INFORMATION TO USERS

This reproduction was made from a copy of a document sent to us for microfilming. While the most advanced technology has been used to photograph and reproduce this document, the quality of the reproduction is heavily dependent upon the quality of the material submitted.

The following explanation of techniques is provided to help clarify markings or notations which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting through an image and duplicating adjacent pages to assure complete continuity.

2. When an image on the film is obliterated with a round black mark, it is an indication of either blurred copy because of movement during exposure, duplicate copy, or copyrighted materials that should not have been filmed. For blurred pages, a good image of the page can be found in the adjacent frame. If copyrighted materials were deleted, a target note will appear listing the pages in the adjacent frame.

3. When a map, drawing or chart, etc., is part of the material being photographed, a definite method of "sectioning" the material has been followed. It is customary to begin filming at the upper left hand corner of a large sheet and to continue from left to right in equal sections with small overlaps. If necessary, sectioning is continued again—beginning below the first row and continuing on until complete.

4. For illustrations that cannot be satisfactorily reproduced by xerographic means, photographic prints can be purchased at additional cost and inserted into your xerographic copy. These prints are available upon request from the Dissertations Customer Services Department.

5. Some pages in any document may have indistinct print. In all cases the best available copy has been filmed.

University Microfilms International
300 N. Zeeb Road
Ann Arbor, MI 48106

HAMBLIN, DAVID LEE

A COMPARATIVE STUDY OF CATEGORY AND MAGNITUDE ESTIMATION SCALING METHODS

THE UNIVERSITY OF ARIZONA M.A. 1983

PLEASE NOTE:

In all cases this material has been filmed in the best possible way from the available copy. Problems encountered with this document have been identified here with a check mark __✓__.

1. Glossy photographs or pages _____
2. Colored illustrations, paper or print _____
3. Photographs with dark background _____
4. Illustrations are poor copy _____
5. Pages with black marks, not original copy _____
6. Print shows through as there is text on both sides of page _____
7. Indistinct, broken or small print on several pages ✓
8. Print exceeds margin requirements _____
9. Tightly bound copy with print lost in spine _____
10. Computer printout pages with indistinct print _____
11. Page(s) _____ lacking when material received, and not available from school or author.
12. Page(s) _____ seem to be missing in numbering only as text follows.
13. Two pages numbered _____. Text follows.
14. Curling and wrinkled pages _____
15. Other _____

University
Microfilms
International

A COMPARATIVE STUDY OF CATEGORY AND MAGNITUDE ESTIMATION SCALING METHODS

by

David Lee Hamblin

A Thesis Submitted to the Faculty of the

DEPARTMENT OF PSYCHOLOGY

In Partial Fulfillment of the Requirements
For the Degree of

MASTER OF ARTS

In the Graduate College

THE UNIVERSITY OF ARIZONA

1 9 8 3

STATEMENT BY AUTHOR

This thesis has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under ruls of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the copyright holder.

SIGNED: David S. Hamblin

APPROVAL BY THESIS DIRECTOR

This thesis has been approved on the date shown below:

G. Domino

G. Domino
Professor of Psychology

8-8-83
Date

ACKNOWLEDGMENTS

I wish to thank my committee members Dr. George Domino, Dr. Terry Daniel and Dr. Bill Ittelson for their guidance and encouragement throughout this project.

This work owes much to Dr. Robert Hamblin, my father, who has made many innovative applications of S. S. Stevens psychophysics into other areas of measurement within psychology and sociology. As he knowledgeably consulted with me, I found even more reasons to respect him.

I received help from two friends in carrying out the experiments; Karen Lawson and Keith James. I thank them here once more.

Finally, I wish to express my admiration for a person who has endured life with me in graduate school: I mean my wife, Roselle.

TABLE OF CONTENTS

	Page
LIST OF ILLUSTRATIONS	v
LIST OF TABLES	vi
ABSTRACT	vii
INTRODUCTION	1
METHODS	14
Experiment 1	14
Subjects	14
Measures	
Procedure	15
Results	17
Experiment 2	19
Subject	19
Measures	19
Procedure	21
Results	23
DISCUSSION	26
REFERENCES	32

LIST OF ILLUSTRATIONS

Figure		Page
1.	Sweetness functions for sucrose and dextrose.	3
2.	Relation between the category confusion scale and the magnitude scale, in semilogarithmic coordinates . .	6
3.	Magnitude estimations of the seriousness of stealing various amounts of money.	8
4.	Relation between the geometric means of judged seriousness of 21 offenses and the maximum penalty prescribed by the Pennsylvania Penal Code	8
5.	Plot of magnitude estimations	9
6.	Psychophysical responses (X-Axis) for each of the party identification category values (Y-Axis) for both the May and October samples.	11

LIST OF TABLES

Table		Page
1.	Pearson Correlation Coefficients.	18
2.	Summary of Pearson Correlations	24

ABSTRACT

Though widely used, the category scaling format yields only an ordinal level of data. Other suspected limitations include loss of information and misclassification of respondents. Research in sensory psychophysics suggests Magnitude Estimation Scaling Procedures to be a relatively direct means of getting accurate exhaustive, responses expressed in ratio level data. Two experiments were designed to test the predictive power of the two procedures against criterion measures. Experiment 1 compared the category and magnitude estimation responses to an achievement test with course grade scores. The second experiment compared correlation coefficients of the two methods between two related attitude inventories. Category measures achieved relatively higher correlations hence better predictions than did the magnitude estimation procedures. High correlations were found between the two scaling methods suggesting a relationship.

INTRODUCTION

The measurement technique most often used by psychologists in researching the direction and intensity of attitude is the category scale. The category scaling method has a number of virtues. In addition to being easy to understand and to respond to, it is relatively easy to score as well. Even so it has a number of disadvantages. Category scales yield only an ordinal level of data which restricts the appropriate use of more powerful statistical procedures. Other important but less well known limitations include loss of information and misclassification of respondents (Lodge and Turskey, 1979). Methodological innovations within the field of psychophysics have provided an alternative which remedies problematic features inherent in category measures. These new psychophysical procedures give a direct means of obtaining an accurate and exhaustive response that is expressed in interval data. This study is methodological in intent and compares category with magnitude scales in order to provide a fair test to directly determine their relative merits and deficits.

Two questions will be addressed: 1) Do magnitude measures, when compared with the category scaling technique, yield routinely higher correlations as is the case in sensory psychophysics; and 2) If magnitude estimation methods are more accurate, will they make a significant difference in predicting a second variable compared with predictions derived from using category scaling?

In addition to responding to these questions, this study will introduce a variation of the traditional magnitude estimation scaling procedure for use in attitude measurement.

Prior to discussing these points more fully, an overview of magnitude estimation and of applications leading to measurement of social stimuli is in order.

S. S. Stevens (1957) discovered the principle governing the relationship between human sensation of a stimulus, and the actual physical properties of that stimulus. Described as the psychophysical law, the principle simply states that equal stimulus ratios produce equal subjective ratios. Stevens found that sensation (ψ) grows in proportion to the stimulus (ϕ) raised to a power $\psi = k\phi^{\beta}$ where

is the exponent of the power function and K is a constant. A convenient aspect of the power function is that when transformed to log-log coordinates it describes a straight line whose slope is the exponent.

$$\text{Log } \psi = \beta \text{ Log } \phi + \text{Log } K$$

Magnitude estimation is the most simple, direct and widely used method for constructing scales having ratio-preserving properties (Stevens, 1969). It essentially consists of matching numbers to stimuli and/or stimuli to numbers. In such an experiment subjects are presented with a series of stimuli, such as varying weights, tastes, tones or lights randomly, one at a time. They are then asked

to match numbers they feel are proportional to the perceived magnitude of each stimuli as compared to an initial reference tone, light or weight.

The estimated numbers (magnitudes) for each stimuli are averaged (by taking geometric means) across subjects and then plotted against the range of stimulus intensities. On ratio-ruled or log-log coordinates the points characteristically fall along (or describe) a straight line. Again the slope of the line is the exponent of the power function (Figure 1). Some examples of exponents of sensory

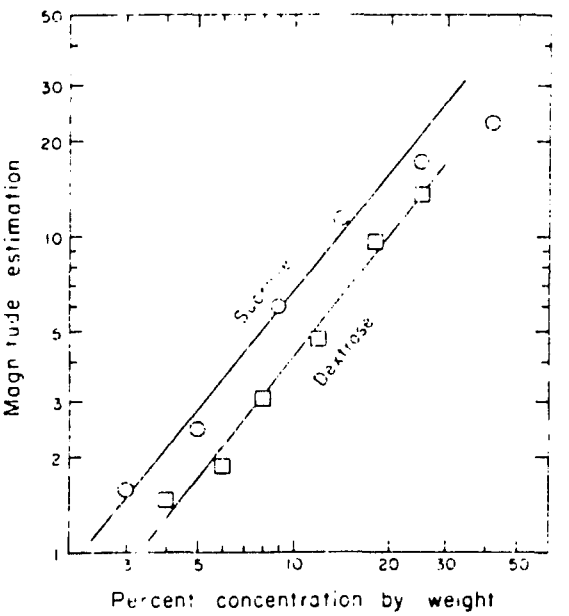

Figure 1. Sweetness functions for sucrose and dextrose.

The symbols are geometric means for 14 Os who judged solutions of sucrose and of dextrose interspersed in the same experiment. The vertical displacement of the two lines through the data suggests that sucrose is about 1.5 times as sweet as dextrose (Stevens, 1969).

continua are loudness of sound (0.67), brightness of light (0.5), line length (1.0), force of hand grip (1.7), and number production (1.0).

Cross-modality matching is a new way of testing the internal validity of magnitude scales (Stevens, 1969). Sensations in one modality may be directly matched with those in another. A subject may adjust the loudness of a sound to match or equal the apparent intensity of a 60 cps vibration on his arm. When the vibration is changed by the experimentor, the subject changes the loudness to equal the new apparent intensity.

The matching function is also a straight line power function (on Log-Log coordinates) but the slope (exponent) is given by (or consists of) the ratio between the exponent for loudness (0.67) and the exponent for vibration (0.95). This means that the stimulus-response relationship of sense modalities can be mapped out without relying on numerical estimation by subjects.

In 1929 Thurstone and Chave used a version of category scaling to scale the strength of attitudes concerning religion. Finnie and Luce (1959) replicated Thurstone's study and method of scaling which correlated highly with the original scale. In addition, subjects made magnitude estimations of attitude strength and the results were compared with the category scale for the same items. The curve obtained by plotting the two methods was highly non-linear. This roughly logarithmic relationship is expected to obtain in sensory psychophysics when these two methods are used to measure prothetic (physical) continuna.

This invariance of results across social and sensory scaling constitutes one test that permits some confidence in using magnitude estimation procedures in the scaling of non-metric stimuli.

Lodge (1976) demonstrated that the logic, method and criteria for validating magnitude scales in sensory psychophysics could also apply for social variables. He used the cross-modality matching paradigm to show that the relationships between measures of non-metric subjective attitudes are consistent with traditional, independently measurable data. Lodge employed three magnitude measures consisting of number production, hand dynamometer and sound pressure responses to estimate strength of support for political statements. He found that the exponent obtained from matching any pair of these measures closely approximated the ratio of the exponents of the given measures.

Thus not only does measurement of attitudes express a power function, but the empirical function obtained matches the theoretical rato between exponents. Both of these studies suggest that although the power law cannot be confirmed directly (due to the fact that the stimuli being measured are subjective), findings are consistent with what one would expect if subjective magnitude is a power function of stimulus magnitude.

A number of studies have demonstrated the potential of magnitude estimation in gauging consensus of attitudes for more than a dozen social variables such as moral judgment (Ekman, 1962), esthetic value of musical selection (Koh, 1963), prestige of occupation (Kunnapas

and Wikstrom, 1963), and strength of expressed attitudes (Ekman and Kunnapas, 1963).

Sellin and Wolfgang (1964, cited in Stevens, 1966) conducted an extensive study that illustrates the utility psychophysics may have in addressing practical problems. They made comparisons of 38 juvenile court judges rating the seriousness of offenses. One half used category scales while the other half made magnitude estimations. When the variability among category assignments were plotted against the magnitude scale, the result was a linear function of the logarithm of the magnitude estimation. This relationship obtains in traditional psychophysical experiments when these two scales are used with sensory continua (Figure 2).

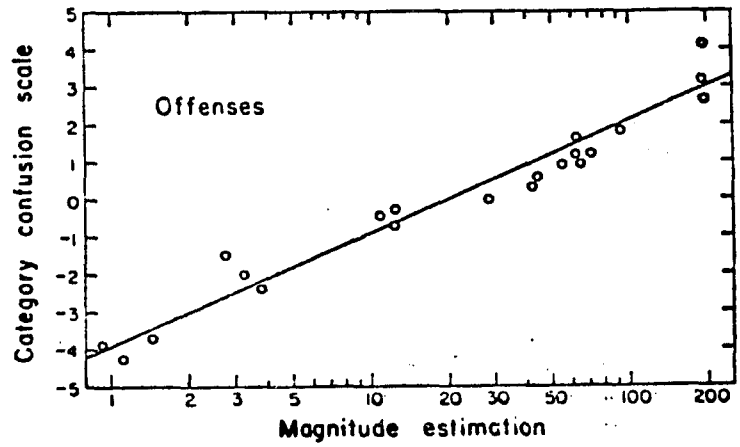

Figure 2. Relation between the category confusion scale and the magnitude scale, in semilogarithmic coordinates. The ordinate represents the Thurstonian scale derived from the confusion or variability among the judgments made on the category scale of seriousness of offenses (Stevens, 1966).

An interesting finding was that age of offenders made no significant difference in determining seriousness. Even more important,

there was an impressive invariance across groups of raters. Judges produced scales comparable with police officers and college students.

Another feature of interest was the additivity of the offense. That there were additive components in magnitude estimations came out only after the analysis of the magnitude estimations. Stealing $5.00 was given the (rounded) number one. Breaking and entering also had the value of one. Breaking into a building and stealing $5.00 was given a value of two. This additive relationship held for more serious and complex crimes.

The value of money was indirectly rated as the judged magnitude of seriousness of a crime involving money grew as a very precise power function of the amount of money stolen (Figure 3).

Finally, when the ratings of seriousness of crimes were matched with the state penal code, a correlation (Pearson) of 0.88 was found for college students and a correlation of 0.94 was obtained for the magnitude-of-offense estimations made by police officers (Figure 4).

Lodge and Turskey (1979) conducted a scaling study confronting category with magnitude scaling of political opinions. They developed a paper and pencil technique which incorporated magnitude estimations in a cross-modal matching paradigm for surveying political opinion in the field. The instrument also contained a variety of category formats which were compared directly with magnitude estimations of opinion for each survey item.

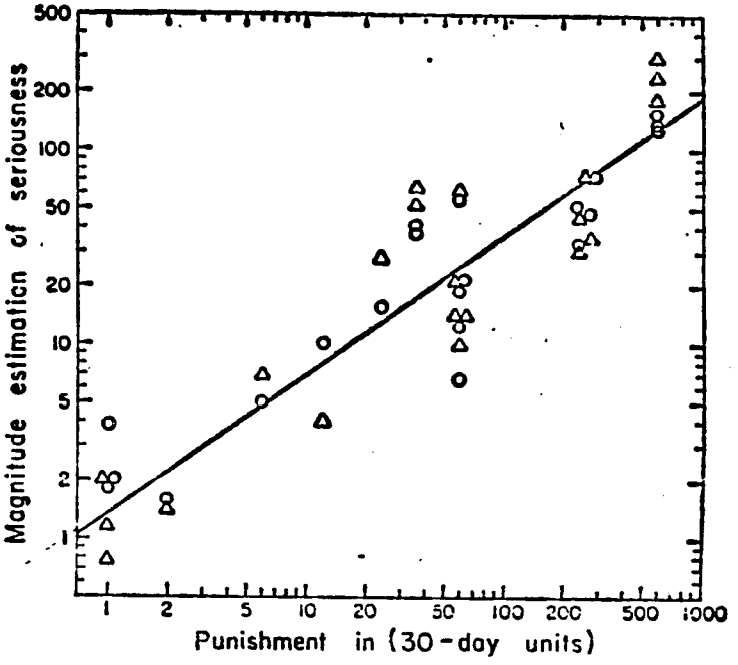

Figure 3. Magnitude estimations of the seriousness of stealing various amounts of money.

The values on the ordinate are the geometric means of estimates made by 105 university students. The line, in log-log coordinates, defines a power function with an exponent of 0.17.

(Figs. 3 and 4: S. S. Stevens, 1960)

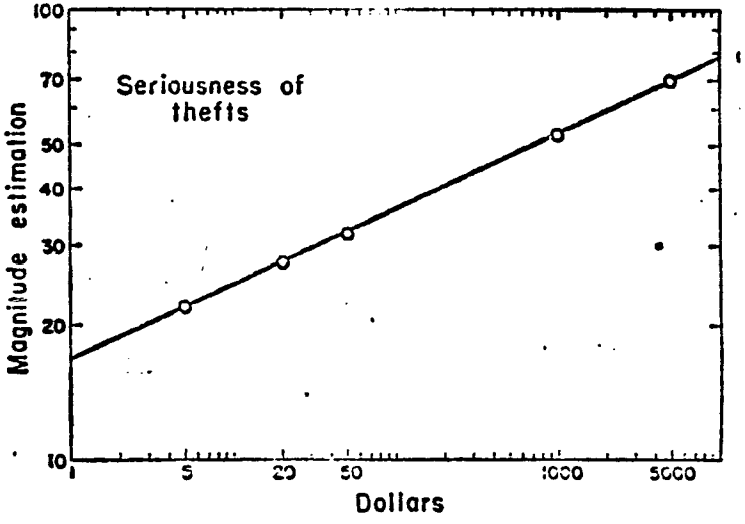

Figure 4. Relation between the geometric means of judged seriousness of 21 offenses and the maximum penalty prescribed by the Pennsylvania Penal Code.

The raters were police officers (circles) and university students (triangles). For plotting purposes the police ratings were multiplied by 0.5; the line through the data has a slope of 0.7.

Lodge and Turskey (1979) plotted magnitude estimations using line production with estimations involving number production. The relationship between the two psychophysical measures was found to be strong with Pearson product-moment correlations ranging from 0.81 to a not common high of 0.96 (Figure 5).

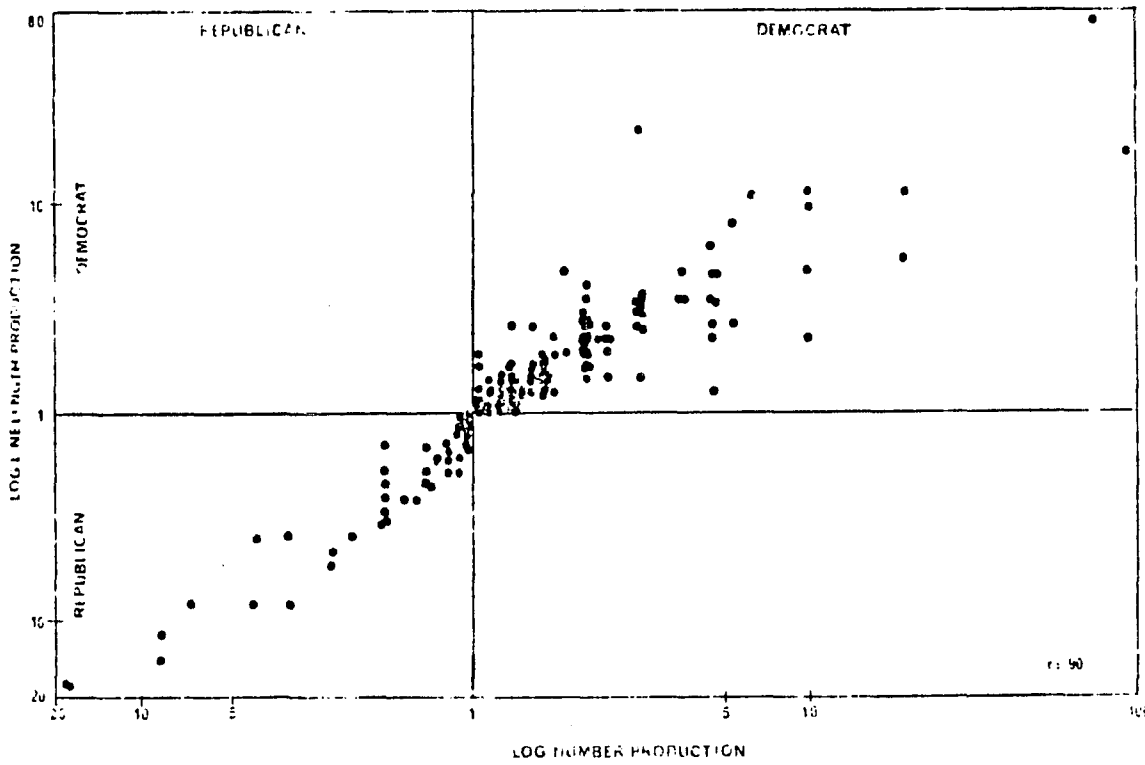

Figure 5. Plot of magnitude estimations.

The variable is the party identification question. Again, the strong relationship between the two measures is obvious. The relationship is shown in logarithmic coordinates (Lodge and Turskey, 1979).

Lodge and Turskey failed to give correlations of category measures matched with magnitude estimation scales, but the results as they are portrayed indicate such a correlation would be relatively

low. Figure 6 depicts systematic misclassification of respondents and loss of information. In this example, the greatest amount of variance occurs in the end most categories. This indicates that the respondents with the strongest feeling are constrained, and thereby unmeasurable within the category format. (This was the case with each item.) Lodge and Turskey found opinion range to exceed 150:1 on some of the questions surveyed. Such intensity information is lost when category scales are used. Also note that the leaning and moderate categories are psychophysically indistinguishable, suggesting an arbitrary classification of respondents. Another interesting feature of category measurement, visually displayed in Figure 6, is that the variance in each category includes respondents who substantially vary in their estimated intensity of opinion.

The limitations illustrated by Figure 6 suggest that correlation coefficients obtained using category scales would be arbitrarilly affected by uncontrollable error, both random and systematic. A number of investigators have found correlations of category scales matched with magnitude estimations to be consistently lower than magnitude scales paired together when measuring sensory continua (Curtis, 1970; Marks, 1972, 1973).

While Lodge and Turskey present evidence that magnitude scales may provide better discriminators of opinion and hence behavior, they provided no separate criterion by which the predictive validity of magnitude over category methods may be empirically demonstrated.

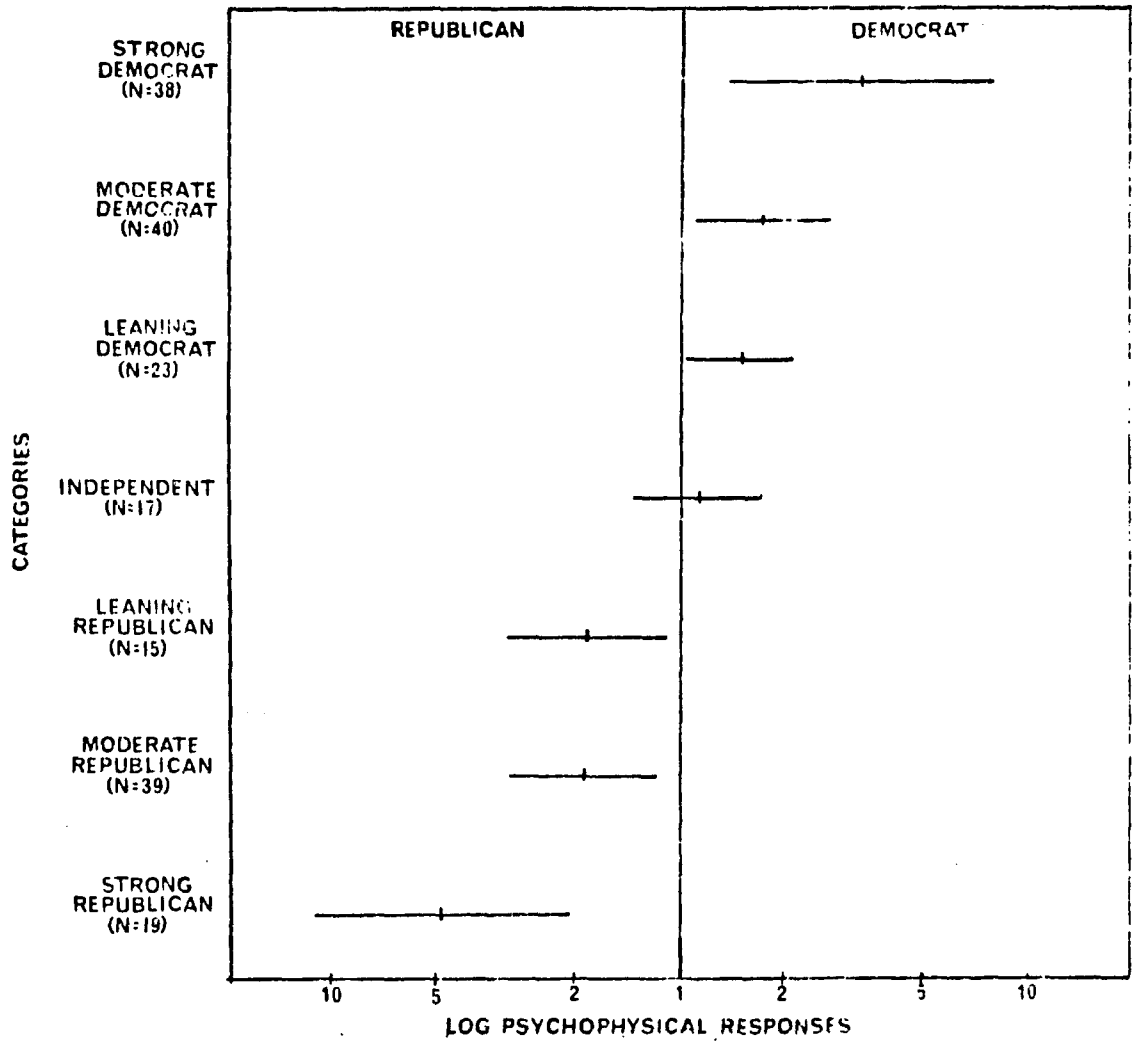

Figure 6. Psychophysical responses (X-Axis) for each of the party identification category values (Y-Axis) for both the May and October samples.

Another criticism arises from the application of traditional psychophysical instructions to the measurement of social (subjective) stimuli. Subjects are instructed to produce a line or number to represent ambivalence which is then used as a referent in producing other estimations. This is a meaningless task. It would be similar to comparing heaviness of weights when the referent for comparison is imagined and an arbitrary number is assigned rather than having an actual weight in hand.

Ambivalence refers to having mixed feelings, to feel both positively and negatively about the same stimulus and will not be constant from item to item. Ambivalence seems to be a common feature of attitude; for example, one may both like and dislike aspects of one's car, vocation, political party, or relationship with others.

A technique called bi-valent magnitude scaling (developed by Dr. Robert Hamblin in discussion with the author) is proposed in order to remedy this problem and will be utilized in one of the experiments in this study. A bi-valent magnitude scale is obtained by soliciting two estimations per item; the first an estimation of the stimulus' positive valence and the second an estimation of its negative value. The larger magnitude indicates the direction of the response. The shorter magnitude is proportionate with the amount of ambivalence the respondent feels toward a given stimuli. The referent, used to insure additional responses are relative in magnitude, will be the first item given. The first bi-valent magnitudes will be arbitrary

in length, but not in proportion to each other, while the following items will be estimated relative to these.

Thus the subject will be free to express both the overall direction and intensity of response as well as any ambivalence he or she may feel toward a given stimulus. This may provide useful information since to say that one endorses an item fully is not equivalent as, say a positive endorsement of equal magnitude coupled with a negative response indicating some ambivalence.

This study proposes to test the following hypotheses:

1. Magnitude scales will yield measures providing significantly better predictions than will category scales when plotted with a second variable,
2. Pearson product-moment correlations will be significantly higher when magnitude scales are plotted together, accounting for more variance than correlations of category with magnitude scales. This hypothesis is consistent with comparison using sensory continua.

METHODS

Experiment 1

Subjects

In this experiment subject volunteers were restricted to those introductory psychology students who are enrolled in discussion sections supervised by a single Teaching Assistant. Since about a quarter of each student's grade is contributed by the graduate instructor, it is desirable to keep this factor uniform across all subjects (a typical teaching assistant has well over 100 students). Fifty-two students were recruited with the specific cooperation of their graduate instructor (and his professor), with extra credit offered as an incentive for participation.

Measures

The subjects were administered a single scale from the California Psychological Inventory. The scale, Achievement via Independence (Ai) was created by Gough (1953b) to predict achievement in college undergraduate courses generally, and undergraduate psychology courses in particular. It appears that the Ai scale predicts achievement in settings where intellectual independence, tendencies toward self-actualization and creativity are valued (Megargee, 1972). The Ai scale consists of 32 items which are rated on a two point category scale, three of which are keyed true and 29 false.

A second, unobtrusive measure was taken at the semester's end. This consisted of the total points accumulated by each subject. These points, which typically range between 100 and 200 represent a magnitude scale which more precisely defines a subject's achievement in the course than does the final letter grade, which represents a category scale.

Finally, subjects were asked to respond to a scale preference questionnaire indicating the scale they preferred, which scaling format allowed the most accurate response and which method, category or magnitude was most difficult to use.

Procedure

Fifty subjects were tested, approximately 17 at a time, in three separate groups. The subjects were told that they were participating in a study comparing methods of representing subjective attitudes and beliefs. They were informed that although their particular results are not the focus of the investigation, that the care and accuracy they employed to portray their attitudes when using the different scales was crucial to the methodological test.

The subjects were presented with category and magnitude versions of the Ai scale in a counter-balanced design. One group took the standard category response format first followed by the magnitude measures. The other groups received them in reverse order so as to match any order effects.

Prior to responding to the magnitude versions of the Ai scale, subjects were given the task of matching numbers to lines. This

can be used as a test as well as a method of providing training and practice.

Subjects responded to the Ai scale using two magnitude estimation procedures, number and line production (Lodge, 1979), to indicate direction and strength of their response to each item. The first response served as a referent for the following items.

Subjects were asked to produce numbers and draw lines to indicate how true or false each item was for them. The response format was uni-valent, that is, only the direction of choice was estimated. Subjects did this by examining the first item on the test. They were then asked to choose a number (or line in the line production condition) which indicated how true or how false that item was for them. They were told that though the number they selected was arbitrary, it now represented a standard to be considered in making proportional judgments on the remaining items.

The instructions for making proportional judgments were as follows. "If, for example, the number 10 represented how true the first item was for you, what number represents how true or how false the next item is for you. If, after looking at it, you find it is twice as true, the correct answer is 20, if half as true the right answer is 5. If it is 10 times as true for you the right answer would be 100, and so on. If it is false, how false is it compared to item number one."

Subjects could place numbers in colums labeled true or false so that negative numbers were not used. Subjects had already been

instructed in the magnitude estimation procedure by making proportional judgments of lines using numbers.

The training using physical stimulii was helpful in having subjects conceptualize making proportional judgments of subjective attitudinal stimuli. All subjects responded to the scale preference questionnaire prior to debriefing.

Student anonimity was maintained by the use of individual matriculation numbers for identification on their test forms. These numbers were submitted to the teaching assistant who matched each number with the total points earned by that student. No permanent record of student numbers was retained.

Results

Three separate responses were obtained from an achievement test (Gough, 1956) consisting of the estalished (true/false) category scale and two magnitude estimation procedures; number and line production. The category responses to the achievement test were scored according to established conventions. Both of the magnitude estimation procedures required standardization as the numerical ranges as well as the range of line lengths used by subjects varied. As in classical psychophysics, subjects had individually selected a range of numbers which expressed their subjective response to the items. The standardization of range was achieved for each magnitude condition by dividing individual subject test scores by the highest score (number or line) produced by that subject. This transformed the range of each subject to 1 and roughly -1. All scores were multiplied by 100. These

standardized scores were transformed into logarithms and then summed to obtain a total line production and a total number production score for each subject. These two scores resulting from magnitude estimation and the category score were correlated with student overall numerical grade and with the letter grade given in the introductory psychology course. The resulting Pearson correlation coefficients are summarized in Table 1.

Table 1. Pearson Correlation Coefficients

	Category	Number	Line
Number est.	-.88		
Line est.	-.66	.74	
Number grade	.36	-.30	-.12
Letter grade	.36	-.31	-.13

The category version of the achievement via independence scale achieved the highest correlation with numberical grade .36, and hence is the best predictor of achievement.

This finding disconfirms hypothesis 1 which stipulated that the magnitude estimation procedure would achieve significantly higher correlations than category measures. The two magnitude measures of number and line production achieved correlation coefficients of -.30 and -.12 respectively. The negative value of the coefficients

result from the use of items scored "false" as the best predictor for grade points.

Hypothesis number 2 was also disconfirmed. Rather than having the correlations be significantly higher when magnitude scales were plotted together, the category method when plotted with number production achieved the highest correlation of the study: -.88. This contrasts with the correlation of .74 found between the two magnitude procedures. Line production gave a considerably weaker correlation of -.61 with the category scale. Letter grade achieved essentially the same correlations as did the numerical grade points. The category correlation with numerical grade, .36, is virtually identical to .35 reported by Gough (Megargee, 1972). The correlation of .30 between number production and numerical grade is significant at the .05 level.

Experiment 2

Subject

The subjects in this experiment were undergraduate students attending the University of Arizona. Thirty-eight students were recruited from the introductory courses with extra credit offered as an incentive for participation.

Measures

One of the two attitude instruments used in this experiments was the Dogmatism Scale (Rokeach 1954). The final version, Form E, contains 40 items. The category scale employed requires subjects to indicate their agreement or disagreement on a range from -3 to

3 with the category of zero excluded in order to force responses toward agreement or disagreement (Rokeach, 1956).

The Dogmatism Scale was designed to measure political and religious dogmatism. The scale was constructed to that individuals adhering dogmatically to such diverse systems as capitalism, communism, Judaism, Catholicism, etc., would all score at the postive end of the category range and opposite in direction to other relatively low in dogmatism.

The components of dogmatism, as conceptualized by Rokeach include a closed belief system, an authoritarian outlook on life, and intolerance toward those with opposing beliefs. Rokeach theorized that a dogmatic person polarizes his beliefs and rejects those whose beliefs extend toward the other pole (Goldstein and Blackman, 1978).

A second well known attitude test, the Authoritarian Scale or California F (for Facist) scale was used (Adorno, 2950). Like the dogmatism scale, the F scale measures attitudes relative to politics, in this case of the extreme right. A convenient feature of this test is that it uses the same category response format as does the Dogmatism scale (described above). An additional advantage lies in the fact that these two attitude measures have been shown to correlate relatively highly with each other at .66 (Rokeach, 1956).

As in Experiment 1, subjects responded to a brief scale preference questionnaire. They were asked to indicate the scale they preferred to use and why. They were also asked which scale, category or magnitude elicited the most complete information in regard to their

attitudes. Finally, they were asked to specify which procedure was most difficult to use.

Procedure

The subjects were tested in groups of about 10 to 15 at one time in order to allow for better supervision and control. The subjects were told that they were participating in a study of methods for measuring attitudes. As in Experiment 1, participants were given to understand that although the focus is not on their particular results that careful and accurate responses on their part would provide the best test for comparing the scales.

Each subject responded to the dogmatism scale a total of 3 times using the standard category scale and two separate versions of magnitude scales. Each subject also responded to the F scale twice using the standard category format and a magnitude estimation procedure. The category and magnitude measures were presented to groups in alternating, counter-balanced fashion to prevent an order effect.

Prior to the introduction of the magnitude measures, a brief five-six minute task of matching numbers to stimuli and stimuli to numbers was given. This was the cross-modality matching paradigm and constitutes a test of accuracy as well as providing training and practice in using lines and numbers in making proportional judgments. Lodge and Turskey (1979) found that typical respondents obtained 97% accuracy but that from 2-5% of their sample were unable to make reliable estimations.

The two magnitude estimations of the dogmatism scale were administered by having subjects make estimations by producing lines to match the subjective intensity of opinion and by utilizing number production. The subjects used the bi-valent method, estimating both the extent of agreement and disagreement they may feel toward a given item. The magnitude estimation procedure used as a comparison measure on the F scale, was number production, also in the bi-valent format. Subjects were given the standard instructions that accompany the use of the category scale.

Subjects were instructed in the use of magnitude estimations initially when resonding to the number and line matching task. They were now told they were to use the same procedure, but this time in estimating the strength of their attitudes or opinions. Subjects were then asked to read the first item of the Dogmatism scale and then to indicate using numbers they felt comfortable with, how much they agreed with the item and how much they disagreed with it. The proportional size of the numbers they choose should match the subjective strength of their attitudes. In responding to the rest of the test items (and in responding to the F scale) subjects should use the first item as a standard so that responses representing strength of attitudes for those items are in proportion to the responses to the first item. For example, if agreement on the first item was represented by 20 then what number would be used to represent agreement on the following item? If you agree half as much, the right answer

is forty, and so on. The scale preference questionnaire was administered prior to debriefing.

Results

Results from Experiment 2 included three responses to the Dogmatism scale (Rokeach, 1956) and two responses to the California F scale (Adorno, 1950) made by each of 38 subjects. The conventional category formats, identical for the two scales, were scored according to established procedures.

This experiment employed bi-valent magnitude estimations as an alternative to the forced choice format used in the category procedure. The bi-valent response obtained estimations of agreement and of disagreement for each item. A convenient feature of the bi-valent format is that the ratio of the two responses preserves proportional judgments regardless of the range used by subjects. This solved the problem of range that occurred in Experiment 1. Since standardization was not required the ratios of the responses to each item were transformed directly into logarithms and summed. This process was performed on both the number and line production responses to the Dogmatism scale, and to the number production response to the California F scale.

The total scores of these three conditions were compared to the category measures of each of these two attitude tests.

Table 2 summarizes the Pearson correlations between the three response formats of the Dogmatism scale and two response formats obtained from the F scale.

The category versions of the attitude scales achieved the highest inter-scale correlation of .64. This finding again refutes hypothesis 1 where magnitude estimation measures were expected to best predict a second variable, in this case a second attitude test. Number production responses to the two tests correlated at .57.

Table 2. Summary of Pearson Correlations.

	Dogmatism Category	Dogmatism Number	Dogmatism Line	F Scale Category
Dogmatism number est.	.80			
Dogmatism line est.	.86	.79		
F Scale category	.64	.49	.54	
F Scale number est.	.44	.57	.48	.71

Hypothesis 2 was contradicated again, this time with both magnitude estimation procedures correlating higher with the category procedure than they did together. Line and number production responses correlated with the category response on the Dogmatism scale with coefficients of .86 and .80 respectively. As in Experiment 1, the correlation between the category responses of the Dogmatism and F scales replicated earlier findings (Rokeach, 1956).

The results of the scaling format questionnaire indicate how respondents feel about category and magnitude scaling. Subjects were asked three questions: 1) which of the methods they found more difficult, 2) which allowed them to provide the most complete and accurate information, and 3) which was the most fun to use

(preference). The category format was assigned the number 5 as a standard and the other procedures were rated proportionally using numerical estimation.

In comparison, number production was perceived as the most difficult with a median response of 7. Line production was second in difficulty with a median of 6.5. The two magnitude estimation procedures were perceived as providing the most complete and accurate response. The median for number production was 9 and for line estimation, 8. Thus, magnitude estimation was, on the average, judged to be 35 percent more difficult than the category format, but 70 percent more accurate.

In response to the question, which procedure was most fun to use, both line and number estimation were judged overall to be equal to the category format; the medians all being 5. However, 63 percent of the respondents prefered one of the two magnitude estimation procedures over the category format. Line production was favored over number production, 1.7 to 1.

DISCUSSION

The results of these two experiments are interesting and useful, but run counter to expectations based on previous research. Both the category scales predicted criterion variables better than did companion magnitude estimation scales. Another even more surprising result was that magnitude estimation scales correlated higher with category measures than they did with other magnitude measures. Again previous research directly comparing category methods with magnitude measures in sensory psychophysics using metric stimuli found opposite results (Marks, 1968, 1974).

In these experiments, magnitude procedures correlated higher when plotted together with category scales, the results of the survey of scaling formats are consistent with findings in sensory psychophysics.

As expected, subjects found number and line production procedures to be 40% and 30% (respectively) more difficult to use than category measures. However, when estimating which procedures obtained the most information accurately expressing respondents' attitudes, subjects reported number and line production to do this 80% and 60% (respectively) better than did category scales. Why would more complete and presumably more accurate data yield lower correlations?

It may simply be that consistent and precise magnitude estimations of subjective stimuli are hard for subjects to make because attitudinal stimuli are so ambiguous (social psychologists cannot

even decide on a definition of "attitude"). Another plausible explanation is that perhaps subjects were not adequately trained in magnitude estimation procedures. This was true in some cases and yet if magnitude estimation procedures are to compete with easier category tasks as alternative scaling formats for self-report instruments, protracted training could be a drawback. As it is the procedures and time involved in obtaining the estimations and scoring them are lengthy and cumbersome compared with the category methods used. However, these difficulties could be resolved by aid of computer should a satisfactory system of scoring be worked out along with higher prediction values to make it worthwhile.

Another possible explanation of the result deals with the nature of category scales themselves. Category scales constrain responses in effect compressing the data to a more constricted range as in the case of the six point scale employed by Rokeach.

In contrast, magnitude estimation procedures preserve the extreme scores. Subjects using the line production method on the Dogmatism scale routinely produced ratios of 1:160 (even here subjects were constrained by width of the answer form). One can see that compressing a range of 1:160 into the category range of 1:6 greatly reduces the variability of scores (by a factor in this case of over 26), while enhancing correlations. There are simply less chances to vary using the category procedure. This reasoning suggests that category measures may produce artificially high correlations by limiting the information they measure. In such a case, it is conceivable

that a correlation could be at least partially an artifact of the scaling method and not due entirely to the quality of the attitude measure itself.

One known problem with Experiment 1, which was not resolved satisfactorily, was with the standardization procedure used to set the variable ranges of subject responses. While the standardization procedure resolved the imminent problem of spurious correlations, the procedure constrained the data in much the same category methods do. For example, one of the subjects in ths experiment used a range of 12 to -11. Let us suppose he initially intended to use a 10 to -10 range but found in the course of answering items that higher proportions more exactly expressed his attitudes towards some items. In such a case, this subject would have appropriately used magnitude estimation to specifiy how true, or how false, the various items were for him. With the standardization procedure, however, this subject's scores were compressed to match the high and low scores of other subejcts. His and others individual variability were lost; the very thing magnitude estimation is designed to preserve was "washed out."

This problem occurred because the range was standardized after, rather than before, the responses were made. There does not appear to be any satisfactory methods to standardize the range after the fact. Even standardizing before responding to items is problematic. Since subjects vary considerably in their responses to items, assigning an arbitrary number such as 10 to the first item as a standard would be meaningless. The "standard" would thereafter represent

the low end of the range for some and the high end for others. A solution which may be used in future studies would be to use a standard, say 10, to represent moderate agreement. Although moderate agreement represents a category (as opposed to a magnitude), proportional judgments could be made using moderate areement as an anchor point.

The bi-valent format proved successful in avoiding the problem of standardization. A methodological problem occurred, however, when subjects used 0 when assigning values to the agreement or disagreement portion of the bi-valent response format. Where a subject gave a ratio of 10 to 0, 10 representing strong agreement and 0 representing no disagreement to an item, the ratio would equal infinity (0). Infinity not only is not computable, but 10 to 0 represents a dogmatic endorsement which when resulitng in 0 does not contribute to the total dogmatism score. This problem was solved by changing all the disagree socres of 0 to 1. This convention gives weight to the dogmatic scores while disregarding the occurrence of non-dogmatic endoresements.

One of the findings of interest regard the phenomenon of ambivalence where the bi-valent measures of agree and disagree were matched by equally proportioned values. A visual inspection of respondents' line and number production records showed that subjects generally expressed ambivalence in mid-range magnitudes. Hence, a subejct employing a range from 50 to -50, ambivalence would tend to be expressed in the ratio of 25/25. This response would indicate moderate agreement and disagreement towards the item.

This finding contrasts with assumptions regarding ambivalence made by both Rokeach (1956) and Lodge and Turskey (1979). Rokeach equated ambivalence with indecision and removed the zero point from his forced choice category scale to eliminate ambivalent responses. Lodge also assumed ambivalence to be near the zero point. He used "ambivalence," represented by a short line or a small numerical value as a standard against which respondents estimated their political party loyalty.

The data from Experiment 2 suggests ambivalence is a more complex phenomenon where indecision may stem from moderate strong attitudes competing rather than having no opinion one way or the other.

While the problematic results of the study are disappointing if inconclusive for magnitude estimation procedures, category measures on the other hand performed relatively well. The category methods used in both studies correlated highly with magnitude procedures as well as with criterion measures. This is especially seen in Experiment 2 where the Dogmatism scale category format, when plotted together with number production scores, correlated at .88. This comparatively high correlation between magnitude estimation and category scales suggest what the underlying relationship between the two might be. Apparently, subjects are very skillful in using categories to approximate subjective ratios of agreement to disagreement.

A visual inspection of the number production ratios on the Dogmatism scale compared with the category measure confirms the general

invariance indicated by the high correlation. Category measures yield ordinal levels of data which means we do not know what the mathematical relationships between the ranking of categories are. This data suggests that there is an underlying metric imposed on the rank order categories in this study, a metric that appears to be shared by most of the respondents. The underlying metric of category scales might vary from format to format; however, the possibility that ratios are being preserved even though they are epxressed through categories gives a rationale for employing the more powerful statistical procedures on at least some category scales.

The fact that the category procedures performed realtively well is encouraging in a discipline where they are so widely used. The magnitude estimation procedures did less well but various problems make these results less conclusive. A number of developments such as the procedure for standardizing the range and the success of the bi-valent magnitude estimation format may be useful in further investigations. An unforseen but instructive aspect of the study was the discovery of a relationship between magnitude estimation and category methods. This is an important finding and needs to be followed up in future research.

REFERENCES

Adorno, T. W. The authoritarian Personality, Harper, New York, New York, 1950.

Curtis, D. W. Magnitude Estimations and Category Judgments of Brightness and Brightness Intervals: A Two State Interpretation. Journal of Experimental Psychology 1970, 83, 201-208.

Ekman, G. Measurement of Moral Judgment. Perceptual and Motor Skills, 1962, 15, 3-9.

Ekman, G., and T. Kunnapas. Scales of Conservatism. Perceptual and Motor Skills 1963, 16, 329-334.

Finnie, B., and R. D. Luce. "Magnitude-estimation, Pair-comparison and Successive-interval Scales of Attitude Items." Department of Psychology, University of Pennsylvania Memorandum, MP-9, 1960.

Goldstein, K., and S. Blackman. Cognitive Style, Wiley, New York, New York, 1978.

Gough, H. G. A Nonintellectual Intelligence Test. Journal of Experimental Psychology, 1153b, 42, 242-246.

Hamblin, R. L. Ratio Measurement for the Social Sciences. Social Forces, 191b, 50, 79-206.

Koh, S. D. Scaling Musical Preferences. Journal of Experimental Psychology 1965, 70, 79-82.

Kunnapas, T., and I. Wilkstrom. Measurement of Occupational Preferences. A Comparison of Scaling Methods. Psychology Laboratory, University of Stockholm Report Number 156, 1963.

Lodge, M., D. Cross, B. Turskey, M. Foley and H. Foley. The Calibration and Cross-modal Validation of Ratio Scales of Political Opinion in Survey Research. Social Science Research, 1976, 5, 325-347.

Lodge, M., and B. Turskey. Comparisons between Category and Magnitude Scaling. American Political Science Review, 1979, 73, 50-

Marks, L. E. Stimulus-Range, Number of Categories, and the Form of the Category Scale. American Journal of Psychology, 1968, 81, 467-749.

Marks, L. E. Sensory Processes: The New Psychophysics, Academic Press, New York, New York, 1974.

Megargee, E. I. The California Psychological Inventory Handbood, Jossey-Bass, San Francisco, CA, 1972.

Rokeach, M. The Nature and Meaning of Dogmatism. Psychological Review, 1954, 61, 194-204.

Rokeach, M. Political and REligious Dogmatism: An Alternative to the Authoritarian Personality. Psychological Monographs, 1956, 70, (18, Whole No. 425).

Sellin, T., and M. E. Wolfgang. The Measurement of Delinquency, Wiley, New York, 1964.

Shinn, A. An Application of Psychophysical Scaling Techniques to the Measurement of National Power. Journal of Politics, 1969, 31, 932-951.

Stevens, S. S. On the Theory of Scales of Measurement. Science, 1957, 108, 677-680.

Stevens, S. S. Sensory Scales of Taste Intensity. Perception and Psychophysics, 1969, 6, 251-256.

Stevens, S. S. A Metric for the Social Conscience. Science, 1966, 151, 530-541.

Thurstone, L. L., and E. J. Chave. The Measurement of Attitude, University of Chicago Press, Chicago, Illinois, 1929.