Perception & Psychophysics
1975, Vol. 17 (4), 351-357

Quantifying private events: A functional measurement analysis of equisection

DAVID J. WEISS

California State University, 5151 State University Drive, Los Angeles, California 90032

Functional measurement theory was applied to bisection, trisection, and quadrisection of grayness. Theoretically, these judgments should obey an averaging model. But the overt responses are not a valid measure of subjective magnitude (since they are made on the physical stimulus continuum), and so they cannot be used directly to test the model. However, scaling and model testing can both be accomplished simultaneously using functional measurement theory. If the subject is indeed averaging, then there exists a monotone transformation that makes the data additive; and this transformation can be computed with FUNPOT, Weiss's (1973a) computer program which finds polynomial transformations that reduce selected effects. Further, determination of this transformation also reveals the psychophysical function, because it gives the relation between subjective magnitude and overt response. For bisection, the averaging model was successful; it was possible to find a monotone transformation that made the data additive. This psychophysical function differed somewhat in form from the Munsell scale. It gained cross-task validity from its agreement with a grayness scale obtained from rating data (Weiss, 1972). For trisection and quadrisection, the averaging model was not accepted; it was not possible consistently to find transformations which induced additivity.

A classical issue in psychology is Fechner's problem: What is the function relating sensations to stimuli? Because sensations are inaccessible, the psychophysicist's problem must be decomposed into two questions: (1) What is the relation between the physical stimulus and the private sensation? (2) How does the private sensation get translated into public response?

One kind of experimental task is uniquely suited to answering these two questions. When the response is made on the stimulus continuum, then the subject must use the psychophysical function in two ways, to process the stimulus and to produce the response. This dual use of the link between stimulus and sensation can provide sufficient leverage to solve the psychophysicist's problem.

Equisection, one of the oldest of the direct scaling methods, is such a task. The subject is instructed to find those stimuli which partition a given interstimulus interval into several subjectively equal portions. This method has one especially valuable property. Because responses are made on the stimulus continuum, the relation between stimulus and response is natural and meaningful to the subject. One need not worry about an artificial, possibly misunderstood, mapping of subjective values onto an arbitrary response system. If it can be assumed that the subject is indeed equisecting, then a psychophysical function can be immediately determined. An example of this scaling method is given in Newman, Volkmann, and Stevens (1937). For the function to be theoretically acceptable, however, the equisection assumption itself must be verified.

This paper is based on a doctoral dissertation submitted to the University of California, San Diego, December 1973. The invaluable assistance and patience of Norman H. Anderson is gratefully acknowledged. Requests for reprints should be sent to David J. Weiss, Department of Psychology, California State University, Los Angeles, California 90032.

Equisection is a generalization of bisection, a technique first employed by Plateau (1872), who asked artists to paint a gray exactly intermediate between black and white. Although bisection has since seen many applications, an experimental critique by Gage (1934) decreased the popularity of the method. Gage called bisection inconsistent because ascending and descending stimulus presentation yielded different bisecting values.

Garner (1954) used equisection in an attempt to develop a loudness scale. Two loudness functions were determined, one based on fractionation and the other on equisection. In making successive fractionations, it was assumed that the subject would set the same ratio, even though this ratio might not be the one prescribed by the instructions. Equisection judgments, producing an interval scale with arbitrary zero, were used to find the value of the ratio and thus determine the scale. Garner's reasoning was clever, but he did not dwell upon the central question of equisection, namely, how to determine whether the subject is in fact partitioning the interval in a consistent way. Garner's analysis is unsatisfactory because it rested on the untested assumption that equisection judgments did yield equal loudness intervals.

Fagot and Stewart (1970) tested a bisection version of Pfanzagl's (1968) axiomatic measurement system. Five axioms generate a representation theorem identical to Equation 1, which is presented below. A sixth axiom, commutativity or lack of response bias, was determined to be violated by the data. Given this violation, Fagot and Stewart claim that it is necessary to assume some specific form for the psychophysical function in order to construct a scale. Several forms of biased power functions were compared with respect to fitting the data, but no goodness-of-fit tests for the proposed models were employed. The approach of Fagot and Stewart is unsatisfactory, because it rested on untested assumptions about the form of the psychophysical function.

Functional measurement provides a new approach to the study of equisection. The basic assumption is that equisection obeys an averaging model. However, this assumption is tested directly in the analysis, and other untested assumptions are unnecessary. If the model is correct, then the analysis reveals the psychophysical function.

Functional Measurement Analysis

The equisection task involves first integrating the stimulus information, and then producing the response. It is assumed that an averaging model describes the integration process:

$$\Psi(R_{ij}) = w_L \Psi(S_i + w_R \Psi(S_j). \quad (1)$$

Here, R_{ij} is the overt response to the stimulus pair S_i, presented in the left spatial position, and S_j, presented in the right; Ψ is the psychophysical function, giving the relation between a physical stimulus and its internal sensation value. The weights, w_L and w_R, depend on the sectioning required (for bisection they would presumably be equal) and possibly on spatial position, but they are assumed not to depend on stimulus context. Since this is an averaging model, the weights are assumed to add to 1, although this assumption is not used in the analysis. In other experimental situations, particularly when stimuli are presented serially, the weight parameter can reflect temporal order (Weiss & Anderson, 1969). Interpretation of the weighting in Equation 1 as an order parameter allows the dismissal of Gage's (1934) objection to bisection (Weiss, 1973b, Appendix 1).

According to functional measurement methodology, additive models such as Equation 1 can be evaluated using analysis of variance to test goodness of fit (Weiss & Anderson, 1969). A factorial design is set up, with the presumed additive components defining the factors. In equisection, the factors are the left and right stimuli, as already noted. The additive model is supported if the factors do not interact statistically.

This analysis does not require that the two weights in Equation 1 be equal, nor is there any need to do so. As long as the weights are constant, the model is additive, and it will reveal the psychophysical function. For present purposes, therefore, bisection will be used in this more general sense. As a consequence, the present analysis automatically allows for position or order effects.

It is not the overt response that is additive in Equation 1, but rather the unobservable subjective value associated with it. But if the model is correct, then a transformation exists that renders the data additive, and that transformation is linearly related to Ψ. Hence, the problem of finding the psychophysical function is reduced to finding a transformation that makes the data additive. This transformation problem can be solved using the FUNPOT computer program developed by the writer (Weiss, 1973a). The present paper applies this functional measurement analysis to the continuum of grayness.

METHOD

Apparatus

The stimuli were 1.59 x 2.22 cm neutral-value Munsell chips, matte finish, ranging from very black to very white in quarter steps on the Munsell scale. Each chip was glued to the head of a 2.54-cm tack, and the tacks were mounted on a continuous belt. Five such belts, each with a complete series of 31 chips, were mounted behind a brown screen that had a circular hole in front of each belt so that one chip was visible at a time through each hole. The stimuli to be judged were always presented on the two outer belts. The subject chose his responses by turning the interior belts. One response belt was used for bisection, two for trisection, and all three for quadrisection; the holes for belts not used were blocked. Between trials, while the experimenter adjusted the two outer belts, a shutter blocked all of the holes.

Design

The present experiment employed a 3 by 3 by 3 design. The first factor was section; its three levels were bisection, trisection, and quadrisection. The other two factors were stimulus factors whose levels were reflectances of 3%, 20%, 59%, and 9%, 36%, 90%, respectively. The levels were chosen so that when the factors were crossed, the resulting interval would be wide enough to allow quadrisection responses.

Subjects

Each of nine paid subjects was run individually in six 1-h sessions. The first session consisted of two replications of the bisection sets, the second and third of one replication of trisection and quadrisection, respectively. The next three sessions were a repetition of the first three.

Instructions

The subjects were told that for each pair of graynesses to be presented to them they were to select graynesses that broke up the grayness interval into two, three, or four equal parts. Instructions were read each day for that day's condition. On the first day, it was explained that there was no objectively correct answer, since a personal continuum was involved. The terms "intermediate grayness" and "grayness midway between" were used to help define the judgment required.

Responding was self-paced, with the rate dependent upon the number of equisection responses required. Responses were made rather slowly, as the nature of the task fostered readjustment. A typical bisection judgment required about 20-30 sec; a set of quadrisection responses required about 1 min.

Further details of procedure are in the author's dissertation (Weiss, 1973b).

RESULTS

Bisection

Raw reflectance responses. The mean raw reflectance responses for bisection are shown in the upper left panel of Figure 1. Plots for the individual subjects, given in Weiss (1973b), were all virtually identical to the group plot. The data are distinctly nonparallel, with the lines diverging markedly for the lighter stimuli. Nonparallelism is the graphic analog of nonadditivity, and so the nonparallelism was expected because the physical reflectance scale is not a valid measure of subjective magnitude. For each subject, the F ratio for interaction was significant and sizable.

The raw data were processed according to the proposed model. If a transformation could be found which rendered the data additive, then the model would be supported, and the transformation would lay claim to the title of psychophysical function. Transformation was made on each subject's data individually, since there is no reason that different individuals should have identical psychophysical functions.

Munsell transformation. Since the Munsell (1967) scale is based on an impressive volume of data, reflectance values were replaced by Munsell values and subjected to a similar analysis. If the Munsell scale is valid and an additive model is correct, then these Munsell values should appear additive. Individual-subject analyses uniformly rejected the Munsell transformation, although some reduction in the interaction F ratios did occur. The group plot of the Munsell values, shown in the upper right panel of Figure 1, reveals that the Munsell scale is not extreme enough; the lines diverge for the lighter stimuli. However, the Munsell transformation does provide considerable improvement over reflectance values; the lines are much closer to parallelism than in the raw data.

Polynomial transformation. To find the transformation which induced additivity, the polynomial method of Bogartz and Wackwitz (1971) was implemented via FUNPOT, a computer program written by Weiss (1973a). The routine solves analytically for the coefficients of a power series expansion of specified degree, following a suggestion made by Anderson (1962). Arbitrary components of an experimental design may be reduced. This allows an exact statistical test of the model with the transformed data.

The program worked in an iterature manner. A second-degree polynomial was applied to the data, and the transformed data were analyzed with a reduction of one degree of freedom for the critical interaction term. If the F ratio was significant, a third-order polynomial was applied and the significance test carried out with an additional reduction of one degree of freedom. This iteration continued until the interaction was reduced to nonsignificance or until the degrees of freedom were used up. An additional linear transformation gave the transformed data the same range as the raw data; this had no effect on the F ratios.

FUNPOT was applied to each subject's bisection data separately. The algorithm was extremely successful. With two analyses per subject, the statistical criterion for additivity, nonsignificance of the interaction, was achieved in 17 of 18 instances.

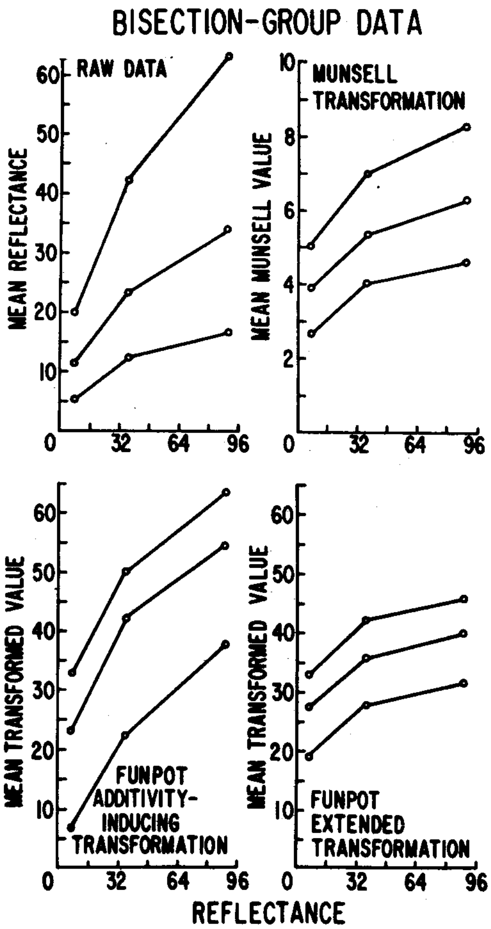

Figure 1. Group mean responses for bisection. Each curve corresponds to one value of the right stimulus. The upper left and upper right panels, respectively, show raw data and Munsell transformed values. The lower left panel shows the data transformed via FUNPOT to the additivity criterion. The lower right panel shows the data transformed via the extended version of FUNPOT.

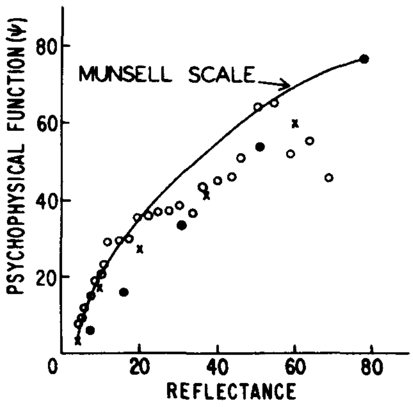

Figure 2. Group-transformation function for bisection, shown as open circles. Filled circles and Xs are two sets of marginal means scale values from a grayness averaging study (Weiss, 1972). Solid line is the Munsell scale of value.

The transformed data were then averaged to yield the lower left panel of Figure 1. The lines appear almost parallel, which reflects the success of the transformation.

Transformation function. In order to determine the purest form of the transformation function, the FUNPOT algorithm was extended. The stopping criterion was suspended and the routine proceeded until no degrees of freedom remained. This extension reveals the best polynomial transformation in the sense that it forces additivity necessarily. The extended transformation produced the parallel lines in the lower right panel of Figure 1.

The success of the bisection model meant that the extended transformations could be considered as true psychophysical functions. The average transformation values are shown as open circles in Figure 2. These points were generated by finding for each reflectance the mean transformed value, averaged over subjects, from the extended transformations of the bisection data. Since the physical values used as responses did not occur equally often, the various points in the graph represent different numbers of responses. Reflectance values above 50% comprised less than 12% of the responses for bisection; thus, the scatter in the parts of the upper end of the reflectance scale reflects unreliable data. There were two reflectances which had fewer than five total occurrences over all nine subjects, and these two points were omitted from the graph.

Cross-task validation. The transformation function from the bisection task appears orderly over most of its range. Also shown in Figure 2 are two sets of marginal means scale values from a previous functional measurement study (Weiss, 1972) which employed grayness stimuli in an averaging task with a rating response. These previously validated scale values, shown as filled circles and Xs, determine a function which is similar in form to the bisection function. This agreement across two judgmental tasks supports the legitimacy of the bisection scaling and justifies the labeling of the transformation function in Figure 2 as a psychophysical function. While the linear relation between the sets of points from the two experiments is not perfect, it is quite strong. A measuring rod for this agreement is the disagreement of both functional measurement scales with the most widely accepted grayness function. The curved line in Figure 2 is the Munsell scale of value, which is not linearly related to the grayness scale as validated in the present work.

Marginal means scaling. Functional measurement studies (e.g., Weiss, 1972) usually use the marginal means of the factorial design as direct estimates of the scale values of the associated stimuli. Weiss (1973b, Appendix 2) has shown that these estimates are linearly related to transformation function values if Equation 1 (or Equation 3, presented below) is true. For the present study, however, these values cannot be compared, since the stimulus design employed only three levels on each factor. As scale values are determined only up to a linear transformation, a larger number of stimuli is required to determine a useful marginal means scale.

Trisection and Quadrisection

The three-factor design which incorporated the higher sectionings did not yield results consistent with the additive model. The basic problem was that additivity-inducing transformations could not be consistently achieved.

The raw data for trisection and quadrisection (given in Weiss, 1973b) were characterized by large stimulus interaction, similar in form to that shown for bisection in Figure 1. As there were two responses generated for each trisection and quadrisection presentation (the central quadrisection response was not used), two separate three-factor subdesigns were available for analysis.

Application of FUNPOT produced limited success. With 18 data sets examined (2 subdesigns for each of 9 subjects), the additivity criterion was met in 11 instances. The Munsell transformation was even less successful; additivity was achieved in only two cases.

A critical test of Equation 1 involves comparing the transformation functions obtained from each subject's subdesigns. If an obtained function is truly the psychophysical function, then the two functions from a given individual should be linearly related. The extended version of FUNPOT was used to generate these functions because it gives the purest estimate of the transformation function, if one exists.

The pair of transformation functions for each subject was examined (they are shown in Weiss, 1973b). The functions seemed to be monotone over the range of most of the data, but the individual pairs of functions were decidedly not linearly related. This meant that Equation 1 did not fit the equisection data.

The locus of the model's failure appears to be in trisection and quadrisection. Extended transformation was applied to each section separately. The pairs of transformation functions for trisection did not agree well and appeared somewhat irregular. The functions for quadrisection were quite irregular, and there was virtually no relation between transformed values for the two subdesigns.

Quadrisection paradox. The failure of the model for quadrisection is especially puzzling, because quadrisection may be considered logically as successive bisection, and was so considered by the subjects. To explore the discrepancy, a direct comparison of the center quadrisection responses with the bisection responses was made. Although these center responses were not used in the primary analyses, a possible source of the difficulty with quadrisection is that errors in these first judgments were compounded as the other responses were made. For each subject, a three-factor design was constructed: the factors were bisection vs. quadrisection and the two stimulus factors. The critical term in this analysis is the three-way interaction, which tests whether the stimulus interaction differed for bisection and the central quadrisection response. This critical term was significant for only two of the nine subjects, and it was not large in either case.

Thus, the locus of the problem for quadrisection must be the exterior responses. These later bisections are of smaller intervals than the original bisections, and a purely speculative explanation of the difficulty is that the discrete steps of the response scale induce computational rather than impressionistic strategies for dividing these smaller intervals. Because the response steps, which were equally spaced in Munsell value, are not subjectively equally spaced, such a strategy would distort the averaging process. There is no direct evidence for this explanation, but it is consistent with the increasingly poor performance of the model for trisection and quadrisection.

Multiplicative model. Anderson (1970) proposed a multiplicative model in which the response is adjusted so that its ratio to one stimulus appears proportional to the ratio of the other stimulus to it. This model can be written,

$$\Psi(R_{ij}) = \sqrt{w\Psi(S_i)\Psi(S_j)}, \qquad (2)$$

where w is a constant of proportionality. This model, although only approximate, is of interest since it is monotonically equivalent to the additive model for bisection, but not for trisection or quadrisection. The model's major prediction is that when the additivity-inducing transformation is applied, the interactions between the stimulus factors and the section factor will vanish, because the section factor which determines w has the same multiplicative relation with the stimulus factors that the stimulus factors have with each other. However, this multiplicative model implies, like the additive model of Equation 1, that the trisection and quadrisection data can be transformed to additivity. The present data, therefore, argue against this multiplicative model.

DISCUSSION

This paper has shown that Fechner's problem can be solved by applying functional measurement theory. The logic of this approach "consists in using the postulated behavior laws to induce a scaling on the dependent variable" (Anderson, 1962). In the present case, the behavior law is the simple averaging model for bisection. Since the overt response is on the physical scale, it requires transformation to the psychological scale of sensation. The averaging model provides the scaling frame for this transformation. The computational basis is provided by the writer's FUNPOT program (Weiss, 1973a).

An important property of functional measurement is that it provides a validational base for the model and its associated measurement scales. The use of factorial design gives the constraints for determining the transformation, as well as degrees of freedom for testing the model. The early attempts to use bisection as a key to the psychophysical law lacked this validational power. Much the same criticism applies to the more recent nonmetric methods, such as conjoint measurement (Luce & Tukey, 1964), since these also fail to provide an adequate test of goodness of fit (Weiss & Anderson, 1972).

The validity of the bisection sc ale is underscored by the good agreement between the present bisection scale and the averaging scale from a previous study in which a direct rating response was used (Weiss, 1972). Both tasks yielded a grayness scale that differed slightly from the Munsell scale. This agreement is important because the two tasks are quite different in their psychological nature. It provides a cross-task validation of the present psychophysical function for grayness.

It should also be noted that the test of the model is powerful. When the same procedure was applied to the higher order sections, the test of fit was not satisfied. This illustrates that the FUNPOT algorithm for processing the data does not automatically produce a scale. This empirical verifiability is an important aspect of functional measurement. To be sure, the failure of the equisection model for the higher sections detracts from the force of the bisection scale; generalization across sections would have enhanced the validational support for the scale. This empirical difficulty should not be held against the theoretical argument that if Equation 1 is true, then a psychophysical scale can be determined.

The present bisection scale for grayness agrees fairly well with the standard Munsell scale. However, the bisection scale appears to be somewhat more concave. Historically, the Munsell scale of value (Munsell, Sloan, & Godlove, 1933) was originally based on just-noticeable-difference data used together with the method of "equal value steps" (successive bisections of a large interval). Later it was modified by using data based primarily on estimates of the ratio of two grayness intervals (Newhall, 1940). This adjusted scale, in current use, is slightly more concave than its predecessor, but the present results suggest that the adjustment did not go far enough. The Munsell determinations were characterized by careful procedures and high intersubject agreement. However, they lack a validity criterion and so leave unanswered the basic issue of whether the subjects could perform as instructed. The advantage of the bisection scale is that it provides a test of whether subjects can indeed partition the interval consistently.

The present grayness scale differs radically from results obtained with magnitude estimation. Stevens and Galanter (1957) report an exponent of 1.2 for a power function fit. Since the present data in Figure 2 are negatively accelerated, a power function fit would have an exponent substantially less than 1. This disagreement is no surprise as Stevens (1971) has repeatedly condemned "partition scales," including bisection and ordinary rating scales. But these partition scales have been able to meet validity criteria, whereas magnitude estimation has not (Anderson, 1974, Section IV.B.4). On this basis, therefore, it would seem that the functional measurement approach provides the better basis for solving Fechner's problem.[1]

REFERENCES

Aczél, J. *Lectures on functional equations and their applications.* New York: Academic Press, 1966.

Anderson, N. H. On the quantification of Miller's conflict theory. *Psychological Review*, 1962, **69**, 400-414.

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

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

Bogartz, R. S., & Wackwitz, J. H. Polynomial response scaling and functional measurement. *Journal of Mathematical Psychology*, 1971, **8**, 418-443.

Fagot, R. F., & Stewart, M. K. Test of a response bias model of bisection. *Perception & Psychophysics*, 1970, **7**, 257-262.

Gage, F. H. An experimental investigation of the measurability of auditory sensation. *Royal Society of London Proceedings, Series B*, 1934, **116**, 103-122.

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

Luce, R. D., & Tukey, J. W. Simultaneous conjoint measurement: A new type of fundamental measurement. *Journal of Mathematical Psychology*, 1964, **1**, 1-27.

Munsell, A. E. O., Sloan, L. L., & Godlove, I. H. Neutral value scales. I. Munsell neutral value scale. *Journal of the Optical Society of America*, 1933, **23**, 394-411.

Munsell, A. H. *A color notation* (12th ed.). Baltimore: Munsell Color Co., 1967.

Newhall, S. M. Preliminary report of the OSA subcommittee on the spacing of the Munsell colors. *Journal of the Optical Society of America*, 1940, **30**, 617-645.

Newman, E. B., Volkman, J., & Stevens, S. S. On the method of bisection and its relation to a loudness scale. *American Journal of Psychology*, 1937, **49**, 134-137.

Pfanzagl, J. *Theory of measurement*. New York: Wiley, 1968.

Plateau, M. J. Sur la mésure des sensations physiques. *Bulletin de L'Academie Royale des Sciences, Brussels, Series 2*, 1872, **33**, 376-388.

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

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

Weiss, D. J. Averaging: An empirical validity criterion for magnitude estimation. *Perception & Psychophysics*, 1972, **12**, 385-388.

Weiss, D. J. FUNPOT, a FORTRAN program for finding a polynomial transformation to reduce any sources of variance in a factorial design. *Behavioral Science*, 1973, **18**, 150. (a)

Weiss, D. J. A functional measurement analysis of equisection. Unpublished doctoral dissertation, University of California, San Diego, 1973. (b)

Weiss, D. J., & Anderson, N. J. Subjective averaging of length with serial presentation. *Journal of Experimental Psychology*, 1969, **82**, 52-63.

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

NOTE

1. Use of the transformation approach introduces a certain indeterminacy into the evaluation of a model. When monotone transformation is routinely incorporated into the data analysis, a class of models is tested rather than a specific model. In the present case, this class of models may be derived from what Aczél (1966) calls the quasilinear weighted mean. The general expression for such models is:

$$\Psi(R_{ij}) = g\left[\frac{w_L g^{-1}(\Psi(s_i)) + w_R g^{-1}(\Psi(S_j))}{w_L + w_R}\right], \tag{3}$$

where g(x) is a function which determines the specific model. If $g(x) = x$, then Equation 1 results. If $g(x) = e^x$, a geometric mean model is produced. Similarly, one may write a harmonic mean model, with $g(x) = 1/x$, or a root-mean-power model, with $g(x) = x^{1/p}$. The analysis-of-variance test of additivity does not distinguish among these models. Once the transformation causes the data to satisfy the additivity requirement, the choice among models is arbitrary. And the psychophysical function is arbitrary, as well, for the right side of Equation 3 is not additive. Additivity is achieved by taking g^{-1} of both sides of the equation; thus the additivity-inducing transformation is the composition $g^{-1}\Psi$ so that the psychophysical function is confounded with the model. When a specific model is assumed, then Ψ can be determined. Although an element of arbitrary convention is logically inherent in the form of any law, practical leverage can be obtained by requiring cross-task generality. For example, the accord of the bisection scale values with those from the instructed averaging study (Weiss, 1972) supports the assumption of arithmetic averaging.

(Received for publication August 12, 1974;
revision accepted November 25, 1974.)