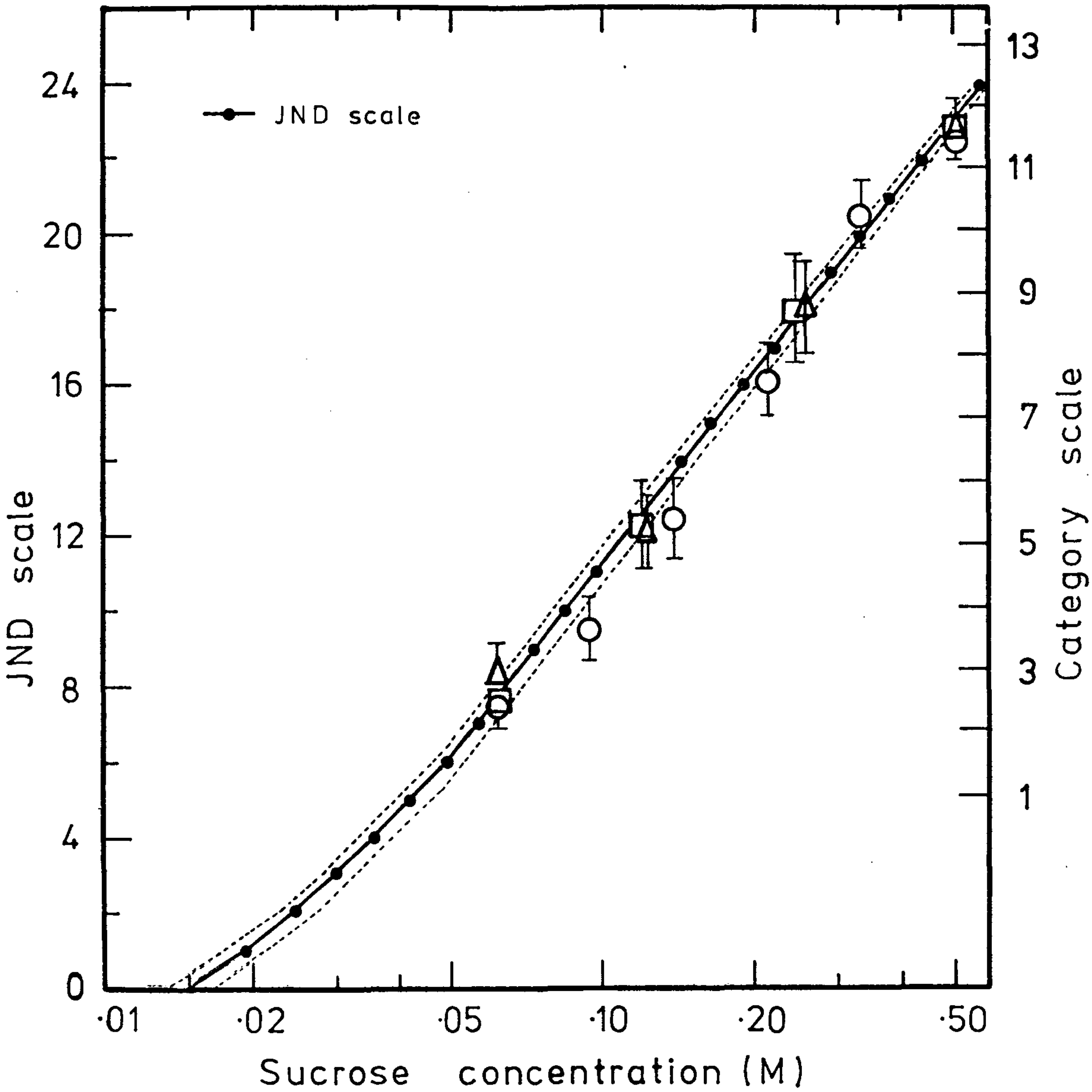

Figure 15. The JND scale (left hand ordinate) and the category scale (right hand ordinate) for the sweetness of sucrose over the concentration range .0625-.5000 Molar (M). The mean ratings (±2 SE) from Experiments Ia (triangles), Ib (circles), and Ic (squares) conform to the JND scale, supporting a JND-category scale congruence (dotted line represents approximately ±2 SE for the JND scale).

between the estimates of slope, $\underline{F}$ (4, 566) = 1.20, indicating good scale reliability.

To facilitate the JND scale-category scale comparison, the category ratings from Experiments Ia, Ib, and the final session of Experiment Ic were then fitted against log sucrose concentration by a single linear regression. Next, the category scale was positioned on the right hand ordinate so that this regression line would coincide with the JND scale (legitimate since the category scale is an interval scale). The mean category ratings (± 2 $\underline{SE}$) from Experiments Ia, Ib, and Ic (squares) are given in Figure 15. (In this Thesis "mean" always refers to arithmetic mean unless otherwise specified.)

JND-Category Scale Isomorphism

The data in Figure 15 offer support for a JND-category scale isomorphism. Most of the mean category ratings actually lie within the error envelope of the JND scale. Moreover, in Experiments Ia, Ib, and Ic the mean category ratings are approximately equidistant on the ordinate, as required of a pure category scale. The mean responses to the final iteration of Experiment Ic provide a valuable cross-check on scale uniqueness: they confirm that obtaining the pure category scale straight off, using geometric or log stimulus spacing (Experiments Ia & Ib), was not mere coincidence. Even when arithmetic stimulus spacing is used initially, the iterative procedure dictates that geometric spacing is eventually necessary to obtain an unbiased sweetness scale for sucrose.

With regard to centering bias, Poulton (1977, 1979) has shown that, for a sensory scale to be free from such influence, the overall mean response should correspond to the midpoint of the rating scale (this bias will be discussed in greater detail in Chapter 10). A check on the 14 mean category ratings in Experiments Ia, Ib, and Ic reveals that this requirement is satisfied: the overall mean rating of 6.96 corresponds closely to 7, the scale midpoint. This result is somewhat fortuitous, since the sucrose concentrations used in the category rating experiments were chosen to match those in the JND determinations, and not specifically selected so as to be exempt from centering bias.

The error bars in Figure 15 are of approximately the same length, consistent with Fechner's assumption that variability on the subjective continuum is constant, irrespective of subjective magnitude. There is, however, a previously observed tendency (Eriksen & Hake, 1957; Parducci, 1974) for response variability to be smaller near the ends of the rating scale.

Figure 16 illustrates operation of the iterative procedure,

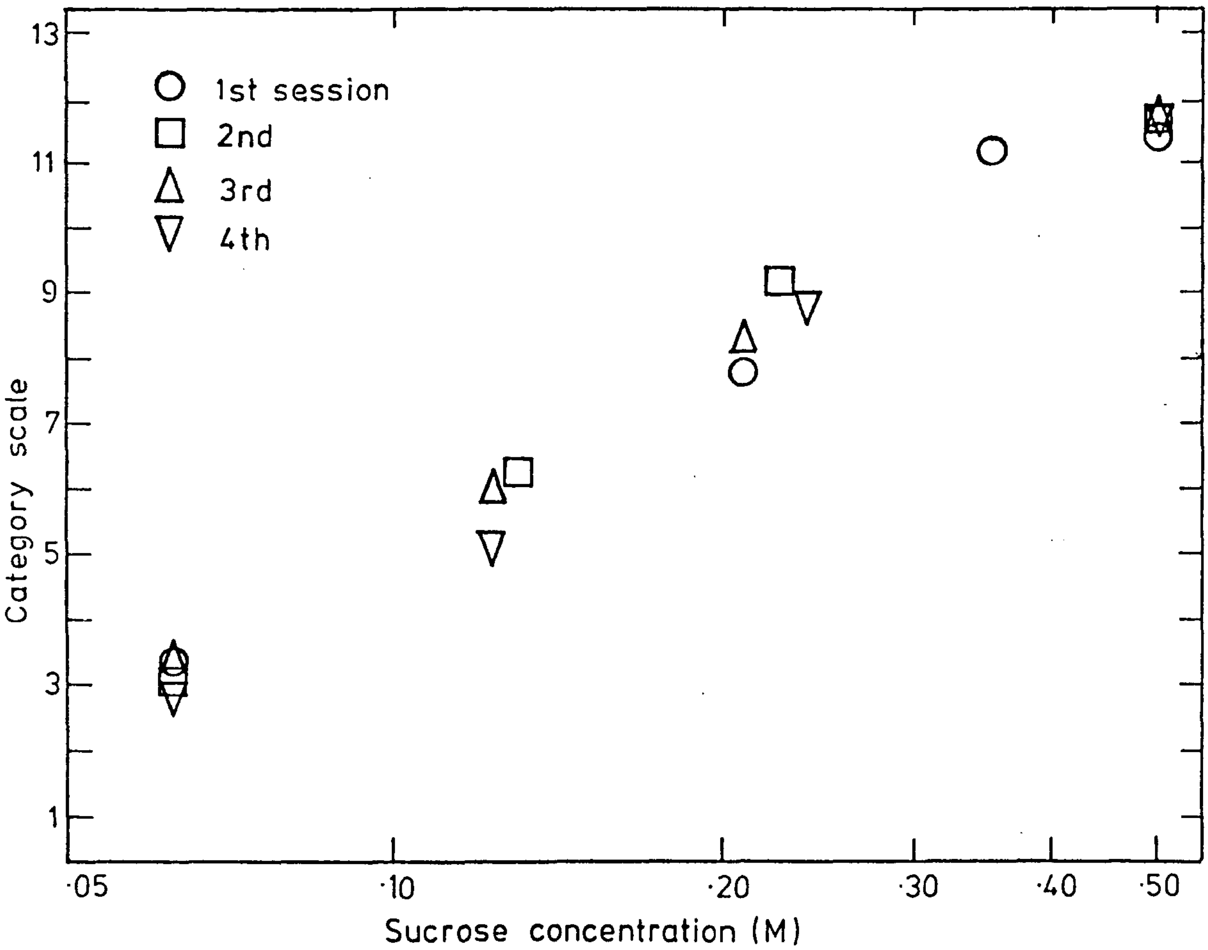

Figure 16. The iterative procedure for removal of possible stimulus spacing bias. The initial arithmetic spacing becomes geometric over the course of four sessions.

the final outcome of which was shown in Figure 15 (squares). Apparently, in this instance, the stimulus spacing bias was not as problematical as is usually thought. While the iterative procedure was effective in producing mean responses which are equidistant on the ordinate, this correction had only a slight effect on the shape of the psychophysical function. The initial arithmetic spacing produced a sweetness function not far removed from the pure category scale, indicating that the subjects were not greatly influenced by the immediate context and made their judgments of sweetness in an absolute, rather than relative, manner.

This first experiment has shown that, when proper regard is paid to methodological factors, the main prediction of the foregoing theory is supported. Does the prediction hold with the other three basic tastes: acid, salty, and bitter?

EXPERIMENT II

Method

The Weber Functions

Experiment I vindicated the Weber function for sucrose obtained by Schutz and Pilgrim (1957b). In view of this confirmation, and taking into account the time that would be required to replicate this work for stimuli representative of the other three basic tastes (at least another 30 weeks of almost daily testing - a veritable endurance test for subjects), it was decided not to repeat testing for the other basic tastes, but to accept the Weber functions already provided by Schutz and Pilgrim (1957b).

Subjects. Schutz and Pilgrim used 10 subjects (five male, five female), all of whom were experienced in psychophysical tasks.

Stimuli. As for sucrose, Schutz and Pilgrim determined the other Weber functions at five concentration levels (these have all been converted to molarity units for consistency with the present work). For citric acid (acid/sour) the levels were .0005, .0016, .0052, .0156, and .0520 Molar; for sodium chloride (salty) .0256, .0684, .1880, .5128, and 1.368 Molar; for caffeine (bitter) .0016, .0031, .0064, .0129, and .0257 Molar.

Procedure. The method of single stimuli was used; Schutz and Pilgrim provide full details of the experimental conditions.

Category Scaling

Subjects. All category scaling in Experiment II was carried out at this laboratory. A panel of 24 subjects was used. Composition of the panel remained constant during testing of a single taste quality but altered slightly between qualities.

Stimuli. Four concentrations of a reagent grade chemical in distilled water were used for each taste. Experiments Ia and Ic showed that four levels could specify the sweetness function just as well as six (Experiment Ib). The efficacy of four stimuli is not altogether surprising considering that there are only 24 JNDs between threshold and .5000 Molar sucrose (Figure 15), a sweetness range commensurate with everyday experience. Besides, with four stimuli there is less chance of interference from adaptation effects.

For citric acid the concentrations were .0010, .0030, .0090, and .0270 Molar; for sodium chloride .0513, .0923, .1641, and .2906 Molar; for caffeine .0031, .0066, .0144, and .0309 Molar. Logarithmic spacing was used in all cases, and the concentration ranges are commensurate with those used by Schutz and Pilgrim in their Weber function determinations.

Response scale. Except for the appropriate name changes on the questionnaire form (e.g., "Extremely salty", "Extremely bitter", etc.), the response scale was as used in Experiment I.

Procedure. As for Experiments Ia and Ib: two replicate sessions for each of the three taste stimuli.

Results and Discussion

The Weber Functions

Overall, the Weber functions for citric acid, sodium chloride, and caffeine, determined by Schutz and Pilgrim (1957b), are very similar to that for sucrose (Schutz & Pilgrim, Tables 1 & 2). Apart from the usual deviation just above threshold, there is reasonable support for Weber's law. Comparing taste qualities, there is little difference between the sizes of the Weber fractions for sucrose, citric acid, and sodium chloride, but the Weber fraction for caffeine is larger than for the other three. With sodium chloride, there is evidence that the Weber fraction increases again at high concentration. Schutz and Pilgrim note, however, that the highest concentration of sodium chloride also produced a stinging (pain) sensation; the apparent drop in taste sensitivity could be due to this phenomenon.

JND Scales

The JND scales for citric acid, sodium chloride, and caffeine, generated by cumulating the Schutz and Pilgrim Weber functions, are given in Figures 17, 18, and 19, respectively. Except for the usual deviaton near threshold, there is once again impressive support for Fechner's law - at least over the stimulus ranges shown.

The threshold value for citric acid was taken as .00014 Molar (Berg, Filipello, Hinreiner, & Webb, 1955; median value from Stahl, 1973), and for sodium chloride .0137 (Stahl, 1973; Pfaffmann, 1959). There is less consensus on the threshold value for caffeine. For instance, the median of the values recorded by Stahl (1973) is .0014 Molar; Schutz and Pilgrim report .0011 Molar; and the median of the values cited by Amerine, Pangborn, and Roessler (1965, p. 106) is considerably lower still at .0002 Molar. It is probable that this variability results, in part at least, from the marked problem of adaptation with bitterness.

A small scale investigation was conducted in an attempt to resolve the confusion. Twelve subjects were each presented with only one pair of solutions, one solution being distilled water, the other .0011 Molar caffeine (the threshold level proposed by Schutz & Pilgrim), and required to select the bitter solution (i.e., the method of constant stimuli, cf. determination of the Weber function in

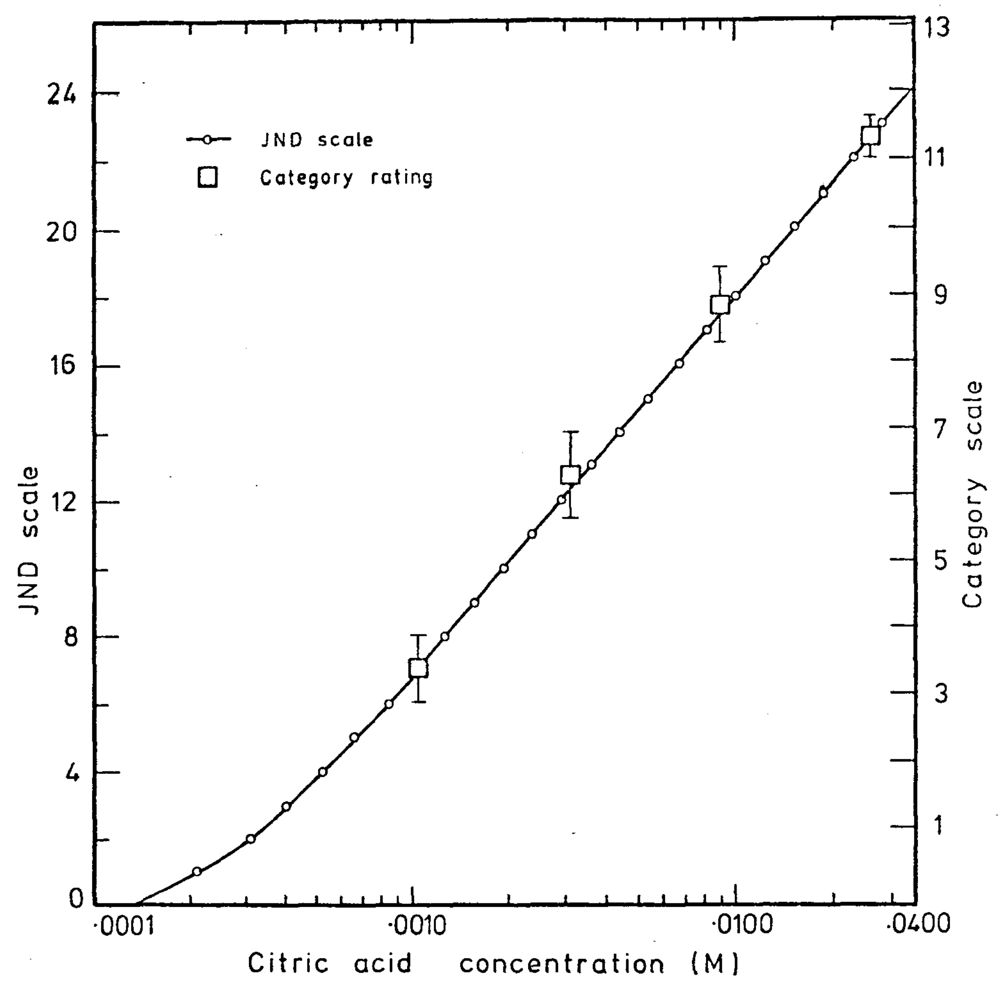

Figure 17. The JND scale (left hand ordinate) and category scale (right hand ordinate) for the perceived acidity of citric acid. Mean category ratings (±2 SE) conform to the JND scale.

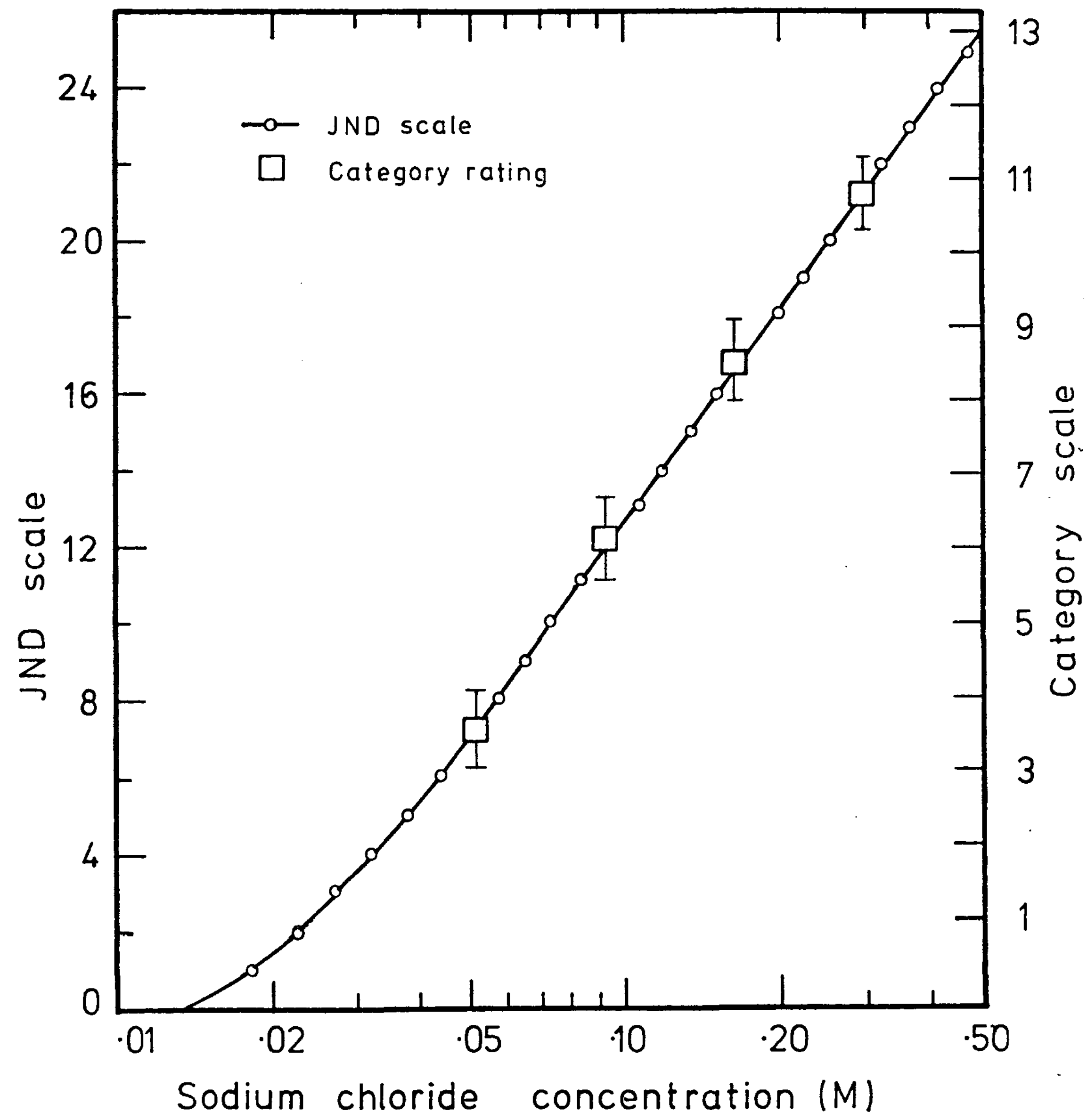

Figure 18. The JND scale (left hand ordinate) and category scale (right hand ordinate) for the perceived saltiness of sodium chloride. Mean category ratings (±2 SE) conform to the JND scale.

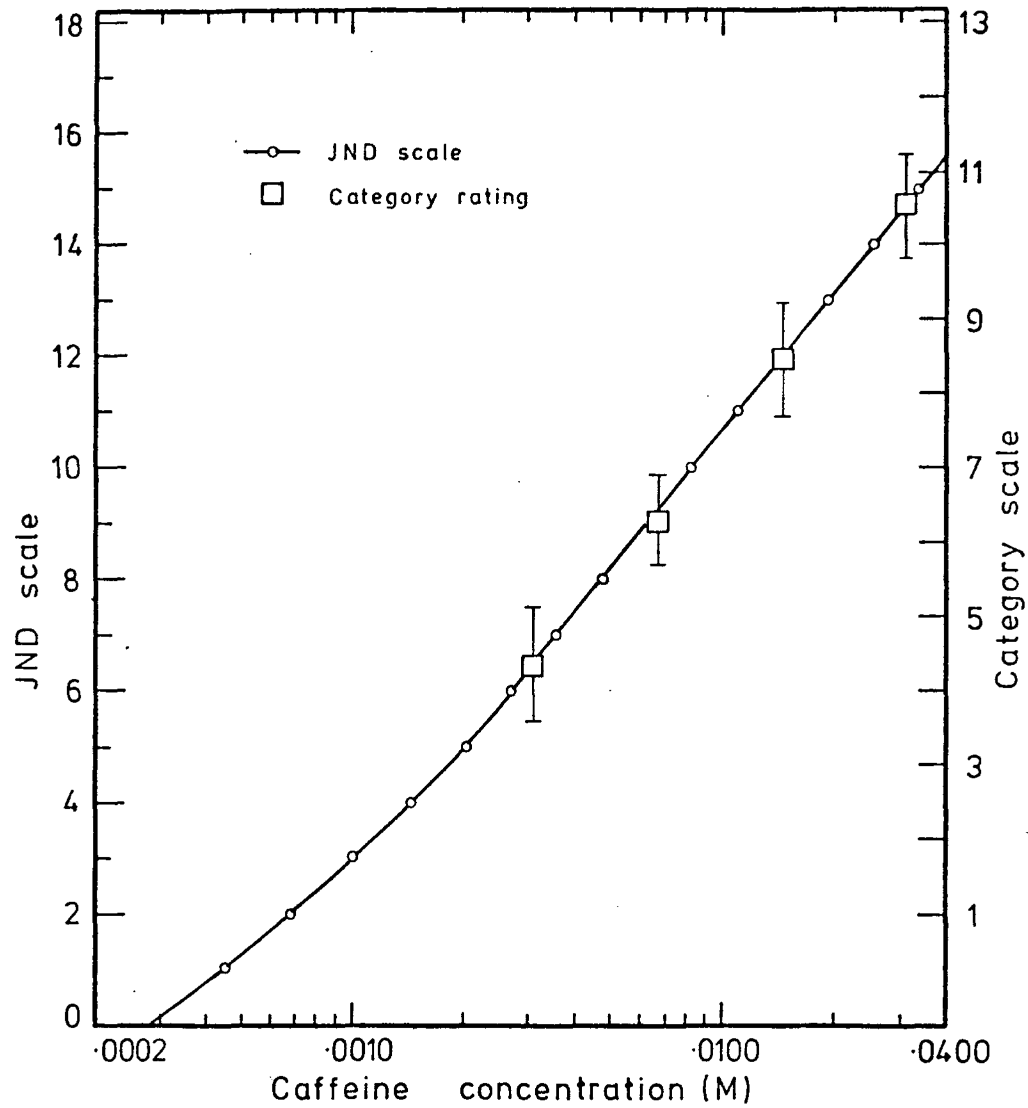

Figure 19. The JND scale (left hand ordinate) and category scale (right hand ordinate) for the perceived bitterness of caffeine. Mean category ratings (± 2 SE) conform to the JND scale.

Experiment I). All 12 easily dectected the bitter solution, confirming the suspicion that the threshold value obtained might depend upon the state of adaptation. The test was repeated the next day with .0005 Molar caffeine, when once again all subjects detected the bitter solution. On the third day, with the concentration further reduced to .00025 Molar, nine of the 12 subjects were correct, suggesting that this value could be taken as threshold level (in this method 75% correct lies halfway between random responding and perfect discrimination). This estimate is close to the median value of Amerine et al. and was used in Figure 19.

Category Scales

For each of the three taste qualities, category ratings were fitted against log stimulus concentration by linear regression, then the category scales were positioned on the right hand ordinate as described in Experiment I. Figures 17-19 contain the mean category ratings, ± 2 SE. Clearly, there is evidence for a JND scale-category scale isomorphism in all three cases, further supporting the proposed model. Moreover, the mean category ratings are equidistant on the ordinate (pure category scale), the overall mean rating in each case corresponds approximately to the midpoint of the rating scale (absence of centering bias), and the error bars indicate that the variability is independent of the subjective magnitude (Fechner's assumption is supported).

Intra-Modality Equality of JNDs?

Comparison of Figures 15, 17, and 18 suggests another intriguing possibility: could JNDs be subjectively equal across taste continua as well within individual continua? For sucrose, citric acid, and sodium chloride, there are approximately 22 JNDs between threshold and a subjective intensity level corresponding to a mean rating of 11 on the category scale. The corresponding number for caffeine is 15 JNDs. This may indicate a genuine difference for caffeine or, as noted above, the influence of adaptation may have resulted in Schutz and Pilgrim obtaining a spuriously large value for the JND, and the difference may not be real at all. This speculation is tantalizing, because it is not readily amenable to experimental check. In a discrimination task with supra-threshold caffeine stimuli, some degree of adaptation is inevitable, even with only one pair of stimuli per session. Indeed, informal tasting of caffeine solutions showed that the order of evaluation is important. A pair of solutions tasted in the order weaker-stronger proved to be more readily discriminable than when assessed the other way around. More work is necessary here.

Ratio Scales of Sensation

Proponents of magnitude scaling have made much of the claim that magnitude scales are ratio scales, whereas category scales are only interval scales. Reverting to category rating would, according to Stevens (1975), "foreclose the measurement of sensation on a ratio scale and relegate sensation measurement to a mere interval scale" (p. 136).

Two points are pertinent here. First, it should be noted that the "ratio" properties of magnitude scales have only been assumed; as noted in Chapter 3, when proper tests of validity are applied these so called ratio properties are usually found not to be substantiated. Second, if detection threshold is taken as zero on the response continuum, then it _is_ possible to obtain a ratio scale of sensory intensity from an interval scale. For instance, the JND scales in Figures 15, 17, 18, and 19 are ratio scales, because they have equal interval properties and a meaningful zero. To convert a category scale of sensory intensity to a ratio scale, the response curve is adjusted vertically so that, with shape retained, it passes through the threshold value on the abscissa (this will usually involve some extrapolation below the lowest category rating). If desired, arbitrary units of sensation may then be attached to the ordinate. Deriving a ratio scale of sensation in this manner bypasses the problem of nonlinear subjective number.

10

BIAS IN CATEGORY RATING OF TASTE INTENSITY

The present model argues in favour of the rating scale, but this response technique is susceptible to many types of methodological bias. Poulton (e.g., 1979) has argued persistently for better recognition of such biases, but his warnings have largely gone unheeded. Parducci (1965, 1968, 1974) has also explored methodological effects, but from a slightly different orientation. Parducci has used human performance with rating scales as a means of devising a general theory of judgment (the range-frequency theory, e.g., Parducci, 1968), which appears to account for the data more successfully than adaptation level theory (Helson, 1964). However, considering the substantial research literature in psychophysics, these are isolated efforts; in general, psychophysicists have ignored methodological bias.

Study of methodological bias is handicapped by the confusing assortment of terms used. For consistency, the terms suggested by Poulton (1979) have been adopted in the following analysis. Poulton (1979, Table 1) claims that rating scales are affected by one nonlinear bias, the stimulus spacing bias, and four types of range bias: centering bias, stimulus equalizing bias, response equalizing bias, and contraction bias.

Stimulus spacing bias. This term refers to the tendency for subjects to regard all stimuli as equally probable. If the stimuli are not, in fact, equally probable, or are bunched at one part of the scale, the shape of the psychophysical function can be affected. Investigators have been aware of this bias for some time (e.g., Stevens & Galanter, 1957); indeed, the "frequency effect" (Parducci, 1965) is a special case of this bias. If a rectangular stimulus distribution is used, iterative techniques can remove the bias, as shown in the previous Chapter.

Centering bias. This term describes the tendency for subjects to centre their responses on the stimulus range presented, regardless of the actual physical intensities of the stimuli. This bias is certainly troublesome in those sensory investigations where the aim is to determine an "ideal" or cutoff point. Poulton (1977) demonstrates that the estimate of what constitutes a just tolerable noise level depends upon the range of noise levels presented in the experiment. Along similar lines, McBride (in press, Appendix A of this Thesis) has shown that the same effect holds in the determination of the optimum sweetness level for a beverage - the estimate of optimum sweetness level varies with the range of sweetness levels presented for evaluation. However, these are hedonic type assessments: as Poulton (1979) notes, psychophysicists are concerned with the shape of psychophysical functions rather than the establishment of cutoff points, consequently the centering bias should not pose a problem to psychophysical research.

Stimulus equalizing bias. This term describes the tendency for the subject to use all of the response range, regardless of the actual physical intensities of the stimuli presented. Poulton notes that, in category rating, this effect is not usually seen as a "bias"; on the contrary, subjects are often specifically instructed to use all of the response range available (this is particularly true when stimulus end anchors are employed). The stimulus equalizing bias is more of a problem in magnitude estimation, where it has led to some spurious hypotheses (see Poulton, 1979).

Response equalizing bias. The converse of the previous bias, here the subject has a tendency to use the full range of responses available (e.g., 13 points on a 13 point scale, 7 points on a 7 point scale), even when the same stimulus range is presented. Since, in category rating, subjects are actually expected to use all of the available response range, the term "response equalizing bias" might be better relabelled "response equalizing effect".

Contraction bias. Perhaps more commonly known as the "regression effect" (Stevens & Greenbaum, 1966), this term describes the tendency for subjects to select a response too close to the centre of the response range (or to avoid the extremes of the response range). The contraction bias is similar to the centering bias inasmuch as it is symmetrical (Poulton, 1979, Figure 1C), and should not affect the relative positioning of category ratings.

Transfer bias. Essentially, this term describes carry-over effects which occur as a result of using within-subjects experimental designs (Poulton, 1973, 1975), e.g., influence of previous experience, previous instructions, experimental conditions, etc.

Just how serious are these biases in psychophysical research on taste? Experiment III explores the stimulus spacing bias with the iterative technique used in Experiment Ic.

EXPERIMENT III

Method

Subjects. The same 24 subjects served in each part of the experiment.

Stimuli. For Experiment IIIa the stimuli were reagent grade D-fructose (fruit sugar) in distilled water. The concentrations were .0971, .1943, .3885, and .7771 Molar (geometric spacing). For Experiment IIIb the stimuli were reagent grade D-glucose and the initial concentrations were .1500, .3000, .6000, and 1.200 Molar (geometric spacing).

Response scale. As for Experiment I (Figure 13).

Procedure. The iterative technique was employed for both fructose and glucose separately, and the procedural details were as for Experiment Ic. The responses from the first session suggested the concentrations to be used at the next session on the following day, and, in turn, the responses from this second session suggested the concentrations to be used at the third and final session a day later. Testing was completed in two weeks.

Results and Discussion

Figures 20 and 21 summarize the outcomes of the iterations applied to fructose and glucose, respectively. As it happened, Experiment IIIa proved not to be an investigation of the stimulus spacing bias at all, since the geometric stimulus spacing produced the pure category scale for fructose at the very first session, necessitating no change in stimulus concentrations, i.e., the two successive sessions were merely replications of the first. The mean ratings in Figure 20 indicate that, as for sucrose, the sweetness of fructose conforms closely to Fechner's law.

The situation was different for glucose, however. Figure 21 shows that the initial geometric stimulus spacing produced mean category ratings which were not equidistant on the ordinate, as required of a pure category scale. Nevertheless, the next iteration had the desired effect, and it was not necessary to further adjust the concentrations for the third session. The curvelinearity (deviation from Fechner's law) at low concentration is more pronounced for the glucose sweetness function than for either sucrose or fructose, implying that the decrease in the Weber fraction for glucose away from threshold (cf. curve 2 in Figure 5) occurs over a wider concentration range.

One notable feature of both Experiments IIIa and IIIb was

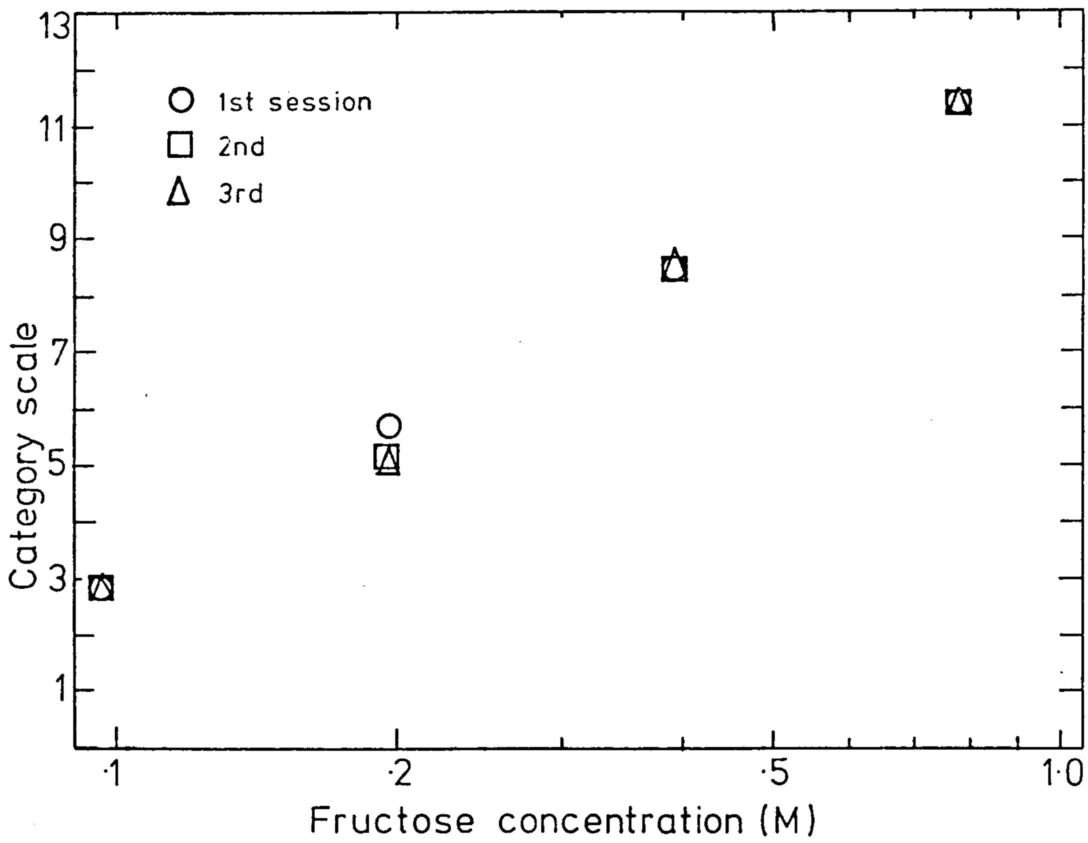

Figure 20. The iterative technique applied to geometrically spaced fructose concentrations. The pure category scale was achieved at the first session.

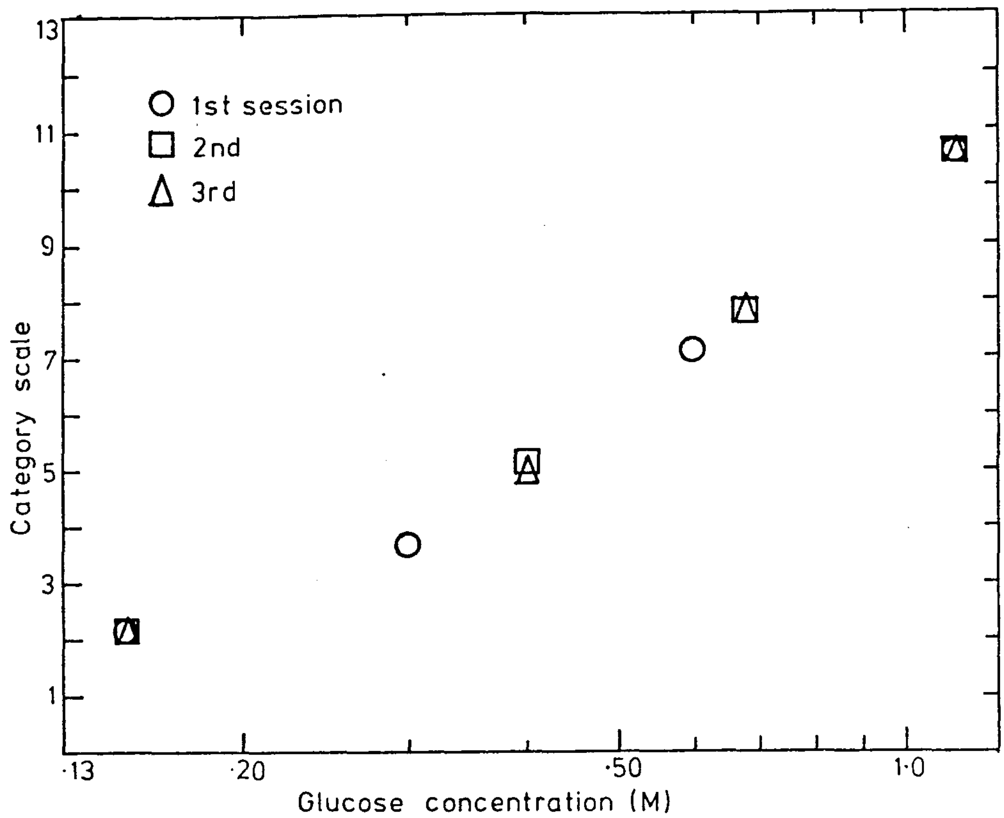

Figure 21. The iterative technique applied to geometrically spaced glucose concentrations. No change in concentrations was necessary after the second session.

the consistency exhibited by subjects. For fructose, the psychophysical function obtained after only one session was extremely similar to that achieved on the third replication. What is even more remarkable, the shape of the psychophysical function for glucose barely changed from the first to the third session - even though the stimulus spacing did. This suggests, as did Experiment Ic, that subjects made their judgments of sweetness in an absolute, rather than relative, manner unaffected by context. Perhaps this occurred in the present experiments because there actually was less "context" than in most other psychophysical investigations of context effects (e.g., Parducci, 1965): with only four stimuli presented at each session there was less chance of the stimulus distribution itself influencing judgment.

Cross-Task Validation

Cameron (1947, Table VIII & Figure 10) investigated the relative sweetness of several different sugars. Given a number of concentrations of one sugar (say sucrose), subjects were required to select concentrations of another (say fructose) of equivalent sweetness. Cameron found the sweetness of sucrose and fructose to be linearly related, but on a weight basis fructose was always slightly sweeter than sucrose. Glucose, on the other hand, was found to be less sweet on a weight basis than sucrose, and its sweetness nonlinearly related to that of sucrose: whereas 2.0%

sucrose was of equivalent sweetness to 3.8% glucose, at higher concentrations the relative discrepancy diminished, and 20% sucrose was equivalent to 25% glucose.

To investigate these claims, data from Experiments Ia (sucrose), IIIa (fructose) and IIIb (glucose) were replotted using % w/v (weight for volume) as the stimulus unit, and these are shown in Figure 22. The claims of Cameron are supported: the sweetness of fructose is indeed linearly related to that of sucrose (slopes are nearly parallel). These curves predict that fructose should have a lower threshold value than sucrose (i.e., the fructose curve should cross the abscissa at a lower concentration). This, too, is confirmed. As noted earlier, the threshold value for sucrose is approximately .015 Molar (.5% w/v), while for fructose, the median of the estimates of Stahl (1973) is .21% w/v. At low concentrations the glucose curve in Figure 22 has a shallower slope than the other two, but this discrepancy diminishes at higher concentration in accord with Cameron's finding.

In a similar paradigm, Pangborn (1963) used a highly trained panel to perform sucrose/fructose/glucose matching. The same pattern resulted. For example, Pangborn found the sweetness of 5% sucrose = 4.2% fructose = 8.3% glucose (Pangborn, Table 3). Figure 22 suggests that 5% sucrose = 4.1% fructose = 9.0% glucose, a close correspondence

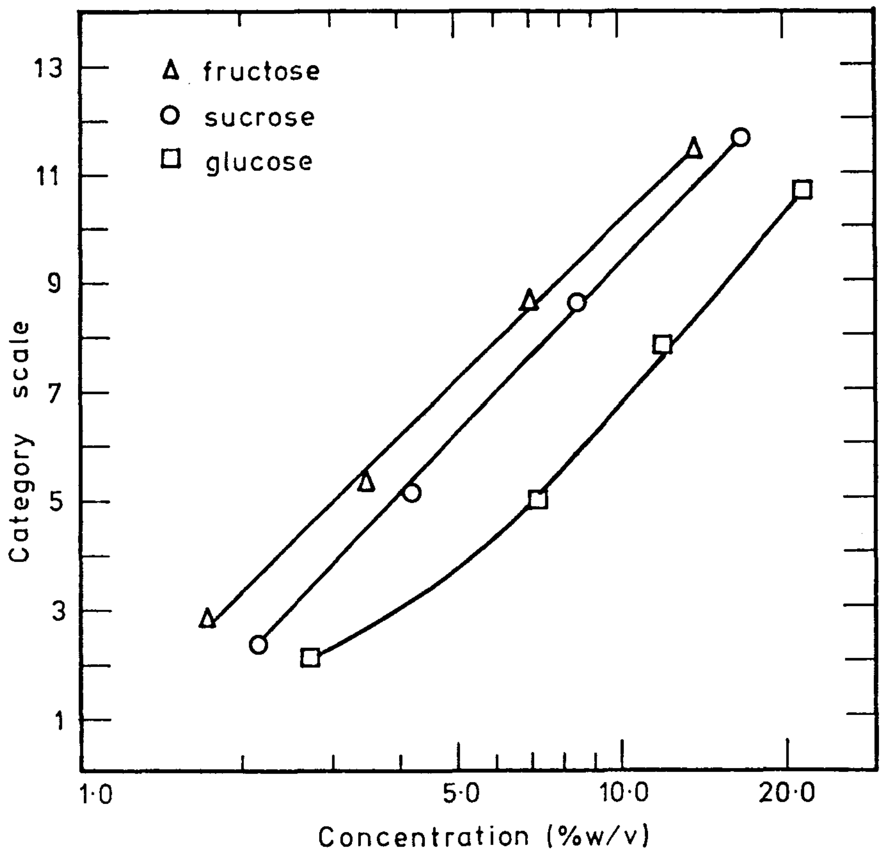

Figure 22. Comparison of the category scales for sucrose, fructose, and glucose as determined in Experiments Ia, IIIa, and IIIb.

considering the vastly different methods used. This type of cross-check on validity is especially valuable. Both Cameron and Pangborn used a "scale-free" matching paradigm, therefore the concordance also indirectly supports the linearity of the category scale. The result also provides further evidence of the absolute nature of the category ratings.

Still on this issue, the present results are entirely consistent with the findings of Schutz and Pilgrim (1957a). Using a 9 point category scale, Schutz and Pilgrim found the sweetness curves of sucrose and fructose (levulose) to be linearly related, with fructose sweeter than sucrose at all concentrations; the sweetness curve for glucose (dextrose), on the other hand, is nonlinear in a semilog plot, as shown in Figure 22. Dahlberg and Penczek (1941) report similar results, as does Lichtenstein (1948), for sucrose/glucose matching only.

The next experiment investigates the stimulus and response equalizing biases by varying the stimulus and response ranges.

EXPERIMENT IV

Method

Subjects. Twelve subjects were used in Experiment IVa; 24 in Experiment IVb.

Stimuli. Stimuli were reagent grade sucrose in distilled water. In Experiment IVa the concentrations were .0625, .1250, .2500, and .5000 Molar (i.e, same as for Experiment Ia); in Experiment IVb .0625, .1250, and .2500 Molar (.5000 Molar concentration not presented).

Response scale. The response scale was similar to Figure 13, but in this case it comprised 11 categories, numbered 0 to 10, with only two verbal anchors, "Extremely sweet" and "No sweetness at all", attached to the top (10) and bottom (0) categories respectively (see Figure 23).

Procedure. Experiments IVa and IVb each consisted of four replicate sessions, and the replicate sessions within each Experiment were held on consecutive days. Other procedural details were as for Experiment Ia. Testing was completed in two weeks.

Results and Discussion

The mean responses from Experiment IVa (circles, ± 2 SE) are shown in Figure 24. The sweetness function is as obtained previously (Figure 15), confirming the reliability of the category scale: a numbered 11 point scale with only two verbal descriptors produces the same result as an unnumbered 13 point scale with five verbal descriptors. Furthermore, the overall mean rating of 4.9 in Figure 24 is close to 5, the midpoint, indicating absence of centering bias.

Comparison of Figures 15 and 24 also confirms the existence of the response equalizing bias (effect). With the 13 point scale in Experiment Ia, the mean responses for the smallest and largest stimuli were 2.9 and 11.8 respectively - approximately 1.5 category points from the end of the scale in each case. The same holds for the 0-10 scale, where the corresponding responses in Experiment IVa were 1.5 and 8.6, also approximately 1.5 category points from their respective ends. This result suggests that the response equalizing bias does in fact induce subjects to use all of the response range available, but at the same time the contraction bias dissuades them from using a constant amount at each end of the scale. The contraction bias would become of greater concern if the number of categories on the response scale were further reduced, since the proportion of response scale effectively available to subjects would also be reduced.

Sweetness Intensity

Date___________Time_________Name____________________________________

Please taste these sucrose solutions in the order specified below. After you have tasted each sufficiently, spit it out in the paper cup provided, then rinse thoroughly with water before proceeding to the next sample.

		Order of tasting			
Extremely sweet	10				
	9				
	8				
	7				
	6				
	5				
	4				
	3				
	2				
	1				
No sweetness at all	0				

Figure 23. The response scale used in Experiment IV.

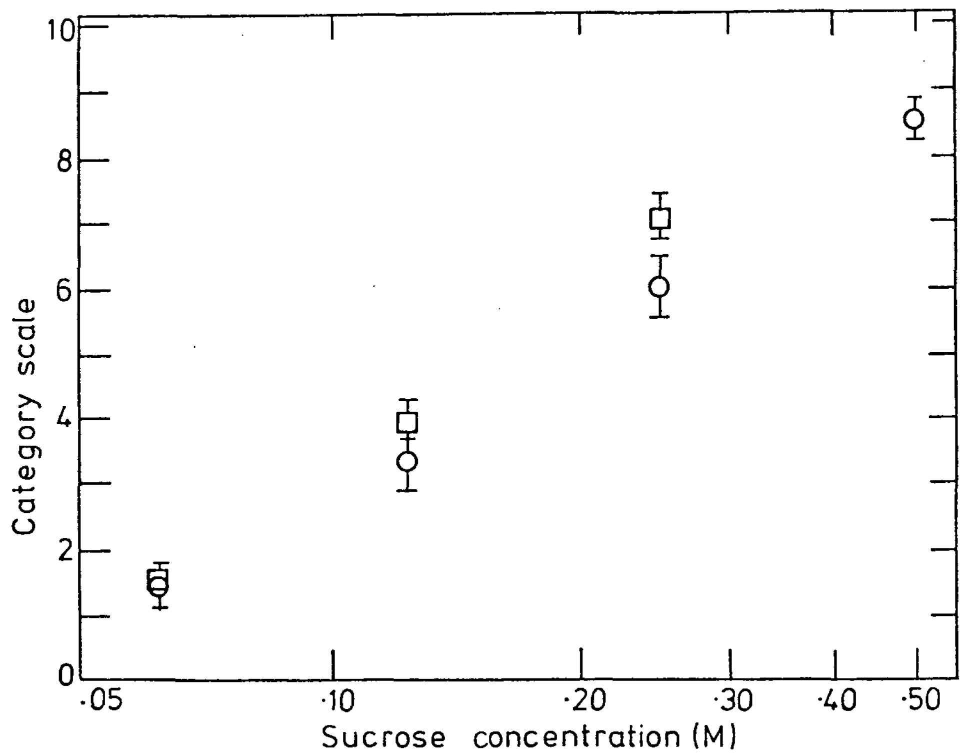

Figure 24. Category rating of four (circles, ± 2 SE) and three (squares, ± 2 SE) sucrose concentrations in Experiments IVa and IVb respectively.

Figure 24 also contains the three mean ratings (squares, ±2 SE) from Experiment IVb. Comparison of the data from Experiments IVa and IVb shows evidence of a stimulus equalizing bias; in fact, the situation is almost perfectly summarized by the schematic Figure 1B of Poulton (1979). When the strongest stimulus (.5000 Molar) was not presented, the responses to the weaker three stimuli increased. The mean response to the weakest stimulus increased slightly, the response to the now medium stimulus a little more, and the response to the now strongest stimulus increased most of all. However, since the size of the change is, in each case, approximately proportional to subjective magnitude, the shape of the sweetness function has not been materially affected (though there is evidence of slight nonlinearity). The taste modality is, therefore, just as susceptible to the stimulus and response equalizing biases as other more commonly measured continua.

A Compromise Procedure

Poulton (1977, 1979) claims that the only way of avoiding all methodological bias in category rating is to have separate groups of unpracticed subjects judge each stimulus once only. Logistically, this is a rigorous and somewhat unrealistic requirement; progress in psychological research would be severely checked if it were to be insisted upon.

Recent research at this laboratory (McBride, 1980, in press) has suggested a procedure which goes halfway toward meeting the requirements suggested by Poulton. Given that most contextual bias is due to the customary presentation of more than one stimulus at a session (multiple presentation), the proposed procedure is a compromise: the same subjects judge all stimuli, but each subject judges only one stimulus per session. This sequential monadic design has been termed single presentation. Provided there is a sufficient time interval between sessions (e.g., 24 hours), and overall the order of presentation is controlled, direct comparison between stimuli is precluded and contextual effects cannot occur (McBride, in press).

In the next two experiments, the efficacy of the single presentation procedure is checked by employing it to generate the psychophysical function for sucrose (Experiment V); and the psychophysical functions for fructose, glucose, citric acid, and sodium chloride (Experiment VI).

EXPERIMENT V

Method

Subjects. Forty eight subjects were used in both parts of the experiment.

Stimuli. Reagent grade sucrose in distilled water. For Experiment Va the concentrations were as for Experiments Ia and IVa, viz., .0625, .1250, .2500, and .5000 Molar (geometric spacing); for Experiment Vb the concentrations were the same as those used for the initial iteration in Experiment Ic, viz., .0625, .2083, .3542, and .5000 Molar (arithmetic spacing).

Response scale. As shown in Figure 13, except only one response column was necessary.

Procedure. To facilitate running of the experimental work, the subjects were arbitrarily split into two panels, 24 in each. Both Experiments consisted of four sessions, and the sessions within each were held on consecutive days. At each session subjects were required to taste and expectorate one stimulus only, and to rate its sweetness on the response scale. For the first panels of 24 subjects in each Experiment, all possible permutations of order of evaluation were used over sessions; for the second panels, order of presentation was based on a latin square design, thereby allowing investigation of sessions and order effects (i.e., the panel was divided into four subpanels of six, and each subpanel received one of the four possible orders in the 4 x 4 latin square design). Testing was completed in 4 weeks.

Results and Discussion

The mean ratings (±2 SE) from Experiments Va (circles) and Vb (squares) are shown in Figure 25. To once again provide a comparison with the JND scale, ratings from Experiment V were fitted by a single linear-log regression, then the category scale was positioned on the right hand ordinate so that this regression line would coincide with the JND scale (i.e., the same procedure as used earlier in Experiment I).

The first point to note is the excellent further support for the JND scale-category scale isomorphism (cf. Figure 15) obtained here with a different experimental design and a different panel. Once again there is a close approximation to Fechner's law. The absolute value of the ratings given the same stimuli are slightly different in Figures 15 and 25, but this is irrelevant; the category scale is an interval scale, and the only important consideration here is that the shape of the response curve remain unchanged.

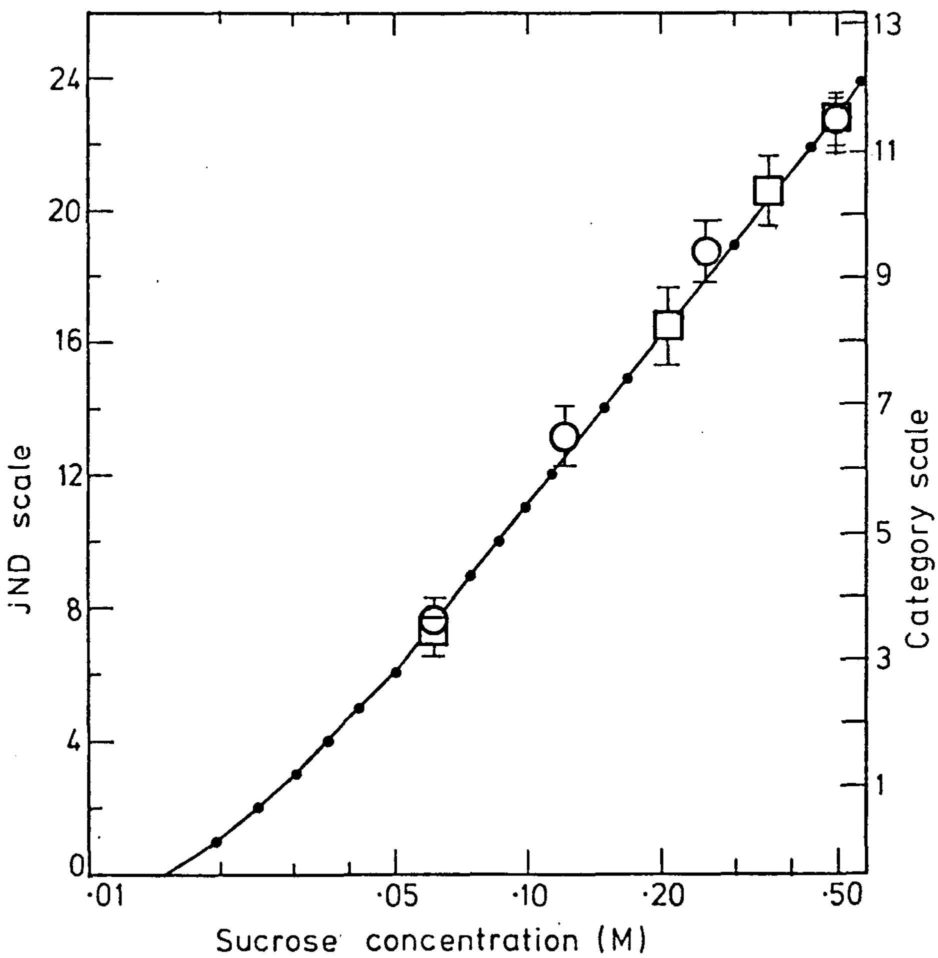

Figure 25. JND scale (left hand ordinate) and category scale (right hand ordinate) for the sweetness of sucrose. The mean ratings (±2 SE) from the single presentation procedure conform closely to the JND scale.

The mean ratings from Experiment Vb lie along the same line as that specified by Experiment Va, indicating that subjects made their judgments of sweetness in an absolute manner, unaffected by context. This was further confirmed by separate latin square analyses of variance on data from the second subpanels of Experiments Va and Vb. For Experiment Va there was the expected effect of sucrose concentration, $\underline{F}$ (3, 6) = 90.59 $\underline{p}<.001$, but no sessions effect, $\underline{F}$ (3, 6) = .69, or order effect, $\underline{F}$ (3, 6) = .15 (order effect taken as difference between groups of subjects). Likewise, for Experiment Vb there was a significant concentrations effect, $\underline{F}$ (3, 6) = 58.06 $\underline{p}<.001$, but no sessions effect, $\underline{F}$ (3, 6) = .03, and no order effect, $\underline{F}$ (3, 6) = .54. The lack of any systematic sessions or order effect implies that judgments were free from between sessions transfer bias.

Comparison of the error bars in Figures 15 and 25 reveals them to be of similar length, indicating that the inter-subject variability in the single presentation paradigm (48 subjects, one assessment of each stimulus) was no greater than with the traditional multiple presentation (24 subjects, two assessments). Also, the error bars are much the same length regardless of subjective magnitude, again supporting Fechner's assumption.

EXPERIMENT VI

Method

Subjects. A panel of 24 subjects was used in each part of the experiment. The panel composition remained constant within each part, but changed slightly between parts.

Stimuli. Reagent grade chemicals in distilled water. Four different stimuli were used: two were sweet (fructose, glucose), one acid (citric acid), and one salty (sodium chloride). The concentrations of these stimuli were the same as in Experiments II and III; geometric spacing was used in each case. For fructose the concentrations were .0971, .1943, .3885, and .7771 Molar; for glucose .1500, .3000, .6000, and 1.200 Molar; for citric acid .0010, .0030, .0090, and .0270 Molar; and for sodium chloride .0513, .0923, .1641, and .2906 Molar.

Response scale. As shown in Figure 13, except for changes in the verbal descriptors for citric acid and sodium chloride. Only one response column was necessary.

Procedure. As for Experiment V; there were four sessions for each stimulus, held on consecutive days. Testing was completed in 4 weeks.

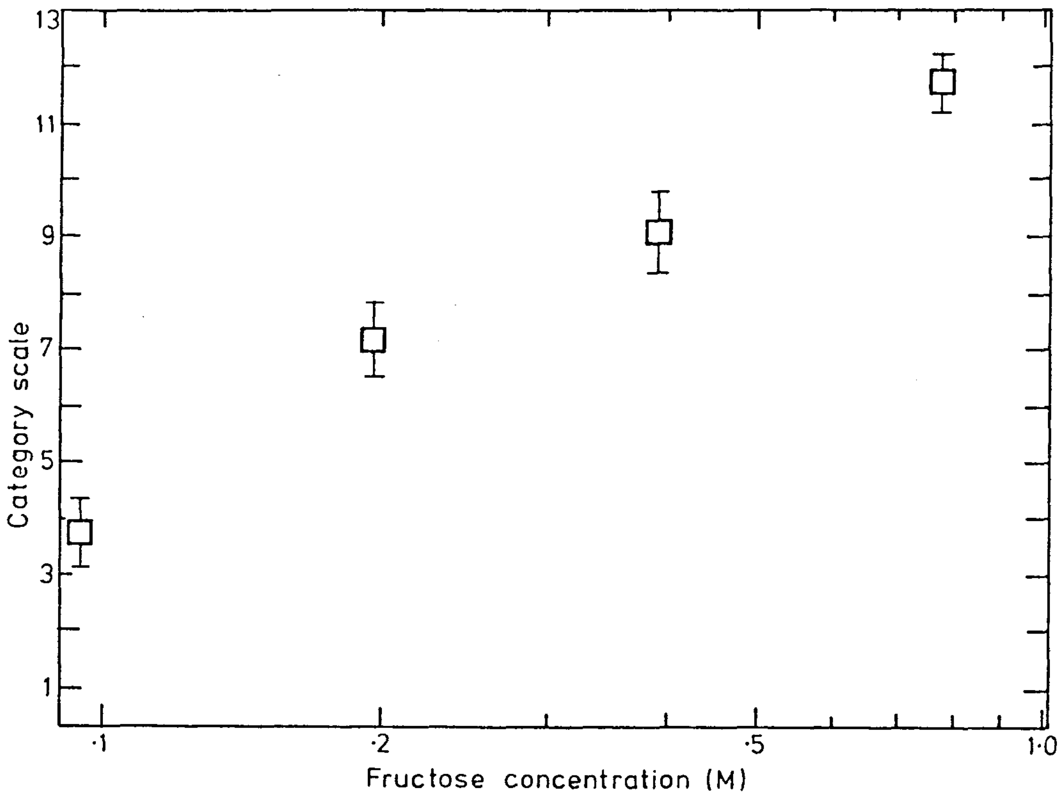

Figure 26. Category ratings for fructose (±2 SE) in the single presentation paradigm.

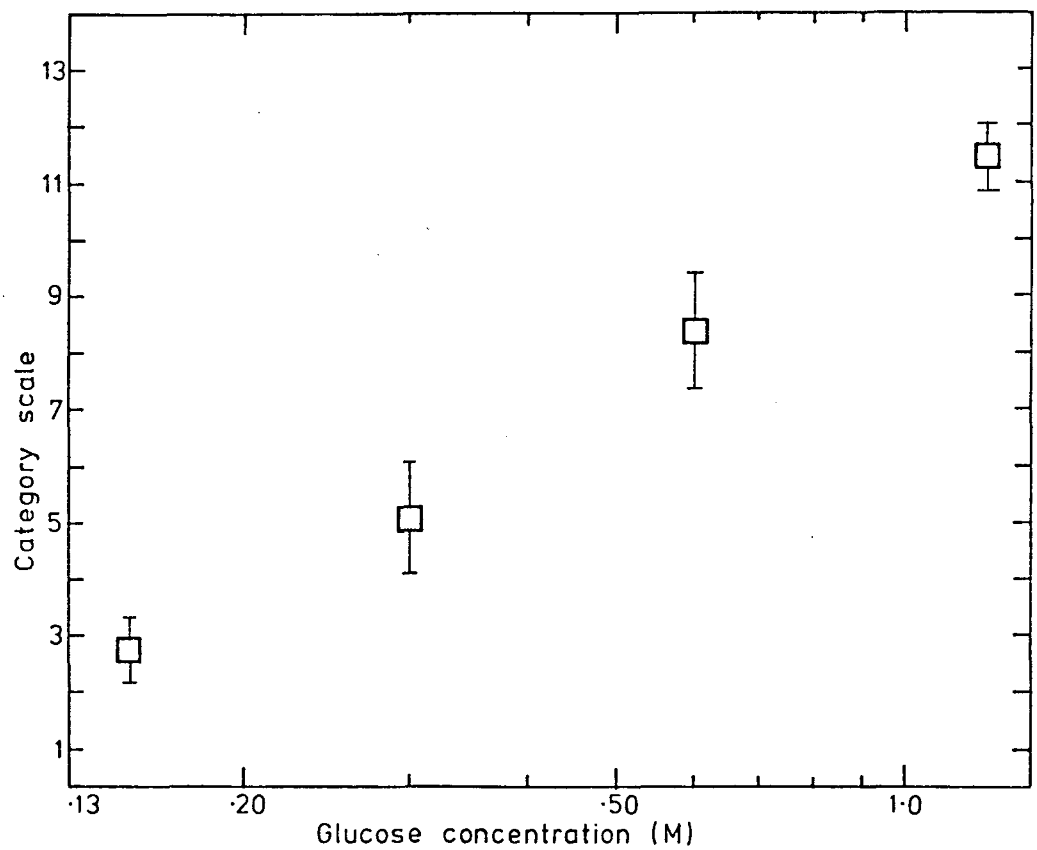

Figure 27. Category ratings (±2 SE) for glucose in the single presentation paradigm.

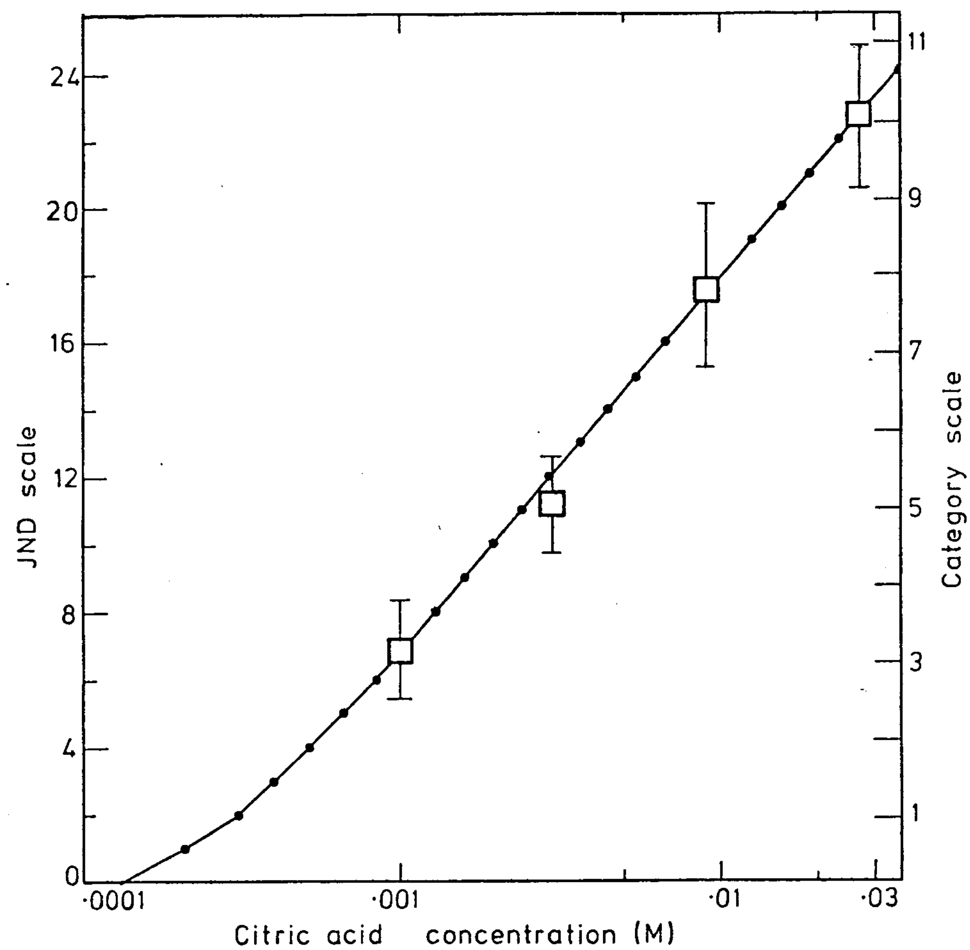

Figure 28. JND scale (left hand ordinate) and the category scale (right hand ordinate) for citric acid. The mean category ratings (±2 SE) from the single presentation correspond to the JND scale.

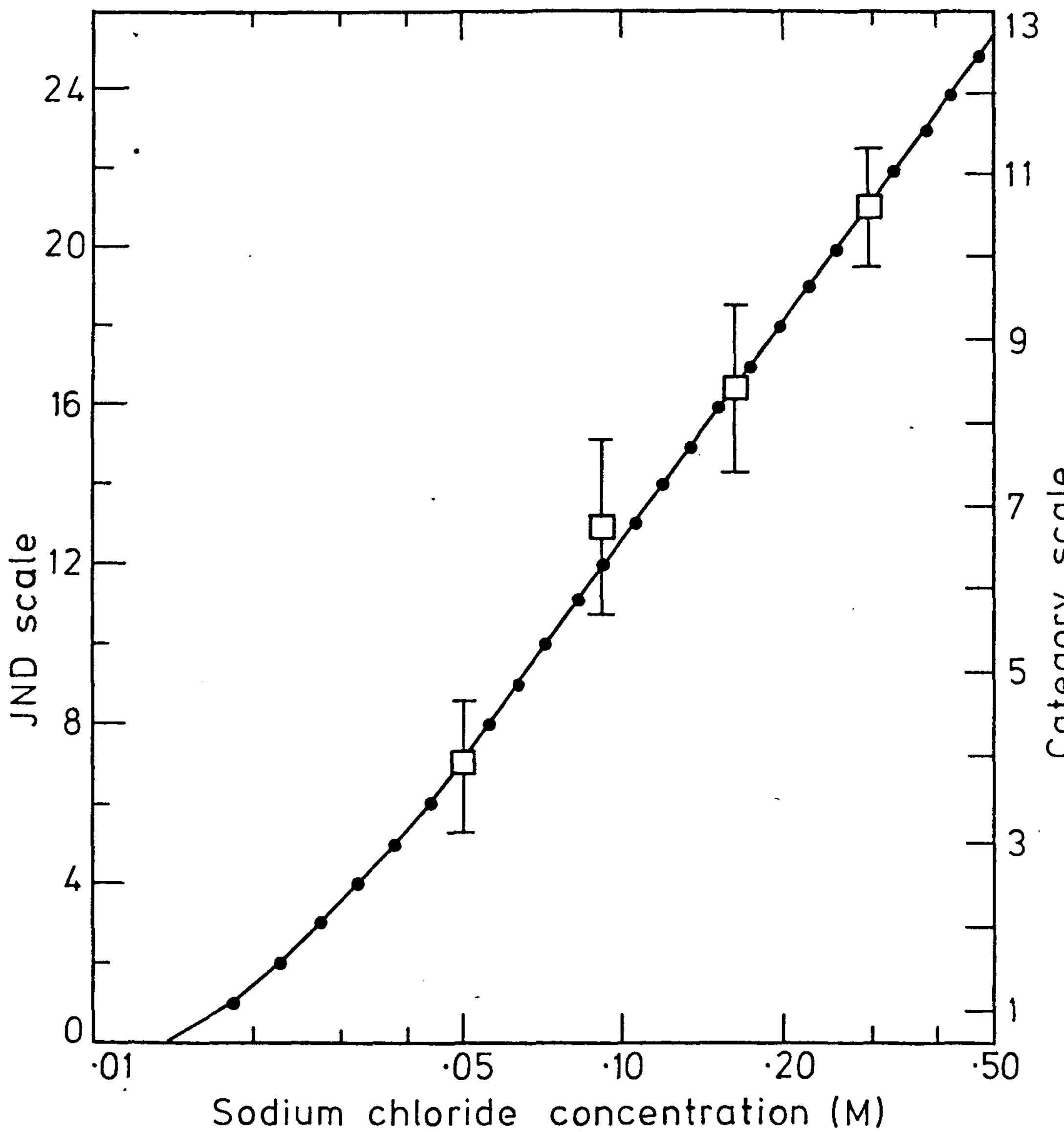

Figure 29. JND scale (left hand ordinate) and category scale (right hand ordinate) for sodium chloride. The mean ratings (±2 SE) from the single presentation conform to the JND scale.

Results and Discussion

The mean ratings (± 2 SE) for fructose, glucose, citric acid, and sodium chloride are given in Figures 26, 27, 28, and 29, respectively. The category scales for citric acid and sodium chloride have been fitted to their respective JND scales by the same method as in Experiments I and V.

The absolute ratings have changed slightly, but the shapes of the sweetness functions for fructose and glucose are similar to those specified earlier by the iterated multiple presentation design (Figures 20 & 21; the rating for the second lowest fructose concentration in Figure 26 is inexplicably discrepant). And, the category ratings for citric acid and sodium chloride coincide with their respective JND scales, once again supporting the JND-category scale isomorphism shown earlier with conventional multiple presentation (Figures 17 & 18). The estimates in Figures 26-29 have larger error than in the previous corresponding multiple presentation condition; however, it should be noted that they are based on fewer (24) responses.

An Overview

In his concluding remarks on methodological bias, Poulton (1979) wrote: "unfortunately, at the present time most investigators simply collect biased data without attempting to correct for the biases or even to measure and report them" (p. 801). This state of affairs cannot be condoned, especially since there are methods available to counteract bias. However, it is also fair to say that methodological bias did not have a drastically misleading effect on the foregoing category scales of taste intensity - even when the conventional multiple presentation paradigm was used. As noted earlier, the presentation of only a small number of stimuli with replication between, not within, sessions may have precluded the usual contextual influence.

The present experiments, then, were hardly an adequate test of the efficacy of the single presentation paradigm as a means of preventing bias, since there was little bias to begin with. The single presentation approach is very time consuming (Experiments V and VI took 8 weeks to complete) and is probably more appropriate for hedonic type scaling where the experimental context has been shown to be a problem. Nevertheless, Experiments V and VI do provide further evidence of the reliability and versatility of category rating.

11

RATING AS POSITIONAL JUDGMENT

The only common feature of rating scales is that they consist of a closed response continuum, as distinct from the open-ended response continuum used in magnitude scaling. In some rating tasks, subjects are presented with a physical scale on which to respond (as in the present work e.g., Figures 13 & 23). In other cases, a physical scale is not presented; the top and bottom anchors are specified (e.g, 1 and 10) and subjects are required to make their responses on a purely notional category scale (see Figure 31). Matters such as the number of categories to be used (if indeed there are any categories at all), and the presence or otherwise of verbal labels/numbers on the response scale are at the discretion of the investigator.

In a useful study, Montgomery (1975) systematically manipulated the experimental conditions to investigate the effects of several methodological factors on category and magnitude scales of loudness. The first notable result was that the instructions given for the assignment of responses (i.e., "ratio" or "difference" instructions) had no effect on the psychophysical function obtained. This runs contrary to previously held beliefs about the importance of instructions in scaling (Torgerson, 1960, 1961). Montgomery did find, however, that the the "openness" of the response scale (e.g., no upper limit as in magnitude estimation vs. a closed continuum as in category rating), and the range of the response scale (e.g., wide 1-100 vs. narrow 1-10) both influenced the responses, and to approximately the same extent. An open response scale with a wide range produced the magnitude scale of loudness; a closed response scale with a narrow range gave the category scale of loudness; and a closed response scale with a wide range gave a compromise between the two.

The effect of response range has also been noted in experiments on tactile intensity (Gibson & Tomko, 1972). Gibson and Tomko found that, when the number of categories on the category scale was expanded to match the range of numbers used in magnitude estimation, the two scaling methods provided similar results (although, as shown in the re-evaluation by Poulton, 1979, there was still a tendency for the expanded category scale to be perceived as linear).

At this point it is pertinent to return to the theoretical explanation for the difference between category and magnitude scales, postulated earlier in Chapter 6. Here it was held that category scales (and graphic rating scales) are valid because they correspond to the linear psychophysical continuum, position on a line. Provided the subject is presented with a line of finite length on which to mark a response (i.e., a closed response scale), the number of intervening categories should be of no concern. So, the present model is consistent with Montgomery's finding that closure of the scale is crucial, but not with the other conclusion that range (number of categories) is important. Why did Montgomery (1975) and Gibson and Tomko (1972) find the range of the category scale to be important?

One possible explanation is that these investigators used notional category scales, not physical category scales, as specified by the present model. It is not possible to deduce from their reports exactly what form of scale was used, but it would seem more probable (e.g., in the 1-100 category scale of Montgomery) that the scales were notional, rather than physical. The theory in Chapter 6 dictates that, without the physical presence of a scale, there is no guarantee that subjects will perceive the numbers in a linear manner. In fact, when the numerical response range is large (e.g., 1-100), there is a possibility of subjects perceiving the number continuum logarithmically, as in magnitude scaling. This possibility is explored in the next experiment, by comparing physical and notional scales in the measurement of sucrose sweetness.

EXPERIMENT VII

Method

Subjects. A different panel of 24 subjects served in each part of the experiment.

Stimuli. Reagent grade sucrose in distilled water. The concentrations were always .0625, .1250, and .2500 Molar (same as in Experiment IVb).

Response scale. The graphic response scale for Experiment VIIa was 100 mm long and is shown in Figure 30. The response scale for Experiment VIIb was identical, except wherever the number "10" appears, it was replaced by "100". Figure 31 illustrates the response form for Experiment VIIc. The response form for Experiment VIId was as shown in Figure 31, but with "10" replaced by "100". The only difference between Figures 30 and 31 is that, in the first case, the response scale is physically present, while in the second it is purely notional. As is evident from these Figures, the instructions were kept as identical as possible throughout the experiment.

Stimulus end anchors were not presented. Preliminary work indicated that presentation of a top anchor (say .5000 Molar sucrose, or higher) caused complications with adaptation: subjects commented that, after tasting the top anchor, the

Sweetness Intensity

Date____________ Time____________ Name________________________

Instructions: PLEASE READ

Sweetness may be considered as a continuum, bounded at the bottom end by 0 (0 = No sweetness at all) and at the top end by 10 (10 = Extremely sweet).

Taste these solutions in the order specified below and mark each line at the position which best describes the sweetness of each solution on the 0 - 10 sweetness continuum. After you have tasted each sufficiently, spit it out in the paper cup provided, then rinse thoroughly with water before proceeding to the next sample.

Order of tasting

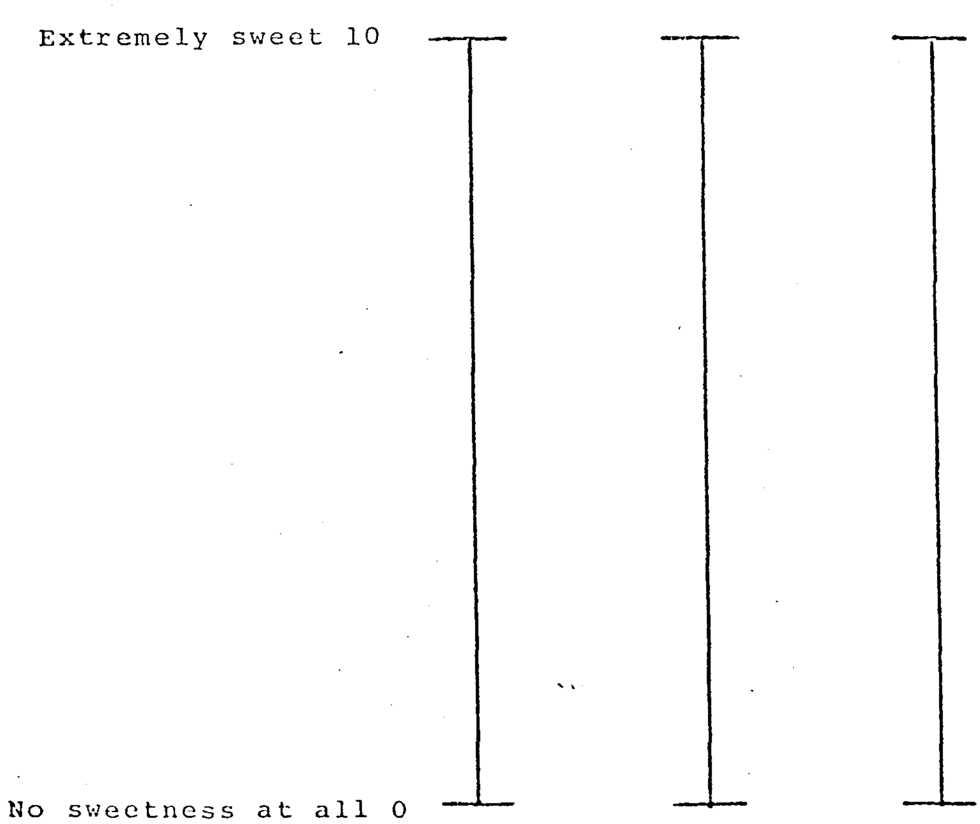

Figure 30. The response scale used in Experiment VIIa.

Sweetness Intensity

Date__________ Time__________ Name______________________________

Instructions: PLEASE READ

Sweetness may be considered as a continuum, bounded at the bottom end by 0 (0 = No sweetness at all) and at the top end by 10 (10 = Extremely sweet).

Taste these solutions in the order specified below and assign each solution a number, such that this number reflects the sweetness of the solution on the 0 - 10 sweetness continuum. After you have tasted each sufficiently, spit it out in the paper cup provided, then rinse thoroughly with water before proceeding to the next sample.

Order of tasting		

Sweetness judgement

(0 = No sweetness at all
10 = Extremely sweet)

Figure 31. The response scale used in Experiment VIIc.

other stimuli seemed weak and difficult to discriminate. Therefore, conceptual anchors were used (e.g., "extremely sweet" = 10), with a narrow range of stimuli. Ratings for the three stimuli used were shown earlier (Experiment IVb) to be well within the range specified by these anchors, precluding end-effects.

Procedure. Each of the Experiments VIIa-VIId was run twice, with a different panel of 12 subjects in each case (i.e., there were 8 series of experimental sessions, conducted in random order). Each series consisted of four replicate sessions with the same 12 subjects, the replicate sessions within each series being held on consecutive days. This gave a total of 96 judgments for each of the three stimuli in each experiment. With this design it was possible to ensure that, for each subject, there was a break of at least four weeks between series, thereby precluding the occurrence of transfer effects (i.e., effectively simulating a separate groups design; cf. Appendix A). Other procedural details were as for Experiment I.

Results and Discussion

The data from Experiment IVb were included in the analyses of Experiment VII. Experiment IVb was in all respects identical to Experiment VIIa, except that the response scale (in this case 95mm long) was subdivided into categories (see Figure 23). For Experiments VIIa and VIIb, the distances between the response marks and the bottom end of the 100mm scale were measured to the nearest millimetre, and these served as the response scores. The response scores from Experiments IVb and VIIc were multiplied by 10 to permit inter-Experiment comparison.

First, an analysis of variance was carried out on the raw data (Appendix B). In the analysis of variance model used, the block structure was sessions within series and the treatment structure was Experiments crossed with sucrose concentrations. There were significant differences between scores for the three sucrose concentrations, F (2, 70) = 3299.76 $p<.001$, between Experiments, F (4, 5) = 14.77 $p<.01$, and there was also a significant Experiment x concentration interaction, F (8, 70) = 2.67 $p<.05$.

However, a subsequent check revealed that the variances were not sufficiently homogeneous to allow this analysis (var_{max}/var_{min}=25.40 $p<.05$; Pearson & Hartley, 1970, Table 31); therefore it was repeated using an arc sine$\sqrt{p}$ transform which did render the variances homogeneous. This time the main effects for concentration, F (2, 70) = 2968.38 $p<.001$, and for Experiments, F (4, 5) = 15.44 $p<.01$, remained much the same, but the Experiment x concentration interaction dropped below significance, F (8, 70) = 1.03 (Appendix B).

Overall mean scores from Experiments IVb, VIIa, VIIb, VIIc, and VIId were respectively: 42.0, 40.4, 43.0, 40.0, and 31.1. Clearly, the only discrepant overall mean is that from Experiment VIId; both analyses of variance revealed the others to be well within the LSD (least significant difference). This discrepancy is also obvious in Table 2, which gives the mean ratings for each concentration level. The ratings from Experiment VIId are significantly different from the others, which are closely grouped within each column.

Examining Table 2 systematically, the first point to note is that the 0-10 category scale (Experiment IVb) and the 0-10 graphic scale (Experiment VIIa) provide the same result: it apparently made no difference whether subjects were restricted to the use of integers or allowed to mark position on a line. This result is consistent with the theoretical explanation advanced in Chapter 6: whether or not categories are used, subjects simply make positional judgments on the scale. The categories themselves merely serve as markers, or calibration points. This interpretation suggests that specifically instructing subjects to regard all categories as representing equal subjective jumps (a carryover from the "method of equal appearing intervals") is superfluous, since the categories themselves do not play an essential role in scale usage.

Table 2

Mean Ratings* from Experiments IVb and VIIa-VIId

Experiment	Molar concentration of sucrose		
	.0625	.1250	.2500
IVb (category scale, 0-10)	15.8	39.7	70.7
SE	1.0	1.5	1.8
VIIa (graphic scale, 0-10)	13.5	37.5	70.3
SE	1.2	1.8	1.8
VIIb (graphic scale, 0-100)	16.7	40.3	71.9
SE	1.6	2.2	1.8
VIIc (notional scale, 0-10)	14.3	36.5	69.4
SE	.8	1.3	1.5
VIId (notional scale, 0-100)	9.0	26.2	58.1
SE	.6	1.3	1.7

* Ratings from the 0-10 scales have been multiplied by 10, to facilitate comparison.

In a number of functional measurement studies, Anderson (e.g., 1979b) has found responses from category scales and graphic rating scales to be closely related. Similarly, in a study on odour perception, Gregson, Mitchell, Simmonds, and Wells (1969) showed that, when subjects were required to respond on a physical category scale, it made little difference whether verbal or numerical labels were attached to the response scale. In fact, even when the numerical labels were deliberately chosen to represent nonlinear increments, a somewhat confusing situation for subjects, the response behaviour was still much the same (see also Stevens & Galanter, Figure 8B). These findings add weight to the hypothesis that subjects are making positional judgments only. The variability of the ratings is much the same in Experiments IVb and VIIa, indicating there is no difference in precision between the two methods. The category scale of Experiment IVb does, however, offer one practical advantage: the experimenter can read the numerical responses directly from the response sheet and there is no measuring involved.

Expanding the range of numbers on the graphic response scale, as in Experiment VIIb, appears to have had no effect on rating behaviour. This result is also consistent with the prediction of the present theory, i.e., provided a response continuum is physically presented, the actual numerical anchors are of secondary consideration. The estimates of variability are of similar magnitude to those in Experiments IVb and VIIa.

The mean ratings from Experiment VIIc show that, when the response range on the notional scale was limited (0-10), subjects performed much the same as when a physical scale was presented. This finding supports the contention of Banks and Hill (1974), who claim that the number continuum from 1 to 10 is perceived as linear, and also Poulton (1979) who states that single digit numbers are perceived as linear because there is no step change in order of magnitude. There was, however, some stereotyping of responses (e.g., 38 of the 96 responses to the .0625 Molar concentration were given as "1"), resulting in a spuriously low estimate of variability. The stereotyping here, and for the .0625 M stimulus in Experiment VIId for which the common responses were "5" and "10", was the main reason for the data failing the test for homogeneity of variance.

The mean ratings from Experiment VIId conform to a different pattern, indicating that a notional 0-100 scale is not perceived in the same way as a notional 0-10 scale. In absolute terms the ratings are lower; furthermore, the significant Experiment x concentration interaction in the initial analysis of raw data suggests they are not linearly related to the other sets of scores in Table 2. Whereas in Experiments IVb and VIIa-VIIc the mean ratings are approximately equidistant, in Experiment VIId there is a tendency for them to be geometrically spaced, consistent with a tendency for subjects to perceive the notional 0-100 scale in a logarithmic manner. There did not, however, appear to be any obvious difference between the response distributions in Experiments VIIb and VIId.

The trend of these results might explain the findings of Montgomery (1975) and of Gibson and Tomko (1972): when the number range becomes large, it is possible the notional category scale takes on some of the properties of a magnitude scale. In practical terms, the best way of avoiding possible bias is to always present a physical response scale. While the numbers 0-10 appeared to work satisfactorily as a notional scale in this instance, there is no guarantee this will always hold.

The claim that the graphic rating scale (or analogue scale) is valid psychophysically is perhaps not surprising. This type of scale has been used for a variety of purposes in everyday life for a long time; and without any knowledge of its underlying psychophysics. All instruments and devices with linear controls (e.g., hi-fi equipment, domestic appliances) utilize this form of psychophysical continuum. "Circular" controls (e.g., knobs, clocks, dials) are even more widely used. The psychophysical continuum underlying these controls is visual angle which, as held in Chapter 6, is also a linear psychophysical continuum.

What Happens in Cross-Modality Matching?

In contrast to the earlier claim of intra-modality equivalence of JNDs in taste (Chapter 9), the present model would suggest that JNDs are not subjectively equal across modalities. For example, with the taste continua investigated earlier there were only 24 JNDs between threshold and a point close to "extremely strong", whereas for loudness, a sound 24 JNDs above threshold corresponds to about 30 dB - a "whisper" (Stevens, 1975, p. 33). Inspection of other continua reveals a wide disparity between the number of JNDs between threshold and a subjectively intense level suggesting that, in cross-modality matching (CMM), subjects do not match sensations on the basis of equal numbers of JNDs.

But the positional matching hypothesis can be applied more generally to CMM. For example, a solution would be considered extremely sweet if its sweetness occupied a position on the subjective sweetness continuum near the position held by the sweetest substance ever tasted (experiental range and frequency effects will determine the endpoints/anchors for each individual). Likewise, a sound will be considered extremely loud if it occupies a corresponding position on the loudness continuum. In a CMM task, these two positions would be matched.

This interpretation is consistent with the results of CMM studies (Stevens, 1959, 1966), where the exponents reflect the ratio of the log physical ranges of the continua matched. It has also been recently proposed in loudness matching (the "proportional-JND" explanation; Lim, Rabinowitz, Braida, & Durlach, 1977; Houtsma, Durlach, & Braida, 1980), following earlier work by Riesz (1933). The concept of sensation having position on a subjective continuum was, of course, proposed by Thurstone more than 50 years ago (Thurstone, 1927a). This explanation of CMM is promising and warrants a more detailed reanalysis of all available data.

Could the subjective size of the JND vary across continua in such a way that there is a constant subjective range, irrespective of the mode of stimulation? Teghtsoonian (1971) takes this position, and at first glance the evidence is impressive. Teghtsoonian goes on to cite evidence for a constancy in Ekman's fraction (Brentano's assumption). In terms of the present model, however, Teghtsoonian's argument is not tenable because it is based around the power function: in the present model there is no Ekman's fraction. In a later contribution, Teghtsoonian (1974) shifts the emphasis from Ekman's law to Weber's law, but he makes the crucial assumption that JNDs are subjectively equal across continua. As demonstrated above, this is not upheld by CMM studies.

In a simple but perceptive reanalysis, Poulton (1979, Figure 10) shows that evidence for a constancy in subjective range is weakened when the subjective ranges of the original data are replotted against their corresponding stimulus ranges: the impressive evidence obtained by Teghtsoonian is an artifact of transforming subjective ranges into exponents. In Figure 10 of Poulton (1979) there is in fact evidence of considerable variation between subjective ranges. Of course, the variation or otherwise of subjective range does not directly bear on the concept of CMM as positional matching, because it is relative positions that are matched. As shown earlier, matching four concentrations of sucrose to a 13 point scale gives the same result as matching the concentrations to an 11 point scale, because the relative positioning of the ratings stays constant. The matching of two continua of unequal subjective range can be regarded as a form of stimulus equalizing bias, illustrated in Figure 1B of Poulton (1979).

12

STIMULUS INTEGRATION OF SWEETNESS

The foregoing work supports the present model: the predicted agreement between JND and category scales has been confirmed. However, the experiments so far have involved traditional univariate designs. To complete the picture, it needs to be shown that the functional measurement paradigm also produces the same psychophysical function for the sweetness of sucrose as was obtained earlier in Chapter 9. Concordance would, of course, add further weight to Fechner's assumption, just as it did in the case of loudness (Chapter 7).

It appears little has been done in the functional measurement analysis of taste. Klitzner (1975) reports one such study, but it was concerned with the integration of hedonic tone of stimuli, not with sensory intensity in the psychophysical tradition (cf. also Shanteau & Anderson, 1969). On this point, mention of the "psychophysical tradition" raises a difference in orientation between the present approach and that normally taken in functional measurement studies. In functional measurement, the integration task is of primary interest (Anderson, 1974a) and whether or not the data meet the parallelism criterion, useful information is still obtained. In the present context, however, the psychophysical law takes precedence, and this relationship can be estimated only if the data meet the parellelism test.

One line of attack might be to explore the additivity (or otherwise) of mixtures of two stimuli of identical taste quality, i.e., a situation where the integration task is "natural". This requirement suggests the simple sugars sucrose and fructose. Both are commonly recognized as pure sweet stimuli, free from side-tastes; in addition, considerable information on these sweeteners has already been obtained in earlier experiments, and this is available as a cross-validational base if required.

There have been several studies on the sweetness of sucrose-fructose mixtures, both in the psychophysical and food laboratory (e.g., Bartoshuk, 1977; Bartoshuk & Cleveland, 1977; Hyvonen, 1980; Moskowitz, 1974; Stone & Oliver, 1969). No clear picture has emerged. Bartoshuk, and Bartoshuk and Cleveland claim that simple additivity (i.e., the perceived intensity of the mixture is equal to the sum of the perceived intensities of the components) cannot occur in mixtures of similar tastes, but these authors appear to be confused on the issue of additivity. They claim that simple additivity can occur only if the psychophysical functions are linear, since only in this case does a single taste "add" to itself in a simple algebraic way. However, this point would seem to have nothing to do with the question of additivity. Simple additivity is logically feasible regardless of the shapes of the individual psychophysical functions in a mixture.

In contrast, Stone and Oliver (1969), Moskowitz (1974), and Hyvonen (1980) claim support for additivity, and also for synergism at certain concentrations, i.e., the perceived intensity of a mixture is greater than the sum of the perceived intensities of the components (there is some confusion over the definition of synergism and this will be discussed later). However, these studies should be treated with caution, since they were conducted with the response technique of magnitude estimation. Furthermore, in all cases the mixture combinations were somewhat arbitrary; factorial designs were not used. The next experiment explores the question of sucrose-fructose additivity using the functional measurement approach.

EXPERIMENT VIII

Method

Subjects. Fourteen subjects were used.

Stimuli. The 16 stimuli consisted of mixtures of sucrose and fructose in distilled water, varied in a 4 x 4 factorial design. Preliminary work suggested the same concentrations for both sucrose and fructose: 1.00, 1.65, 2.73, and 4.50% w/v (log spacing; % w/v units will be used in this case for convenience).

Response scale. As shown in Figure 13.

Procedure. Previous work at this laboratory has shown that 16 stimuli cannot be tasted reliably at a single session without the problem of taster fatigue. A balanced incomplete block design was therefore employed (plan 12.2, Cochran & Cox, 1950, p. 360). This involved subjects tasting four stimuli per session at a total of 20 sessions: one session per day, four sessions per week, for five weeks. Thus, each taster completed five replicate judgments of each of the 16 stimuli. Other procedural details were as for Experiment I.

Results and Discussion

Two analyses of variance were performed on the data (Appendix C). In the model used, the block structure was sessions and the treatment structure was sucrose crossed with fructose concentrations. The first analysis adjusted scores for session to session variation, while in the second the data were treated as if obtained from a conventional complete block design. The mean scores in each case were almost identical, as were the corresponding F ratios in each analysis, indicating there was no sessions or context effect (hardly surprising considering the lack of context effects found in the experiments of Chapter 10). In subsequent analyses, therefore, the design was regarded as a complete block.

Figure 32 gives the mean response scores in a factorial plot typical of functional measurement analysis. At first glance there appears to be support for parallelism; however, the analysis of variance revealed a significant sucrose-fructose interaction, $F\ (9,\ 896) = 6.90\ p<.001$; similarly for the incomplete block design, $F\ (9,\ 877) = 5.38\ p<.001$. This interaction was increased by application of a log transformation to the data, $F\ (9,\ 877) = 10.71\ p<.001$.

Figure 32 shows evidence of a convergence toward the right hand side of the plot, especially at the higher fructose concentrations. This suggested a reanalysis without the top

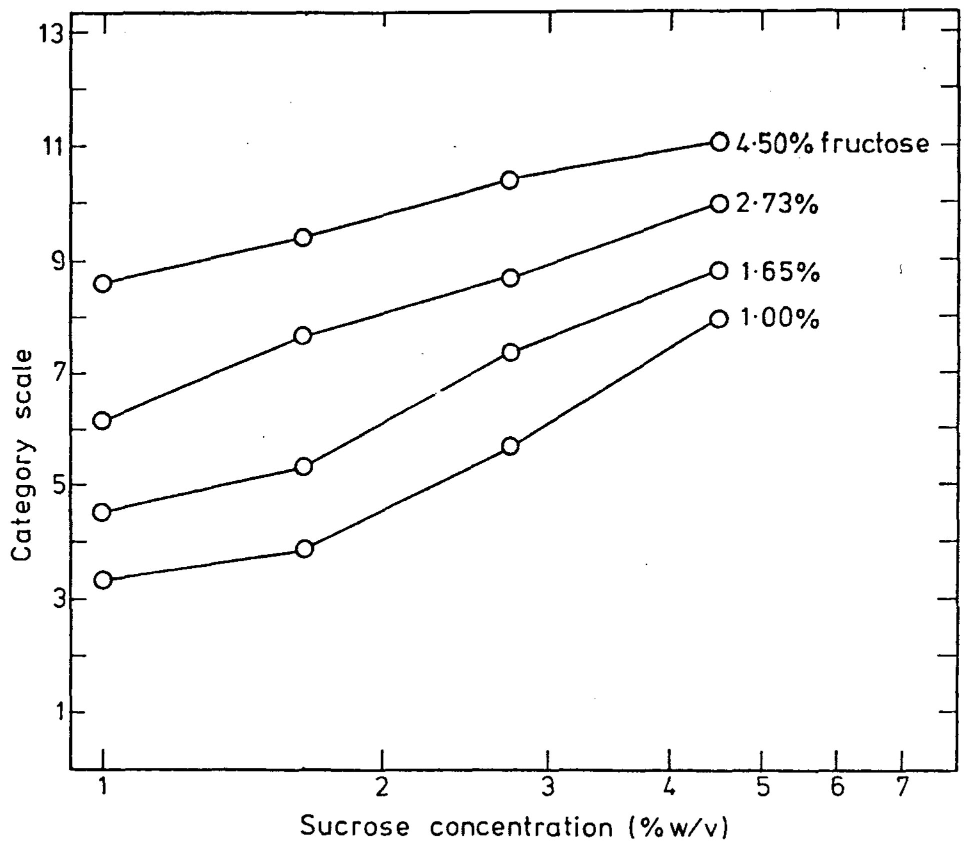

Figure 32. Mean response scores for the 4 x 4 factorial design of Experiment VIII. The four sucrose concentrations are plotted on the abscissa and each curve corresponds to a different level of fructose.

(4.50%) level of fructose. The interaction of this 4 x 3 design dropped, but was still significant, $\underline{F}$ (6, 672) = 3.53 $\underline{p}<.001$. However, when the two top levels of fructose were omitted from the analysis the interaction dropped out completely, implying additivity, $\underline{F}$ (3, 448) = .73. This procedure was then repeated, omitting successively the top, and two top, concentrations of sucrose from the analysis. Once again the interaction in the 3 x 4 design was lower, but still significant, $\underline{F}$ (6, 672) = 3.55 $\underline{p}<.001$; however, it dropped below significance in the 2 x 4 design, $\underline{F}$ (3, 448) = 2.23. Thus the effect was symmetrical: additivity held for all four concentrations of one sweetener when mixed with the lowest two levels of the other.

Individual analyses. Separate analyses of variance were performed on the data from individual subjects. Somewhat surprisingly, these revealed the sucrose-fructose interaction to be significant for only two of the 14 subjects. But, when the group data were reanalysed with these two subjects omitted, the sucrose-fructose interaction dropped only slightly and was still significant, $\underline{F}$ (9, 749) = 4.52 $\underline{p}<.001$. This outcome suggests the nonadditive trend was consistent over subjects, even though it was not pronounced enough to reach significance at the individual level. Further individual analyses for both the 4 x 2 and 2 x 4 subdesigns showed none of the sucrose-fructose interactions to be significant in these cases.

The psychophysical law. The limited parallelism of Figure 32 can provide estimation of the psychophysical law for sucrose and for fructose from the marginal means of the 4 x 2 and 2 x 4 subdesigns, respectively. These functions are plotted in Figure 33, positioned to facilitate comparison with Figure 22. There is good agreement between Figures 22 and 33 over their common concentration range. The curvature in the sucrose function at low concentration is as dictated by the sucrose Weber function, and was evident in the JND scale of Figure 15. The concordance between the JND scale and category scale for sucrose, claimed earlier in Chapter 9, is corroborated by this finding, as is Fechner's assumption of subjective equality of JNDs.

Failure of parallelism in the overall design is equivocal. As Anderson (1979b) comments, interpretation is difficult and often relies upon collateral data. The first explanation is that the results genuinely reflect what is happening: that is, the sweetnesses of sucrose and fructose are additive up to a certain level (or are additive as determined by the parallelism criterion; there may have been a trace of nonadditivity which the design was not capable of detecting), then either a stimulus interaction occurs (unlikely with these two compounds), or else the perceptual system "saturates" and the sweetness of the mixture obeys the law of diminishing returns. However, the effect could also be due to a bias or nonlinearity in the upper end of

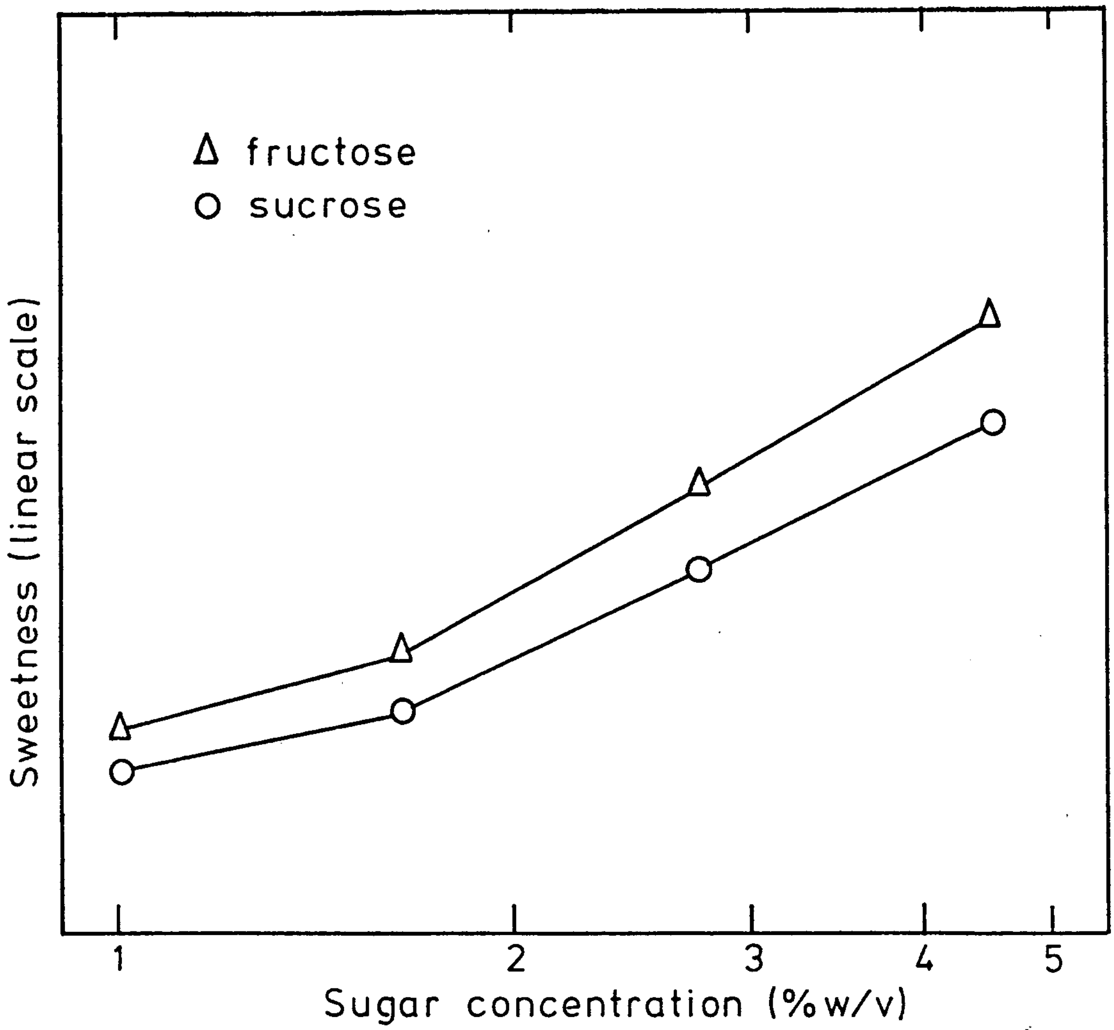

Figure 33. Psychophysical functions for sucrose (circles) and fructose (triangles) obtained from the parallel subdesigns of Experiment VIII. The functions correspond well with those obtained earlier in Figure 22.

the response scale. This would seem unlikely, given the now impressive support for rating scale linearity in many different applications, nevertheless it cannot be dismissed. The overall mean response to the 16 stimuli was 7.44, higher than the midpoint of the response scale, so it is possible a centering bias may have influenced the result. Or, perhaps the finding was simply idiosyncratic, in which case replication is necessary. This indeterminacy suggested another experiment with different subjects, a different experimental design, and slightly lower concentrations.

EXPERIMENT IX

Method

Subjects. Fifteen subjects were used (a different panel from Experiment VIII).

Stimuli. The 6 stimuli consisted of sucrose and fructose in distilled water, varied in a 3 x 2 factorial design. The concentrations of sucrose were 1.0, 2.0, and 4.0% w/v; for fructose 1.0 and 4.0% w/v.

Response scale. As shown in Figure 13, but with two more response columns.

Procedure. Subjects tasted all 6 stimuli at four separate sessions, held on consecutive days (complete block design). Order of tasting was randomized. Other procedural details were as for Experiment I.

Results and Discussion

Mean response scores are shown in Figure 34 and follow the same pattern as Experiment VIII. An analysis of variance on all data (Appendix D) revealed a significant sucrose-fructose interaction, $\underline{F}$ (2, 267) = 5.18 $\underline{p}$<.01, which increased after application of a log transformation, $\underline{F}$ (2, 267) = 36.03 $\underline{p}$<.001. In keeping with the previous experiment, however, the interaction disappeared when the data were reanalysed without the top (4.0%) level of sucrose, $\underline{F}$ (1, 177) = .75, implying once again that simple additivity may hold up to a certain sweetness level. Furthermore, in this experiment the overall mean response score was 6.98, extremely close to the scale midpoint, discounting a centering bias.

Individual analyses. Separate analyses of variance were carried out on data from individual subjects, as in Experiment VIII, and the sucrose-fructose interaction was found to be significant for three of the 15 subjects. This time, however, when the group data were reanalyzed with these three subjects omitted, the sucrose-fructose

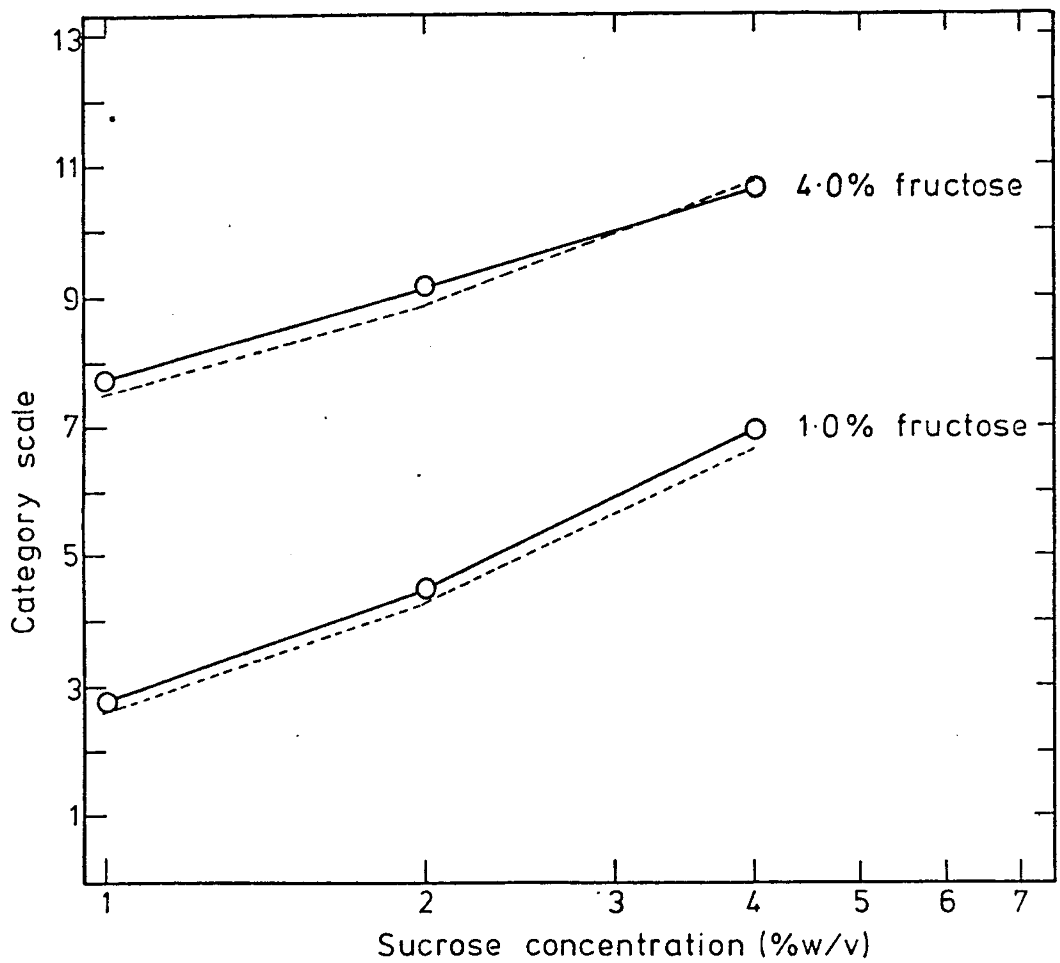

Figure 34. Mean response scores for the 3 x 2 factorial design of Experiment IX. Sucrose concentrations are given on the abscissa, and the curves correspond to the two levels of fructose. The broken lines join the mean scores from those data whose interactions were not significant at the individual level.

interaction dropped below significance, $\underline{F}$ (2, 213) = 1.62. The mean response scores for these 12 subjects are shown by the broken lines in Figure 34. Although this finding is contrary to that obtained earlier in Experiment VIII, inspection of the broken lines shows there is nonetheless still a convergent trend toward the right hand side.

The replication in Experiment IX of a similar pattern of results with a different experimental design and different individuals, tends to suggest that the sweetness of sucrose and fructose is effectively additive up to a certain level, but then the sweetness of the mixture falls below that expected from additivity. However, this conclusion is by no means definitive, and more experimental work is required for confirmation. The use of still higher concentrations may help resolve the issue, since nonadditivity, if it does occur, should then be more obvious.

The Synergism Paradox

Synergism was described earlier in psychological terms. However, researchers in other disciplines, food scientists, pharmacologists, take a different view: for them, synergism is defined in physical units. For instance, if a mixture of 4.0% sucrose and 4.0% fructose produces a sweetness that exceeds that of 8.0% sucrose (or 8.0% fructose for that matter), then synergism is deemed to operate (cf. Hyvonen, 1980).

The above example is paradoxical when viewed in physical (%w/v) units. It is well known, as shown earlier in Figure 22, that fructose is sweeter than sucrose at all concentrations. On this basis, if half of the fructose in an 8% fructose solution were replaced by sucrose, then the sweetness of the solution might be expected to drop. On the contrary, the sweetness increases. This effect was demonstrated with 18 visitors to the laboratory. All declared the mixture of 4% sucrose-4% fructose to be noticeably sweeter than 8% fructose.

The explanation is straightforward when viewed from the psychophysical standpoint. The sweetness functions for both sucrose and fructose approximate to Fechner's law (of diminishing returns), so for both compounds the sweetness at 4.0% is greater than half the sweetness at 8.0% (see Figures 15 & 22). Therefore, even though Experiments VIII and IX suggest that the sweetness of the 4.0% sucrose-4.0% fructose mixture actually fails additivity, its magnitude (score of 10.8 in Figure 34) is greater than that of either 8% sucrose or 8% fructose alone (scores of 8.5 and 9.2 respectively in Figure 22), still sufficient for the mixture to appear synergistic in physical terms.

This paradox highlights the necessity for a psychophysical orientation, and is analogous to the finding that the size-weight illusion is an illusion only when measured in physical terms (Anderson, 1972). In speaking of synergism in physical units, there is always the risk that the term conveys the general impression of "getting more out than went in", and is therefore misleading. Other cases of synergism in taste (Rifkin & Bartoshuk, 1980) might also be explained in this way, and perhaps even cases of synergism in pharmacology, if the dose-response curves are negatively accelerating and parameters are measured in physical units only, e.g., mg/Kg of body weight.

13

A UNIFIED PSYCHOPHYSICS?

Looking back on this Thesis, it is apparent that eclectism has prevailed. As stated at the outset, the present theory is a rearrangement of existing ideas: the empirical Weber function, but not Weber's law; Fechner's assumption, but not Fechner' law; Thurstone's concept of a subjective continuum; Stevens' concept of direct scaling, but none of Stevens' specific techniques. In terms of more recent developments, the present model leans heavily on information integration theory (Anderson, 1981), an approach which could well herald a new era in psychophysics.

Inspiration for the theoretical re-evaluation came from the failure of magnitude scaling techniques in the psychophysics of taste. Such failure is not obvious with, for example, loudness data, because of the excellent (yet deceptive) descriptive ability of the power function.

The advent of functional measurement, with its proper validity criteria, has confirmed the failure of magnitude estimation as a linear response technique. The proposition of logarithmic subjective number, supported empirically by a number of diverse investigations, explains this failure; furthermore, it has fundamental implications for variability on the subjective continuum. The proposition also suggests, ironically enough, that Stevens' law fails because the number continuum, as used in magnitude estimation, approximates to Fechner's law (cf. Ekman, 1964).

But it was inspection of the empirical Weber function for loudness, and the concomitant re-evaluation of the theoretical foundations of psychophysics, which provided the insight necessary for further resolution. From acceptance of Fechner's assumption follows the JND scale; from the JND scale follows a re-interpretation of prothetic and metathetic continua, and a theoretical explanation of category scale validity.

The experiments of this Thesis support the theory. There is good evidence for a JND scale-category scale isomorphism in taste. There is evidence that category rating is a robust and reliable direct scaling technique. There is preliminary support for category rating as positional judgment. Finally, there is a posteriori validation of category rating and the JND-category scale isomorphism by functional measurement analysis - albeit limited to two taste stimuli over a small concentration range.

The reassertion of Fechner's assumption - perhaps the primary claim of this Thesis - has implications for the fundamental measurement issue in psychology. If JNDs can be regarded as valid linear units of subjective magnitude, then, to reverse the fundamental objection of the British committee cited in the first Chapter, a meaning can indeed be given to the concept of addition as applied to sensation.

And what of the psychophysical law? In practice, it would seem that the most straightforward way of obtaining this relationship for a given modality would be to use a factorial design in the functional measurement paradigm. Provided the data can be shown to meet the parallelism criterion on some task, the properly validated psychophysical law will follow (at least over the stimulus range used in the experiment).

From a more theoretical angle, the present model dictates that the shape of a psychophysical function is determined by the shape of its underlying Weber function. While there is some pattern in these shapes (Figures 5 & 7), it is unlikely any single, simple mathematical formulation will be able to account for them (in a recent article Weiss, 1981 also takes this view). There is no point in a descriptive curve-fitting exercise unless the mathematics invoked bear some meaningful relationship to the psychophysics. History has shown that, no sooner are mathematical relationships fitted as purely descriptive aids, than they assume more fundamental theoretical significance.

In the present model, psychophysical functions can be regarded as integrated Weber functions, or integrated sensitivity functions. Where the receptor system is particularly sensitive to changes in the intensity of stimulation, the growth in subjective magnitude will be rapid; where it is less sensitive to such changes, the growth will be slow. To be of theoretical significance, a mathematical relationship would have to reflect these changes in the variety of empirical Weber functions.

The present theory suggests a substantial amount of further work: the experiments reported here represent but one possible strand of research. To begin with, the generality of the present findings must be confirmed by experimental work in other modalities. The tentative prediction that the model might better dovetail with neurophysiological findings should be checked, especially the speculation that discrimination in the sensory system might be absolute near threshold. Taking another angle, the indirect derivation of true ratio scales with a meaningful zero (Chapter 9) may have implications for the current dialogue on "ratios" vs. "differences" (Birnbaum, 1980; Rule & Curtis, 1980; Veit, 1980). It will now be possible to construct a factorial design which is known, a priori, to comprise a wide range of subjective ratios, and this may show whether "ratio" and "difference" judgments are truly monotonic (cf. Rule & Curtis, 1980).

The use of parallelism as a validity criterion opens the way for a more rigorous investigation of bias in scaling. For example, is the notional 0-100 scale (Experiment VIId, Chapter 11) sufficiently nonlinear to cause failure of parallelism in a situation where, with a physical rating scale, parallelism obtains?

In terms of the present work in taste, the next step is to employ factorial designs to explore further sweetener mixtures, acid mixtures, bitter mixtures, and mixtures of different taste qualities. This might shed more light on the operating characteristics of the taste system and should also have some practical value for food science.

In concluding their treatise on psychophysics, Baird and Noma surmise: "it is by no means easy to develop theoretical connections between the various psychophysical models. The total understanding of these connections was, of course, Fechner's goal and hence the goal of the field he founded in 1860" (p. 271). The goal of this Thesis was precisely that of Fechner: to what extent it has been realised, only time will tell. Nevertheless, the present theory does offer a means of cohering magnitude scales, interval scales, and scales derived indirectly by discrimination; and for this it lays tentative claim to unification.

REFERENCES

Amerine, M. A., Pangborn, R. M., & Roessler, E. B. Principles of sensory evaluation of food. New York: Academic Press, 1965.

Anderson, N. H. Functional measurement and psychophysical judgment. Psychological Review, 1970, 77, 153-170.

Anderson, N. H. Cross-task validation of functional measurement. Perception & Psychophysics, 1972, 12, 389-395.

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). San Francisco: Freeman, 1974. (a)

Anderson, N. H. Algebraic models in perception. In E. C. Carterette & M. P. Friedman (Eds.), Handbook of perception (Vol. 2). New York: Academic Press, 1974. (b)

Anderson, N. H. On the role of context effects in psychophysical judgment. Psychological Review, 1975, 82, 462-482.

Anderson, N. H. Note on functional measurement and data analysis. Perception & Psychophysics, 1977, 21, 201-215. (a)

Anderson, N. H. Failure of additivity in bisection of length. Perception & Psychophysics, 1977, 22, 213-222. (b)

Anderson, N. H. Algebraic rules in psychological measurement. American Scientist, 1979, 67, 555-563. (a)

Anderson, N. H. Introduction to cognitive algebra (Tech. Rep. CHIP 85). La Jolla: Center for Human Information Processing, University of California, San Diego, June, 1979. (b)

Anderson, N. H. Foundations of information integration theory. New York: Academic Press, 1981.

Attneave, F. Perception and related areas. In S. Koch (Ed.), _Psychology: A study of a science_ (Vol. 4). New York: McGraw-Hill, 1962.

Baird, J. C. A cognitive theory of psychophysics. II Fechner's law and Stevens' law. _Scandinavian Journal of Psychology_, 1970, _11_, 89-102.

Baird, J. C., & Noma, E. _Fundamentals of scaling and psychophysics_. New York: Wiley, 1978.

Banks, W. P., & Hill, D. K. The apparent magnitude of number scaled by random production. _Journal of Experimental Psychology Monograph_, 1974, _102_, 353-376.

Bartoshuk, L. M. Modification of taste quality. In G. C. Birch, J. G. Brennan & K. J. Parker (Eds.), _Sensory properties of foods_. London: Applied Science Publishers, 1977.

Bartoshuk, L. M., & Cleveland, C. T. Mixtures of substances with similar tastes. _Sensory Processes_, 1977, _1_, 177-186.

Bastian, J. M., McBean, D. McG., & Smith, M. B. _Fifty years of food research_. Melbourne: CSIRO, 1979.

Beebe-Center, J. G., & Waddell, D. A general psychological scale of taste. _Journal of Psychology_, 1948, _26_, 517-524.

Beidler, L. M. Properties of chemoreceptors of tongue of rat. _Journal of Neurophysiology_, 1953, _16_, 595-607.

Bendig, A. W., & Hughes II, J. B. Effect of amount of verbal anchoring and number of rating-scale categories upon transmitted information. _Journal of Experimental Psychology_, 1953, _46_, 87-90.

Berg, H. W., Filipello, F., Hinreiner, E., & Webb, A. D. Evaluation of thresholds and minimum difference concentrations for various constituents of wines. I. Water solutions of pure substances. _Food Technology_, 1955, _9_ (1), 23-26.

Birnbaum, M. H. Using contextual effects to derive psychophysical scales. _Perception & Psychophysics_, 1974, _15_, 89-96.

Birnbaum, M. H. Comparison of two theories of "ratio" and "difference" judgments. _Journal of Experimental Psychology: General_, 1980, _109_, 304-319.

Birnbaum, M. H., & Elmasian, R. Loudness "ratios" and "differences" involve the same psychophysical operation. Perception & Psychophysics, 1977, 22, 383-391.

Birnbaum, M. H., & Veit, C. T. Scale convergence as a criterion for rescaling: Information integration with difference, ratio, and averaging tasks. Perception & Psychophysics, 1974, 15, 7-15.

Borg, G., Diamant, H., Strom, L., & Zotterman, Y. The relation between neural and perceptual intensity: A comparative study on the neural and psychophysical response to taste stimuli. Journal of Physiology, 1967, 192, 13-20.

Box, G. E. P., Hunter, W. G., & Hunter, J. S. Statistics for experimenters. New York: Wiley, 1978.

Bronowski, J. A sense of the future. Cambridge, Mass.: MIT Press, 1977.

Cameron, A. T. The taste sense and the relative sweetness of sugars and other sweet substances (Sci. Rep. No. 9). New York: Sugar Research Foundation, 1947.

Carterette, E. C., & Anderson, N. H. Bisection of loudness. Perception & Psychophysics, 1979, 26, 265-280.

Christie, E. M. Tasting tests in the C.S.I.R.O., Australia. Laboratory Practice, 1964, 13, 630-637.

Christie, E. M. Practical aspects of tasting tests. Food Technology in New Zealand, 1966, 1, 175-180, 186.

Christman, R. J. Sensory experience. Scranton, N.J.: Intext, 1971.

Cobb, P. W. Weber's law and the Fechnerian muddle. Psychological Review, 1932, 39, 533-551.

Cochran, W. G., & Cox, G. M. Experimental designs. New York: Wiley, 1950.

Comrey, A. L. A proposed method for absolute ratio scaling. Psychometrika, 1950, 15, 317-325.

Corso, J. F. The experimental psychology of sensory behavior. New York: Holt, 1967.

Curtis, D. W. Magnitude estimations and category judgments of brightness and brightness intervals: A two stage interpretation. Journal of Experimental Psychology, 1970, 83, 201-208.

Curtis, D. W., Attneave, F., & Harrington, T. L. A test of a two-stage model of magnitude judgment. Perception & Psychophysics, 1968, 3, 25-31.

Curtis, D. W., & Fox, B. E. Direct quantitative judgments of sums and a two-stage model for psychophysical judgments. Perception & Psychophysics, 1969, 5, 89-93.

Dahlberg, A. C., & Penczek, E. S. The relative sweetness of sugars as affected by concentration (Tech. Bull. No. 258). Geneva, N.Y.: New York State Agricultural Experiment Station, 1941.

Eisler, H. Empirical test of a model relating magnitude and category scales. Scandinavian Journal of Psychology, 1962, 3, 88-96.

Eisler, H. Magnitude scales, category scales, and Fechnerian integration. Psychological Review, 1963, 70, 243-253.

Eisler, H. The connection between magnitude and discrimination scales and direct and indirect scaling methods. Psychometrika, 1965, 30, 271-289.

Ekman, G. Discriminal sensitivity on the subjective continuum. Acta Psychologica, 1956, 12, 233-243.

Ekman, G. Weber's law and related functions. Journal of Psychology, 1959, 47, 343-352.

Ekman, G. Is the power law a special case of Fechner's law? Perceptual and Motor Skills, 1964, 19, 730.

Ekman, G., & Hosman, B. Note on subjective scales of number. Perceptual and Motor Skills, 1965, 21, 101-102.

Engen, T. Psychophysics. In J. W. Kling & L. A. Riggs (Eds.), Experimental psychology. New York: Holt, 1971.

Eriksen, C. W., & Hake, H. W. Anchor effects in absolute judgments. Journal of Experimental Psychology, 1957, 53, 132-138.

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). San Francisco: Freeman, 1974.

Fechner, G. [Elements of psychophysics] (D. H. Howes & E. G. Boring Eds., H. E. Adler Trans.). New York: Holt, 1966. (Originally published, 1860.)

Final Report. Quantitative estimates of sensory events. Advancement of Science, 1940, No. 2, 331-349.

Galanter, E. Contemporary psychophysics. New Directions in Psychology, 1962, 1, 89-156.

Galanter, E., & Messick, S. The relation between category and magnitude scales of loudness. Psychological Review, 1961, 68, 363-372.

Garner, W. R. A technique and a scale for loudness measurement. Journal of the Acoustical Society of America, 1954, 26, 73-88.

Garner, W. R. Advantages of the discriminability criterion for a loudness scale. Journal of the Acoustical Society of America, 1958, 30, 1005-1012.

Gescheider, G. A. Psychophysics method and theory. Hillsdale, N. J.: Erlbaum, 1976.

Gibson, R. H., & Tomko, D. L. The relation between category and magnitude estimates of tactile intensity. Perception & Psychophysics, 1972, 12, 135-138.

Graham, C. H., & Bartlett, N. R. The relation of size of stimulus and intensity in the human eye: III. The influence of area on foveal intensity discrimination. Journal of Experimental Psychology, 1940, 27, 149-159.

Gregson, R. A. M., Mitchell, M. J., Simmonds, M. B, & Wells, J. E. Relative olfactory intensity perception as mediated by ratio-range category scale responses. Perception & Psychophysics, 1969, 6, 133-136.

Guilford, J. P. Psychometric methods (2nd ed.). New York: McGraw-Hill, 1954.

Hagstrom, E. C., & Pfaffmann, C. The relative taste effectiveness of different sugars for the rat. Journal of Comparative and Physiological Psychology, 1959, 52, 259-262.

Hardy, J. D., Wolff, H. G., & Goodell, H. Studies on pain: Discrimination of differences in intensity of a pain stimulus as a basis of a scale of pain intensity. Journal of Clinical Investigation, 1947, 26, 1152-1158.

Harper, R. S., & Stevens, S. S. A psychological scale of weight and a formula for its derivation. American Journal of Psychology, 1948, 61, 343-351.

Harris, J. D. Pitch discrimination. Journal of the Acoustical Society of America, 1952, 24, 750-755.

Harrison, S., & Harrison, M. J. A psychophysical method employing a modification of the Müller-Urban weights. Psychological Bulletin, 1951, 48, 249-256.

Hecht, S. The visual discrimination of intensity and the Weber-Fechner law. Journal of General Physiology, 1924, 7, 235-267.

Helson, H. Adaptation level theory: An experimental and systematic approach to behavior. New York: Harper & Row, 1964.

Herrick, R. M. Foveal luminance discrimination as a function of the duration of the decrement or increment in luminance. Journal of Comparative and Physiological Psychology, 1956, 49, 437-443.

Holway, A. H., & Hurvich, L. M. Differential gustatory sensitivity to salt. American Journal of Psychology, 1937, 49, 37-48.

Holway, A. H., & Pratt, C. C. The Weber-ratio for intensitive discrimination. Psychological Review, 1936, 43, 322-340.

Houtsma, A. J. M., Durlach, N. I., & Braida, L. D. Intensity perception XI. Experimental results on the relation of intensity resolution to loudness matching. Journal of the Acoustical Society of America, 1980, 68, 807-813.

Hyvonen, L. Synergism between sweeteners. In P. Koivistoinen & L. Hyvonen (Eds.), Carbohydrate sweeteners in food and nutrition. London: Academic Press, 1980.

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.

Keller, M. The relation between the critical duration and intensity in brightness discrimination. Journal of Experimental Psychology, 1941, 28, 407-418.

Kendall, M., & Stuart, A. The advanced theory of statistics (Vol. 1, 4th ed.). London: Griffin, 1977.

Klitzner, M. D. Hedonic integration: Test of a linear model. Perception & Psychophysics, 1975, 18, 49-54.

Krueger, L. E. Apparent combined length of two-line and four-line sets. Perception & Psychophysics, 1970, 8, 210-214.

Laming, D. Mathematical psychology. London: Academic Press, 1973.

Lewis, D. R. Psychological scales of taste. Journal of Psychology, 1948, 26, 437-446.

Lichtenstein, P. E. The relative sweetness of sugars: Sucrose and dextrose. Journal of Experimental Psychology, 1948, 38, 578-586.

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.

Luce, R. D. Thurstone's discriminal processes fifty years later. Psychometrika, 1977, 42, 461-489.

Luce, R. D., & Edwards, W. The derivation of subjective scales from just noticeable differences. Psychological Review, 1958, 65, 222-237.

Lundgren, B., Pangborn, R. M., Barylko-Pikielna, N., & Daget, N. Difference taste thresholds for sucrose in water and in orange juice: An interlaboratory study. Chemical Senses and Flavor, 1976, 2, 157-176.

MacLeod, S. A construction and attempted validation of sensory sweetness scales. Journal of Experimental Psychology, 1952, 44, 316-323.

Marks, L. E. Stimulus range, number of categories, and form of the category scale. American Journal of Psychology, 1968, 81, 467-479.

Marks, L. E. On scales of sensation: Prolegomena to any future psychophysics that will be able to come forth as science. Perception & Psychophysics, 1974, 16, 358-376.

Marks, L. E. What (good) are scales of sensation? Behavioral and Brain Sciences, 1981, 4, 199-200.

McBride, R. L. Can shelf life be measured? CSIRO Food Research Quarterly, 1980, 40, 149-152.

McBride, R. L. Range bias in sensory evaluation. Journal of Food Technology, in press.

McBride, R. L., & Laing, D. G. Threshold determination by triangle testing: effects of judgemental procedure, positional bias and incidental training. Chemical Senses and Flavour, 1979, 4, 319-326.

McBurney, D. H., & Gent, J. F. On the nature of taste qualities. Psychological Bulletin, 1979, 86, 151-167.

Meiselman, H. L. Human taste perception. CRC Critical Reviews in Food Technology, 1972, 3, 89-119.

Miller, G. A. Sensitivity to changes in the intensity of white noise and its relation to masking and loudness. Journal of the Acoustical Society of America, 1947, 19, 609-619.

Montgomery, H. Direct estimation: effect of methodological factors on scale type. Scandinavian Journal of Psychology, 1975, 16, 19-29.

Moskowitz, H. R. Ratio scales of sugar sweetness. Perception & Psychophysics, 1970, 7, 315-320.

Moskowitz, H. R. The sweetness and pleasantness of sugars. American Journal of Psychology, 1971, 84, 387-405.

Moskowitz, H. R. Models of additivity for sugar sweetness. In H. R. Moskowitz, B. Scharf & J. C. Stevens (Eds.), Sensation and measurement. Dordrecht: Reidel, 1974.

Moskowitz, H. R., & Vaisy Genser, M. Sensory response to food. Zurich: Forster, 1977.

Mueller, C. G. Frequency of seeing functions for intensity discrimination at various levels of adapting intensity. Journal of General Physiology, 1951, 34, 463-474.

Mueller, C. G. Sensory psychology. Eaglewood Cliffs, N.J.: Prentice-Hall, 1965.

Newman, E. B. The validity of the just noticeable difference as a unit of psychological magnitude. Transactions of the Kansas Academy of Sciences, 1933, 36, 172-175.

Nihm, S. D. Polynomial law of sensation. American Psychologist, 1976, 31, 808-809.

Oberlin, K. W. Variation in intensitive sensitivity to lifted weights. Journal of Experimental Psychology, 1936, 19, 438-455.

Ono, H. Difference threshold for stimulus length under simultaneous and nonsimultaneous viewing conditions. Perception & Psychophysics, 1967, 2, 201-207.

Pangborn, R. M. Relative taste intensities of selected sugars and organic acids. Journal of Food Science, 1963, 28, 726-733.

Parducci, A. Category judgment: A range-frequency model. Psychological Review, 1965, 72, 407-418.

Parducci, A. The relativism of absolute judgments. Scientific American, 1968, 219, 84-90.

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.

Parker, S., & Schneider, B. Nonmetric scaling of loudness and pitch using similarity and difference estimates. Perception & Psychophysics, 1974, 15, 238-242.

Parker, S., & Schneider, B. Loudness and loudness discrimination. Perception & Psychophysics, 1980, 28, 398-406.

Parker, S., Schneider, B., & Kanow, G. Ratio scale measurement of the perceived lengths of lines. Journal of Experimental Psychology, 1975, 104, 195-204.

Pearson, E. S., & Hartley, H. O., (Eds.). Biometrika tables for statisticians (Vol. 1, 3rd ed.). Cambridge, England: Cambridge University Press, 1970.

Pfaffmann, C. An experimental comparison of the method of single stimuli and the method of constant stimuli in gustation. American Journal of Psychology, 1935, 37, 470-475.

Pfaffmann, C. The sense of taste. In J. Field (Ed.), Handbook of physiology (Vol. 1). Washington, D.C.: American Physiological Society, 1959.

Piggott, J. R., & Harper, R. Ratio scales and category scales of odour intensity. Chemical Senses and Flavor, 1975, 1, 307-316.

Pollack, I. Iterative techniques for unbiased rating scales. Quarterly Journal of Experimental Psychology, 1965, 17, 139-148.

Poulton, E. C. Population norms of top sensory magnitudes and S. S. Stevens' exponents. Perception & Psychophysics, 1967, 2, 312-316.

Poulton, E. C. The new psychophysics: Six models for magnitude estimation. Psychological Bulletin, 1968, 69, 1-19.

Poulton, E. C. Unwanted range effects from using within-subjects experimental designs. Psychological Bulletin, 1973, 80, 113-121.

Poulton, E. C. Range effects in experiments on people. American Journal of Psychology, 1975, 88, 3-32.

Poulton, E. C. Quantitative subjective assessments are almost always biased, sometimes completely misleading. British Journal of Psychology, 1977, 68, 409-425.

Poulton, E. C. Models for biases in judging sensory magnitude. Psychological Bulletin, 1979, 86, 777-803.

Reese, T. W. The application of the theory of physical measurement to the measurement of psychological magnitudes, with three experimental examples. Psychological Monographs, 1943, 55(3, Whole No. 251).

Riesz, R. R. Differential intensity sensitivity of the ear for pure tones. Physical Review, 1928, 31, 867-875.

Riesz, R. R. The relationship between loudness and the minimum perceptible increment of intensity. Journal of the Acoustical Society of America, 1933, 4, 211-216.

Rifkin, B., & Bartoshuk, L. M. Taste synergism between monosodium glutamate and disodium 5′-guanylate. Physiology & Behavior, 1980, 24, 1169-1172.

Rubin, D. C. Frequency of occurrence as a psychophysical continuum: Weber′s fraction, Ekman′s fraction, range effects, and the phi-gamma hypothesis. Perception & Psychophysics, 1976, 20, 327-330.

Rule, S. J., & Curtis, D. W. Conjoint scaling of subjective number and weight. Journal of Experimental Psychology, 1973, 97, 305-309.

Rule, S. J., & Curtis, D. W. Ordinal properties of subjective ratios and differences: Comment on Veit. Journal of Experimental Psychology: General, 1980, 109, 296-300.

Savage, C. W. The measurement of sensation. Berkeley, Ca.: University of California Press, 1970.

Schiffman, S. S., & Erickson, R. P. The issue of primary tastes versus a taste continuum. Neuroscience & Biobehavioral Reviews, 1980, 4, 109-117.

Schutz, H. G., & Pilgrim, F. J. Sweetness of various compounds and its measurement. Food Research, 1957, 22, 206-213. (a)

Schutz, H. G., & Pilgrim, F. J. Differential sensitivity in gustation. Journal of Experimental Psychology, 1957, 54, 41-48. (b)

Shanteau, J. C., & Anderson, N. H. Test of a conflict model for preference judgment. Journal of Mathematical Psychology, 1969, 6, 312-325.

Stahl, W. H., (Ed.). Compilation of odor and taste threshold values data. Philadelphia: American Society for Testing and Materials, 1973.

Stevens, S. S. A scale for the measurement of a psychological magnitude: Loudness. Psychological Review, 1936, 43, 405-416.

Stevens, S. S. On the brightness of lights and the loudness of sounds. Science, 1953, 118, 576. (Abstract)

Stevens, S. S. Pitch discrimination, Mels, and Kock's contention. Journal of the Acoustical Society of America, 1954, 26, 1075-1077.

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.

Stevens, S. S. The psychophysics of sensory function. American Scientist, 1960, 48, 226-253.

Stevens, S. S. To honor Fechner and repeal his law. Science, 1961, 133, 80-86.

Stevens, S. S. A metric for the social consensus. Science, 1966, 151, 530-541.

Stevens, S. S. Sensory scales of taste intensity. Perception & Psychophysics, 1969, 6, 302-308.

Stevens, S. S. Issues in psychophysical measurement. Psychological Review, 1971, 78, 426-450.

Stevens, S. S. Psychophysics. New York: Wiley, 1975.

Stevens, S. S., Carton, A. S., & Shickman, G. M. A scale of apparent intensity of electric shock. Journal of Experimental Psychology, 1958, 56, 328-334.

Stevens, S. S., & Galanter, E. H. Ratio scales and category scales for a dozen perceptual continua. Journal of Experimental Psychology, 1957, 54, 377-411.

Stevens, S. S., & Greenbaum, H. B. Regression effect in psychophysical judgment. Perception & Psychophysics, 1966, 1, 439-446.

Stevens, S. S., & Guirao, M. Subjective scaling of length and area and the matching of length to loudness and brightness. Journal of Experimental Psychology, 1963, 66, 177-186.

Stevens, S. S., & Stone, G. Finger span: Ratio scale, category scale, and JND scale. Journal of Experimental Psychology, 1959, 57, 91-95.

Stevens, S. S., & Volkmann, J. The relation of pitch to frequency: A revised scale. American Journal of Psychology, 1940, 53, 329-353.

Stevens, S. S., Volkmann, J., & Newman, E. B. A scale for the measurement of the psychological magnitude pitch. Journal of the Acoustical Society of America, 1937, 8, 185-190.

Stone, H., & Bosley, J. J. Olfactory discrimination and Weber's law. Perceptual and Motor Skills, 1965, 20, 657-665.

Stone, H., & Oliver, S. M. Measurement of the relative sweetness of selected sweeteners and sweetener mixtures. Journal of Food Science, 1969, 34, 215-222.

Teghtsoonian, R. On the exponents in Stevens' law and the constant in Ekman's law. Psychological Review, 1971, 78, 71-80.

Teghtsoonian, R. On facts and theories in psychophysics: Does Ekman's law exist? In H. R. Moskowitz, B. Scharf & J. C. Stevens (Eds.), Sensation and measurement. Dordrecht: Reidel, 1974.

Teghtsoonian, R., & Teghtsoonian, M. Two varieties of perceived length. Percption & Psychophysics, 1970, 8, 389-392.

Thurstone, L. L. Psychophysical analysis. American Journal of Psychology, 1927, 38, 368-389. (a)

Thurstone, L. L. A law of comparative judgment. Psychological Review, 1927, 34, 273-286. (b)

Thurstone, L. L. A mental unit of measurement. Psychological Review, 1927, 34, 415-423. (c)

Thurstone, L. L. Equally often noticed differences. Journal of Educational Psychology, 1927, 18, 289-293. (d)

Thurstone, L. L. Three psychophysical laws. Psychological Review, 1927, 34, 424-432. (e)

Thurstone, L. L. The phi-gamma hypothesis. Journal of Experimental Psychology, 1928, 11, 293-305.

Thurstone, L. L. Stimulus dispersions in the method of constant stimuli. Journal of Experimental Psychology, 1932, 15, 284-297.

Torgerson, W. S. Theory and methods of scaling. New York: Wiley, 1958.

Torgerson, W. S. Quantitative judgment scales. In H. Gulliksen & S. Messick (Eds.), Psychological scaling. New York: Wiley, 1960.

Torgerson, W. S. Distances and ratios in psychophysical scaling. Acta Psychologica, 1961, 19, 201-205.

Treisman, M. Sensory scaling and the psychophysical law. Quarterly Journal of Experimental Psychology, 1964, 16, 11-22. (a)

Treisman, M. Noise and Weber's law: The discrimination of brightness and other dimensions. Psychological Review, 1964, 71, 314-330. (b)

Troland, L. T. The principles of psychophysiology (Vol. 2). New York: Greenwood Press, 1969. (Originally published, 1930).

Uttal, W. R. The psychobiology of sensory coding. New York: Harper & Row, 1973.

Van der Wel, H. Psychophysical studies of thaumatin and monellin. In J. Solms & R. L. Hall (Eds.), Criteria of food acceptance. Zurich: Forster, 1981.

Veit, C. T. Analyzing "ratio" and "difference" judgments: A reply to Rule and Curtis. Journal of Experimental Psychology: General, 1980, 109, 301-303.

Volkmann, J. A quantal model for psychological magnitude and differential sensitivity. In H. R. Moskowitz, B. Scharf & J. C. Stevens (Eds.), Sensation and measurement. Dordrecht: Reidel, 1974.

Warren, R. M. Measurement of sensory intensity. Behavioral and Brain Sciences, 1981, 4, 175-223.

Weiss, D. J. The impossible dream of Fechner and Stevens. Perception, 1981, 10, 431-434.

Weiss, D. J. & Anderson, N. H. Use of rank order data in functional measurement. Psychological Bulletin, 1972, 78, 64-69.

Wenzel, B. M. Differential sensitivity in olfaction. Journal of Experimental Psychology, 1949, 39, 129-143.

Wier, C. C., Jesteadt, W., & Green, D. M. Frequency discrimination as a function of frequency and sensation level. Journal of the Acoustical Society of America, 1977, 61, 178-184.

Due to copyright laws, the following articles have been omitted from this thesis.
Please refer to the following citations for details.

McBride, R . L. (1984) Range bias in sensory evaluation, *Journal of food technology*, Vol. 17, Issue 3, p. 1-6, DOI: **http://doi.org/10.1111/j.1365-2621.1982.tb00195.x**

APPENDIX B.

SOURCE OF VARIATION	DF	SS	SS%	MS	VR
EXPT STRATUM					
INSTRUCT	4	2184.373	3.46	546.093	14.765
RESIDUAL	5	184.929	0.29	36.986	4.110
TOTAL	9	2369.302	3.75	263.256	29.256
EXPT.SESSION STRATUM	30	526.328	0.83	17.544	1.950
EXPT.SESSION.*UNITS* STRATUM					
CONC	2	59384.343	94.11	29692.172	3299.756
INSTRUCT.CONC	8	192.415	0.30	24.052	2.673
RESIDUAL	70	629.880	1.00	8.998	
TOTAL	80	60206.639	95.41	752.583	
GRAND TOTAL	119	63102.269	100.00		

(i) Analysis of variance table for raw data from Experiment VII.

SOURCE OF VARIATION	DF	SS	SS%	MS	VR
EXPT STRATUM					
INSTRUCT	4	0.290265	3.85	0.072566	15.436
RESIDUAL	5	0.023505	0.31	0.004701	3.948
TOTAL	9	0.313771	4.16	0.034863	29.277
EXPT.SESSION STRATUM	30	0.069329	0.92	0.002311	1.941
EXPT.SESSION.*UNITS* STRATUM					
CONC	2	7.069522	93.69	3.534761	2968.381
INSTRUCT.CONC	8	0.009774	0.13	0.001222	1.026
RESIDUAL	70	0.083356	1.10	0.001191	
TOTAL	80	7.162652	94.92	0.089533	
GRAND TOTAL	119	7.545752	100.00		

(ii) Analysis of variance table for transformed (arc sine$\sqrt{p}$) data from Experiment VII

APPENDIX C

SOURCE OF VARIATION	DF	SS	SS%	MS	VR
SESSION STRATUM					
FRUCTOSE	3	449.835	4.22	149.945	102.872
SUCROSE	3	231.607	2.17	77.202	52.966
FRUCTOSE.SUCROSE	9	90.067	0.84	10.007	6.866
RESIDUAL	4	5.830	0.05	1.458	0.541
TOTAL	19	777.339	7.29	40.913	15.196
SESSION.*UNITS* STRATUM					
FRUCTOSE	3	2711.235	25.43	903.745	335.668
SUCROSE	3	1873.969	17.57	624.656	232.009
TASTER	13	2101.625	19.71	161.663	60.045
FRUCTOSE.SUCROSE	9	130.294	1.22	14.477	5.377
FRUCTOSE.TASTER	39	271.339	2.54	6.957	2.584
SUCROSE.TASTER	39	176.568	1.66	4.527	1.682
FRUCTOSE.SUCROSE.TASTER	117	260.039	2.44	2.223	0.825
RESIDUAL	877	2361.217	22.14	2.692	
TOTAL	1100	9886.286	92.71	8.988	
GRAND TOTAL	1119	10663.625	100.00		

(i) Analysis of variance table for Experiment VIII - incomplete block design

SOURCE OF VARIATION	DF	SS	SS%	MS	VR
UNITS STRATUM					
FRUCTOSE	3	3141.161	29.46	1047.054	376.227
SUCROSE	3	2046.532	19.19	682.177	245.120
TASTER	13	2101.625	19.71	161.663	58.089
FRUCTOSE.SUCROSE	9	172.761	1.62	19.196	6.897
FRUCTOSE.TASTER	39	271.339	2.54	6.957	2.500
SUCROSE.TASTER	39	176.568	1.66	4.527	1.627
FRUCTOSE.SUCROSE.TASTER	117	260.039	2.44	2.223	0.799
RESIDUAL	896	2493.600	23.38	2.783	
TOTAL	1119	10663.625	100.00	9.530	
GRAND TOTAL	1119	10663.625	100.00		

(ii) Analysis of variance table for Experiment VIII - complete block design

APPENDIX D

SOURCE OF VARIATION	DF	SS	SS%	MS	VR
SESSION STRATUM	3	6.411	0.16	2.137	1.087
SESSION.*UNITS* STRATUM					
SUCROSE	2	795.267	19.61	397.633	202.191
FRUCTOSE	1	1831.511	45.16	1831.511	931.297
TASTER	14	674.650	16.63	48.189	24.504
SUCROSE.FRUCTOSE	2	20.356	0.50	10.178	5.175
SUCROSE.TASTER	28	65.233	1.61	2.330	1.185
FRUCTOSE.TASTER	14	77.072	1.90	5.505	2.799
SUCROSE.FRUCTOSE.TASTER	28	60.311	1.49	2.154	1.095
RESIDUAL	267	525.089	12.95	1.967	
TOTAL	356	4049.489	99.84	11.375	
GRAND TOTAL	359	4055.900	100.00		

Analysis of variance table for Experiment IX