Psychological Bulletin
1972, Vol. 78, No. 1, 64–69

USE OF RANK ORDER DATA IN FUNCTIONAL MEASUREMENT[1]

DAVID J. WEISS [2]
California State College, Los Angeles

NORMAN H. ANDERSON
University of California, San Diego

Subjects estimated average value of angle pairs using magnitude estimations or graphic ratings. These numerical response data follow a simple averaging model. Functional scaling yields a linear relation between subjective and objective angle. The numerical data are then reduced to rank orders, and J. B. Kruskal's monotone analysis of variance (MONANOVA) procedure is applied. This allows a reconstruction of the original metric information from the strictly ordinal information, illustrating the power of MONANOVA in scaling. Limitations of MONANOVA in testing the underlying model are discussed.

Functional measurement procedures have been useful in the study of stimulus integration (e.g. Anderson, 1971; Shanteau[3]; Weiss & Anderson, 1969). The subject is required to integrate several informational stimuli into a single judgment, and a primary theoretical problem is the nature of the integration rule. The integration rule itself involves two scaling problems: measuring the subjective values of the response and also of the stimuli. Functional measurement provides a simple basis for the simultaneous solution of these three problems (Anderson, 1970, Figure 1).

All this work has used numerical response measures which, in the quantitative tests, need to be on equal interval scales. Simple rating scales have worked surprisingly well, but some response measures will not ordinarily be interval scales. Functional measurement allows for a monotone transformation to rectify the response scale (Anderson, 1962b; Bogartz & Wackwitz, 1971). This procedure utilizes whatever metric information is in the measured response.

An extreme case arises when the overt response is in the form of ranks. Of themselves, the ranks contain no metric information, and it might seem impossible to derive any. However, fundamental work by Shepard (1962, 1966) and by Kruskal (1964, 1965) has demonstrated the contrary: with sufficient constraints and assumptions, rank order techniques can, in certain respects, approach the power of interval methods. The present report supports the Shepard-Kruskal approach by demonstrating an equivalence between functional scales obtained from numerical responses and scale values derived from strictly ordinal properties of the same data.

Judgmental Task and Model

A simple task of psychophysical integration was used: Subjects were presented with two angles and asked to estimate their average inclination. The obvious model assumes that the subjective response is just an average of the subjective values of the inclinations. With a factorial stimulus design, the overt response to the two angles in Cell *ij* would then be

$$R_{ij} = ws_i + (1 - w)t_j \qquad [1]$$

where s_i and t_j are the subjective values of the row and column stimuli, and w and $1 - w$ their relative weights. Unequal weighting can handle order or position effects.

Two properties of this simple model have fundamental importance (Anderson, 1970, pp. 155ff). First, if Equation 1 holds, then the data will plot as a set of parallel lines. This parallelism prediction, testable by analysis of variance, constitutes a joint test of the model and of the overt response scale. In this way, therefore, the model can provide a validational base for the response scale. Magnitude

[1] This work was supported by National Science Foundation Grants GB-6666 and GB-21028.

[2] We wish to thank J. B. Kruskal for making his MONANOVA program available and for his helpful comments on some of these analyses. We also wish to thank James C. Shanteau for good advice, as well as Carlyn Joergensen who ran the subjects in this experiment.

Requests for reprints should be sent to David J. Weiss, Department of Psychology, California State College, Los Angeles, California 90032.

[3] J. C. Shanteau. Component processes in risky decision judgments. Unpublished doctoral dissertation, University of California, San Diego, 1970.

estimation, as has been noted (e.g., Garner & Creelman, 1967; Shepard, 1966), lacks such a validational base. The simple psychophysical integration task, despite its apparent triviality, has a basic theoretical role to play (Weiss & Anderson, 1969). Simple integration tasks, it should be added, do not always follow simple models. Thus, Birnbaum, Parducci, and Gifford (1971) and Kreuger (1970) found discrepancies from simple models for averaging and adding two lengths, respectively.

Second, the averaging model, if successful, provides interval scales of the stimuli simply and directly. Indeed, the row means of the factorial design estimate the subjective values of the row stimuli on an equal interval scale (Anderson, 1962a).

Experimental Procedure

The subject was presented with cards containing two angles, constructed according to a 5 × 5 factorial design. His task was to judge their average inclination using two response modes, graphic ratings, and magnitude estimation.

Graphic rating response. Each response was a mark across a horizontal line, 20 millimeters below and parallel to the long edge of a 137 × 72 millimeter sheet of paper. Two anchor stimuli were present throughout this part of the experiment. One had 10-degree and 20-degree angles, the other had 160-degree and 165-degree angles. The subject was told that the anchors corresponded to positions "about here," 2 centimeters from the left and right ends of the scale, respectively.

Magnitude estimation. A standard stimulus with the angles 60 degrees and 120 degrees was always present in this part of the experiment. The subject was told that the average inclination of the standard was to be called 100, and that each stimulus was to be judged in terms of its ratio to the standard.

Stimuli. The 5 × 5 design yielded 25 angle pairs. The left angle had the values 15, 45, 75, 105, and 135 degrees; the right angle had the values 30, 60, 90, 120, and 150 degrees. Each angle was defined by two segments of 30-millimeter length, with vertex toward the left. One segment of each angle was horizontal, 42 millimeters from the lower edge of the long side of the 127 × 203 millimeter stimulus card.

Procedure. The data reported here are from the second and fourth sessions of a five-session experiment, the other sessions of which were concerned with similar judgments of grayness. In each session, the subject judged four replications of a group of single angles, followed by four replications of the 25 angle pairs under one response mode. The eight subjects received $1.87 per hour. Trials took approximately 15 seconds for magnitude estimation, about 5 seconds longer for the graphic rating since the experimenter read the response to the nearest millimeter on each trial.

Results

Averaging model. Figure 1 plots the raw data means for the two response conditions. The right-hand angle is on the horizontal axis, and the five curves correspond to values of the left-hand angle. The averaging model predicts that these curves should be parallel, and this is approximately the case. The graphic ratings are a bit more regular than the magnitude estimations, a recurrent finding in the writers' research program. The plot of geometric means of the magnitude estimations gave virtually the identical picture.

The statistical test of parallelism is given by the two-way interactions in the analysis of variance. These were nonsignificant, with $F = .91$ and 1.31 ($df = 16/112$), for graphic and magnitude ratings, respectively. Overall, the data follow the model quite well, though deviations in particular subjects or in particular components of the interaction cannot be ruled out.

Nonmetric rank analysis. Since the raw data appear reasonably additive, it is appropriate to ask whether they can be reconstructed from knowledge of the rank orders alone. Accordingly, the 25 responses within each replicate for each subject were ranked and fed into Kruskal's (1965) MONANOVA computer program. This program finds that monotone transformation which, in effect, makes factorial data maximally additive. Fitted values were computed for each replicate for each subject.

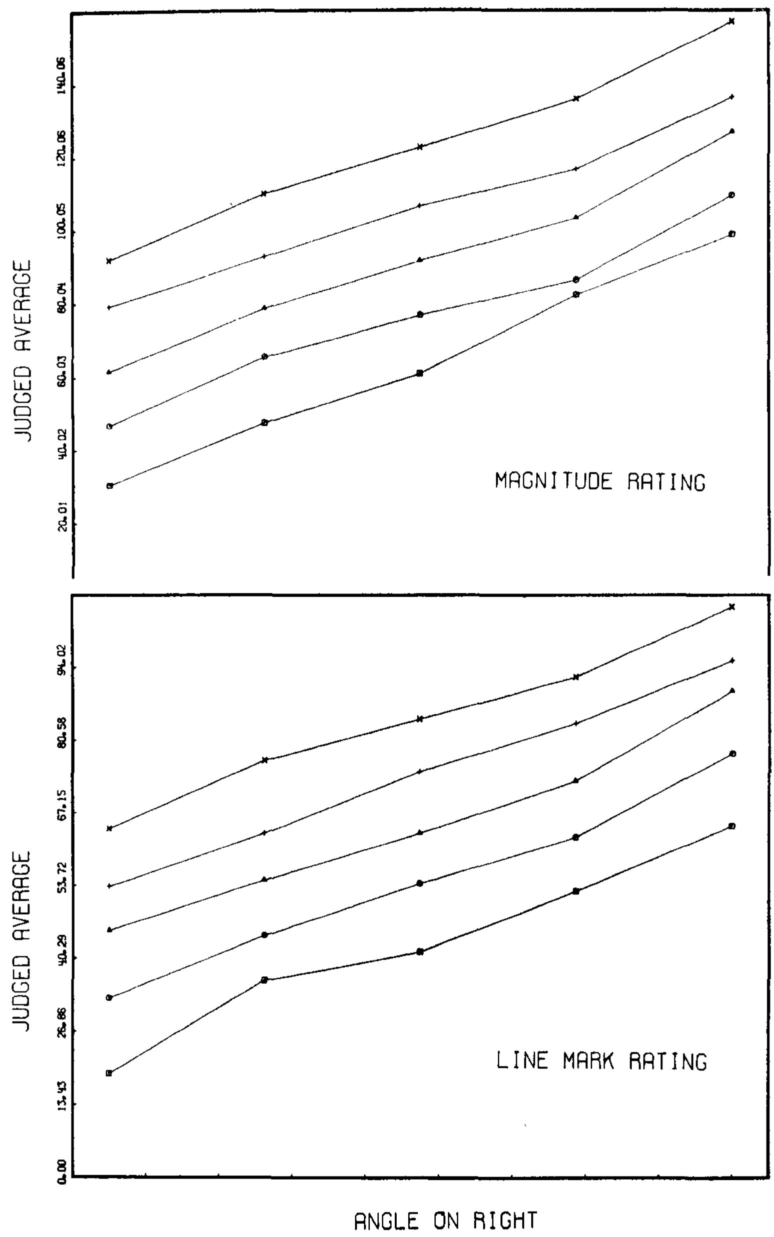

FIG. 1. Raw data means for the two response conditions.

The means of these fitted values are in Figure 2, plotted in the same manner as Figure 1. These reconstructed data appear even more parallel than the raw data, reflecting the success of the additivity-inducing transformation.

Stimulus scaling. According to functional measurement theory, the marginal means of the factorial design constitute interval scales of the subjective values of the stimuli. The linearity of the curves in Figure 1, and their equidistant vertical spacing, mean that the subjective and physical values of the angle stimuli are linearly related. Because of the averaging model, this linearity reflects a corresponding linear relation between the aver-

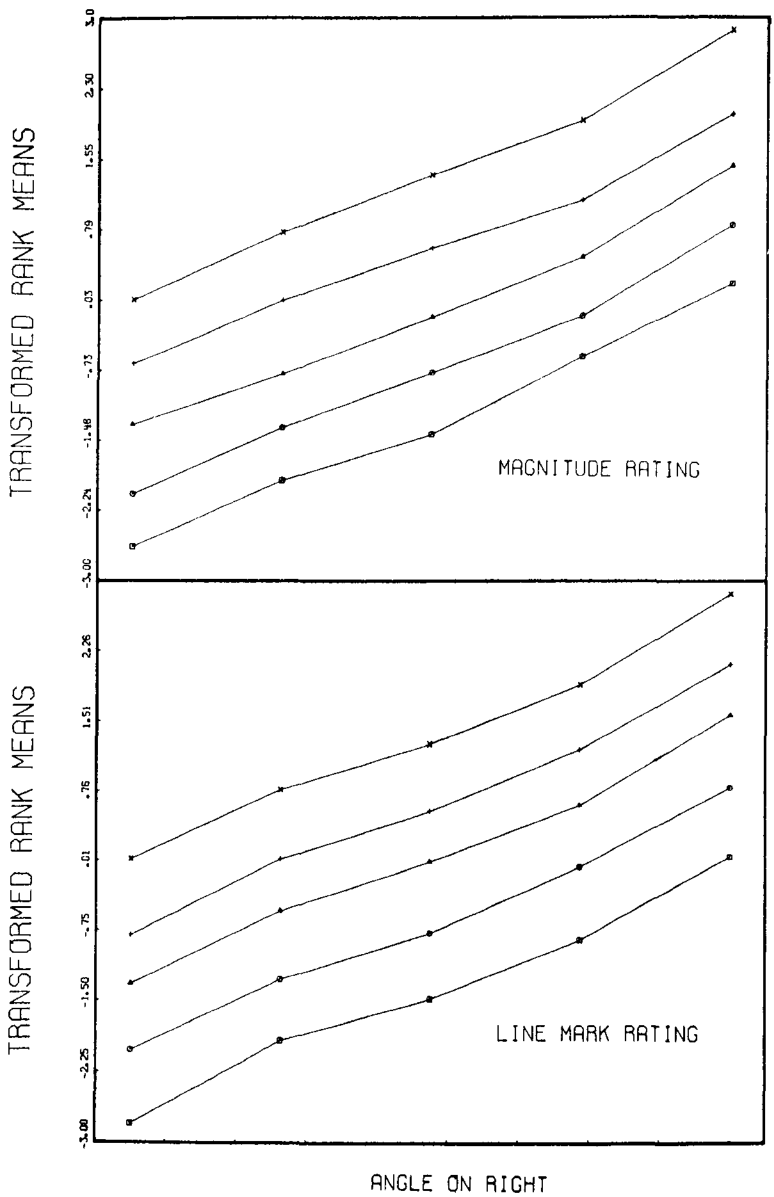

FIG. 2. Fitted data means for two response conditions.

aging response and the mean physical angle, consistent with the work of Miller and Sheldon (1969) on average inclination.

It is important to know whether the transformed ranks actually recover the original metric information. The additivity in the reconstructed data does not guarantee this with a small stimulus design. Accordingly, scale values were estimated from the reconstructed data of Figure 2 in the same way as for the original data of Figure 1. These two sets of scale values are compared in Figure 3. The left-hand angle was chosen arbitrarily, and the marginal raw means plotted on the horizontal, with the transformed rank means on the vertical. Except for a reversal in the magnitude estimation data for one subject, the curves are approximately linear. That means

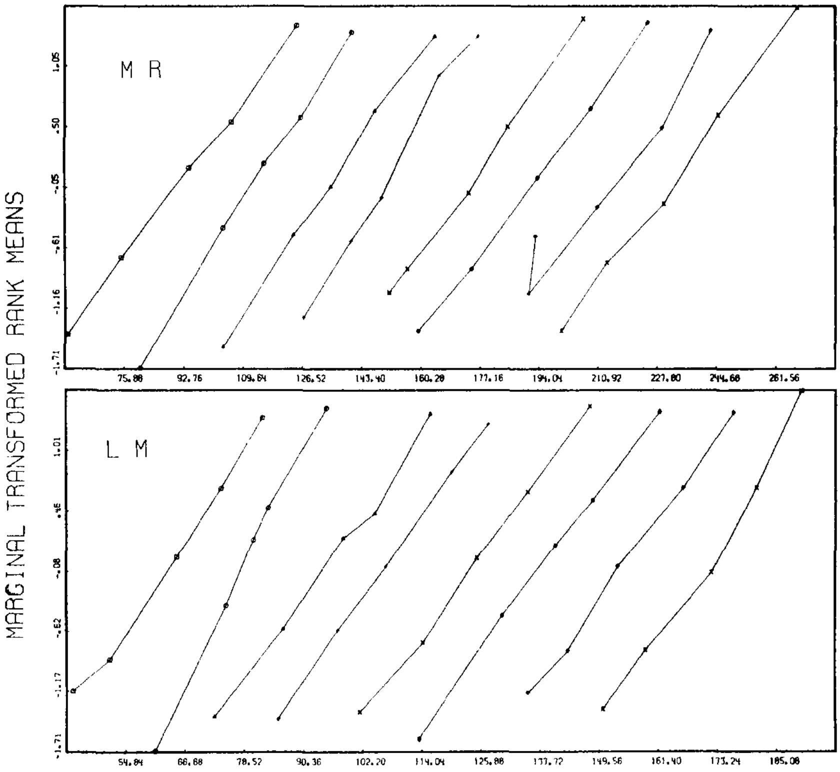

FIG. 3. Comparison of two sets of scale values.

that the rank analysis has recovered the original metric information.

DISCUSSION

The present results illustrate both positive and negative features of nonmetric MONANOVA analysis. On the positive side is the impressive reconstruction, from the ranks alone, of the metric information of the raw data. This is no surprise, of course, in view of the extensive work on multidimensional scaling by Shepard and by Kruskal. However, it does validate their method for a case in which the criterion was set by the initial metric analysis based on functional measurement.

This last consideration brings out certain limitations of MONANOVA, both technical and substantive. At a substantive level, the analysis is completely dependent on the assumption of additivity. If the additive model does not apply, then the MONANOVA scaling will not in general be valid; interval scaling from ordinal data requires knowledge of the appropriate model. And at the technical level, unfortunately, the nonmetric analysis makes it difficult to test whether the basic model is correct. Two main cases require brief comment.

If only rank orders are available, no reasonable assessment of response variability seems possible. With suitably spaced stimuli, for instance, different replications must be expected to yield identical rank orderings. No variability estimate is then available. More generally, any estimate of variability would depend on stimulus spacing as well as true response variability.

MONANOVA is not limited to rank order data, but may also be applied to numerical responses. A transformation based on cell means would then allow a valid estimate of within-cell response variability. However, an unknown number of degrees of freedom are used up so that a valid test would still not be available.

Valid tests could be obtained if MONANOVA were used with preliminary data to define the form of the transformation. In particular, one random half of the data could be used to define the transformation, the other half to test goodness of fit. Alternatively, a power series (Anderson, 1962b; Bogartz & Wackwitz, 1971) allows at least an approximately valid test.

In short, the MONANOVA technique is very powerful for the purpose of scaling, but only if it is known that an additive model applies. It is less effective at determining whether or not such a model does apply. This is a serious problem because there is considerable evidence for averaging processes in judgment, and averaging models are additive only under certain restrictions.

The functional measurement approach has been primarily concerned with the judgment model itself, additive or nonadditive. Scaling, both of stimulus and response, is important, but mainly because it is functional in the elucidation of the integration model.

REFERENCES

Anderson, N. H. Application of an additive model to impression formation. *Science,* 1962, **138**, 817–818. (a)

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

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

Anderson, N. H. Integration theory and attitude change. *Psychological Review,* 1971, **78**, 171–206.

Birnbaum, M., Parducci, A., & Gifford, R. K. Contextual effects in information integration. *Journal of Experimental Psychology,* 1971, **88**, 149–157.

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

Garner, W. R., & Creelman, C. D. Problems and methods in psychological scaling. In H. Helson & W. Bevan (Eds.), *Contemporary approaches to psychology.* Princeton, N. J.: Van Nostrand, 1967.

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

Kruskal, J. B. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika,* 1964, **29**, 1–27.

Kruskal, J. B. Analysis of factorial experiments by estimating monotone transformations of the data. *Journal of the Royal Statistical Society* (Series B), 1965, **27**, 251–263.

Miller, A. L., & Sheldon, R. Magnitude estimation of average length and average inclination. *Journal of Experimental Psychology,* 1969, **81**, 16–21.

Shepard, R. N. The analysis of proximities: Multidimensional scaling with an unknown distance function. I. *Psychometrika,* 1962, **27**, 125–140.

Shepard, R. N. Metric structures in ordinal data. *Journal of Mathematical Psychology,* 1966, **3**, 287–315.

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

(Received November 9, 1970)